جامعة الجزائر3

كلية العلوم الاقتصادية و العلوم التجارية وعلوم التسيير

قسم العلوم الاقتصادية

مطبوعة مقدمة لطلبة السنة ثانية ماستر تخصص: اقتصاد كمي

بعنوان:

دروس وتطبيقات على برمجية STATA

إعداد:

الأستاذ: بوعراب رابح

أستاذ محاضر قسم " أ "

السنة الجامعية: 2021/2020

الفهـــــرس

مقدمة..... 1

الفصل الأول: التعريف ببيئة العمل الخاصة البرنامج Stata..... 2

1. التعريف ببرمجية Stata..... 2

2. المتغيرات وخصائصها..... 4

1.2- القيم المفقودة:..... 5

2.2- الرموز والدوال في Stata:..... 5

3. تعبئة البيانات الى البرنامج:..... 6

1-3. التعبئة المباشرة في متصفح البيانات:..... 6

2-3. استيراد البيانات من ملف Excel:..... 8

4. بعض الأوامر الخاصة بالعمليات على البيانات:..... 10

5. بعض الأوامر الخاصة بانشاء وتحويل المتغيرات:..... 11

الفصل الثاني: إدارة البيانات باستخدام برنامج Stata..... 13

1. تحويل البيانات الشهرية الى بيانات سنوية باستخدام الأمر collapse: 13

2. تحويل الجداول العرضية الى جداول طولية والعكس:..... 14

1-2. التحول من الجداول العرضية (Wide)الى الجداول الطولية (long) 14

2-2. التحول من الجداول الطولية (long) الى الجداول العرضية (Wide) 16

3. ترتيب ملف البيانات..... 19

4. تحديد فئات فرعية من البيانات باستخدام المحددات if و in..... 21

1-4. المحدد in..... 21

2-4. المحدد if..... 22

5. حذف عينة من البيانات او الاحتفاظ بعينة محددة..... 23

1-5. الأمر drop..... 24

2-5. الأمر keep..... 24

6. تبويب البيانات:..... 24

الفصل الثالث: التحليل الاحصائي للبيانات باستخدام برنامج Stata..... 27

1. الاحصائيات الوصفية..... 27

2. الجداول التكرارية والجداول التقاطعية..... 31

اعداد الأستاذ نوري رابح

3. اختبار اعتدالية التوزيع وتجانس التباين..... 36
الفصل الرابع: المخططات والرسومات البيانية باستخدام برنامج Stata 40
1. شكل الانتشار..... 40
2. مصفوفة الانتشار:..... 44
3. المخططات البيانية الخطية..... 45
4. المدرج التكراري..... 49
5. الأعمدة التكرارية والدوائر النسبية..... 52
6. المخطط البياني في شكل صندوق..... 55
7. التمثيل البياني لدالة الارتباط الذاتي والارتباط الذاتي الجزئي للسلسلة الزمنية Correlogram: 57
الفصل الخامس: تحليل التباين والمقارنات المتعددة باستخدام برنامج Stata..... 60
1. اختبار متوسط العينة الواحدة:..... 60
2. اختبار متوسطي عينتين مستقلتين..... 61
3. اختبار تساوي المتوسط للعينات المزدوجة 65
4. اختبار تساوي المتوسطات لأكثر من عينتين مستقلتين 67
5. تحليل التباين الثنائي..... 71
الفصل السادس: تحليل الارتباط والانحدار باستخدام برنامج Stata 74
1. تحليل الارتباط باستخدام STATA..... 74
2. تحليل الانحدار..... 79
2-1. تحليل الانحدار البسيط..... 79
2-2. الانحدار المتعدد..... 80
الفصل السابع: مدخل لتحليل بيانات البانل باستخدام برنامج Stata..... 85
1. الشكل العام للنموذج:..... 85
2. اختبار التجانس لنماذج المعطيات الطولية باستخدام البرنامج Stata 87
3. نموذج الأثر الثابت ونموذج الاثر العشوائي 93
3-1. نموذج الأثر الثابت..... 93
3-2-. نموذج الأثر العشوائي..... 93
4. تقدير نماذج البانل باستخدام Stata..... 94
4-1. تقدير النموذج التجميعي..... 94

2-4. التعريف بمعطيات البانل (البعد الفردي والبعد الزمني) 95
3-4. تقدير نموذج المربعات الصغرى للمتغيرات الصورية LSDV 95
4-4. تقدير نموذج الآثار الثابتة FEM..... 97
5-4. تقدير نموذج الآثار العشوائية REM..... 98
6-4. المفاضلة بين النماذج..... 98
5. بعض الاختبارات حول بيانات البانل:..... 108
1-5. اختبار التوزيع الطبيعي للبواقي (الأخطاء العشوائية): 108
2-5. اختبار الاثر العشوائي للافراد:..... 109
خاتمة..... 111
قائمة المراجع:..... 112

إعداد الأستاذ بوعزاب رابح

مقدمة

يعتبر البرنامج الجاهز Stata من بين البرامج الأكثر استعمالا من طرف الطلبة الباحثين في ميدان القياس الاقتصادي. ففي الوقت الذي اصبح فيه التحليل الاحصائي للبيانات أمرا ضروريا وأداة اساسية لتحليل مختلف الظواهر، فإن الاستعانة بالبرنامج كذلك اصبح ضروريا نظرا لأن حجم البيانات التي يجب على الباحث أن يتعامل معها كبير جدًا، ومن الصعب اجراء التحليلات الاحصائية اللازمة بالطرق البسيطة (كإعداد الجداول الاحصائية ومختلف الحسابات بالاستعانة بالآلة الحاسبة)، بالإضافة إلى تطور الاساليب الكمية الممكن استعمالها والتي تتميز بنوع من التعقيد. ومن هذا المنطلق برزت الحاجة لاستخدام الحاسوب لإعداد الدراسات الاحصائية لاختصار الوقت والجهد.

وقد عرفت الآونة الاخيرة تطوير عدة برمجيات جاهزة تساعد الباحثين والطلبة في التعامل مع البيانات ذات الحجم الكبير، والمبرمجة لإجراء التحليلات الاحصائية المتقدمة، لكن تعتبر بمرمجية Stata من بين أهم البرمجيات المتطورة بشكل خاص للتعامل مع البيانات الإسمية والكمية على حد سواء، بطرق بسيطة وسهلة في متناول الطلبة والباحثين خاصة في مجال العلوم الاقتصادية والتي تتميز بهذا النوع من البيانات.

وارتأينا ان نضع بين ايدي طلبتنا في العلوم الاقتصادية تخصص الاقتصاد الكمي هذه المطبوعة، التي تتناول شرح بعض الطرق الاساسية للتعامل مع البرنامج Stata في اعداد الدراسات الاحصائية والقياسية في ميدان العلوم الاقتصادية وفق البرنامج المقرر من طرف الوزارة الوصية، من خلال استخدام بعض الأمثلة البسيطة التي تساعد الطالب في التدريب على البرنامج والتعود على استخدام الاوامر الأساسية للبرنامج.

واضع بين ايديكم عنوان البريد الالكتروني: rabahbouarab@gmail.com للادلاء بالملاحظات الممكن ملاحظتها من أجل اثراء هذا العمل وحتى الاخطاء المحتمل وجودها فيه.

الفصل الأول: التعريف ببيئة العمل الخاصة البرنامج Stata

1. التعريف ببرمجية Stata

برمجية Stata هي احدى الحزم الاحصائية متنوعة المهام وذات الاستخدامات العامة، تم استحداثها بواسطة الشركة Stata Corp سنة 1985، ويتم استخدام هذه البرمجية من قبل المؤسسات العلمية والتجارية في مخلف بلدان العالم، وتتركز استخداماتها في الابحاث والدراسات الاقتصادية والاجتماعية والسياسية. تعمل هذه البرمجية في بيئة النوافذ ويندوز، كما تتوفر لأنواع وأنظمة التشغيل الاخرى مثل لينكس وماكنتوش.

يتميز البرنامج بوجود طريقتين لإعطاء الاوامر وتنفيذ عمليات التحليل المختلفة، حيث يمكن ان يتم تنفيذ مختلف الاوامر من خلال القوائم (مثلها مثل البرامج الجاهزة كوورد واكسل مثلا)، أو بإعطاء الاوامر كتابيا في الفراغ الخاص بالأوامر (Commande).

ومن مميزات هذا البرنامج أنه يتيح امكانية كتابة برامج مخصصة تقوم بحساب متغير أو مقياس أو مؤشر معين مثلا، وذلك من خلال عدة أوامر برمجية توضع معا وتسمى برنامج، يتم تنفيذها بخطوة أو خطوتين لنختصر عدة خطوات. ويمكن استخدام هذه الخاصية لنقوم بعمليات قد لا تكون متضمنة ضمن قوائم البرنامج الاحصائي او أوامره.

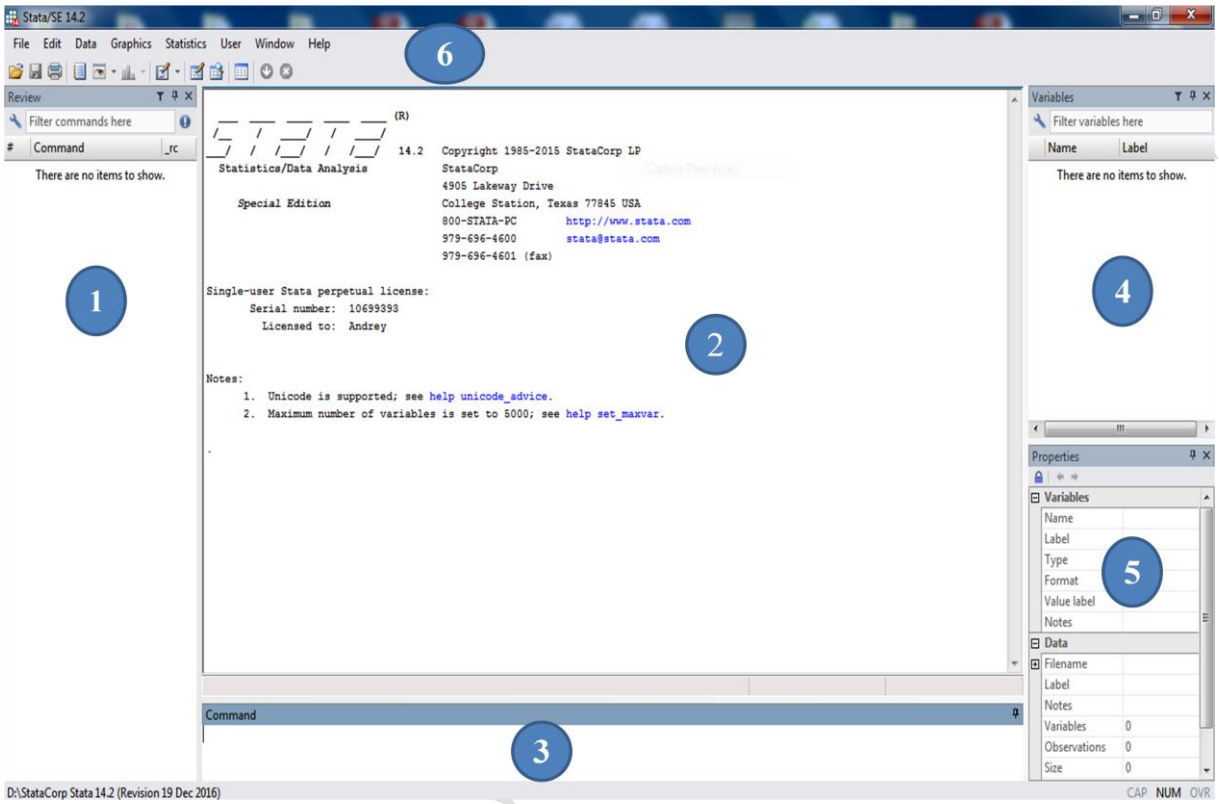
وتقوم الشركة المنتجة للبرنامج Stata بإصدار العديد من النسخ من اهمها ما يلي:

- Stata/IC (Intercooled Stata): وهي النسخة الافتراضية أو الاساسية، يتم استخدامها لأغراض التدريب، يمكنها التعامل مع 2047 متغير.
- Stata/SE: وهي نسخة أخرى من البرنامج قادرة على التعامل مع 32766 متغير، كما انها تستطيع استيعاب مصفوفات أكبر ومعلومات أكبر بداخل المتغيرات. ويعتمد عدد المشاهدات التي يمكن استيعابها على مقدار الذاكرة الموجودة على جهاز الكمبيوتر المستعمل على ان لا تتعدى 2 مليار مشاهدة.
- Stata/MP (Multi Processors): وهي نسخة مدعمة للمعالجات المتعددة (وحدات معالجة مركزية متعددة)، وتتعامل مع نفس المقدار من المتغيرات كسابقتها ولكنها اسرع في عمليات التحليل والمعالجة، والجدير بالذكر ان هذه النسخ الثلاثة متوفرة لأنظمة التشغيل 32 و 64 bite، تمتد قدرات هذه البرمجية لشمل المهام التالية:

- ادارة قواعد البيانات؛
- التحليل الاحصائي؛
- المحاكاة (التنبؤ)؛
- المخططات البيانية؛

- البرمجة.

بعد تثبيت البرنامج يتم تشغيل البرمجية وتظهر الشاشة الرئيسية للبرنامج حسب ما هو مبين في الصورة التالية:



وهي تتكون من 6 أجزاء رئيسية هي:

1. Review: تُظهر هذه القائمة آخر أوامر قمت بتنفيذها، حيث تنتقل الأوامر الى هذه القائمة مباشرة بعد تنفيذها، وتكمن الفائدة من هذه القائمة في امكانية مراجعة الأوامر التي تم تطبيقها في أي وقت، كما تفيد في اختصار كتابة أوامر طويلة من خلال اختيارها من هذه القائمة حيث يقوم البرنامج بكتابتها مباشرة من مكان سطر الأوامر رقم (3) بالنقر على Entrer.
2. Output Screen: وفيها تظهر مخرجات الأوامر، ففيها يظهر البرنامج نتائج التحليل المراد أو المطلوب منه، كما يظهر من خلال هذه الشاشة الاعلان عن وجود خطأ في الصياغة أو غيرها من خلال رسالة توضيحها.
3. Command: هي صفحة بيضاء تسمح بكتابة الأوامر وتلقينها للبرمجية من أجل تنفيذها.
4. Variables: تظهر هذه القائمة مجموعة المتغيرات التي قمت بتعريفها، أو الموجودة في قاعدة البيانات التي قمنا بفتحها، وتجدر الاشارة هنا الى النسخ الأولى للبرنامج (حتى النسخة 9) تظهر فقط اسم المتغير، بينما في النسخ الحديثة فتظهر اسم المتغير ونوعه و طبيعته...

5. Proprties: يظهر فيها مختلف صفات المتغيرات عند النقر على متغير معين يظهر من خلال هذا الجزء الصفات الاساسية الخاصة به، بالاضافة الى خصائص ملف البيانات Data من اسم وحجم وعدد المتغيرات الذي يحتويه الملف.
6. شريط الأدوات والقوائم: وهو شريط يحتوي على مجموعة من القوائم والنوافذ مشابهة لمختلف برمجيات الويندوز تسهل عملية التعامل مع البرنامج، دون الرجوع الى البحث في قوائم البرنامج.

ملاحظة:

تستخدم البرمجية الأحرف الصغيرة (minuscules) في صيغ الأوامر، لكن يمكن تسمية المتغيرات بالأحرف الكبيرة (majuscules) مع مراعاة أن البرنامج يفرق بين المتغيرات المسماة بأحرف كبيرة وصغيرة، فمثلا: كتابة الأمر LIST لا يعمل بل يجب يكون بأحرف صغيرة list كما أن التعريف بالمتغير X list ليس نفسه x list، فالبرنامج يعتبر أن المتغير x يختلف عن المتغير X، فيمكن للباحث تسمية متغيرين بنفس الرمز x ويكون الاختلاف في حجم الحرف (بأحرف صغيرة وأخرى كبيرة) على عكس البرامج الاحصائية الأخرى.

2. المتغيرات وخصائصها

يميز برنامج Stata بين نوعين من المتغيرات، متغيرات كمية ومتغيرات اسمية أو نوعية (كيفية). ويتخصص في تخزين كل منها حسب خصائص معينة، فالمتغيرات الكمية تخزن وفق 5 صيغ مختلفة حسب الحاجة، فالمتغيرات التي تأخذ فقط القيم العددية الصحيحة تخزن وفق الصيغ التالية:

نوع المتغيرة	القيم المقبولة
byte	± 100
int	± 3200
long	$\pm 2\times 10^{9}$

أما المتغيرات التي تحمل قيم صحيحة وعشرية فيتم تخزينها وفق الصيغ التالية:

نوع المتغيرة	القيم المقبولة	تدقيق
float	$\pm 10^{36}$	10^{-8}
double	$\pm 10^{308}$	10^{-16}

لكن على العموم يقوم البرنامج اتوماتيكيا بتخزين المتغيرات الكمية (الرقمية) بصيغة Float.

بينما يتم تخزين البيانات الاسمية بصيغة strX، بحيث X يمثل عدد الرموز المخصصة لكتابة قيم المتغيرة. وفي حالة اخذ المتغير الاسمي قيم عددية (رقمية) لا يمكن اجراء أي حسابات عليها حتى وان كانت تحمل قيم عددية.

1.2- القيم المفقودة

بالنسبة للقيم المفقودة يرمز البرنامج للقيم المفقودة بالرمز (.) في حالة المتغيرات العددية (الرقمية)، بينما يرمز للمتعيرات الاسمية بالرمز ( ) أي فراغ. ومعرفة القيم المفقودة أمر مهم لأنه في بعض الحالات يركز البرنامج لبعض القيم غير المحتملة بالرمز (.) ويأخذها بعين الاعتبار عند حساب المؤشرات الاحصائية. فمثلا عند حساب المؤشرات الاحصائية للأفراد الذين تفوق اعمارهم 60 سنة مثلا summarize wage if age>=60 يتم احصاء كل الافراد الذي تفوق اعمارهم 60 سنة اضافة الى القيم المفقودة. وفي هذه الحالة يجب اضافة شرط عدم احتساب الرمز (.). بإضافة العبارة (& <.) اي كتابة العبارة .summarize wage if age>=60 & age<

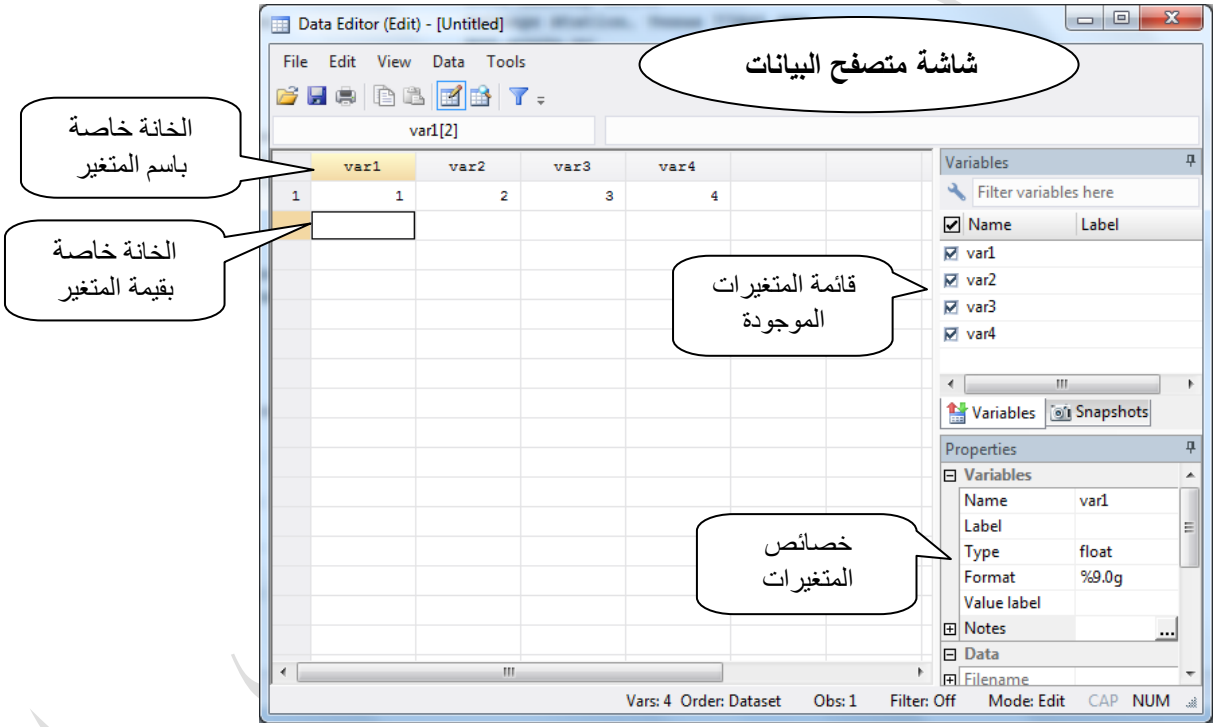
2.2- الرموز والدوال في Stata:

العملية	الرمز	العملية	الرمز
الجمع	+	الطرح	-
الضرب	*	القيمة	/
المساواة	=	عدم المساواة	=~أو =!
القوة (الأس)	^	الجزء الصحيح	int ()
الجذر التربيعي	sqrt ()	الدالة الأسية	exp()
اللوغاريتم	log ()	القيمة المطلقة	abs ()
أكبر من أو أصغر من	<أو >	أكبر من أو تساوي أو أصغر من أو تساوي	=<أو =>
أو	|	و	&
أقل قيمة	min ()	أكبر قيمة	max

3. تعبئة البيانات الى البرنامج

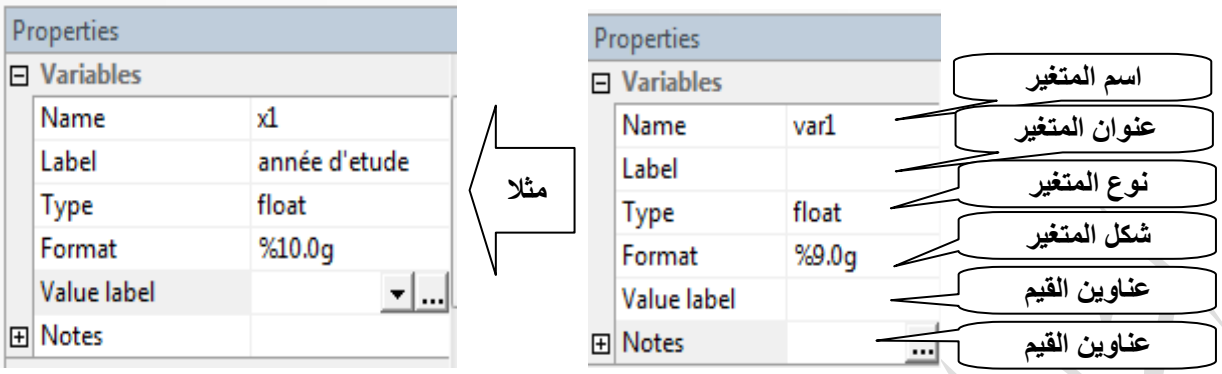
3-1. التعبئة المباشرة في متصفح البيانات

يمكن ادخال المعطيات الى البرنامج من خلال الأمر edit بحيث يسمح بفتح ورقة متصفح البيانات، بحيث يعرض مصفوفة تشبه جداول الاكسل تسمح بكتابة المعطيات وتسمية مختلف المتغيرات، بحيث يسمي البرنامج تلقائيا باسم varX بحيث X يمثل رقم المتغير تلقائيا بمجرد كتابة القيمة في الخانة الأولى في عمود المتغير. ويمكن اعادة تسميتها بالنقر مرتين بزر الفأرة في الخانة الخاصة باسم المتغير واعادة تسميتها وإعطاء عنوان لها وغيرها من الخصائص المتعلقة بالمتغير. ويتم ذلك بكتابة الأمر Edit أو اختيار الأمر Data Editor من قائمة Data، فيظهر لدينا شاشة Data Editor وهي شبيهة بورقة عمل ملف اكسل ومختلف البرمجيات الاحصائية الأخرى، تسمح بكتابة البيانات وتسميتها، وتعديل بيانات تم استرجاعها واستيرادها بصيغ اخرى.

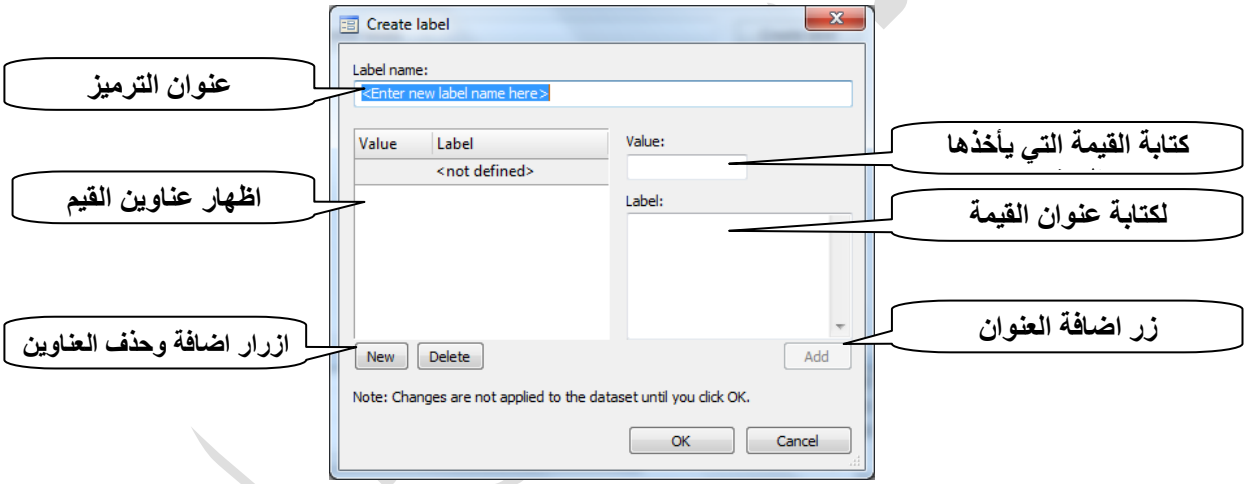


بحيث يعرض البرنامج في الجانب الايمن لمتصفح البيانات مجالات خاصة بالمتغيرات وهي كما يلي:

- Varibles: وفيه يتم عرض كل المتغيرات التي تم تعبئتها؛
- Properties: وهو مجال خاص بخصائص المتغيرات يتم التعديل فيه بالنقر على المتغير المعني في المجال Varibles ثم تغيير اسم المتغير واعطائه عنوان، بالاضافة الى اعطاء عناوين القيم من خلال المجال Value label في حالة المتغيرات الاسمية التي يتم تعبئتها باستعمال ترميز رقمي. بينما لا يمكن تعديل نوع المتغير الذي يتم تحديده تلقائيا من طرف البرنامج. كما يمكن كتابة ملاحظات حول المتغير المعني في الخانة المقابلة لـ Notes.



ولتحديد عناوين القيم الخاصة بالمتغير، يجب النقر على المربع الذي يحتوي على ثلاث نقاط المقابل لـ Value label. بافتراض انه لدينا متغير x1 يعبر عن سنة الدراسة بالنسبة لطلبة الليسانس، في هذه الحالة نقوم بتسمية المتغير بـ x1 ثم نعطيه عنوان « Année d'etude »، وننقر على المربع المقابل لـ Value label ثم نختار الخيار Create label فيظهر صندوق حوار خاص بذلك على الشكل التالي:



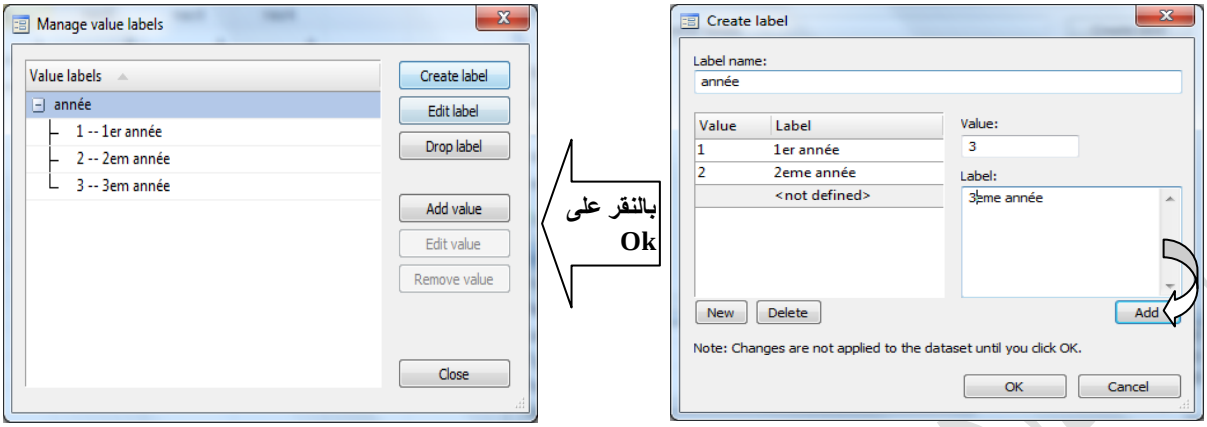
لتحديد عناوين القيم نتبع الخطوات التالية:

1. اعطاء اسم للعنوان في المجال label name وهو الذي يظهر في المجال value label عند اختيار اظهار العناوين بدل القيم.

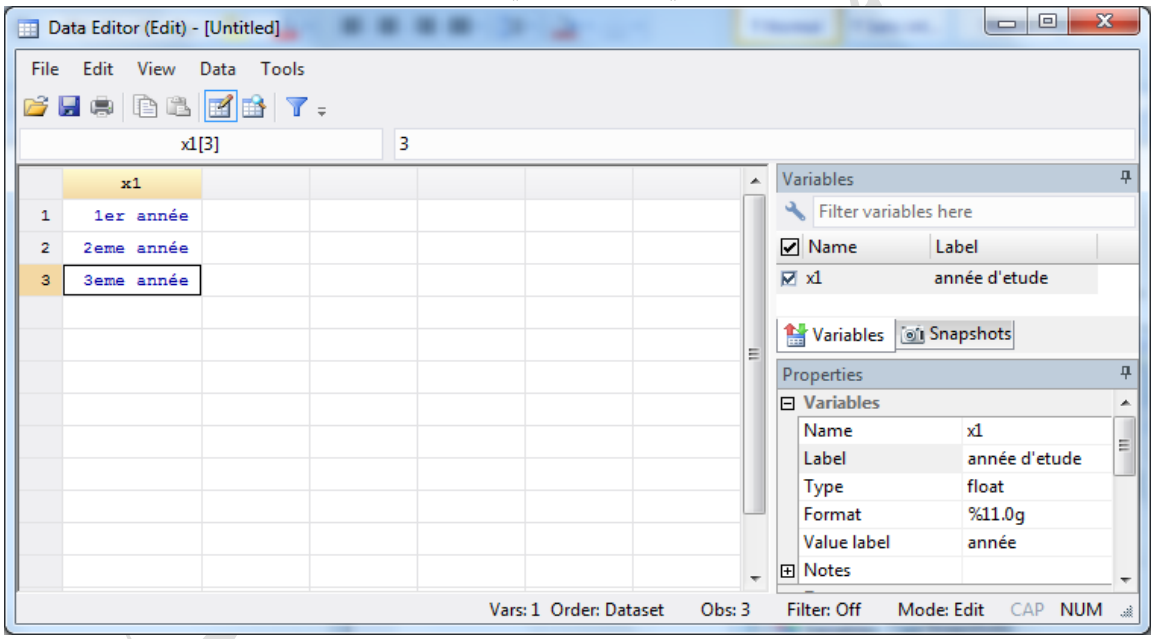
2. كتابة القيم في المجال Value والعناوين في المجال Label الواحدة تلو الأخر مع النقر على زر الاضافة Add في كل مرة.

كما يمكن حذف أو اضافة عنوان جديد بالنقر على الأزرار New و Delete عند الحاجة.

وبالنسة لمثالنا يتم ترميز سنوات الدراسة بتسمية الترميز بـ année ثم كتابة القيمة 1 في Value و1er année في Label ونظيفها بالنقر على الزر Add وهكذا بالنسبة لباقي القيم كما يظهر في الصورة التالية:



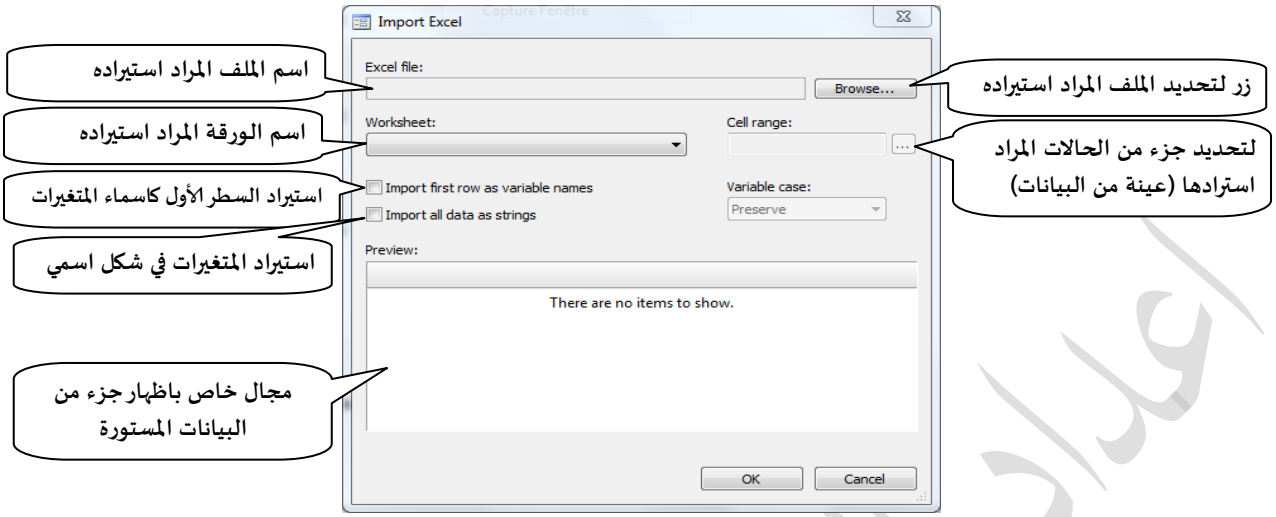
وبذلك يصبح المتغير x1 عبارة عن قيم لها عنوان يمكن اظهار العناوين بالنقر على السهم الموجود في المجال Value label واختيار اسم العنوان année فيظهر البرنامج عناوين القيم بدلا من رموزها وتكون باللون الازرق دليل على انها تأخذ ترميز رقمي. كما يظهر في الصورة التالية:



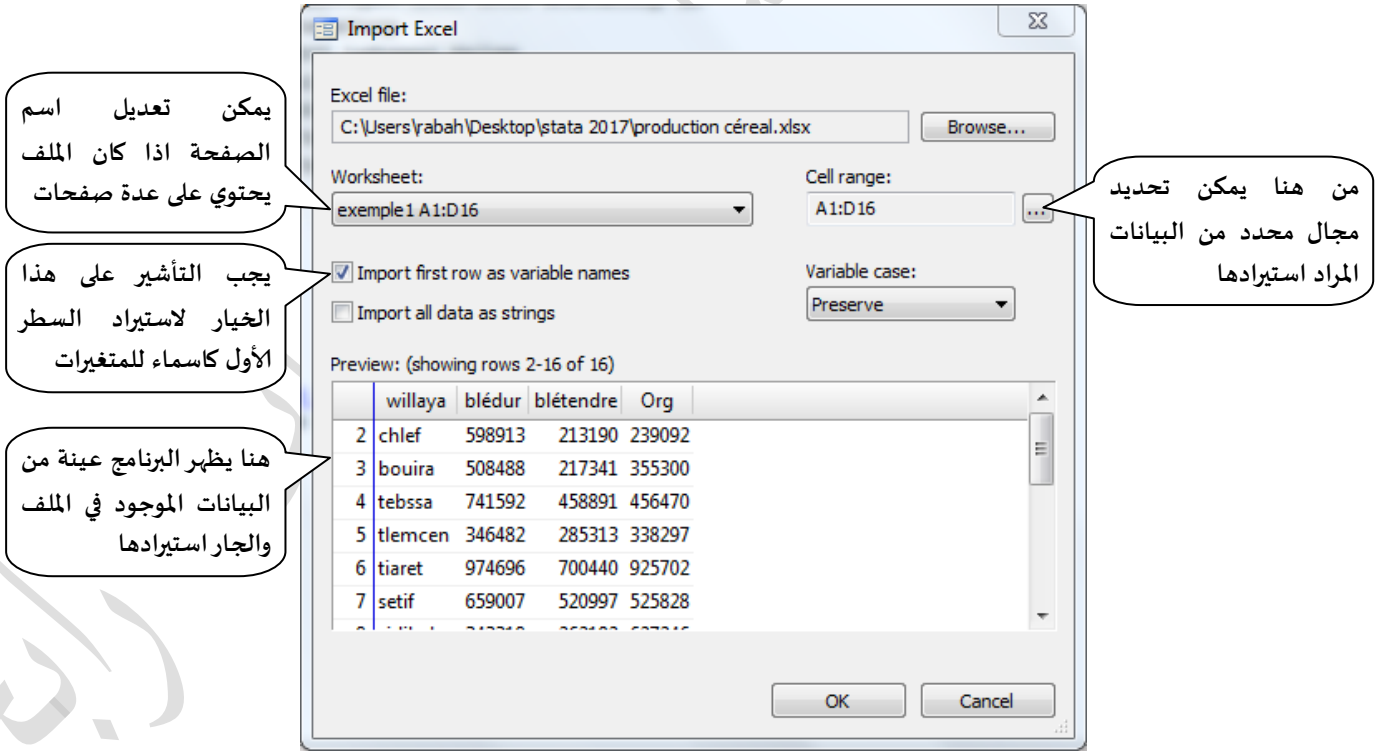
ويتم تخزين الملف بصيغة dta، بحيث يتم تخزين البيانات لإعادة استعمالها لاحقا باستعمال الأمر Open. كما يمكن حفظ مختلف نتائج التحليل التي تم اجراؤها على البيانات في ملف خاص يعرف بملف التسجيل بصيغة SMCL، والتي سنشير الى طريقة تخزينها لاحقا.

2-3. استيراد البيانات من ملف Excel

في كثير من الأحيان يتوفر للمستخدم بيانات مخزنة في ملف Excel، وفي هذه الحالة يمكن للمستخدم استيراد هذه البيانات بطريقة مباشرة دون اللجوء الى تعبئتها أو حتى نسخها ولصقها ربحا للوقت وتجنب الاخطاء في التعبئة. وللقيام بهذه العملية نستعين بالأمر Import من القائمة File للبرنامج وليس في متصفح البيانات، ثم نختار الخيار Excel Speader Sheet(*.xls,*xlsx) فيظهر البرنامج صندوق الحوار التالي:



مثال: نريد استيراد المعطيات الخاصة بانتاج محاصيل الحبوب في بعض الولايات الجزائرية، المخزن بصيغة XLSX باسم production céreals في ملف Stata 2017 في سطح المكتب. نختار الأمر Importe من قائمة File ثم الخيار Excel Speader Sheet(*.xls,*xlsx)، وننقر على الزر Browse لاختيار موقع تخزين الملف المراد ونفتحه باتباع الخطوات اللازمة فيظهر البرنامج صندوق الحوار التالي:



وبالنقر على Ok يتم استيراد البيانات حسب ما تم تحديده من طرف المستخدم.

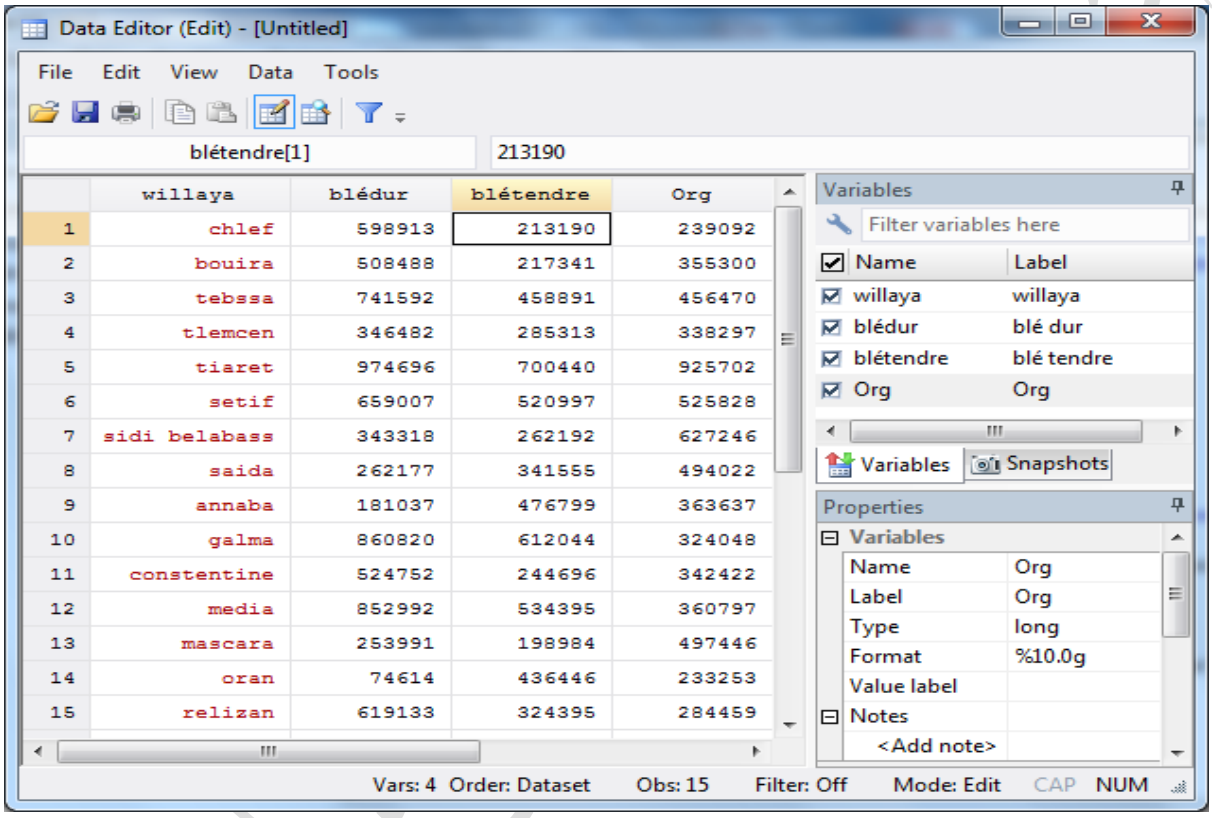
ملاحظة يمكن كتابة الأمر (la commande) التالي في مجال الأوامر لاستيراد الملف:

```
import excel "C:\Users\rabah\Desktop\stata 2017\production céreal.xlsx", sheet("exemple1") firstrow
```

الأمر الخاص باستيراد ملف اكسيل — اسم الملف المراد ومكان تواجده — اسم الورقة المراد استيرادها — لاستيراد السطر الاول كاسم للمتغيرات

ويمكن كتابة الأمر edit لعرض نتائج عملية الاستيراد والتأكد من البيانات المستوردة وتخزينه بصيغة dta الخاصة ببرنامج Stata. أو استخدام الأمر list لعرضها في شاشة المخرجات.

لكن من الأفضل استخدام الأمر edit لعرض متصفح البيانات لتعديل المتغيرات من حيث العناوين والخصائص عند الحاجة. كما ان عملية الاستيراد في بعض الاحيان تستورد متغيرات على انها اسمية بالرغم من كونها رقمية في ملف اكسيل، وعليه يجب اعادة النظر في طبيعة البيانات المستوردة بتأني.



	willaya	blédur	blétendre	Org
1	chlef	598913	213190	239092
2	bouira	508488	217341	355300
3	tebssa	741592	458891	456470
4	tlemcen	346482	285313	338297
5	tiaret	974696	700440	925702
6	setif	659007	520997	525828
7	sidi belabass	343318	262192	627246
8	saida	262177	341555	494022
9	annaba	181037	476799	363637
10	galma	860820	612044	324048
11	constentine	524752	244696	342422
12	media	852992	534395	360797
13	mascara	253991	198984	497446
14	oran	74614	436446	233253
15	relizan	619133	324395	284459

4. بعض الأوامر الخاصة بالعمليات على البيانات:

- Describe: لوصف قاعدة البيانات. بحيث يقوم البرنامج بإعداد جدول ملخص يوصف قاعدة البيانات بكتابة عدد المتغيرات، عدد الملاحظات التي تتوفر في كل متغيرة ونوع المتغيرات والصيغة المخزن بها وغيرها.
- list: تسمح بعرض البيانات في شكل جدول في شاشة المخرجات.
- Summarize: لاستخراج المؤشرات الاحصائية الوصفية للمتغيرات (يمكن استخراجها لكل متغير على حدى بكتابة اسم المتغير أو اسماه المتغيرات المعنية بعد الأمر Summariz. وللحصول على أكثر تفاصيل حول متغير معين نكتب العبارة: su متبوعة باسم المتغير المراد ،(فاصلة) detail، مثلا: sux1 , detail

- codebook: تسمح بعرض مفصل لكل متغير بعرض اسم المتغير، عدد القيم المفقودة، اليقمة الدنيا والعظمى، والوسط الحسابي والانحراف المعياري وبعض المئنيات.
- sort: لترتيب قاعدة البيانات تصاعديا حسب متغير معين يتم كتابة اسمه بعد الأمر sort.
- correlate: يسمح باستخراج مصفوفة الارتباط بين المتغيرات.
- covariance: تسمح باستخراج مصفوفة التباين والتباين المشترك بين المتغيرات.
- tabulate: يسمح باستخراج جدول تقاطع بين متغيرين.
- keep: تعمل على حذف متغيرات والاحتفاظ بتلك المذكورة مثلا: عند كتابة keep x1 يحذف البرنامج كل المتغيرات ويحتفظ فقط بـ x1.

drop: يحذف المتغيرات المختارة، مثلا drop x1 x2 يقوم البرنامج بحذف المتغيرين x1 و x2 من قاعدة البيانات.

5. بعض الأوامر الخاصة بانشاء وتحويل المتغيرات

يتيح البرنامج Stata امكانية انشاء متغيرات جديدة أو تحويل متغيرات عن طريق عمليات حسابية على متغير أو عدة متغيرات موجودة في القاعدة.

ويتوفر البرنامج على أمرين أساسيين يعملان على حساب متغيرات جديدة باستعمال العلاقات الرياضية والإحصائية المختلفة وهما:

- الأمر generate: يعمل هذا الأمر على حساب متغيرات جديدة باستعمال عمليات بسيطة. وذلك بكتابة الأمر generate أو اختصارا gen متبوع باسم المتغير المراد انشاؤه = العبارة الخاصة بالعملية.

مثلا: كتابة الأمر: gen x=log(y) . يقوم بانشاء متغير جديد x يساوي اللوغاريتم النايبيري للمتغير y.

- gen x = 1: لحساب المتغير x الذي يكون ثابت يأخذ القيمة 1 لكل الملاحظات.
- gen x=log(a*b)-sqrt(abs(b)): لحساب قيمة اللوغاريتم النيبيري لحاصل ضرب المتغيرين a و b مطروحا منه الجذر التربيعي للقيمة المطلقة للمتغر b.
- gen x=(sexe== "homme"): لإنشاء متغير جديد x يحمل القيمة واحد اذا كان الفرد ذكر والقيمة 0 لغير ذلك.
- gen x = y [_n-1]: لحساب متغير x المعبر عن القيم المؤخرة للمتغير Y.
- gen x=ln(y) أو gen=log(y) لحساب متغير جديد x يمثل اللوغاريتم النيبيري للمتغير y.

- الأمر egenerate: يقوم هذا الأمر بحساب متغيرات جديدة بعمليات معقدة. مثلا:

- egen x=sd(y)، يسمح بحساب المتغير x المساوي للانحراف المعياري للمتغير y.
- egen x=group(sexe)-، يقوم هذا الأمر بإنشاء متغير جديد لمتغير الجنس بقيم عددية بحيث يقوم البرنامج بترميز فئة الاناث بـ 1 وفئة الذكور بـ 2 .

بالاضافة الى ذلك يتوفر البرنامج على أوامر اخرى من شانها حساب او انشاء متغيرات جديدة حسب رغبة المستخدم نذكر منها ما يلي:

- الأمر xi: يسمح بإنشاء متغيرات صورية للمتغيرات الاسمية مثلا:

xi i.sexe*csp: يقوم بإنشاء متغير صوري يتعلق بمتغير الجنس وتقاطعه بالمهنة. بحيث يظهر البرنامج مهنة الافراد الذين تقابلهم القيمة 1 في المتغير الصوري.

- الأمر replace: يمكنه ان يقوم بنفس العمليات الحسابية التي يقوم بها الأمر generate ولكنه يقوم بتغيير قيم المتغير الموجود (الأصلي) دون انشاء متغير جديد. فمثلا كتابة الأمر replace x=log(x).

- الأمر encode: يقوم هذا الأمر بتحويل البيانات الاسمية الى بيانات رقمية، بحيث يتم اعطاء الرمز 1 للتسمية التي تبدأ بأول حرف ابجدي في البيانات و2 للثاني ...وهكذا، فمثل لترميز متغير الولاية willaya نكتب الأمر التالي:

encode willaya, gen(willayacode)، فيظهر متغير جديد خلاياه مكتوبة باللون الازرق دليل على انها قيم رقمية وليست اسمية، وللاحظة الفرق بين المتغيرين يمكننا التعديل في عناوين القيم مجال variables بحيث نختار Non عوض willayacode. أو بالاستعانة بالأمر list بإضافة العبارة nolabel لإظهار المتغير الجديد في شكله الرقمي ومعرفة الارقام الدالة على كل اسم، فنكتب الأمر:

list willaya willayacode, nolabel، فيظهر البرنامج جدول الترميز لمتغير الولاية.

الفصل الثاني: إدارة البيانات باستخدام برنامج Stata

1. تحويل البيانات الشهرية الى بيانات سنوية باستخدام الأمر collapse

يسمح هذا الأمر بتجميع البيانات في مجموعات بسيطة تتضمن متوسط انحراف البيانات وجعلها صالحة للاستعمال حسب حاجة المستخدم. فمثلا البيانات الشهرية الخاصة بمؤشر اسعار الاستهلاك خلال الفترة جانفي 2012- ديسمبر 2015، يمكن تحويلها من بيانات شهرية الى بيانات سنوية باستخدام احدى مقاييس الاحصاء الوصفي وفي العادة يستخدم المتوسط الحسابي. ويتم استخدام هذا الأمر بكتابة الصيغة التالية: collapse (mean) cpi, by (année)

فيقوم البرنامج بإنشاء بيانات جديدة عبارة عن المتوسط الحسابي لمتغير CPI لكل سنة أي حساب المتوسط الشهري لكل سنة وفق الصيغة: $cpi_t = \frac{\sum_{j=1}^{12} cpi_j}{12}$ لكل سنة t، وتصبح بذلك لدينا البيانات السنوية لمؤشر الاستهلاك.

كما يمكن استخدام مؤشرات أخرى كالانحراف المعياري (sd) بكتابة الأمر: collapse (sd) cpi, by (année)، أو مؤشر الانحراف المتوسط (semean) بكتابة الأمر collapse (semean) cpi, by (année).

مثال: لتكن لدينا البيانات الخاصة بمؤشر الاستهلاك المبينة في الجدول التالي:

الجدول رقم (2-1): قيم مؤشر اسعار الاستهلاك خلال الفترة جانفي 2012- ديسمبر 2015

mois	cpi	mois	cpi	mois	cpi	mois	cpi
janv-12	157,2	janv-13	169,9	janv-14	173,4	janv-15	181,7
févr-12	160	févr-13	170	févr-14	173,8	févr-15	183,8
mars-12	162,3	mars-13	170,8	mars-14	174,6	mars-15	184
avr-12	164,3	avr-13	170,1	avr-14	173,3	avr-15	183,3
mai-12	161,4	mai-13	169,5	mai-14	174,2	mai-15	183,8
juin-12	160,5	juin-13	169,3	juin-14	176,2	juin-15	182,9
juil-12	162,5	juil-13	169,7	juil-14	177,1	juil-15	183,1
août-12	165,2	août-13	170,7	août-14	177,3	août-15	184,8
sept-12	166	sept-13	170,8	sept-14	178,9	sept-15	188,3
oct-12	168,9	oct-13	170,5	oct-14	182,1	oct-15	187,4
nov-12	166,8	nov-13	170,2	nov-14	181,6	nov-15	185,9
déc-12	167,3	déc-13	171,5	déc-14	179,5	déc-15	186,4

المصدر: الديوان الوطني للاحصائيات

والمطلوب تحويل هذه البيانات من بيانات شهرية باستخدام المتوسط الحسابي.

لتنفيذ الأمر يكفي تعبئة هذه البيانات في برنامج Stata مع ضرورة اضافة متغير السنة الى البيانات في عمود على يمين البيانات الخاصة بمؤشر اسعار الاستهلاك ثم كتابة الأمر:

collapse (mean) cpi, by(année)

وبالنقر على الزر Entrer لتنفيذ الأمر يتم تحويل البيانات الى بيانات سنوية باستخدام المتوسط الحسابي كما هي موضحة في الجدول التالي:

	année	cpi
1	2012	170.88847
2	2013	55.226039
3	2014	6.7979475
4	2015	2.8254528

ملاحظة:

ان اضافة متغير السنة الى البيانات ضروري لاجراء هذا الأمر، بحيث يتم اضافة السنة أمام كل ملاحظة خاصة بشهر من تلك السنة.

2. تحويل الجداول العرضية الى جداول طولية والعكس

2-1. التحول من الجداول العرضية (Wide)الى الجداول الطولية (long)

في بعض الاحيان يتوفر الباحث على بيانات احصائية في صورة عرضية، وتحويلها الى شكل طولي يتناسب مع متطلبات الدراسة يكلفه وقتا وجهدا كبيرين، الشكل العرضي المقصود يكون حسب الجدول التالي:

الجدول رقم (2-2): كميات انتاج القمح في بعض الولايات الجزائرية خلال الفترة 2009-2012

pays	2009	2010	2011	2012
CHLEF	809440	671790	752700	1163340
O.E.BOUAGHI	1196273	591274	841400	644127
BATNA	679194	600695	827400	710540
BOUIRA	651793	746345	1060500	1007615
TEBESSA	899030	136493	633600	275000
TIARET	1376500	1659742	1151400	2280600
SETIF	1443300	1444950	1974900	1479608
SKIKDA	414500	451065	614000	572480
GUELMA	850930	1264313	1493600	1509540
CONSTANTINE	700157	882540	1065700	1028692
MEDEA	986890	1161260	1297600	1475790

المصدر: تم اعداده بناء على معطيات وزارة الفلاحة والتنمية الريفية (سلسلة "B" لعدة سنوات)

يتوفر البرنامج على أمر (command) من شأنه تحويل الجدول من شكله العرضي الذي يتم استيراده من ملف اكسيل الى شكل طولي، وذاك بعد اضافة اسم المتغير في كل سنة (في السطر الأول) ليصبح متكون من المتغير والسنة في كل خانة على النحو التالي: blé 2009، blé 2010،...، blé 2012.

ولتنفيذ الأمر يتم استيراد البيانات من ملف اكسيل بالطريقة المذكورة سابقا، ثم كتابة الأمر reshape في المجال Command متبوع بشكل الجدول المراد وفي هذه الحالة (long) ثم اسم المتغيرة (blé) فاصلة، ثم نكتب i(paye) للتعبير عن عمود البلد، وj(année) للتعبير عن عمود السنة كما هو في الصياغة التالية:

.reshape long blé, i(pays) j(année)

وبتنفيذ الأمر ينتج لدينا في متصفح البيانات الجدول التالي:

الجدول رقم (2-3): الشكل الطولي لبيانات الجدول (2-2) بعد التحويل:

	pays	année	blé
1	AIN-DEFLA	2009	1246000
2	AIN-DEFLA	2010	859255
3	AIN-DEFLA	2011	1373000
4	AIN-DEFLA	2012	1561966
5	B.B.ARRERIDJ	2009	797180
6	B.B.ARRERIDJ	2010	863422
7	B.B.ARRERIDJ	2011	1120000
8	B.B.ARRERIDJ	2012	947035
9	BATNA	2009	679194
10	BATNA	2010	600695
11	BATNA	2011	827400
12	BATNA	2012	710540
13	BOUIRA	2009	651793
14	BOUIRA	2010	746345
15	BOUIRA	2011	1060500
16	BOUIRA	2012	1007615
17	CHLEF	2009	809440
18	CHLEF	2010	671790
19	CHLEF	2011	752700
20	CHLEF	2012	1163340
21	CONSTANTINE	2009	700157
22	CONSTANTINE	2010	882540
23	CONSTANTINE	2011	1065700
24	CONSTANTINE	2012	1028692
25	GUELMA	2009	850930
26	GUELMA	2010	1264313
27	GUELMA	2011	1493600
28	GUELMA	2012	1509540
29	KHENCHELA	2009	574800
30	KHENCHELA	2010	251570
31	KHENCHELA	2011	674600
32	KHENCHELA	2012	383800
33	MASCARA	2009	522800
34	MASCARA	2010	338700
35	MASCARA	2011	102800
36	MASCARA	2012	374560
37	MEDEA	2009	986890
38	MEDEA	2010	1161260
39	MEDEA	2011	1297600
40	MEDEA	2012	1475790
41	MILA	2009	856496
42	MILA	2010	788285
43	MILA	2011	989300
44	MILA	2012	1113467
45	O.E.BOUAGHI	2009	1196273
46	O.E.BOUAGHI	2010	591274
47	O.E.BOUAGHI	2011	841400
48	O.E.BOUAGHI	2012	644127
49	RELIZANE	2009	904355
50	RELIZANE	2010	861000
51	RELIZANE	2011	942700
52	RELIZANE	2012	1202420
53	SETIF	2009	1443300
54	SETIF	2010	1444950
55	SETIF	2011	1974900
56	SETIF	2012	1479608
57	SKIKDA	2009	414500
58	SKIKDA	2010	451065
59	SKIKDA	2011	614000
60	SKIKDA	2012	572480
61	SOUK-AHRAS	2009	1299027
62	SOUK-AHRAS	2010	937902
63	SOUK-AHRAS	2011	1109400
64	SOUK-AHRAS	2012	1040979
65	TEBESSA	2009	899030
66	TEBESSA	2010	136493
67	TEBESSA	2011	633600
68	TEBESSA	2012	275000
69	TIARET	2009	1376500
70	TIARET	2010	1659742
71	TIARET	2011	1151400
72	TIARET	2012	2280600
73	TISSEMSILT	2009	692852
74	TISSEMSILT	2010	540230
75	TISSEMSILT	2011	631200
76	TISSEMSILT	2012	809465

ويظهر البرنامج من خلال شاشة المخرجات أنه تم تنفيذ الأمر بوضع الجدول الذي يوضح طريقة تحويل البيانات كما توضحه الصورة التالية:

```
. reshape long blé, i(pays) j(année)
(note: j = 2009 2010 2011 2012)

Data                               wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                       19   ->       76
Number of variables                   5   ->        3
j variable (4 values)                     ->   année
xij variables:
         blé2009 blé2010 ... blé2012      ->   blé
-----------------------------------------------------------------------------
```

بحيث يتضح انه تم تحويل الجدول من الشكل العرضي wide الى الشكل الطولي long بتحويل عدد المشاهدات من 19 الى 76، وعدد المتغيرات من 5 إلى 3، وتعويض j متغيرة (أربعة قيم) الى متغير السنوات، المتغير xij تم تحويلها الى متغير واحد هو blé.

2-2. التحول من الجداول الطولية(long) الى الجداول العرضية (Wide)

كما يمكن تحويل الجداول من الشكل الطولي الى الشكل العرضي أو اعادة الجدول السابق الى شكله الطبيعي باعادة كتابة نفس العبارة تقريبا والاختلاف يكون العبارة الخاصة بشكل الجدول، بحيث نكتب wide عوض long كما يلي:

reshape wide blé, i(pays) j(année)

كما يوضحه الجدول التالي:

الجدول رقم (2-4): طريقة تعبئة بيانات الجدول (2-2) من أجل التحويل:

	pays	blé2009	blé2010	blé2011	blé2012
1	AIN-DEFLA	1246000	859255	1373000	1561966
2	B.B.ARRERIDJ	797180	863422	1120000	947035
3	BATNA	679194	600695	827400	710540
4	BOUIRA	651793	746345	1060500	1007615
5	CHLEF	809440	671790	752700	1163340
6	CONSTANTINE	700157	882540	1065700	1028692
7	GUELMA	850930	1264313	1493600	1509540
8	KHENCHELA	574800	251570	674600	383800
9	MASCARA	522800	338700	102800	374560
10	MEDEA	986890	1161260	1297600	1475790
11	MILA	856496	788285	989300	1113467
12	O.E.BOUAGHI	1196273	591274	841400	644127
13	RELIZANE	904355	861000	942700	1202420
14	SETIF	1443300	1444950	1974900	1479608
15	SKIKDA	414500	451065	614000	572480
16	SOUK-AHRAS	1299027	937902	1109400	1040979
17	TEBESSA	899030	136493	633600	275000
18	TIARET	1376500	1659742	1151400	2280600
19	TISSEMSILT	692852	540230	631200	809465

ويظهر البرنامج من خلال شاشة المخرجات أنه تم تنفيذ الأمر بوضع الجدول الذي يوضح طريقة تحويل البيانات كما توضحه الصورة التالية:

```
. reshape wide blé, i(pays) j(année)
(note: j = 2009 2010 2011 2012)

Data                               long   ->   wide
-----------------------------------------------------------------------------
Number of obs.                       76   ->      19
Number of variables                   3   ->       5
j variable (4 values)           année    ->   (dropped)
xij variables:
                                    blé    ->   blé2009 blé2010 ... blé2012
-----------------------------------------------------------------------------
```

لتحويل جدول يشمل مجموعة من المتغيرات من الشكل العرضي الى الشكل الطولي، يجب ان تكون بنفس الطريقة المذكورة سابقا في الجدول العرضي تضاف المتغيرات الأخرى على يمين المتغير الأول، بحيث يشمل كل عمود قيم المتغيرة في السنة المعنية ويأخذ اسم Xi année .

مثال:

نعتبر جدول كميات انتاج القمح والشعير في بعض الولايات الجزائرية خلال الفترة 2009-2012، في هذه الحالة يكون الجدول على الشكل التالي:

الجدول رقم(2-5): كميات انتاج القمح والشعير في بعض الولايات الجزائرية خلال الفترة 2009-2012

	المتغير الأول				المتغير الثانى			
pays	blé 2009	blé 2010	blé 2011	blé 2012	org 2009	org 2010	org 2011	org 2012
CHLEF	809440	671790	752700	1163340	326760	229000	209180	324500
O.E.BOUAGHI	1196273	591274	841400	644127	2171753	705304	1210087	1151280
BATNA	679194	600695	827400	710540	1154495	743610	967410	914447
BOUIRA	651793	746345	1060500	1007615	515714	522723	563120	527935
TEBESSA	899030	136493	633600	275000	1273900	24554	782700	170000
TIARET	1376500	1659742	1151400	2280600	2213417	1516939	444344	1650000
SETIF	1443300	1444950	1974900	1479608	997900	716951	866500	590845
SKIKDA	414500	451065	614000	572480	89950	74000	72000	77625
GUELMA	850930	1264313	1493600	1509540	200200	289767	271951	299200
CONSTANTINE	700157	882540	1065700	1028692	126850	108500	112775	127744
MEDEA	986890	1161260	1297600	1475790	531647	606835	374681	652788
MASCARA	522800	338700	102800	374560	1298500	898000	353850	1076300
B.B.ARRERIDJ	797180	863422	1120000	947035	197792	255700	242720	264000
TISSEMSILT	692852	540230	631200	809465	266910	274338	191909	227697
KHENCHELA	574800	251570	674600	383800	824200	226900	415000	230000
SOUK-AHRAS	1299027	937902	1109400	1040979	536419	344240	440450	468432
MILA	856496	788285	989300	1113467	418262	354607	340270	416150
AIN-DEFLA	1246000	859255	1373000	1561966	359845	257536	295537	329267
RELIZANE	904355	861000	942700	1202420	642820	391000	368550	574780

المصدر: وزارة الفلاحة والتنمية الريفية (سلسلة "B" لعدة سنوات)

وان كان هناك متغيرات اخرى تضاف على اليمين بنفس الطريقة.

لتحويل الجدول الي الشكل العرضي يكفي كتابة الأمر: reshape long blé org, i(pays) j(année)
فَيُظهر البرنامج في المخرجات الجدول التالي:

```
. reshape long blé org, i(pays) j(année)
(note: j = 2009 2010 2011 2012)

Data                                    wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                            19   ->     76
Number of variables                        9   ->      4
j variable (4 values)                          ->   année
xij variables:
          blé2009 blé2010 ... blé2012          ->   blé
             org2009 org2010 ... org2012       ->   org
-----------------------------------------------------------------------------
```

دليل على تحويل بينات الجدول من الشكل العرضي الى الشكل الطولي بانشاء متغير الزمن الذي يحتوي على 4 وحدات، والولاية 19 وحدة، وبالتالي يكون عدد المشاهدات 76 مشاهدة وهو حاصل ضرب 19×4 لمتغيرتين هما الكميات المنتجة من محصولي القمح blé والشعير Org.

ويظهر جدول البيانات في شاشة data edutor بعد عملية التحويل في الصورة التالية:

الجدول رقم (2-6): الشكل الطولي لبيانات الجدول (2-5) بعد التحويل:

	pays	année	blé	org
1	AIN-DEFLA	2009	1246000	359845
2	AIN-DEFLA	2010	859255	257536
3	AIN-DEFLA	2011	1373000	295537
4	AIN-DEFLA	2012	1561966	329267
5	B.B.ARRERIDJ	2009	797180	197792
6	B.B.ARRERIDJ	2010	863422	255700
7	B.B.ARRERIDJ	2011	1120000	242720
8	B.B.ARRERIDJ	2012	947035	264000
9	BATNA	2009	679194	1154495
10	BATNA	2010	600695	743610
11	BATNA	2011	827400	967410
12	BATNA	2012	710540	914447
13	BOUIRA	2009	651793	515714
14	BOUIRA	2010	746345	522723
15	BOUIRA	2011	1060500	563120
16	BOUIRA	2012	1007615	527935
17	CHLEF	2009	809440	326760
18	CHLEF	2010	671790	229000
19	CHLEF	2011	752700	209180
20	CHLEF	2012	1163340	324500
21	CONSTANTINE	2009	700157	126850
22	CONSTANTINE	2010	882540	108500
23	CONSTANTINE	2011	1065700	112775
24	CONSTANTINE	2012	1028692	127744
25	GUELMA	2009	850930	200200
26	GUELMA	2010	1264313	289767
27	GUELMA	2011	1493600	271951
28	GUELMA	2012	1509540	299200
29	KHENCHELA	2009	574800	824200
30	KHENCHELA	2010	251570	226900
31	KHENCHELA	2011	674600	415000
32	KHENCHELA	2012	383800	230000
33	MASCARA	2009	522800	1298500
34	MASCARA	2010	338700	898000
35	MASCARA	2011	102800	353850
36	MASCARA	2012	374560	1076300
37	MEDEA	2009	986890	531647
38	MEDEA	2010	1161260	606835
39	MEDEA	2011	1297600	374681
40	MEDEA	2012	1475790	652788
41	MILA	2009	856496	418262
42	MILA	2010	788285	354607
43	MILA	2011	989300	340270
44	MILA	2012	1113467	416150
45	O.E.BOUAGHI	2009	1196273	2171753
46	O.E.BOUAGHI	2010	591274	705304
47	O.E.BOUAGHI	2011	841400	1210087
48	O.E.BOUAGHI	2012	644127	1151280
49	RELIZANE	2009	904355	642820
50	RELIZANE	2010	861000	391000
51	RELIZANE	2011	942700	368550
52	RELIZANE	2012	1202420	574780
53	SETIF	2009	1443300	997900
54	SETIF	2010	1444950	716951
55	SETIF	2011	1974900	866500
56	SETIF	2012	1479608	590845
57	SKIKDA	2009	414500	89950
58	SKIKDA	2010	451065	74000
59	SKIKDA	2011	614000	72000
60	SKIKDA	2012	572480	77625
61	SOUK-AHRAS	2009	1299027	536419
62	SOUK-AHRAS	2010	937902	344240
63	SOUK-AHRAS	2011	1109400	440450
64	SOUK-AHRAS	2012	1040979	468432
65	TEBESSA	2009	899030	1273900
66	TEBESSA	2010	136493	24554
67	TEBESSA	2011	633600	782700
68	TEBESSA	2012	275000	170000
69	TIARET	2009	1376500	2213417
70	TIARET	2010	1659742	1516939
71	TIARET	2011	1151400	444344
72	TIARET	2012	2280600	1650000
73	TISSEMSILT	2009	692852	266910
74	TISSEMSILT	2010	540230	274338
75	TISSEMSILT	2011	631200	191909
76	TISSEMSILT	2012	809465	227697

لاحظ أن البيانات مرتبة في شكل سلسلتين زمنيتين للمتغيرين blé و org حسب الولايات، وهي ما يعرف ببيانات السلاسل الزمنية المقطعية (panel data). والبرنامج Stata يسمح بتحويل البيانات من هذا الشكل مهما كان عدد البلدان والفترة الزمنية في وقت وجيز.

3. ترتيب ملف البيانات

ان القراءة الاولية للبيانات في الغالب يستدعي ان تكون مرتبة تصاعديا او تنازليا حسب متغير معين من أجل تسهيل عملية التحليل، خاصة في حال الجداول الصغيرة، ويوفر البرنامج Stata الأمر sort والذي يعني ترتيب ملف البيانات، فيقوم بترتيب الملف حسب متغير معين من اختيار المستخدم. فمثلا الأمر sort blédur يعني ترتيب الملف الخاص بانتاج الحبوب حسب متغير القمح الصلب. لكن الأمر sort يعمل على ترتيب الملف تصاعديا تلقائيا، أما ترتيب الأمر تنازليا فيجب كتابة الأمر gsort متبوعا بالاشارة (-) ثم اسم متغير الترتيب. مثلا كتابة الأمر gsort -blédur.

ملاحظة: إذا كان الملف يحتوي على قيم مفقودة فإنها تظهر في أعلى الملف المرتب تنازليا وفي أسفل الملف المرتب تصاعديا كون البرنامج يعتبر القيم المفقودة أكبر عدد ممكن.

مثال: لدينا البيانات الخاصة بانتاج بعض المحاصيل الزراعية في بعض ولايات الجزائر كما هي موضحة في الجدول التالي:

الجدول رقم (2-7): كميات انتاج محاصيل الحبوب في بعض الولايات الجزائرية

willaya	blédur	blétendre	Org
chlef	598913	213190	239092
bouira	508488	217341	355300
tebssa	741592	458891	456470
tlemcen	346482	285313	338297
tiaret	974696	700440	925702
setif	659007	520997	525828
sidi belabass	343318	262192	627246
saida	262177	341555	494022
annaba	181037	476799	363637
galma	860820	612044	324048
constentine	524752	244696	342422
media	852992	534395	360797
mascara	253991	198984	497446
oran	74614	436446	233253
relizan	619133	324395	284459

المصدر: بيانات وزارة الفلاحة والتنمية الريفية

لترتيب ملف البيانات المبينة في الجدول يكفي كتابة الأمر: **sort blédur** أو الأمر: **gsort blédur** في شريط الأوامر. وبتنفيذ الأمر يظهر البرنامج في متصفح البيانات الملف المرتب تصاعديا كما يلي:

الجدول (2-8): بيانات الجدول (2-7) مرتبة تنازليا حسب متغير blédur:

willaya	blédur	blétendre	Org
oran	74614	436446	233253
annaba	181037	476799	363637
mascara	253991	198984	497446
saida	262177	341555	494022
sidi belabass	343318	262192	627246
tlemcen	346482	285313	338297
bouira	508488	217341	355300
constentine	524752	244696	342422
chlef	598913	213190	239092
relizan	619133	324395	284459
setif	659007	520997	525828
tebssa	741592	458891	456470
media	852992	534395	360797
galma	860820	612044	324048
tiaret	974696	700440	925702

اما في حال الترتيب تنازليا فيجب اضافة الاشارة (-) الى الأمر السابق ليصبح: **gsort –blédur** ، وبتنفيذه يصبح الملف مرتب تنازليا حسب المتغي blédur كما هو موضح في الصورة التالية:

الجدول (2-9): بيانات الجدول (2-7) مرتبة تنازليا حسب متغير blédur:

willaya	blédur	blétendre	Org
tiaret	974696	700440	925702
galma	860820	612044	324048
media	852992	534395	360797
tebssa	741592	458891	456470
setif	659007	520997	525828
relizan	619133	324395	284459
chlef	598913	213190	239092
constentine	524752	244696	342422
bouira	508488	217341	355300
tlemcen	346482	285313	338297
sidi belabass	343318	262192	627246
saida	262177	341555	494022
mascara	253991	198984	497446
annaba	181037	476799	363637
oran	74614	436446	233253

تجدر الاشارة هنا الى ان هذا الأمر لا يؤثر في العمل على البيانات كون هذا الترتيب شكلي فقط على خلاف عميلية تجزئة ملف البيانات.

4. تحديد فئات فرعية من البيانات باستخدام المحددات if و in

احيانا يرغب الباحث اجراء التحليل على مجموعة معينة من من مجموعة كبيرة من البيانات المتوفرة في ملف البيانات، وكغيره من البرامج الاحصائية الأخرى كـ SPSS مثلا يتوفر على الأمر select cases. فالبرنامج stata يتوفر على الخيارين if و in في مختلف الأوامر الخاصة بالبرنامج، لتحديد فئات معينة للتحليل بدلا من القيام بالتحليل على كل البيانات المتوفرة في الملف. ولهذا نجد ان صناديق الحوار الخاصة باوامر البرنامج.

وفي ما يلي سنشرح كيفية استخدام هذين المحددين:

4-1. المحدد in

يستخدم لتحديد عدد معين من المشاهدات من بين مجموعة من المشاهدات باختيار رقم الحالة بكتابة n أو بتحديد رقم الحالة الاخيرة بكتابة هى 1/n. فمثلا لعرض الخمس ملاحظات الأولى من البيانات الخاصة بإنتاج الحبوب في الجزائر أي الولايات الأولى الأكثر انتاجا لمحصول القمح مثلا من أجل معرفة أي الولايات المعنية، نختار اولا ترتيب الملف تنازليا حسب متغير انتاج القمح الصلب بكتابة الأمر **gsort -blé dur** لترتيب الملف تنازليا حسب متغير القمح الصلب، ومن ثم نطلب استخراج الخمس ملاحظات الأولى بالأمر

list willaya blédur in 1/5 بحيث نطلب من البرنامج عرض الخمس ولايات الأولى من حيث انتاج القمح، والرقم 5 يدل على آخر مشاهدة مطلوب عرضها. ويستعمل هذا المحدد مع مختلف الأوامر بكتابة الأمر متبوعا بـ **in 1/n**. مثلا: **summarize blédur in 1/5**.

لتنفيذ الأمر نكتب الاوامر الواحد تلو الاخر مع التنفيذ بالنقر على الزر entrer على الشكل التالي:

ترتيب الملف تنازليا : **gsort -blédur**

عرض عينة البيانات المحددة وهي الولايات الخمسة الأولى (حسب الرتبة): **list willaya blédur in 1/5**

عرض المؤشرات الاحصائية والوصفة الخاصة بالعينة المحددة: **summarize blédur in 1/5**

والنتائج تظهر في شاشة المخرجات كما يلي:

	willaya	blédur
1.	oran	74614
2.	annaba	181037
3.	mascara	253991
4.	saida	262177
5.	sidi belabass	343318

```
. summarize blédur in 1/5
```

Variable	Obs	Mean	Std. Dev.	Min	Max
blédur	5	223027.4	100934.4	74614	343318

2-4. المحدد if

يستخدم المحدد if بغض النظر عن ترتيب البيانات، لاختيار حالات معينة حسب قيم متغير معين، باضافة العبارة if متبوعة بشرط يحدده المستخدم بعد كتابة الأمر المطلوب. فمثلا عرض البيانات الخاصة بإنتاج الحبوب في حالة كمية الانتاج من القمح الصبلب تتجاوز 500 ألف قنطار، نكتب: **list if blédur>500000**

أو حساب مختلف المؤشرات الاحصائية لإنتاج القمح للولايات التي يفوق انتاجها 500000 قنطار، فنكتب: **summariz blédur if blédur>500000**

Variable	Obs	Mean	Std. Dev.	Min	Max
blédur	9	704488.1	162827.9	508488	974696

ونشير هنا إلى ان stata يعتمد في استخدام المحددة if على مختلف العلامات المعروفة كالمساواة (=) وأكبر (>) أو أقل (<) لكن يتميز بنوع من الخصوصية فيمكن التمييز في ذلك ما يلي:

العلامة (=) لوحدها لا تدل على المساواة بل لها معنى اخر ولا تستعمل لوحدها في المحدد if، فلتحديد المساواة في هذه الحالة يجب كتابة الرمز = مرتين أي (==) للتعبير على شرط الطرف الايمن يساوي الطرف الايسر. العلامة (=!) و تدل على عدم مساواة الطرف الايمن للطرف الايسر أي لا يساوي أو يختلف عن. بينما يستعمل الرمزين (=<) و (=>) للتعبير على أقل أو يساوي و أكبر أو يساوي على التوالي.

كما يستعمل الرمز & (and) للمزج بين عدة شروط للتعبير على الشرط الأول والثاني بشرط أن يكون بمتغيرين مختلفين وليس بنفس المتغير مثلا: list if blédur<300000 & Org<400000 لعرض البيانات في حالة كمية الانتاج من القمح الصلب أقل من 300000 والشعير أقل من 400000.

	willaya	blédur	bléten~e	Org
1.	oran	74614	436446	233253
2.	annaba	181037	476799	363637

ويستعمل الرمز (|) للتعبير عن الاتحاد في المجموعات أي على العبارة (أو)، مثلا:

list if blédur<300000 | Org<400000

	willaya	blédur	bléten~e	org
1.	oran	74614	436446	233253
2.	annaba	181037	476799	363637
3.	mascara	253991	198984	497446
4.	saida	262177	341555	494022
6.	tlemcen	346482	285313	338297
7.	bouira	508488	217341	355300
8.	constentine	524752	244696	342422
9.	chlef	598913	213190	239092
10.	relizan	619133	324395	284459
13.	media	852992	534395	360797
14.	galma	860820	612044	324048

ان البرنامج يعرض فقط البيانات التي تتوفر فيها الشروط وفي هذه الحالة هي: كميات القمح تقل عن 300000 قنطار أو كميات الشعير تقل عن 400000 قنطار. ويمكن اضافة العبارة if متبوعة بشرط محدد لجعل البرنامج يعمل فقط على البيانات التي يتوفر فيها الشرط بنفس الطريقة المتبعة بالأمر list.

ملاحظة:

في حال احتواء ملف البيانات على قيم مفقودة، فإن البرنامج يعتبر القيمة المفقودة أكبر عدد ممكن وبالتالي تظهر في الترتيب التنازلي في بداية الجدول وفي الترتيب التصاعدي في نهاية الجدول، وعليه ففي استخدام المحدد if يطرح عرض القيم المفقودة أو اخذها بعين الاعتبار في الحساب مشكلا. ولتفادي ذلك يمكن الاستعانة بالعبارة: **!messing(var)** الذي يقوم باستثناء القيم المفقودة في المتغير المذكور بين وقوسين.

5. حذف عينة من البيانات او الاحتفاظ بعينة محددة

ان المحددين in و if يساعدان في التعامل مع عينات مختارة من البيانات مؤقتا دون غيرها، وبدون حذف للبيانات بحيث يمكن التعامل مع كامل البيانات باستخدام نفس الاوامر بعد حذف العبارة الواردة فيها المتعلقة بالمحددتين in و if . ولكن في بعض الاحيان وخاصة في حال استيراد البيانات من ملفات جاهزة مقدمة من اطراف معينة، يجد المستخدم نفسه امام مجموعة كبيرة من المتغيرات والبيانات لا يحتاج الى التعامل معها، فهنا يحتاج المستخدم لحذفها أو الاحتفاظ بتلك التي يحتاج اليها فقط، فيمكن للمستخدم استخدام محرر البيانات للحذف بنفس طريقة العمل في اكسيل، أو الاستعانة بالاوامر التي يتوف عليها البرنامج Stata، وفيما يلي نعرض طريقة الحذف او الاحتفاظ بعينة من البيانات.

5-1. الأمر drop

يساعد الأمر drop على حذف البيانات الخاصة بالمتغيرات غير المرغوب فيها، فان تم تخزين الملف بعد تنفيذه سوف يتم حذفها نهائيا من ملف البيانات. ويتم الاستعانة بهذا الأمر في حالة احتواء الملف على عدد كبير من المتغيرات ويريد المستخدم حذف البعض منها، فيكتب الأمر: **drop var1 var2 var3** مثلا.

بالنسبة لحذف المتغيرات يتم كتابة الأمر drop متبوعا باسماء المتغيرات المراد حذفها،

اما بالنسبة للمشاهدات فيتم كتابة الأمر drop in متبوعا بالرمز n/m بحيث n هو رقم المشاهدة الأولى (نقطة البداية) و m هو رقم المشاهدة الاخيرة (نقطة النهاية) وبالتالي تكون العينة ضمن المجال [n-m]، مثلا حذف المشاهدات من 25 الى 80، نكتب: **drop in 25/80**

5-2. الأمر keep

يعمل الأمر keep على الاحتفاظ بعينة مختارة من البيانات وحذف تلك المذكورة بعد الأمر keep مثلا: keep var1 var2 var3 فهذا يعني حذف كل المتغيرات الموجودة في ملف البيانات والاحتفاظ بالمتغيرات var1، var2 و var3 فقط. وتخزين الملف يؤدي الى فقدان باقي المتغيرات التي كانت موجودة في الملف. ويتم التعامل مع هذا الأمر بنفس طريقة التعامل مع drop بالنسبة للمتغيرات والمشاهدات، فمثلا للاحتفاظ بعينة من المشاهدات يكفي كتابة الأمر: **keep in n/m**

6. تبويب البيانات:

في بعض الاحيان يحتاج الباحث الى تحويل بيانات متغير معين الى شكل فئات، بغرض تسهيل عملية التحليل الاحصائي أو استعمالها لتفسير متغيرات أخرى، مثل استعمال متغير السن، الأجر، الخبرة المهنية ...الخ، لتفسير مستوى الدخل أو الاجر.

ولاجراء هذا التحويل باستعمال برنامج Stata، يمكن كتابة الأمر recod متبوع باسم المتغير المراد تحويله ونضع بين قوسين الفئات المراد انشاؤها وفي الاخير اعطاء الأمر generat (المتغير الجديد).

مثلا: تبويب متغير السن في بيانات الأجور (paye)، بافتراض اننا نريد تقسيم متغير الخبرة الى فئات طول كل فئة يساوي 3 سنوات.

مثال:

لدينا البيانات الخاصة بعينة من الموظفين على مستوى المؤسسة (Matrex) المبينة في الجدول التالي:

الجدول رقم (2-10): بيانات خاصة بالموظفين في المؤسسة Matrex.

fiche	salaire	expérience	sexe	évaluation
1	38000	2	1	5
2	43000	6	1	7
3	41000	5	2	8
4	35000	1	1	5
5	37000	1	2	4
6	58000	12	1	8
7	47000	9	1	5
8	45000	10	2	3
9	39000	4	1	4
10	46000	8	1	6
11	43000	9	2	3
12	39000	2	2	4
13	36000	1	1	5
14	46000	5	2	5
15	49000	9	2	5
16	35000	1	1	5
17	42000	6	2	6
18	41000	7	2	6
19	45000	6	1	5
20	41000	8	2	3
21	39000	4	2	4
22	40000	5	2	5

المصدر: Gérald Baillargeon, Fernando ouellet : «Analyse des données avec SPSS pour Windows» ; Les Edition SMG, Canada, 2008, p 20.

يبين الجدول اعلاه المعلومات الخاصة بـ 22 موظف على مستوى مؤسسة **Matrex**، وهي تتمثل في الاجر ، عدد سنوات الخبرة، الجنس وتقييم داخلي خاص بمردودية كل موظف. والذي تم تعبئته الى البرنامج بكتابة الأمر التالي:

```
import excel "C:\Users\rabah\Desktop\stata 2017\Matrex.xlsx", sheet("Feuil3") firstrow clear
```

لاستعمال متغير عدد سنوات الخبرة (Experience) في مختلف العمليات الاحصائية غالبا ما يتم تبويبه الى شكل فئات طول كل فئة يحدده المستخدم حسب الحاجة. واذا افترضنا ان طول الفئة يساوي 3سنوات، يمكننا تبويب المتغير بكتابة الأمر التالي:

```
recode expérience (min/3=1) (3/6=2) (6/9=3) (9/max=4), generate(exp1)
```

بحيث ان الأمر مقسم الى شطرين، الشطر الاول يتعلق بطريقة تجزئة قيم المتغير expérience والشطر الثاني يتعلق بانشاء المتغير الجديد الذي يتضمن القيم الجديدة للمتغير expérience وليكن exp1، وتجدر الاشارة الى أن العبارة min تعبر عن 3 وأقل، وmax تعبر عن 9 فأكثر.

وبتطبيق هذا الأمر يقوم البرنامج بإنشاء متغير جديد باسم exp1 يشمل فئات سنوات الخبرة المهنية.

no	salaire	expérience	sexe	évaluation	exp1
1	38000	2	1	5	1
2	43000	6	1	7	2
3	41000	5	2	8	2
4	35000	1	1	5	1
5	37000	1	2	4	1
6	58000	12	1	8	4
7	47000	9	1	5	3
8	45000	10	2	3	4
9	39000	4	1	4	2
10	46000	8	1	6	3
11	43000	9	2	3	3
12	39000	2	2	4	1
13	36000	1	1	5	1
14	46000	5	2	5	2
15	49000	9	2	5	3
16	35000	1	1	5	1
17	42000	6	2	6	2
18	41000	7	2	6	3
19	45000	6	1	5	2
20	41000	8	2	3	3
21	39000	4	2	4	2
22	40000	5	2	5	2

ملاحظة:

ان عملية تبويب البيانات يستدعي نوع من الدقة، فمثلا في حالة البيانات المتقاربة في ما بينها كالأجور أو نقاط الطلبة مثلا يجب التدقيق في كتابة الأمر لأن الفئة ما بين 10 و 15 نقطة مثلا تأخذ بعين الاعتبار كل النقاط الأكبر تماما من القيمة 10 و الأقل أو تساوي 15، والفئة المحصورة بين 5 و10 تأخذ بعين الاعتبار القيمة 10، وعليه يجب تحديد الفئات بإضافة ارقام بعد الفاصلة مثلا الفئة بين 5.01 و 9.99 لتجنب احتساب القيمة 5 و 10 في الفئة المطلوبة. وفي مثالنا الفئة 4 تشمل الفئة التي تفوق 9 دون احتساب القيمة 9.

الفصل الثالث: التحليل الاحصائي للبيانات باستخدام برنامج Stata

يتناول هذا الفصل طريقة استخدام البرنامج Stata في التحليل الاحصائي للبيانات، ونخصصه على وجه الخصوص لأهم العمليات الاحصائية المعمول بها في ميدان العلوم الاقتصادية، وفق طبيعة البيانات.

1. الاحصائيات الوصفية

1-1. الأمر Summarize

يستخدم هذا الأمر على المتغيرات الكمية القابلة للقياس، ويتيح للمستخدم امكانية استخراج مختلف المؤشرات الاحصائية الوصفية (مقاييس النزعة المركزية، التشتت والتوزيع) الخاصة بمتغير أو عدة متغيرات بكتابة الأمر summarize متبوع باسم أو أسماء المتغيرات المعنية في لوحة الاوامر، كما يمكن اختصار الأمر في الحرفين الأولين su، . فمثلا لاستخرج المؤشرات الاحصائية الوصفية لمتغير الأجر Salaire نكتب su salaire. وبتطبيق الأمر يعرض البرنامج جدول على الشكل التالي:

```
. su salaire
```

Variable	Obs	Mean	Std. Dev.	Min	Max
salaire	22	42045.45	5304.78	35000	58000

اسم المتغير | عدد المشاهدات | المتوسط الحسابي | الانحراف المعياري | أدنى قيمة | أعظم قيمة

ملاحظة:

اذا قام المستخدم بكتابة الأمر summarize دون اضافة اي متغير أو متغيرات، يقوم البرنامج تلقائيا بعرض المؤشرات الاحصائية لكل المتغيرات الكمية في ملف البيانات.

كما يمكن استخراج باقي المؤشرات باضافة العبارة detail بكتابة الأمر su (name), detail، فيظهر البرنامج النتائج كما يلي:

```
                    salaire
-------------------------------------------------------------
      Percentiles      Smallest
 1%         35000         35000
 5%         35000         35000
10%         36000         36000       Obs                  22
25%         39000         37000       Sum of Wgt.          22

50%         41000                     Mean           42045.45
                        Largest       Std. Dev.      5304.78
75%         45000         46000
90%         47000         47000       Variance       2.81e+07
95%         49000         49000       Skewness       1.137653
99%         58000         58000       Kurtosis       4.777517
```

أدنى خمسة قيم | قيم المئينات | المؤشرات الاحصائية الوصفية | أعلى خمسة قيم

بحيث يعرض البرنامج الى جانب المقاييس المعروضة سابقا، مقاييس التشتت ومقاييس التوزيع التي تتمثل في قيم المئينات percentiles، وأدنى القيم (ادنى اربعة قيم) Smallest وأعظم أربعة قيم Largest موجودة في ملف البيانات، بالإضافة الى قيمة التباين ومعامل الالتواء ومعامل التفرطح.

يمكن للمستخدم من خلال الملخص الاحصائي الناتج باضافة detail الى الأمر summarize معرفة مدى اعتدالية التوزيع من خلال معامل الالتواء ومعامل التفلطح، بالاضافة الى امكانية استنتاج قيم الربيعيات من خلال المئينات بحيث الربيع الأول يقابله المئيني 25% والربيع الثاني يساوي قيمة الوسيط وهو المئيني 50% أما الربيع الثالث فيقابله المئيني 75%.

2-1. الأمر tabstat

يعتبر الأمر tabstat بديل للامر summarize، بحيث يتيح للمستخدم اختيار المؤشرات الاحصائية المراد استخراجها دون الاعتماد على الجداول التلقائية التي يعرضها البرنامج باستخدام الأمر summarize.

فمثلا لاستخراج قيمة المتوسط الحسابي واعلى وادنى قيمة لمتغير salaire، يكفي ان نكتب:

tabstat salaire, stats (mean min max)

```
. tabstat salaire, stats (mean min max)
```

variable	mean	min	max
salaire	42045.45	35000	58000

بحيث يعرض البرنامج المؤشرات الاحصائية المحدد دون غيرها.

كما يمكن الاستعانة بالخيار by لعرض المؤشرات الوصفية بالتجزئة حسب متغير نوعي او اسمي، مثلا عرض المتوسط والقيمة الدنيا والعظمى للاجر حسب الجنس الى للذكور والاناث، بحيث يكفي ان نكتب الأمر: tabstat salaire, stats (mean min max) by(sexe)

فيظهر البرنامج الجدول التالي في المخرجات:

sexe	mean	min	max
homme	42200	35000	58000
femme	41916.67	37000	49000
Total	42045.45	35000	58000

من خلال الجدول يمكن ملاحظة ان متوسط الاجر عند الذكور اكبر من متوسط الاجر لدى الاناث، وهذا النوع من العرض يسهل من المقارنة بين فئات المتغير الاسمي أو النوعي. ويمكن اضافة مختلف المؤشرات الاحصائية الوصفية الي جانب المعروضة في الجدول حسب رغبة المستخدم.

وتجدر الاشارة الى انه يمكن استخراج مختلف المؤشرات الاحصائية الوصفية حسب فئات متغير اسمي باستخدام الأمر summarize لكن بكتابة مختلفة عن حالة الأمر tabstat، بحيث يتم تحديد المتغير الفئوي مسبقا بكتابة الأمر by sexe, sort : summarize salaire، فيكون عرض البرنامج كما يلي:

```
. by sexe, sort : summarize salaire

-> sexe = homme

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     salaire |         10       42200    7192.589     35000      58000

-> sexe = femme

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     salaire |         12    41916.67    3369.875     37000      49000
```

2-1. الأمر tablulate

يقوم هذا الأمر باعداد جداول للمتوسطات والانحراف المعياري على شكل فئات المتغيرات الاسمية أو النوعية، بحيث يقوم البرنامج بحساب المتوسط لكل فئة من فئات هذه الاخيرة دون الحاجة الى تجزئة جدول البيانات. ولتوضيح ذلك نستعين بالمثال التالي:

لدينا نقاط لعينة من طلبة السنة الاولى جذع مشترك في مقياس الاحصاء حسب الجنس والشعبة:

الجدول رقم (3-1): العلامات المحصلة لعينة من الطلبة الجامعيين في مادة الاحصاء

no	sexe	Filiere	note
1	feminin	medcine	13.875
2	masculin	science economique	11.125
3	feminin	technologie	12.375
4	feminin	science economique	12.25
5	feminin	science economique	11
6	feminin	science economique	10
7	feminin	medcine	9.875
8	feminin	technologie	12
9	feminin	medcine	9.75
10	masculin	science economique	9.75
11	feminin	medcine	10
12	masculin	medcine	13.625
13	masculin	science economique	10.125
14	masculin	technologie	11.25
15	feminin	science economique	8.75
16	masculin	medcine	13.75
17	feminin	technologie	13
18	feminin	technologie	10.75
19	feminin	science economique	9.875
20	masculin	medcine	13.875
21	feminin	technologie	9.75
22	feminin	technologie	10
23	feminin	science economique	11.75
24	feminin	medcine	11.125
25	feminin	science economique	10
26	masculin	technologie	9.875
27	feminin	technologie	14.375
28	feminin	technologie	11.25
29	masculin	medcine	12.5
30	feminin	medcine	10.25

المطلوب:

1. حساب متوسط علامة مقياس الاحصاء حسب الجنس،
2. حساب متوسط علامة مقياس الاحصاء حسب الشعبة؛
3. حساب متوسط علامة مقياس الاحصاء حسب الشعبة ضمن فئات الجنس.

الحل:

1. نكتب الأمر التالي: **tabulate sexe , summ(note)** ثم entrer يعرض البرنامج الجدول التالي:

sexe	Summary of note Mean	Std. Dev.	Freq.
masculin	11.763889	1.709857	9
feminin	11.047619	1.484017	21
Total	11.2625	1.5610252	30

2. نكتب الأمر التالي: **tabulate Feliere , summ(note)** ثم entrer يعرض البرنامج الجدول التالي:

Filiere	Summary of note Mean	Std. Dev.	Freq.
science e	10.4625	1.0492226	10
technolog	11.4625	1.4965818	10
medcine	11.8625	1.8318649	10
Total	11.2625	1.5610252	30

3. نكتب الأمر التالي:

tabulate sexe Feliere , summ(note)

ثم entrer يعرض البرنامج الجدول التالي:

sexe	Filiere science e	technolog	medcine	Total
masculin	10.333333	10.5625	13.4375	11.763889
	.71078009	.97227182	.63327851	1.709857
	3	2	4	9
feminin	10.517857	11.6875	10.8125	11.047619
	1.2128405	1.5669579	1.5786664	1.484017
	7	8	6	21
Total	10.4625	11.4625	11.8625	11.2625
	1.0492226	1.4965818	1.8318649	1.5610252
	10	10	10	30

ملاحظة:

يمكن الاكتفاء بحساب المتوسط فقط دون حساب الانحراف المعياري باضافة العبارة mean الى الأمر tabulate var(n),summ(var(q)) مثلا في الحالة 3 بدلا من الحصول على جدول كبير يمكن كتابة الأمر على الشكل : **tabulate sexe Feliere , summ(note)mean**.

sexe	Filiere science e	technolog	medcine	Total
masculin	10.333333	10.5625	13.4375	11.763889
feminin	10.517857	11.6875	10.8125	11.047619
Total	10.4625	11.4625	11.8625	11.2625

3-1. الأمر table

يسمح هذا الأمر بحساب مختلف المؤشرات الاحصائية لمتغير معين أو عدة متغيرات حسب فئات متغير أو عدة متغيرات وذلك بتوفير امكانية تحديد المؤشرات المراد حسابها حسب الحاجة.

- بالنسبة لمتغير واحد يمكن استخدام الأمر التالي:

table Filiere , contents(mean note sd note median note)

بحيث يطلب من البرنامج حساب كل من المتوسط الحسابي والانحراف المعياري والوسيط للمتغير note حسب فئات المتغير Feliere. وبتطبيق الأمر نتحصل على النتائج التالية:

Filiere	mean(note)	sd(note)	med(note)
science economique	10.4625	1.049223	10.0625
technologie	11.4625	1.496582	11.25
medcine	11.8625	1.831865	11.8125

- بالنسبة لمتغيرين أو أكثر: في هذه الحالة يمكن استخراج المؤشرات الاحصائية لأكثر من متغير حسب فئات متغير أو أكثر وذلك بتعديل صيغة الأمر السابق باضافة المتغيرات الاسمية في الجانب الايسر اي الى جانب المتغير Filiere والمؤشرات الاحصائية للمتغيرات داخل قوسين بعد العبارة contents، مثلا حساب المتوسط والانحراف المعياري للمتغيرين note و moyenne، اي نكتب:

table sexe Filiere , contents(mean note sd note mean moy sd moy)

بتطبيق الأمر نتحصل على النتائج التالية:

sexe	Filiere science economique	technologie	medcine
masculin	10.333333	10.5625	13.4375
	.7107801	.9722718	.6332785
	9.466666	10	13.8
	.5033223	.8485287	.9380832
feminin	10.517857	11.6875	10.8125
	1.212841	1.566958	1.578666
	10.02857	12.075	10.53333
	1.397276	2.036629	2.111555

ان هذا النوع من الجداول يسهل عمليات المقارنة حسب فئات المتغيرات النوعية لمتوسطات وانحرافات المتغيرات الكمية دون الحاجة الى عملية التجزئة واستخراج جداول مستقلة للمقارنة.

2. الجداول التكرارية والجداول التقاطعية

2-1. الجداول التكرارية

يعتبر الجدول التكراري من ابسط الملخصات الاحصائية للمتغيرات، ويتعلق الأمر بحالة البيانات ذات المتغيرات النوعية (الاسمية)، كما يمكن استخدامها بالنسبة للمتغيرات الكمية المبوبة (في شكل فئات)، والجداول التكرارية تعد اكثر استخداما في تحليل نتائج الدراسات الاستقصائية بحيث يلائمها هذا النوع من الجداول. و يتم استخراج الجدول التكراري باستخدام البرنامج stata عن طريق كتابة الأمر **tabulate** أو اختصارا **ta** متبوع باسم المتغير في لوحة الأوامر.

وبالعودة الى البيانات الخاصة بالاجور في المؤسسة Matrex (الجدول رقم (2-10))، يمكن استخراج الجدول التكراري للمتغير الاسمي الوحيد وهو الجنس بحيث يمكن معرفة عدد الذكور والاناث من بين الموظفين في المؤسسة، فبكتابة الأمر ta sexe فيعرض البرنامج الجدول التكراري على النحو التالي:

```
. ta sexe
```

sexe	Freq.	Percent	Cum.
homme	10	45.45	45.45
femme	12	54.55	100.00
Total	22	100.00	

يظهر من خلال الجدول عدد مرارت تكرار الجنس ذكر وانثى .Freq، والتكرار النسبي Percent اي نسبة الذكور والاناث بالنسبة الى عدد الموظفين في المؤسسة، بالاضافة الى التكرار المتجمع الصاعد Cum.

➢ **ملاحظة:**

لا يمكن استخراج الجداول التكرارية لعدة متغيرات بكتابة اسماء المتغيرات بعد الأمر tabulate بشكل متتابع، بحيث ان ذلك يؤدى الى استخراج الجدوال التقاطعي لهذه المتغيرات.

2-2. الجداول التقاطعية:

أولا. الجداول التقاطعي الثنائي:

يستخدم هذا النوع من الجداول لتحليل العلاقة بين فئات متغيرين من النوع الاسمي، ويستخدم لتحديد وجود علاقة بين المتغيرين من الناحية الاحصائية، بحيث يعرض تكرارات كل فئة من متغير داخل فئات متغير آخر لاختبار العلاقة بين المتغيرين. ولاستخراج الجدول التقاطعي نكتب الأمر tabulate متبوع باسم المتغيرين المعنيين، ولتوضيح ذلك نستعين بالمثال التلي:

مثال:

لدينا البيانات الخاصة بنتائج استتبيان حول استخدام وسائل النقل في احدى المدن المبينة في الجدول التالي:

الجدول رقم (3-2): نتائج اسبيان لعينة من السكان حول استخدام وسائل النقل

Rép	Q1	Q2	Q3	Rép	Q1	Q2	Q3	Rép	Q1	Q2	Q3
1	M	18-34 ANS	Régulière	27	M	55 ANS ET PLUS	Ne prend pas l'autobus	53	M	35-54 ANS	Occasionnelle
2	F	55 ANS ET PLUS	Régulière	28	M	55 ANS ET PLUS	Occasionnelle	54	F	35-54 ANS	Ne prend pas l'autobus
3	F	18-34 ANS	Régulière	29	M	55 ANS ET PLUS	Ne prend pas l'autobus	55	M	18-34 ANS	Ne prend pas l'autobus
4	F	18-34 ANS	Ne prend pas l'autobus	30	M	18-34 ANS	Ne prend pas l'autobus	56	F	55 ANS ET PLUS	Ne prend pas l'autobus
5	M	35-54 ANS	Occasionnelle	31	F	18-34 ANS	Occasionnelle	57	F	35-54 ANS	Ne prend pas l'autobus
6	F	18-34 ANS	Ne prend pas l'autobus	32	M	35-54 ANS	Régulière	58	M	18-34 ANS	Occasionnelle
7	F	18-34 ANS	Ne prend pas l'autobus	33	F	18-34 ANS	Ne prend pas l'autobus	59	F	18-34 ANS	Ne prend pas l'autobus
8	F	55 ANS ET PLUS	Occasionnelle	34	F	18-34 ANS	Occasionnelle	60	F	35-54 ANS	Ne prend pas l'autobus
9	M	18-34 ANS	Régulière	35	F	18-34 ANS	Occasionnelle	61	M	35-54 ANS	Ne prend pas l'autobus
10	M	35-54 ANS	Ne prend pas l'autobus	36	F	35-54 ANS	Ne prend pas l'autobus	62	M	35-54 ANS	Occasionnelle
11	M	35-54 ANS	Occasionnelle	37	M	55 ANS ET PLUS	Ne prend pas l'autobus	63	M	18-34 ANS	Ne prend pas l'autobus
12	F	55 ANS ET PLUS	Régulière	38	F	18-34 ANS	Ne prend pas l'autobus	64	F	35-54 ANS	Ne prend pas l'autobus
13	F	35-54 ANS	Ne prend pas l'autobus	39	F	35-54 ANS	Occasionnelle	65	F	35-54 ANS	Ne prend pas l'autobus
14	M	35-54 ANS	Ne prend pas l'autobus	40	F	35-54 ANS	Ne prend pas l'autobus	66	F	18-34 ANS	Ne prend pas l'autobus
15	M	18-34 ANS	Occasionnelle	41	M	18-34 ANS	Ne prend pas l'autobus	67	M	55 ANS ET PLUS	Occasionnelle
16	M	18-34 ANS	Ne prend pas l'autobus	42	F	18-34 ANS	Ne prend pas l'autobus	68	F	35-54 ANS	Ne prend pas l'autobus
17	F	35-54 ANS	Ne prend pas l'autobus	43	F	18-34 ANS	Ne prend pas l'autobus	69	M	18-34 ANS	Ne prend pas l'autobus
18	F	18-34 ANS	Régulière	44	F	35-54 ANS	Ne prend pas l'autobus	70	M	18-34 ANS	Ne prend pas l'autobus
19	M	35-54 ANS	Ne prend pas l'autobus	45	M	55 ANS ET PLUS	Occasionnelle	71	F	35-54 ANS	Occasionnelle
20	M	35-54 ANS	Ne prend pas l'autobus	46	M	55 ANS ET PLUS	Ne prend pas l'autobus	72	F	55 ANS ET PLUS	Ne prend pas l'autobus
21	M	35-54 ANS	Occasionnelle	47	F	18-34 ANS	Occasionnelle	73	F	35-54 ANS	Ne prend pas l'autobus
22	F	18-34 ANS	Ne prend pas l'autobus	48	F	55 ANS ET PLUS	Ne prend pas l'autobus	74	M	18-34 ANS	Ne prend pas l'autobus
23	F	55 ANS ET PLUS	Ne prend pas l'autobus	49	F	35-54 ANS	Ne prend pas l'autobus	75	M	35-54 ANS	Ne prend pas l'autobus
24	M	18-34 ANS	Occasionnelle	50	M	35-54 ANS	Ne prend pas l'autobus				
25	F	18-34 ANS	Ne prend pas l'autobus	51	F	35-54 ANS	Ne prend pas l'autobus				
26	F	55 ANS ET PLUS	Ne prend pas l'autobus	52	M	35-54 ANS	Ne prend pas l'autobus				

المصدر: Gérald Baillargeon, Fernando ouellet : «Analyse des données avec SPSS pour Windows» ; Les Edition SMG, Canada, 2008, p 36-38.

بحيث تم استجواب عينة متكونة من 75 شخص حول طريقة استخدامهم لوسائل النقل بالاجابة على الاسئلة التالية:

1. الجنس: ذكر ☐ أنثى ☐

2. العمر أو السن: 1. 18 – 34 سنة ☐ 2. 35 – 54 سنة ☐ 3. 55 سنة فأكثر ☐

3. خلال الـ 12 شهرا الماضية كيف كان استعمالك للحافلة؟

1. دائما ☐ 2. أحيانا ☐ 3. لم استعمل الحافلة ☐

لاحظ أن كل المتغيرات الموجودة في الجدول عبارة عن متغيرات اسمية، مما يسمح لنا باستخراج جداول تقاطعية ثنائية واكثر، فمثلا لاستخراج جدول تقاطع بين متغيري الجنس واستخدام الحافلة، نكتب: tabulat Q1 Q3

فيظهر الجدول على الشكل التالي:

sexe	utilisation de l'autobus			
	Ne pren..	Occasio..	Régulière	Total
F	30	7	4	41
M	20	11	3	34
Total	50	18	7	75

كما يمكن استخراج قيمة اختبار العلاقة بين المتغيرتين بإضافة العبارة chi2 الى العبارة السابقة بكتابة: tabulate Q1 Q3, chi2، فيقوم البرنامج بحساب القيمة الاحصائية والاحتمالية لبيرسون كاي تربيع والتي يتم من خلالها اختبار الاستقلالية بين المتغيرتين. وبتطبيق الأمر يظهر البرنامج النتائج في الجدول التالي:

```
. tabulate Q1 Q3, chi2
```

sexe	utilisation de l'autobus			
	Ne pren..	Occasio..	Régulière	Total
F	30	7	4	41
M	20	11	3	34
Total	50	18	7	75

Pearson chi2(2) = 2.3993 Pr = 0.301

لاحظ أن البرنامج قام بحساب قيمة اختبار كاي تربيع للاستقلالية، وبما ان القيمة الاحتمالية أكبر من 0.05 فإنه يمكن الاقرار بعدم وجود علاقة بين المتغيرين (مستقلين)، أي ان السن ليس له علاقة بطريقة استخدام الحافلة.

كما يتيح البرنامج امكانية حساب قيم مختلف الاختبارات الخاصة بالارتباط بالاضافة الى اختبار كاي تربيع بكتابة الأمر التالي: tabulate var1 var2, missing row all، وتطبيق هذا الأمر على مثالنا السابق يعرض البرنامج النتائج على النحو التالي:

sexe	utilisation de l'autobus Ne pren..	Occasio..	Régulière	Total
F	30	7	4	41
	73.17	17.07	9.76	100.00
M	20	11	3	34
	58.82	32.35	8.82	100.00
Total	50	18	7	75
	66.67	24.00	9.33	100.00

```
          Pearson chi2(2) =   2.3993   Pr = 0.301
likelihood-ratio chi2(2) =   2.3990   Pr = 0.301
                Cramér's V =   0.1789
                     gamma =   0.2511   ASE = 0.212
           Kendall's tau-b =   0.1278   ASE = 0.112
```

بحيث يتم عرض قيم V لكرامر Cramer's V، وقيمة غاما لغودمان وكروسكال (Goodman and Kruskal's gamma) وقيمة تو (τ_b) لكندال Kendall، ويبقى على المستخدم استخدامها حسب الحاجة وطبيعة المتغيرات.

ثانيا. الجدول التاقطعي لأكثر من متغيرين

في هذه الحالة يتم استخراج التكرارات الخاصة بالتقاطع بين متغيرين داخل فئات متغير ثالث، فمثلا تحليل العلاقة بين استعمال الحافلة والسن ضمن فئات الجنس باستعمال معطيات المثال السابق. ولاستخراج هذه العلاقة باستخدام برنامج Stata يكفي كتابة الأمر التالي:

by Q1, sort : tabulate Q2 Q3

بحيث يقوم البرنامج بعرض جدول تقاطعي مقسم حسب فئات متغير التجزئة (الجنس)، ولاحظ انه تم استخدام الأمر by في المقدمة لاعلام البرنامج باعتبار متغير الجنس متغير تجزئة، والأمر sort لترتيب البيانات حسب فئات الجنس، فاذا كان الملف مرتبا يمكن الاستغناء عن sort. وبتطبيق الأمر يعر البرنامج الجدول التالي:

```
-> Q1 = F
```

age	utilisation de l'autobus Ne pren..	Occasio..	Régulière	Total
18-34 ANS	11	4	2	17
35-54 ANS	14	2	0	16
55 ANS ET PLUS	5	1	2	8
Total	30	7	4	41

جدول تقاطعي خاص بفئة الاناث

```
-> Q1 = M
```

age	utilisation de l'autobus Ne pren..	Occasio..	Régulière	Total
18-34 ANS	8	3	2	13
35-54 ANS	8	5	1	14
55 ANS ET PLUS	4	3	0	7
Total	20	11	3	34

جدول تقاطعي خاص بفئة الاناث

وتجدر الاشارة الى ان البرنامج يعرض النتائج بترتيب فئات متغير التجزئة تصاعديا حسب الحروف الابجدية اذا كان المتغير اسمي (F يسبق M)، وتصاعديا اذا كان متغير رقمي (الفئات عبارة عن ارقام).

كما يمكن حساب القيمة الاحصائية لاختبار الدقة لفيشر باضافة كلمة exact الى الأمر السابق كما يلي:

by Q1, sort : tabulate Q2 Q3, exact

بحيث يعرض البرنامج قيمة اختبار الدقة لفيشر لكل جدول فرعي، كما هو موضح في الجدول التالي:

```
-> Q1 = F

Enumerating sample-space combinations:
stage 3:  enumerations = 1
stage 2:  enumerations = 10
stage 1:  enumerations = 0

                |      utilisation de l'autobus
            age | Ne pren..  Occasio..  Régulière |     Total
----------------+---------------------------------+----------
     18-34 ANS |        11          4          2 |        17
     35-54 ANS |        14          2          0 |        16
55 ANS ET PLUS |         5          1          2 |         8
----------------+---------------------------------+----------
          Total |        30          7          4 |        41

           Fisher's exact =                 0.273

-> Q1 = M

Enumerating sample-space combinations:
stage 3:  enumerations = 1
stage 2:  enumerations = 3
stage 1:  enumerations = 0

                |      utilisation de l'autobus
            age | Ne pren..  Occasio..  Régulière |     Total
----------------+---------------------------------+----------
     18-34 ANS |         8          3          2 |        13
     35-54 ANS |         8          5          1 |        14
55 ANS ET PLUS |         4          3          0 |         7
----------------+---------------------------------+----------
          Total |        20         11          3 |        34

           Fisher's exact =                 0.841
```

3. اختبار اعتدالية التوزيع وتجانس التباين

3-1. اختبار اعتدالية التوزيع

إن عملية التحليل الاحصائي للبيانات واختبار الفروض يشترط توفر بعض الشروط في عينات الدراسة، خاصة في حالة العينات الصغيرة، ومن أهم هذه الشروط: ان تكون عينة الدراسة مسحوبة من مجتمع يتبع التوزيع الطبيعي، بينما يمكن أن يُستغنى عن هذا الشرط في حالة تجاوز حجم العينة 30 مشاهدة ويفترض انها تتبع التوزيع الطبيعي وفق نظريات النزعة المركزية. وشرط ثاني يتمثل في شرط تجانس التباين في حالة تعدد العينات، وكذا استقلاليتها.

فمثلا: اجراء تحليل التباين بمعيار واحد يشترط تجانس تباين العينات وان تكون مسحوبة من مجتمع يتبع التوزيع الطبيعي.

وسنتناول من خلال هذا المثال طريقة اختبار التوزيع الطبيعي لعينة تتكون من أقل من 30 مشاهدة.

مثال:

نفرض انه لدينا البيانات الخاصة بالمعدل المحصل في مقياس السلاسل الزمنية لعينة متكونة من 15 طالب تخصص: اقتصاد كمي في كلية العلوم الاقتصادية المبينة في الجدول التالي:

الجدول رقم (3-3): المعدلات المحصلة لعينة من الطلبة في مقياس السلاسل الزمنية

étud	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
moyenne	12	10	8	11	15	15	11	3	14	16	15	10	7	16	10

بحيث يتم اختبار الفرضية:

H0: ان العينة مسحوبة من مجتمع تتبع بياناته التوزيع الطبيعي؛

H1: ان العينة مسحوبة من مجتمع لا تتبع بياناته التوزيع الطبيعي.

- يتوفر برنامج Stata على مجموعة من الاوامر لاختبار التوزيع الطبيعي للعينات على خلاف برنامج Spss الذي يتوفر على اختبارين فقط. فيمكن اختبار التوزيع الطبيعي لعينة معينة باستعمال Stata كما يلي:

اختبار التوزيع من خلال اختبار نسبة الالتواء skewness والتفلطح kurtosis، بحيث نقوم بكتابة الأمر sktest متبوع باسم المتغيرة، لاختبار الفرضية الصفرية التي تقر بأن العينة مسحوبة من مجتمع يتبع التوزيع الطبيعي.

- اختبار شابيرو-ويلك (Shapiro-Wilk) بكتابة الأمر swilk متبوع باسم المتغير.

- اختبار شابيرو-فرانسيا (Shapiro-Francia) بكتابة الأمر sfrancia متبوع باسم المتغير.

ملاحظة: يشترط اجراء اختبار الالتواء 8 مشاهدات على الاقل، بينما يشترط اختبار شابيرو-ويلك ما بين 4 و200 مشاهدة، في حين يشترط اختبار شابيرو فرانسيا ما بين 5 و5000 مشاهدة.

كما يسمح البرنامج Stata باختبار التوزيع بيانيا من خلال التمثيل البياني Q-plot أو P-plot من أجل مقارنة توزيع العينة بالتوزيع الطبيعي، وذلك باستعمال الأوامر التالية:

- qnorm متبوعا باسم المتغير، للتمثيل البياني لربيعيات التوزيع الطبيعي وربيعيات توزيع العينة المراد اختبارها. بحيث كلما اقتربت نقاط الانتشار الى الخط المستقيم فان توزيع العينة يقترب الى التوزيع الطبيعي.

- pnorm متبوعا باسم المتغير، للتمثيل البياني للمقارنة بالقيم المركزة للتوزيع N(0,1).

ولاختبار ما اذا كانت عينة المعدلات تتبع التوزيع الطبيعي أم لا نتبع الخطوات التالية:

أولا: اختبار نسبة الالتواء

بعد ادخال المعطيات الخاصة بالمتغير moyenne الى البرنامج، نقوم بكتابة الأمر:

Sktest moyenne

فيظهر البرنامج نتائج الاختبار كما يلي:

```
. sktest moyenne

                    Skewness/Kurtosis tests for Normality
                                                      ------ joint ------
    Variable |   Obs  Pr(Skewness)  Pr(Kurtosis)  adj chi2(2)  Prob>chi2
-------------+-------------------------------------------------------------
    moyenne |    15        0.2074        0.6438         2.07        0.3560
```

بحيث يقوم البرنامج بعرض نتائج اختبار تساوي قيمة الالتواء والتفلطح بقيمهما في التوزيع الطبيعي (0 و 3 على التوالي)، بالاضافة الى القيمة الاحتمالية لكلاهما معا من خلال اختبار كاي تربيع.

من خلال هذه النتائج نقبل فرضية العدم القائلة ان العينة مسحوبة من مجتمع يتبع التوزيع الطبيعي، لان كل القيم الاحتمالية المعروضة في الجدول أكبر من 0.05 دليل على ان قيمة الالتواء والتفلطح لا تختلف معنويا عن 0 و 3 على التوالي. وكذلك بالنسبة لاختبار كاي تربيع بحيث يتم الاقرار بقبول فرضية العدم أي إتباع العينة للتوزيع الطبيعي لأن القيمة الاحتمالية للإحصائية المحسوبة أكبر من 0.05.

ويمكن تأكيد ذلك من خلال استخراج المؤشرات الاحصائية الوصفية كما يوضحها الجدول التالي:

```
. su moyenne, detail

                            moyenne
-------------------------------------------------------------
      Percentiles      Smallest
 1%            3              3
 5%            3              7
10%            7              8       Obs                  15
25%           10             10       Sum of Wgt.          15

50%           11                      Mean           11.53333
                      Largest         Std. Dev.      3.739111
75%           15             15
90%           16             15       Variance       13.98095
95%           16             16       Skewness      -.6384943
99%           16             16       Kurtosis       2.804831
```

قيمة الالتواء قريبة من 0

قيمة الالتواء قريبة من 3

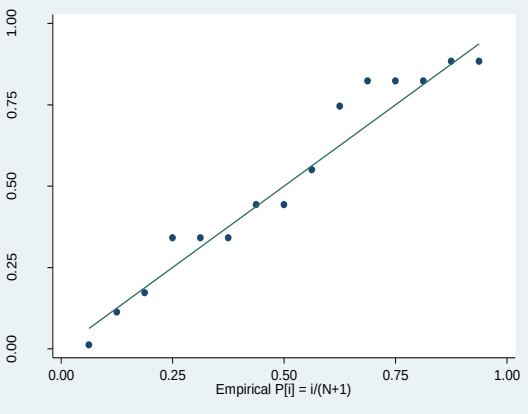
ثانيا: اختبار شابيرو-ويلك (Shapiro-Wilk) بكتابة الأمر swilk moyenne، فيعرض البرنامج نتائج الاختبار في الجدول التالي:

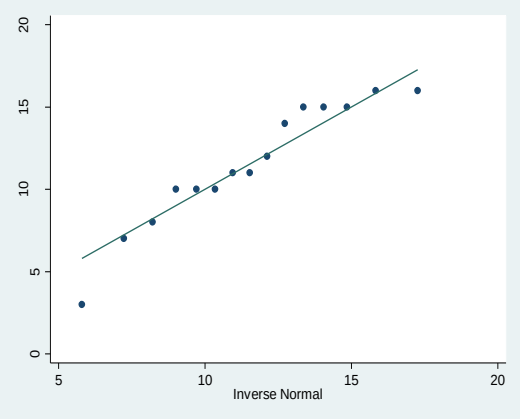
```
                Shapiro-Wilk W test for normal data

    Variable |   Obs        W          V          z       Prob>z
-------------+--------------------------------------------------
    moyenne |    15      0.92727    1.410      0.680    0.24827
```

وفي هذه الحالة كذلك يتم مقارنة القيمة الاحتمالية للاختبار بمستوى المعنوية المفترض (5%)، بحيث أن القيمة المحسوبة تساوي 0.24 وهي أكبر من 0.05 وبالتالي يمكن الاقرار بان العينة مسحوبة من مجتمع يتبع التوزيع الطبيعي.

ولتحليل النتائج بطريقة افضل يمكن الاستعانة ببعض المخططات البيانية الخاصة بتوزيع بيانات العينة كمخطط P (p-plot) ومخطط Q (q-plot). بكتابة الأوامر التالية:

pnorm moyenne و **qnorm moyenne**





بحيث يتم الاقرار باعتدالية التوزيع كلما كانت معظم نقاط الانتشار تقترب من الخط المستقيم الموضح في البيان.

ملاحظة: يشترط اجراء اختبار الالتواء Sktest 8 مشاهدات على الاقل، بينما يشترط اختبار شابيرو-ويلك ما بين 4 و200 مشاهدة، في حين يشترط اختبار شابيرو فرانسويا ما بين 5 و5000 مشاهدة.

الفصل الرابع: المخططات والرسومات البيانية باستخدام برنامج Stata

يتبح البرنامج Stata كغيره من البرامج الاحصائية المتعددة، امكانية اعداد مختلف المخططات البيانية حسب حاجة المستخدم، وذلك من خلال الأيقونة Gtaphics من شريط الأدوات الخاصة بالبرنامج، أو بطريقة كتابة الأوامر في لوحة الأوامر وهي افضل طريقة كون ان المستخدم بتعوده عليها في القيام بمختلف العمليات الاحصائية تصبح العملية سهلة وبسيطة، ويخرج البرنامج Stata على خلاف باقي البرامج الأشكال البيانية في شاشة خاصة (شاشة Graph) تفتح تلقائيا عندما يطلب انجاز التمثيل البياني المعني، وسنتاول من خلال هذا الفصل عرض طريقة اعداد بعض الاشكال البيانية التي يوفرها البرنامج.

1. شكل الانتشار

شكل الانتشار من بين أهم المخططات البيانية لتحليل العلاقة بين المتغيرات، ويتوفر البرنامج على عدة خيارات لإعداد هذا النوع من المخططات. والأمر الاساسي يكون بكتابة العبارة التالية في لوحة الأوامر:

graph twoway scatter y x أو twoway (scatter y x)

وهو الأمر الذي من خلال يتم رسم نقاط الانتشار للعلاقة بين المتغيرين y وx.

مثال:

لدينا المعطيات الخاصة بمتوسط الدخل الشهري 15 عائلة والنفقات الخاصة بالأطفال المتمدرسين:

الجدول رقم (4-1): متوسط الدخل الشهري والانفاق على الأطفال المتمدرسين لعينة من العائلات

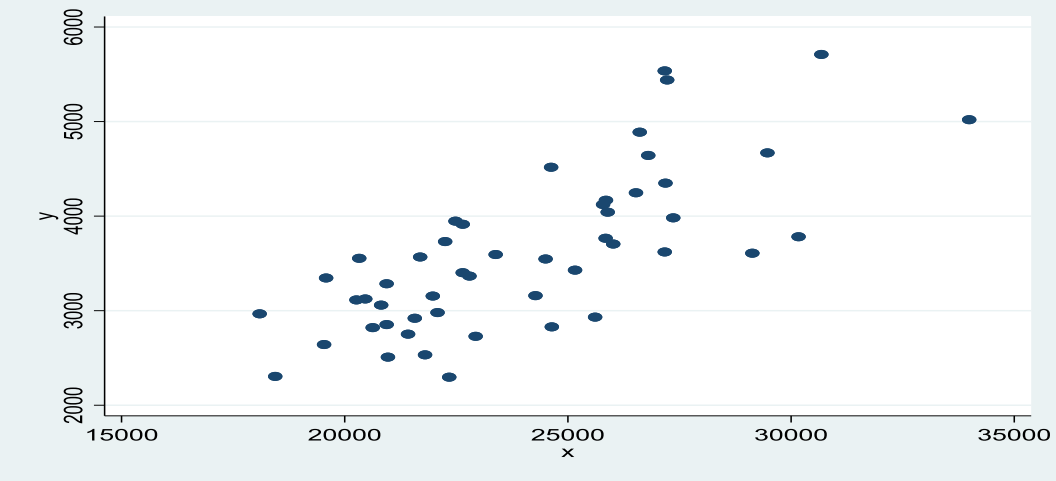
ind	y	x	ind	y	x	ind	y	x
1	3346	19583	18	3059	20816	35	2642	19538
2	3114	20263	19	2967	18095	36	3124	20460
3	3554	20325	20	3285	20939	37	2752	21419
4	4642	26800	21	3914	22644	38	3429	25160
5	4669	29470	22	4517	24624	39	3947	22482
6	4888	26610	23	4349	27186	40	2509	20969
7	5710	30678	24	5020	33990	41	5440	27224
8	5536	27170	25	3594	23382	42	4042	25892
9	4168	25853	26	2821	20627	43	3402	22644
10	3547	24500	27	3366	22795	44	2829	24640
11	3159	24274	28	2920	21570	45	2297	22341
12	3621	27170	29	2980	22080	46	2932	25610
13	3782	30168	30	3731	22250	47	3705	26015
14	4247	26525	31	2853	20940	48	4123	25788
15	3982	27360	32	2533	21800	49	3608	29132
16	3568	21690	33	2729	22934	51	3766	25845
17	3155	21974	34	2305	18443			

المصدر: Damodar N. Gujarati: « Basic Econometrics », Fourt Edition, The McGraw Hill Companies, 2004, p 154

من أجل تحليل العلاقة بين المتغيرين نستعين بالتمثيل البياني في شكل انتشار بكتابة الأمر:

graph twoway scatter y x

وبتنفيذه بالنقر على الزر Entre يظهر لنا الشكل التالي:

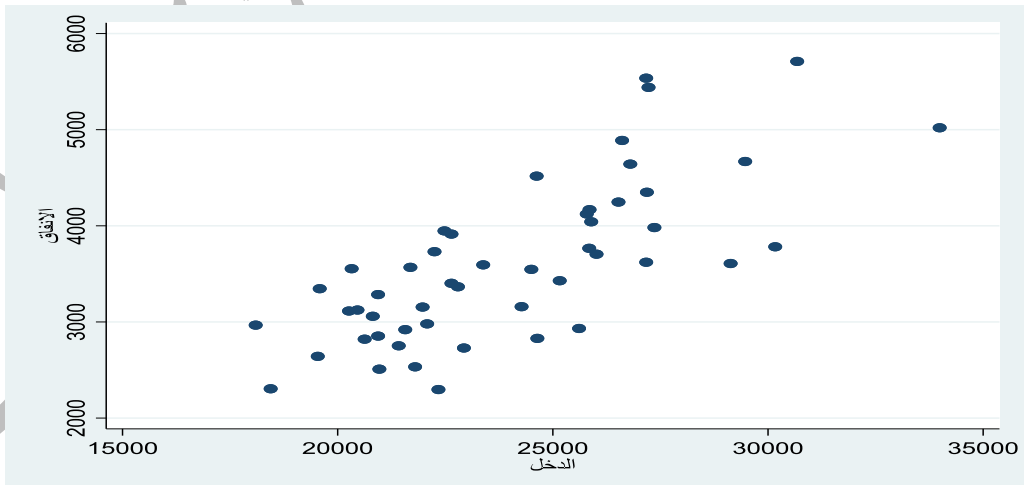


وهو الشكل التلقائي الذي يعرضه البرنامج، بحيث ليس للمستخدم اي دخل في القيم المعروضة في محاور المعلم، بل البرنامج يقوم باظهار قيم معينة دون غيرها كما هو موضح في الشكل، مثلا محور الفواص يحتوي على 16 قيمة لكن يظهر محور الفواصل في شكل فئات، وكذلك الأمر بالنسبة لمحور التراييب. وعليه يتيح البرنامج امكانية اجراء تعديلات في طريقة اعداد الرسم بالاستعانة بالإضافات التالية[(*)]:

- **xtitle(titre)** ، و **ytitle(titre)** لإعطاء عنوان لمحور الفواصل ومحور التراتيب على التوالي. وذلك باضافة العبارتين الى الأمر السابق، مثلا:

graph twoway scatter y x, ytitle(الانفاق) xtitle(الدخل)

فيعرض البرنامج العناوين المذكورة عوض x و y ، كما يوضحه الشكل التالي:

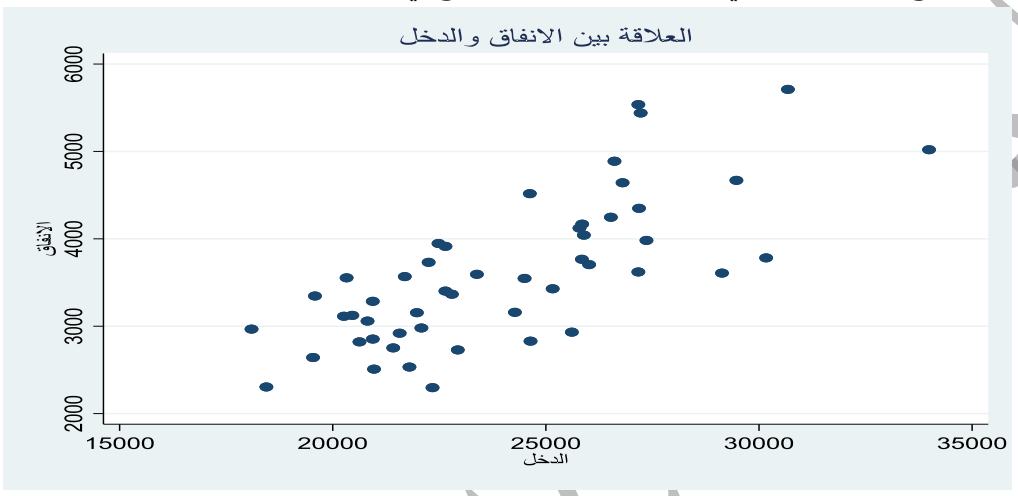


[(*)] هذه الاضافات عبارة عن تقنيات لتوضيح أكثر للرسم البياني يمكن استعمالها مع مختلف الأوامر الخاصة بالاشكال البيانية، وسنفصل فيها في هذا النوع من الأشكال ومن ثم يمكن تعميمها على باقي الاشكال دون الاشارة اليها.

-**title("titre de graphique")** : لإعطاء عنوان للمخطط البياني. بحيث يتم اضافة هذه العبارة الى الأمر السابق لاظهار عنوان الشكل، وهنا نكتب الأمر التالي:

graph twoway scatter y x, ytitle(الانفاق) xtitle(الدخل) title(العلاقة بين الانفاق والدخل)

فيعرض البرنامج التمثيل البياني بالعنوان كما هو موضح في الصورة التالية:



-

xlabel(1(2)20) لتعديل محور الفواصل (الوحدات) بحيث (1) هي نقطة البداية، (2) هي المسافة بين وحدات المتغير في محور الفواصل، 20 اقصى قيمة في محور الفواصل.

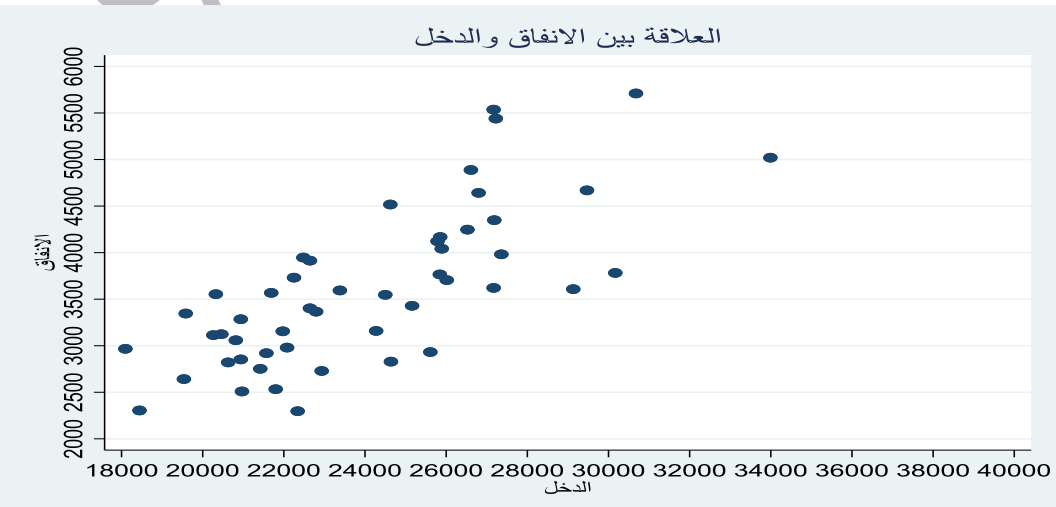
- **ylabel(0,(10) 200)**، لتعديل محور التراتيب بنفس الطريقة المعتمدة لمحور الفواصل.

وهنا نشير الى انه يجب ان يكون المستخدم على دراية بمدى المعطيات المتوفرة لكل متغيرة حتى يتسنى له تقسيم المحاور الى عدد معين من الفئات بتحديد القيمة الأولى والأخيرة والمدى بين كل فئة من الفئات التي يتم عرضها في الشكل البياني. فمثلا اضافة العبارة التالية:

xlabel(18000(2000)40000) و ylabel(2000,(500) 6000)

ليصبح الأمر كاملا:

graph twoway scatter y x, ytitle (الانفاق)xtitle (الدخل)title (العلاقة بين الانفاق والدخل)
xlabel(18000(1500)40000) ylabel(2000(500)6000)



- كما يسمح البرنامج بتغيير اشكال نقاط الانتشار من دوائر صغيرة ملونة الى دوائر غير ملونة أو مثلثات أو مربعات أو علامة (+) بإضافة العبارة: **msymbol(oh)** أو **msymbol(th)** أو **msymbol(s)** أو **msymbol(+)**. وغيرها.

بحيث ان استخدام هذه الرموز:

Oh: دوائر صغيرة

Th: مثلثات

(s): مربعات

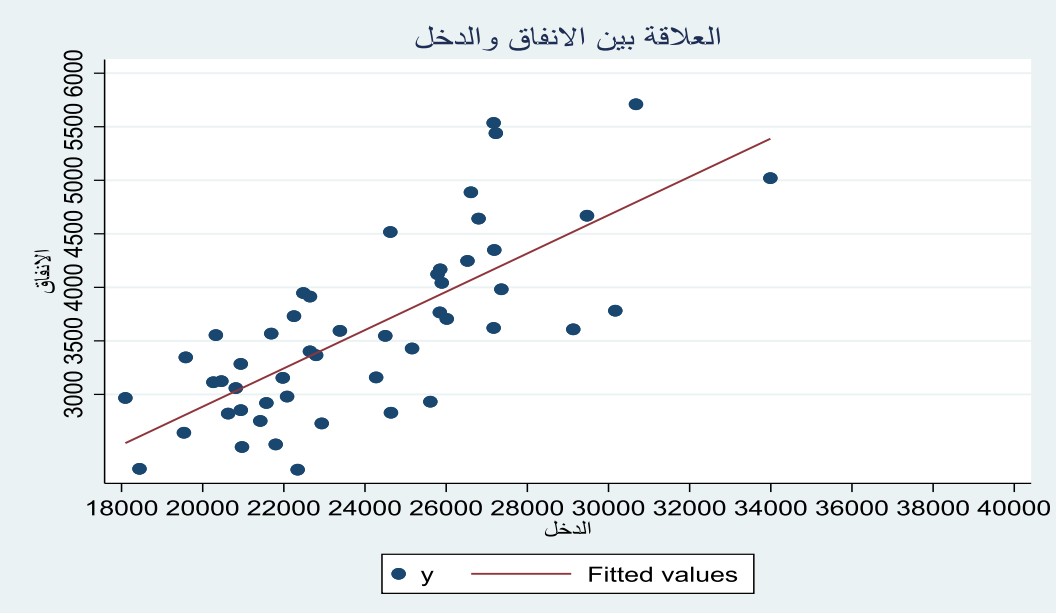
(+): علامة الجمع

يعرضها البرنامج عوض الدوائر الملونة وهي العرض التلقائي للبرنامج.

- بما أن هذا النوع من الأشكال البيانية يستخدم غالبا لتحليل العلاقة بين متغيرين كميين، يتيح البرنامج امكانية رسم خط الانحدار الى جانب نقاط الانتشار بإضافة الرمز ||[*] للتعبير على رسم الشكلين في شكل واحد، واضافة العبارة **lfit** الخاصة بخط الانحدار متبوعة باسماء المتغيرين أي كتابة الأمر كاملا وفق الشكل العام التالي: **graph twoway scatter y x, title(« titre") || lfit y x**

وفي مثالنا نكتب

graph twoway scatter y x, ytitle (الانفاق) xtitle (الدخل)title (العلاقة بين الانفاق والدخل) xlabel(18000(2000)40000) ylabel(3000(500)6000)|| lfit y x



[*] لكتابة هذا الرمز باستخدام لوحة المفاتيح يكفي النقر على الزر alt gr والزر 6 معا بشرط ان تكون اللغة المستخدمة هي اللغة الفرنسية أو الانجليزية.

2. مصفوفة الانتشار:

يسمح البرنامج بإعداد رسم بياني في شكل مصفوفة الانتشار بين متغير تابع ومجموعة من المتغيرات وبين كل زوج من المتغيرات في آن واحد، لاظهار علاقة الارتباط بين المتغيرات (مصفوفة الارتباط)، وذلك من خلال الأمر:

graph matrix y x1 x2 x3 x4..... -

ويستحسن اضافة العبارة **half** لعرض الجزء السفلي للمصفوفة فقط لأنها مصفوفة متناظرة. وكذلك اضافة بعض العبارات التي يراها المستخدم ملائمة من بين العبارات المستخدمة في شكل الانتشار. فيصبح الأمر على الشكل:

graph matrix y x1 x2 x3, half msymbol(oh)

ولتوضيح ذلك نستعين بالبيانات التالية:

الجدول (4-2): عدد ساعات العمل ومتغيرات أخرى باحدى المؤسسات

Obs	y	x1	x2	x3	Obs	y	x1	x2	x3
1	2,905	2157	1121	291	19	1,423	1985	553	381
2	2,97	2174	1128	301	20	3,636	2184	1091	291
3	2,35	2062	1214	326	21	2,983	2084	1327	331
4	2,511	2111	1203	49	22	2,573	2051	1194	279
5	2,791	2134	1013	594	23	3,262	2127	1226	314
6	3,04	2185	1135	287	24	3,234	2102	1188	414
7	3,222	2210	1100	295	25	2,28	2098	973	364
8	2,493	2105	1180	310	26	2,304	2042	1085	328
9	2,838	2267	1298	252	27	2,912	2181	1072	304
10	2,356	2205	885	264	28	3,015	2186	1122	30
11	2,922	2121	1251	328	29	3,01	2188	990	366
12	2,499	2109	1207	347	30	1,901	2077	350	209
13	2,796	2108	1036	300	31	3,009	2196	947	294
14	2,453	2047	1213	297	32	1,899	2093	342	311
15	3,582	2174	1141	414	33	2,959	2173	1116	296
16	2,909	2067	1805	290	34	2,971	2179	1128	312
17	2,511	2159	1075	289	35	2,98	2200	1126	204
18	2,516	2257	1093	176					

المصدر: Damodar N. Gujarati: « Basic Econometrics », Fourt Edition, The McGraw Hill Companies, 2004, p 385.

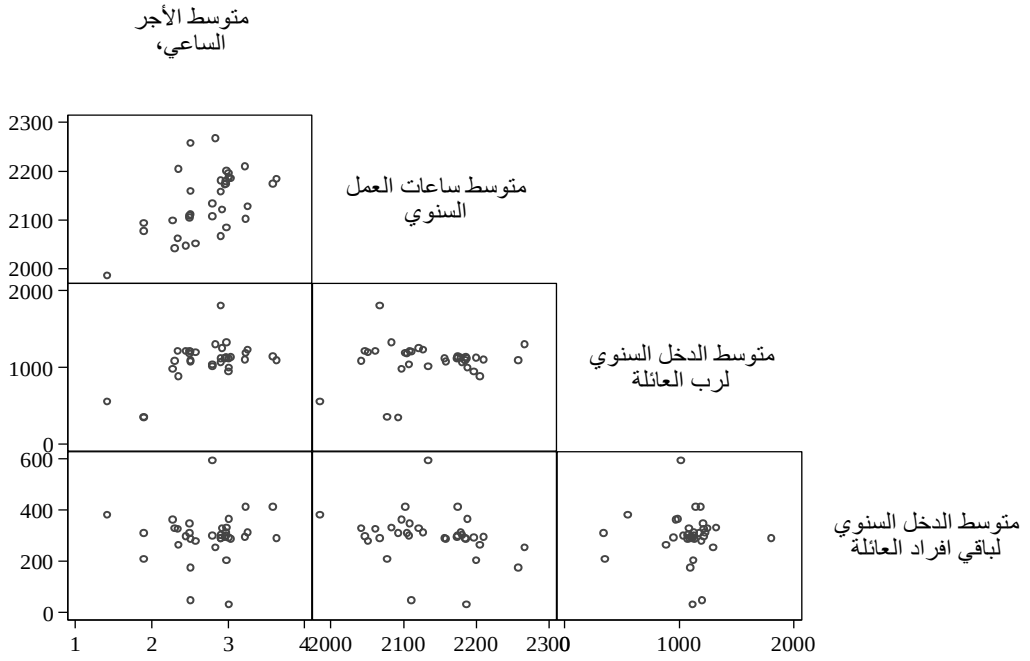
بحيث تمثل: y: متوسط الأجر الساعي، X1: متوسط ساعات العمل السنوي، X2: متوسط الدخل السنوي لرب العائلة، X3: متوسط الدخل السنوي لباقي افراد العائلة الآخرين.(*)

وبكتابة الأمر:

(*) يتم اعطاء عناوين المتغيرات في متصفح البيانات لاظهارها في الشكل عوض الرموز y، x1،...

graph matrix y x1 x2 x3, half msymbol(oh)

يعرض البرنامج مصفوفة الانتشار على النحو التالي:



يوضح الشكل مجموع من اشكال نقاط الانتشار بين كل زوج من المتغيرات المدرجة، ويفيد في الكشف عن امكانية وجود تعدد خطي في نموذج الانحدار المتعدد، بحيث في هذا النوع من النماذج يشترط عدم ترابط المتغيرات المفسرة خطيا مع وجود علاقة خطية بين المتغيرات المفسرة والمتغير التابع، والتي تظهر في العمود الأول من المصفوفة.

3. المخططات البيانية الخطية

هذا النوع من الاشكال البيانية تتمثل في الربط بين مختلف نقاط الانتشار بقطع مستقيمة، ليظهر الرسم في شكل خطي. ويتم اعدادها باستعمال الأمر:

graph twoway connect y x

أو

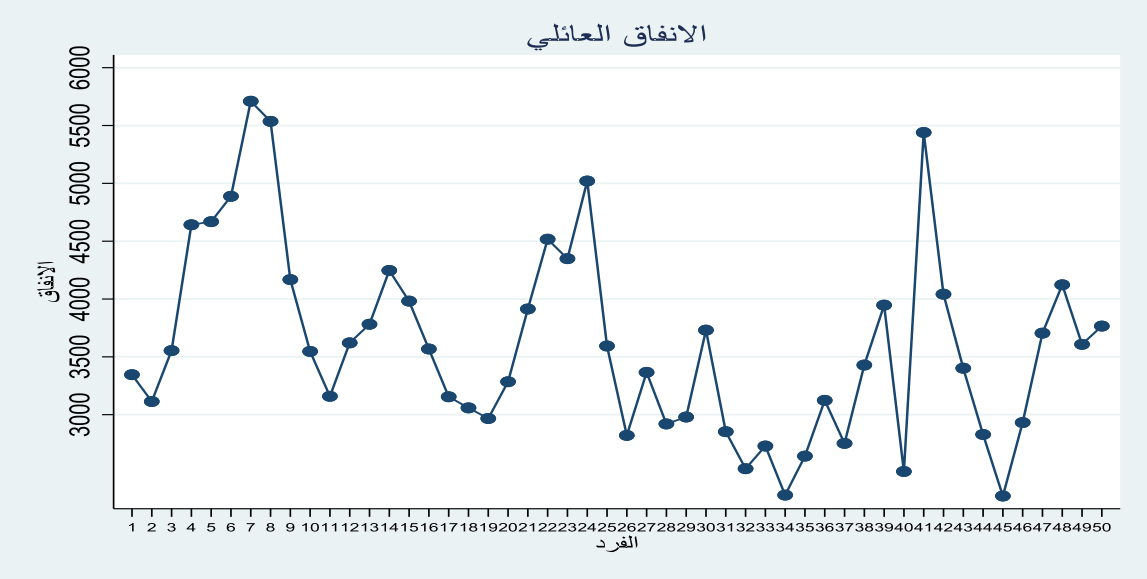
graph twoway line y x

مع امكانية التحكم في محور الفواصل والتراتيب والعنوان بنفس الطريقة المعمول بها في اشكال الانتشار. فمثلا بالاستعانة ببيانات الجدول رقم (4-1) نقوم الانفاق على التمدرس حسب الافراد، وذلك بكتابة الأمر التالي:

graph twoway connect y ind, ytitle (الانفاق)xtitle (الفرد)title (الانفاق العائلي)xlabel(1 (1)50)

ylabel(3000(500)6000)

فينتج لدينا التمثيل البياني من الشكل التالي:



كما يشتهر هذا النوع من الأشكال البيانية في تحليل تطور متغير معين أو عدة متغيرات عبر الزمن، في شكل سلسلة زمنية، ففي هذه الحالة يتم تعريف متغير الزمن ورسم المخطط بالأمر:

في حالة المتغير الواحد: **graph twoway line y année**

وفي حالة عدة متغيرات: **graph twoway line x1 x2 année**

ويمكن تحديد حدود محور الفواصل بإضافة العبارة **xlabel(t1(5)T)** أي تحديد السنة الأولى (نقطة البداية)، (مجال معين)، السنة الأخيرة (نقطة النهاية).

مثال: لدينا البيانات المبينة في الجدول التالي:

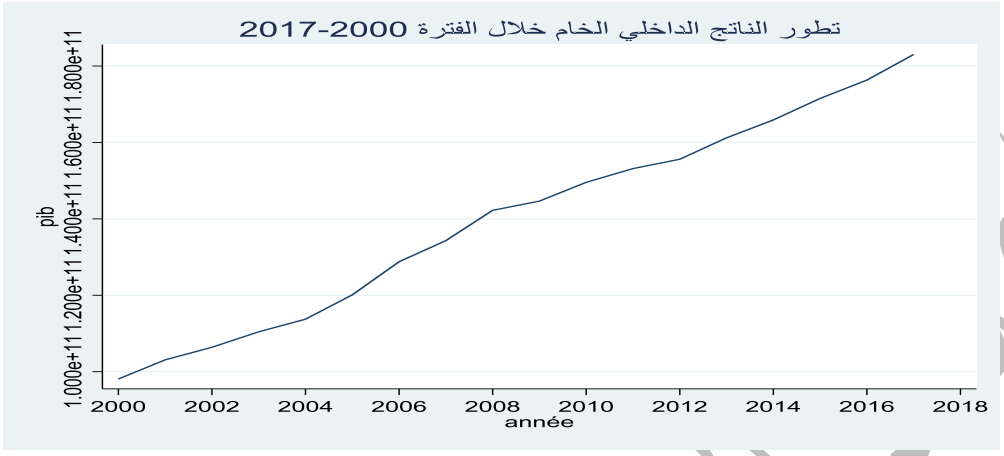
الجدول رقم (4-3): الناتج الداخلي الخام ورأس المال الثابت في الجزائر خلال الفترة 2000-2017

année	pib	gfcf	année	pib	gfcf
2000	98092030277	22207173479	2009	144647409921	34965906473
2001	103094727361	22940009850	2010	149565421858	38392565258
2002	106393760237	23559390522	2011	153154991982	42155036686
2003	110436723126	24124817199	2012	155605471854	54664621343
2004	113749824820	24727935998	2013	161207268841	58491144839
2005	120119815010	26607260180	2014	165882279637	60187388032
2006	128768441691	27964230109	2015	171522277145	64520879998
2007	134305484683	30229331986	2016	176324900905	70069675664
2008	142229508280	32617450053	2017	183025247139	74554134892

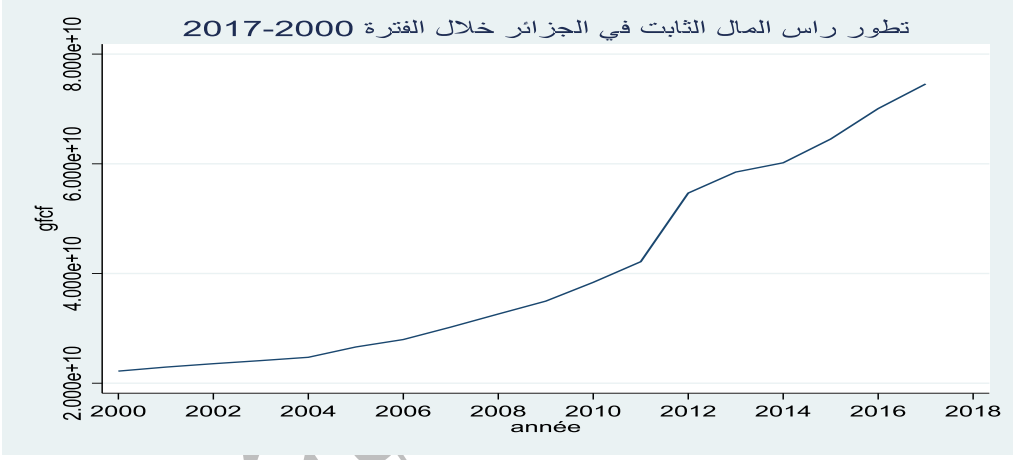
المصدر: تم اعداده بناء على بيانات البنك الدولي لسنة 2019.

باستخدام بيانات الجدول (4-3) أعلاه، يمكننا تمثيل مختلف المتغيرات وتطورها عبر الزمن بكتابة الأمر التالي:

graph twoway line pib année, xlabel(2000(2)2018) title(تطور الناتج الداخلي الخام خلال الفترة 2000-2017)

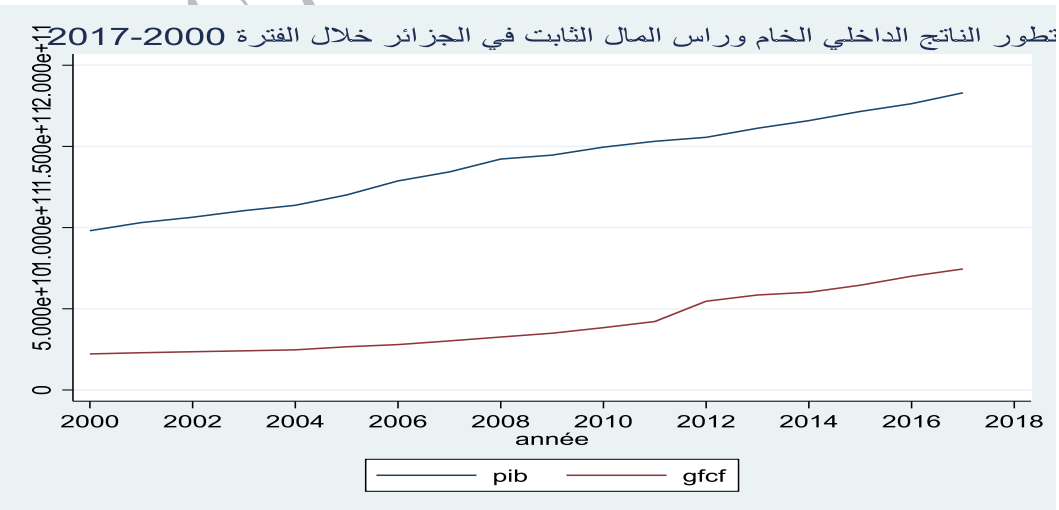


graph twoway line gfcf année, xlabel(2000(2)2018) title(تطور راس المال الثابت في الجزائر خلال الفترة 2000-2017)



كما يمكن تمثيل السلسلتين في نفس المعلم بكتابة الأمر التالي:

graph twoway line pib gfcf année, xlabel(2000(2)2018) title(تطور الناتج الداخلي الخام وراس المال الثابت في الجزائر خلال الفترة 2000-2017)



بحيث تم الاستعانة بالعبارة xlabel لتعديل عنوان محور الفواصل ويمكن استخدامه ايضا لتعديل عنوان محور التراتيب حسب الحاجة، وذلك وفقا لما تمت الاشارة اليه سابقا في شكل الانتشار.

مثال:

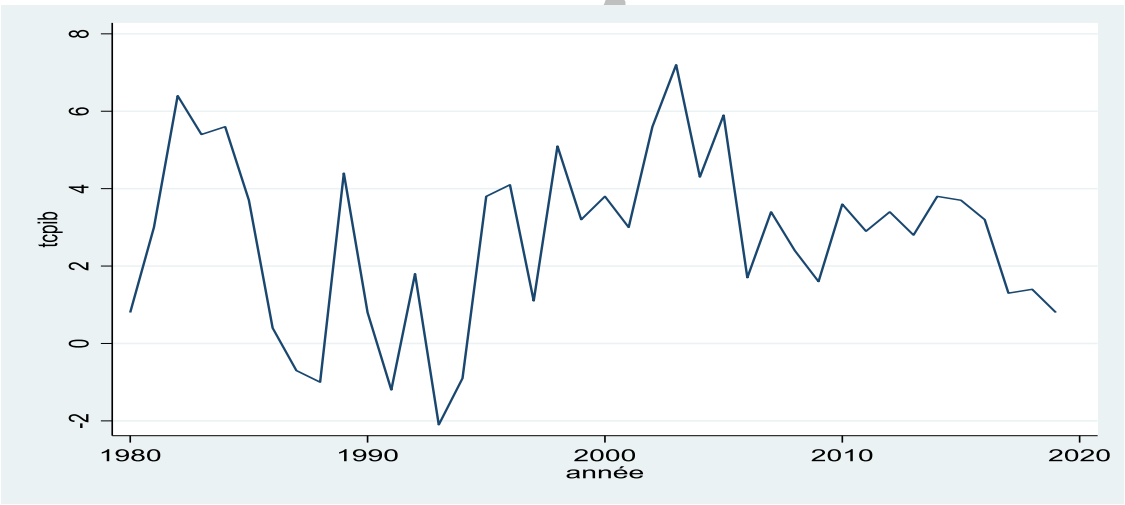
لتكن لدينا السلسلة الزمنية الخاصة بمعدل نمو الناتج الداخلي الخام في الجزائر خلال الفترة 1980-2016، ونريد تمثيل هذه السلسلة بيانيا.

الجدول رقم (4-4): تطور معدل نمو الناتج الداخلي الخام في الجزائر خلال الفترة 1980-2016

année	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
tpib	0,8	3	6,4	5,4	5,6	3,7	0,4	0,7-	1-	4,4
année	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
tpib	0,8	1,2-	1,8	2,1-	0,9-	3,8	4,1	1,1	5,1	3,2
année	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
tpib	3,8	3	5,6	7,2	4,3	5,9	1,7	3,4	2,4	1,6
année	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
tpib	3,6	2,9	3,4	2,8	3,8	3,7	3,2	1,3	1,4	0,8

المصدر: تم اعداده بناء على بيانات البنك الدولي 2020.

بتطبيق الأمر: **twoway line tpib année** نتحصل على الشكل التالي:



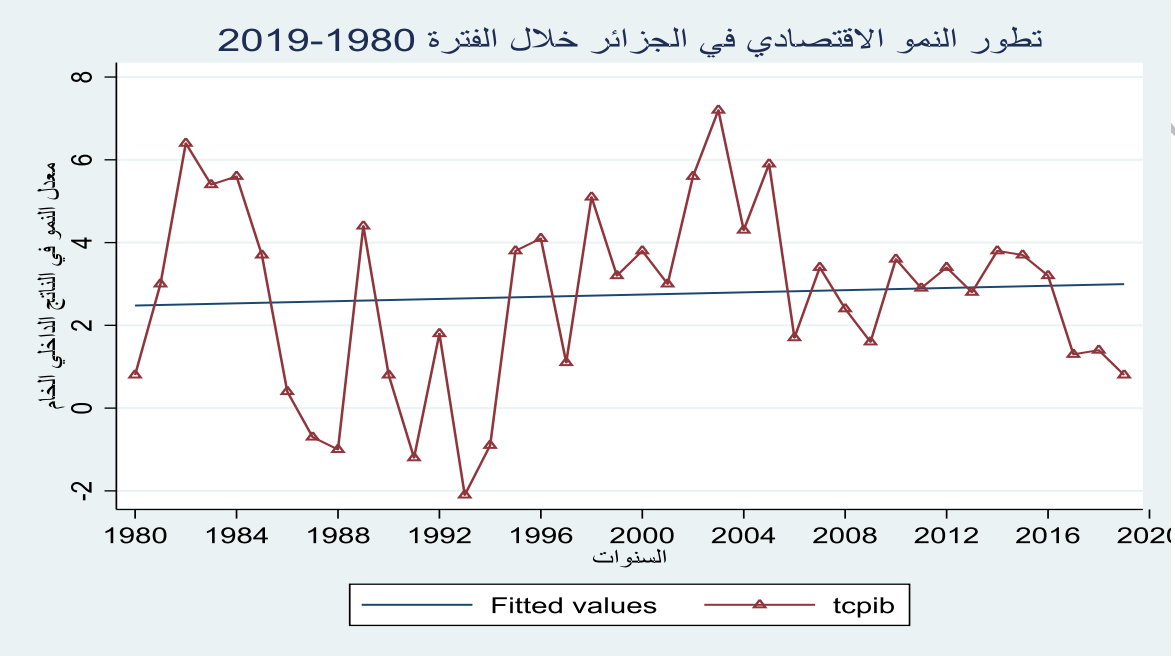
وهو الشكل التلقائي الذي يعرضه البرنامج، وسنقوم بتعديل عبارة الأوامر من أجل تعديل العرض باظهار نقاط الانتشار الخاصة بمعدل النمو ورسم خط الاتجاه عبر الزمن كما يلي:

graph twoway lfit tpib année ||connect tpib année, msymbol(th) ytitle(معدل نمو الناتج) title(تطور معدل نمو الناتج الداخلي الخام خلال الفترة 1980-2019) xlabel(1980(4)2019) (الداخلي الخام 2019-1980)

بحيث تم اضافة العبارة lfit لرسم خط الاتجاه، || للربط بين مختلف نقاط الانتشار، msymbol(th) لرسم نقاط الانتشار في شكل مثلثات ملونة، ytitle() لاعطاء عناون لمحور التراتيب، xlabel() لتعديل محور الفواصل وأخيرا اعطاء عنوان للشكل باضافة العبارة title().

كما يمكن اضافة عبارات اخرى حسب الحاجة، فمثلا يمكن اضافة « ci » للامر lfit ونكتب lfitci لتحديد مجال الثقة 95%.

وبتطبيق الأمر نتحصل على الشكل التالي:



4. المدرج التكراري

يستخدم هذا النوع من المخططات البيانية في دراسة التوزيع التكراري لمتغير معين، فهو يعطينا صورة عن شكل التوزيع الذي يمكن مقارنته بشكل توزيع معين كالتوزيع الطبيعي مثلا، ويمكن رسم المدرج التكراري لمتغير معين باستخدام البرنامج Stataبعدة طرق نذكرها فيما يلي:

أ. الأمر: **histogram x, frequency** يسمح بعرض المدرج التكراري للمتغير x مع توضيح التكرارات في حور التراتيب.

ب. الأمر **histogram x, start (0) width(10) norm fraction**: لعرض المدرج التكراري للمتغير x بعرض 10 وحدات مع البداية من الصفر ورسم منحنى التوزيع الطبيعي.

جـ. الأمر **histogram x, by(y,total) percent**: لعرض عدة مدرجات تكرارية للمتغير x لكل قيمة من المتغير y، والمدرج الأخير يعطي المجموع للعينة ككل، مع عرض النسب المئوية على محور التراتيب.

مثال:

لتكن لدينا البيانات الخاصة بالمعدلات المحصلة (moyen) لـ 90 طالب اجتازوا امتحان البكالوريا الى جانب المعلومات الخاصة بشعبة البكالوريا (Filière) والجنس (Sexe).

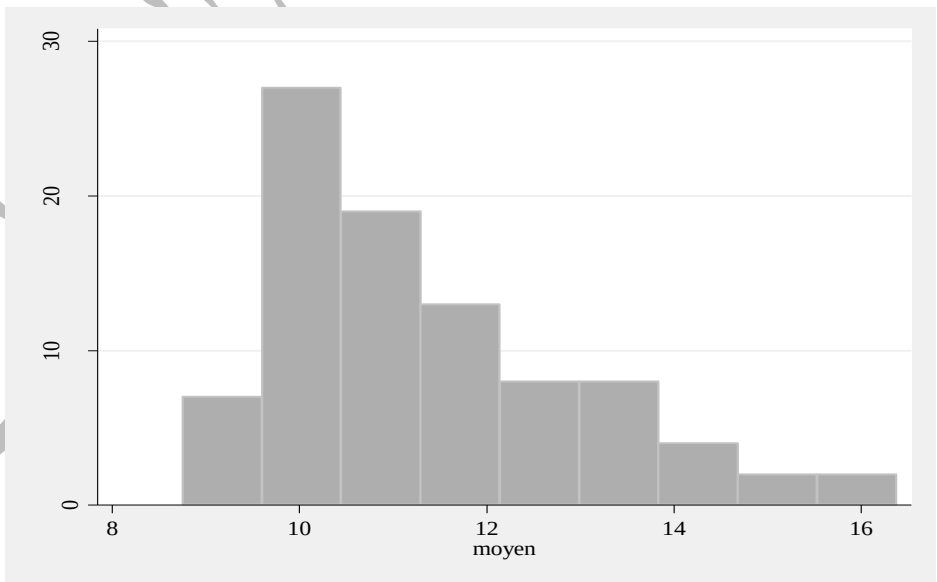
الجدول رقم (4-5): بعض البيانات الخاصة بنتائج امتحان البكالوريا لسنة 2006.

	sexe	Filiere	moyen
1	Femme	LLE	13.875
2	homme	LLE	11.125
3	Femme	LLE	12.375
4	Femme	LLE	12.25
5	Femme	LLE	11
6	Femme	LLE	10
7	Femme	LLE	9.875
8	Femme	LLE	12
9	Femme	LLE	9.75
10	homme	LLE	9.75
11	Femme	LLE	10
12	homme	LLE	13.625
13	homme	LLE	10.125
14	homme	LLE	11.25
15	Femme	LLE	8.75
16	homme	LLE	13.75
17	Femme	LLE	13
18	Femme	LLE	10.75
19	Femme	LLE	9.875
20	Femme	LLE	13.875
21	Femme	LLE	9.75
22	Femme	LLE	10
23	Femme	LLE	11.75
24	Femme	LLE	11.125
25	Femme	LLE	10
26	Femme	LLE	9.875
27	Femme	LLE	14.375
28	Femme	LLE	11.25
29	homme	LLE	12.5
30	Femme	LLE	10.25
31	Femme	SNV	12.375
32	homme	SNV	10.375
33	homme	SNV	11
34	homme	SNV	15.75
35	Femme	SNV	13
36	homme	SNV	11.375
37	homme	SNV	11.25
38	Femme	SNV	10.375
39	homme	SNV	9.375
40	Femme	SNV	10.375
41	homme	SNV	16.375
42	Femme	SNV	15.375
43	homme	SNV	10.75
44	homme	SNV	11.375
45	homme	SNV	11.75
46	homme	SNV	10.5
47	Femme	SNV	13.25
48	homme	SNV	12.625
49	Femme	SNV	10.75
50	homme	SNV	14.875
51	Femme	SNV	9.5
52	homme	SNV	9.625
53	homme	SNV	11.125
54	homme	SNV	11.25
55	homme	SNV	10.75
56	homme	SNV	11.5
57	Femme	SNV	12.125
58	Femme	SNV	13.375
59	homme	SNV	12.125
60	homme	SNV	13.375
61	homme	GE	10.125
62	homme	GE	14
63	Femme	GE	12.25
64	Femme	GE	9.125
65	Femme	GE	11.5
66	homme	GE	9.125
67	Femme	GE	13.25
68	Femme	GE	12.125
69	homme	GE	11.375
70	homme	GE	12.375
71	Femme	GE	10.125
72	homme	GE	10.25
73	homme	GE	10.875
74	homme	GE	9.625
75	homme	GE	11.125
76	Femme	GE	12.25
77	Femme	GE	10.125
78	homme	GE	10
79	Femme	GE	12.125
80	Femme	GE	10.5
81	Femme	GE	8.875
82	homme	GE	11.75
83	Femme	GE	10.375
84	Femme	GE	10.875
85	homme	GE	10
86	Femme	GE	9.25
87	homme	GE	10.125
88	Femme	GE	10
89	Femme	GE	11.125
90	homme	GE	10.125

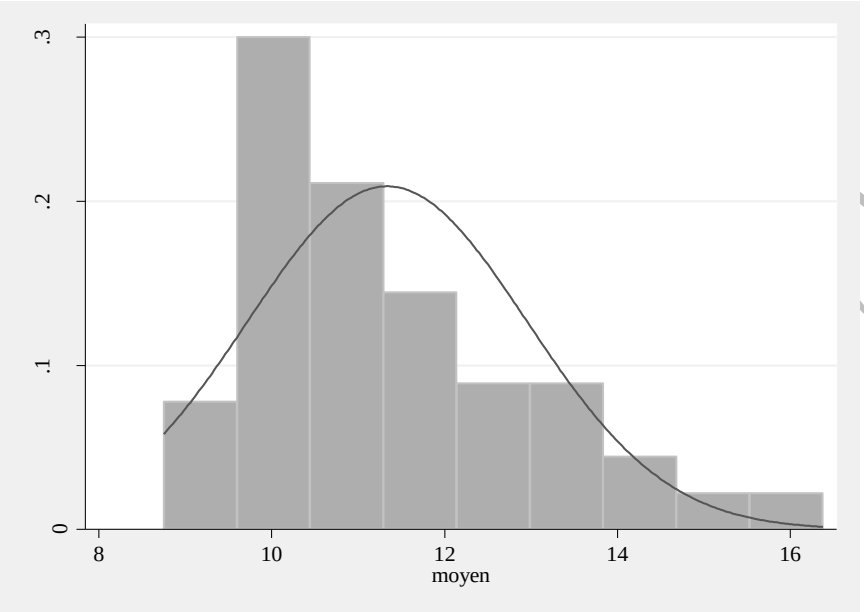
المصدر: تم اعداده بناء على بيانات قاعدة البيانات للديوان الوطني للامتحانات والمسابقات

المطلوب: اعداد المدرج التكراري لمتغير معدل البكالوريا لعينة الدراسة.

للاجابة على المطلوب، نكتب الأمر: **histogram x, frequency** فيظهر البرنامج من خلال شاشة المخطط البياني المدرج التكرار كما هو موضح في الصورة التالية:

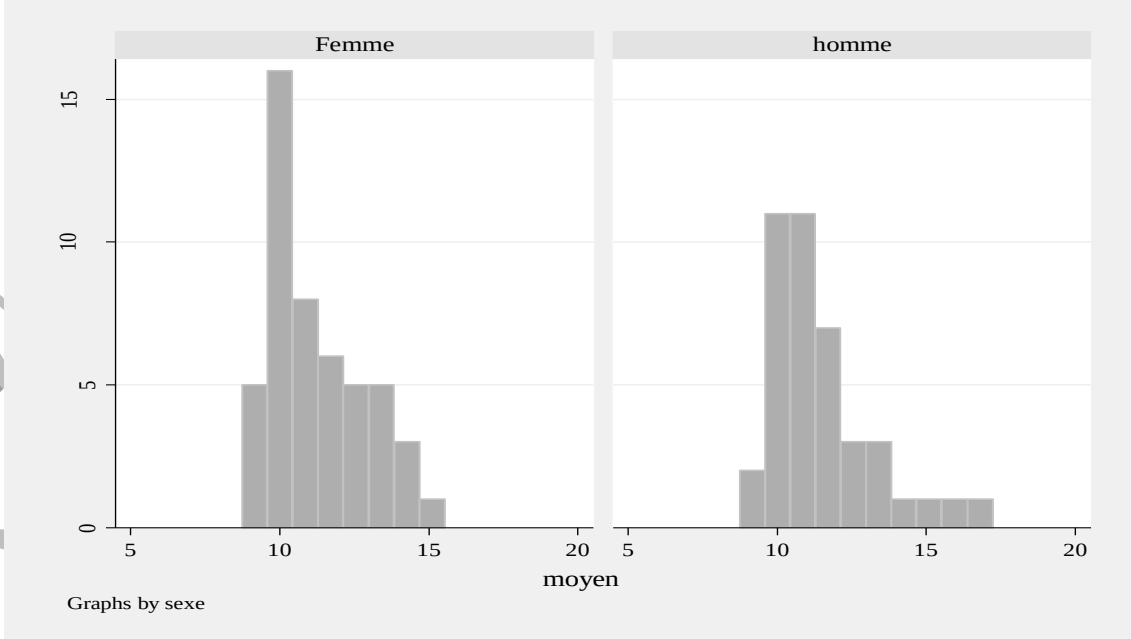


ومن اجل اضافة منحنى التوزيع الطبيعي الى المدرج التكراري يكفي اضافة العبارة norm fraction بعد وضع فاصلة، أي كتابة الأمر: histogram moyen , norm fraction فنتحصل على الشكل التالي:



وهو يوضح ان التوزيع التكراري لمعدل البكالوريا ملتوي نحو اليمين.

كما يمكن اعداد المدرج التكراري حسب فئات متغير معين، مثلا حسب فئات الجنس. في هذه الحالة يتم اعداد مخطط مزدوج للمدرج التكراري، فبكتابة الأمر: histogram moyen , by(sexe)frequency يعرض البرنامج المخطط التالي:

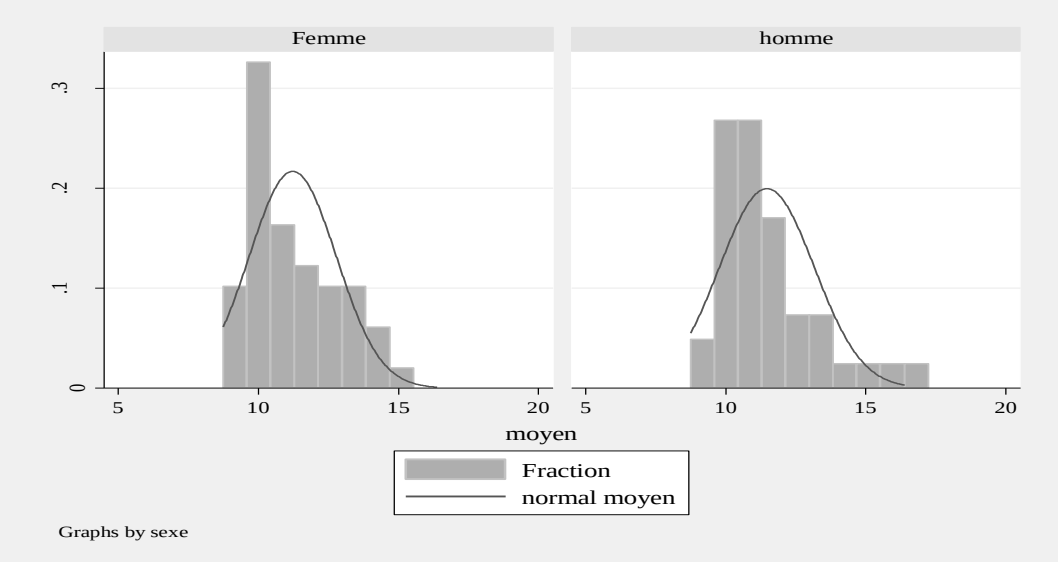


تجدر الاشارة الى انه يمكن عرض المدرج التكراري للعينة ككل الى جانب المدرج الخاص بكل فئة من متغير الجنس بكتابة الأمر: histogram moyen , by(sexe ,total) frequency، او تعديل العبارة frequency بكتابة percent لعرض التكرار النسبي عوض التكرار المطلق في محور التراتيب.

ويمكن استخدام التقنيات المذكورة سابقا لاضافة منحنى التوزيع الطبيعي بكتابة الأمر:

histogram moyen , by(sexe)norm fraction

فنتحصل على المخطط التالي:



5. الأعمدة التكرارية والدوائر النسبية

ان هذا النوع من المخططات يستخدم بكثرة في تحليل نتائج الدراسات الاستقصائية يتم اعداد المخططات البيانية في شكل أعمدة أو دوائر باستعمال الأمرين:

- الأعمدة التكرارية لمتغير كمي حسب متغير اسمي:

graph bar blédur, over(willaya)

بحيث يجب ان يكون المتغير المستهدف كميا وهي كمية الانتاج من القمح، بينما متغير التجزئة فيجب ان يكون اسمي في شكل فئات ومعبر عنها باحرف.

- الدائرة النسبية لمتغير معين:

graph pie blédur, over(willaya)

مثال:

لدينا البيانات الخاصة بانتاج الحبوب حسب المناطق الجغرافية في الجزائر والمعبر عنها بالولايات كما هي مبينة الجدول التالي:

الجدول رقم (4-6): كميات انتاج القمح في بعض الولايات الجزائرية للفترة 2009-2012

willaya	blé 2009	blé 2010	blé 2011	blé 2012
CHLEF	809440	671790	752700	1163340
TEBESSA	899030	136493	633600	275000
TIARET	1376500	1659742	1151400	2280600
SETIF	1443300	1444950	1974900	1479608
CONSTANTINE	700157	882540	1065700	1028692
MEDEA	986890	1161260	1297600	1475790
MASCARA	522800	338700	102800	374560
AIN-DEFLA	1246000	859255	1373000	1561966
RELIZANE	904355	861000	942700	1202420

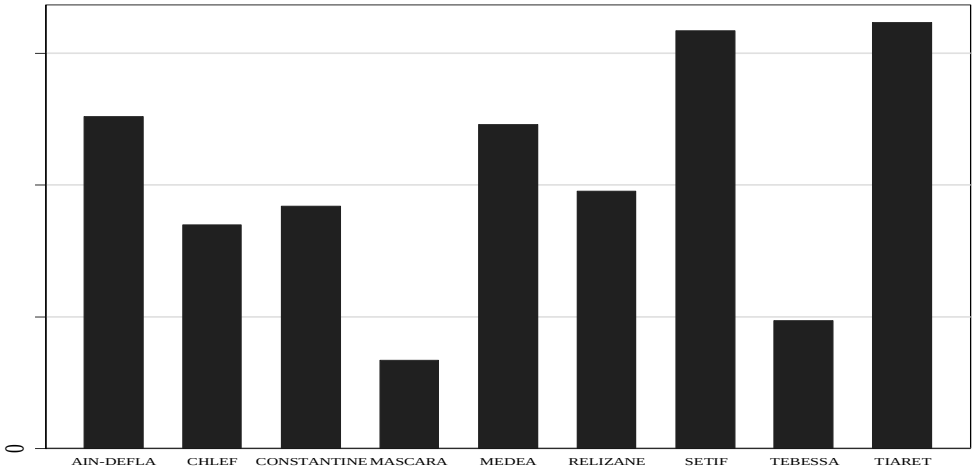
المصدر: بيانات وزارة الفلاحة والتنمية الريفية

يتم تحويل الجدول من شكله العرضي الى الشكل الطولي باستخدام الأمر resshape بكتابة:

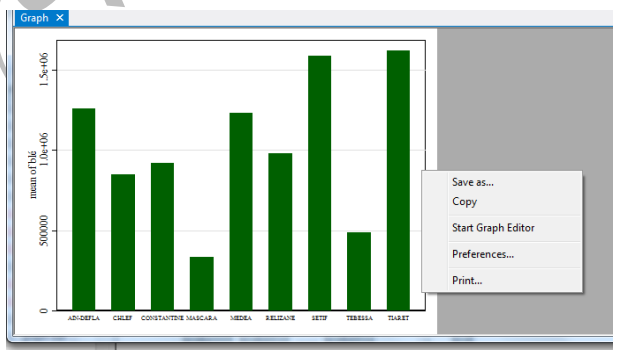
reshape long blé, i(willaya) j(année)

ثم نقوم بتمثيل متوسط كميات الانتاج من القمح حسب الولايات بكتابة الأمر:

graph bar blé ,over(willaya)



وتجدر الاشاة الى انه تم تعديل حجم الخط الخاص بمحور الفةاصل لكي يظهر اسم الولاية بوضوح، وذلك بالنقر على الز الايمن للفارة فوق البيان واختيار الأمر Start Graph Editor كما هو موضح في الصورة المقابلة:

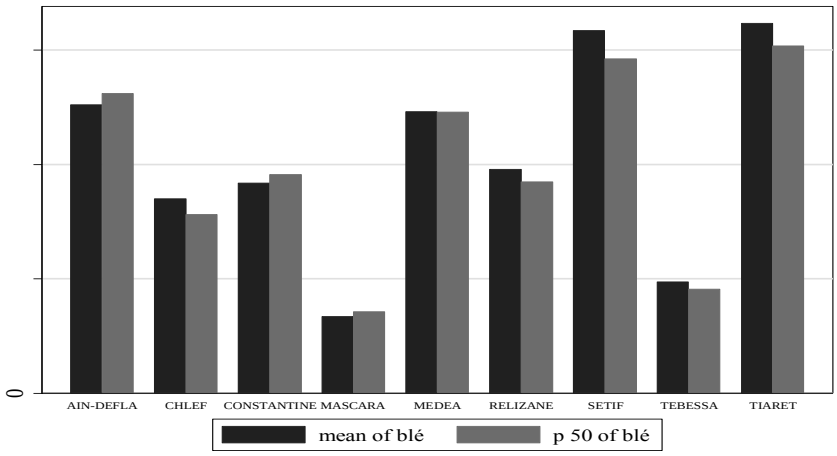


وتعديل حجم الخط من خلال الخيارات المتوفرة حسب الحاجة.

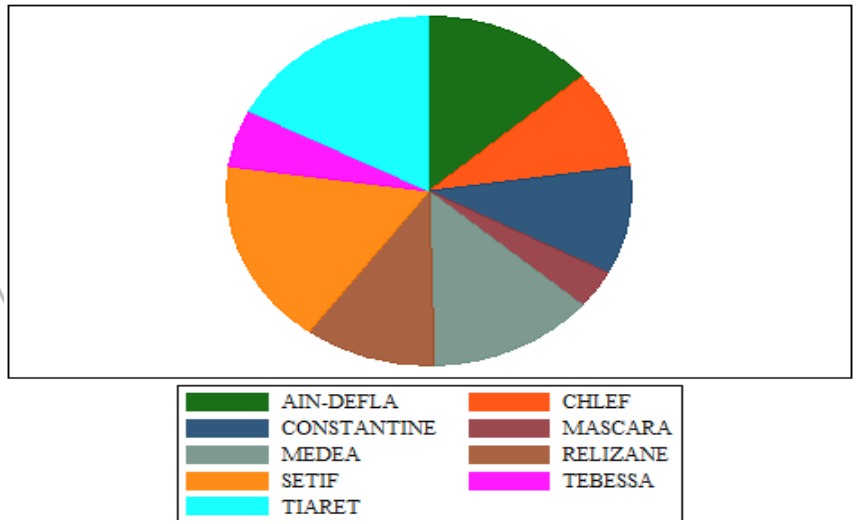
ويمكن عرض مخطط بياني في شكل اعمدة باستخدام مختلف المؤشرات الاحصائية دون اللجوء اى حسابات، من خلال تحديد نوع المؤشر بين قوسين قبل اسم المتغيرة في الأمر السابق، فالبرنامج يعرض تلقائيا قيمة المتوسط الحسابي دون كتابة الأمر (mean)، لكن عند استخراج المتوسط والوسيط مثلا يجب كتابته كما يلي:

graph bar (mean) blé (median) blé,over(willaya)

وبتطبيق الأمر نتحصل على الشكل التالي:



- لاعداد مخطط الدائرة، نستعين بالأمر الاساسي graph pie بنفس الطريقة، ففي بيانات المثال السابق، نكتب الأمر: graph pie blé, over(willaya) فنتحصل على الشكل التالي:

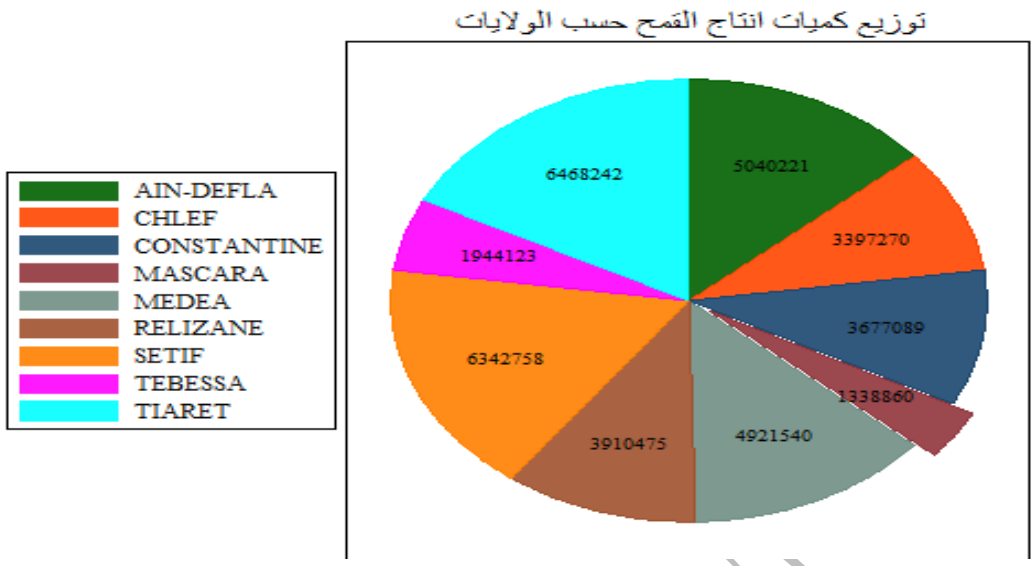


ويمكن تعديل خصائص الدائرة من الوان وغيرها من خلال النقر على الزر الايمن للفارة فوق المخطط واختيار الأمر Start Graph Editor وتعديل ما نراه مناسبا.

لكن تجدر الاشارة الى انه يمكن الاستعانة ببعض الأوامر الاساسية مثلا كتابة الأمر على الشكل التالي:

graph pie blé, over(willaya) pie(4, explode) plabel(_all sum,format(%7.0f))

title (توزيع كميات انتاج القمح حسب الولايات)legend(col(1) position(9))



الخيار pie(4, explode) يقوم بابراز الشريحة الرابعة خارجا للتأكيد على أهميتها، أما الخيار plabel(_all sum,format(%7.0f)) فيحدد من خلاله اظهار قيمة كل فئة من الفئات في الدائرة وهي كميات الانتاج من القمح لكل ولاية، أم الخيار format(%7.0f) فهو يعني التنسيق الرقمي الثابت مع سبعة أرقام بدون وجود اي رقم بعد الفاصلة. وللاطلاع أكثر على مختلف الخيارات المتوفرة مع الرسم البياني الدائري، اكتب العبارة help graph pie حيث ان يوفر للمستخدم العديد من الخيارات التي تساعد في استخراج رسم الدائرة باشكال دقيقة.

6. المخطط البياني في شكل صندوق

يعتبر مخطط الصندوق من بين اهم الاشكال البيانية المستخدمة في حال المتغيرات الكمية، كونها تعطي صورة حول مدى تمركز البيانات حول المركز وتشتتها وتناسقها وقيمها المتطرفة والشاذة ان وجدت، وهو يستخدم بكثرة الى جانب اختبارت التوزيع الطبيعي واستكشاف البيانات.

ويتم ستخراج هذا الشكل البياني باستخدام البرنامج Stata بكتابة الأمر:

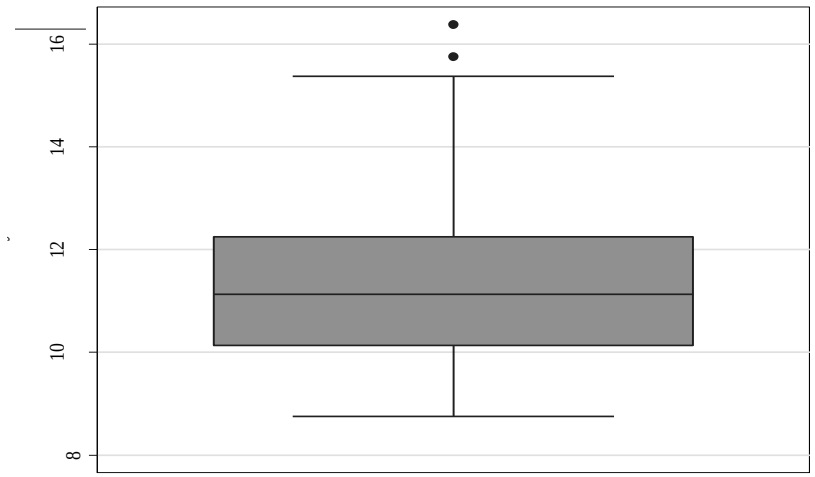
graph box variable

مثال:

استخراج مخطط الصندوق لمتغير معدل البكالوريا في المثال الخاص بالمدرج التكراري.

نقوم بكتابة الأمر:

graph box moyen

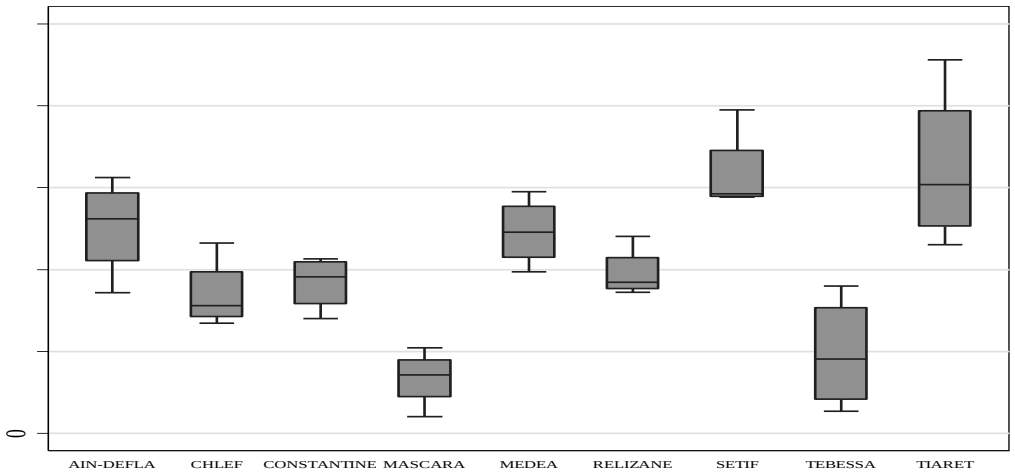


بحيث تمثل حافتي الصندوق قيمة الربيع الأول من الأسفل، وقيمة الربيع الثالث من الأعلى، بينما القيمة الدنيا فهي عبارة عن القيمة الأولى المشاهدة بعد ترتيب البيانات تصاعديا، وهي عبارة قيمة الحاصل: $Q_1 - 1.5 \times (Q_3 - Q_1)$ ، أما القيمة العظمى فهي القيمة الأخيرة المشاهدة (بعد استبعاد القيم المتطرفة والشاذة) وهي عبارة عن الحاصل: $Q_3 + 1.5 \times (Q_3 - Q_1)$. بينما تظهر القيم الشاذة والمتطرفة بعيدا عن الصندوق وتظهر في شكل نقاط انتشار بعيدة عن الصندوق. ويمكن الحكم على اعتدالية التوزيع والتوائه من خلال وضع الخط الذي يظهر في داخل الصندوق وهو قيمة الوسيط، فكلما كانت بعيدة عن وسط الصندوق كان توزيع البيانات ملتوي.

نلاحظ أمن خلال الشكل الالتواء الاجابي لهذا التوزيع. كما يعرض قيمتين متطرفتين، والصندوق يمتد من المدى الربيعي ويتضمن 50% من البيانات.

كما يتيح البرنامج استخراج مخطط الصندوق لعدة عينات في آن واحد باستخدام متغير تجزئة. فمثلا في البيانات الخاصة بكميات انتاج القمح عبر الولايات الجزائرية، نستخرج مخطط الصندوق بكتابة الأمر التالي:

graph box blé, over(willaya)



كما يمكن كتابة العبارة marker(1,mlabel(willaya)) لاظهار الولاية صاحبة القيم المتطرفة ان وجدت، فبافتراض وجود قيم متطرفة في ولايات معينة يظهر البرنامج مخطط الصندوق للعينة ككل [*] التالي:

graph box blé, marker(1,mlabel(willaya)

ملاحظة:

يمكن للطالب أو الباحث الاستعانة بالايقونة الخاصة بالتمثيل البياني من شريط الادوات، فهي تحتوي على كل انواع المخططات البيانية وبتقنيات مختلفة، لكن التعامل معها يكون معقد نوعا ما ولهذا قدمنا بعض التقيات البسيطة التي تساعد المستخدم على اعداد هذه المخططات بطريقة بسيطة تفيد في ربح الوقت.

7. التمثيل البياني لدالة الارتباط الذاتي والارتباط الذاتي الجزئي للسلسلة الزمنية Correlogram:

في اطار تحليل السلاسل الزمنية يحتاج الباحث الى رسم بياني لدالة الارتباط الذاتي والذاتي الجزئي، التي تساعده على تحديد سلوك السلسلة في الماضي وتحديد فترات الابطاء في نماذج الانحدار الذاتي والمتوسطات المتحركة على حد سواء، وتتوفر معظم البرامج الاحصائية الجاهزة على هذا النوع من المخططات البيانية على غرار البرنامج Stata. ولاستخراج هذا الاخير يكفي كتابة الأمر التالي:

Corrgram var(), lags(p)

بحيث يتم تحديد اسم السلسلة ()var بعد الامر corrgram ثم تحديد فترات الابطاء p المختارة من طرف المستخدم امام العبارة lags.

مثال:

لدينا البيانات الخاصة بمعدل التضخم في الجزائر خلال الفترة 1989-2019، المبينة في الجدول التالي:

الجدول رقم (4-7): تطور معدل التضخم في الجزائر خلال الفترة 1989-2019

année	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
inf	9,30	16,65	25,89	31,67	20,54	29,05	29,78	18,68	5,73	4,95	2,65
année	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
inf	0,34	4,23	1,42	4,27	3,96	1,38	2,31	3,68	4,86	5,74	3,91
année	2011	2012	2013	2014	2015	2016	2017	2018	2019		
inf	4,52	8,89	3,25	2,92	4,78	6,40	5,59	4,27	1,95		

المصدر: بيانات البنك الدولي بتصرف[1]

[*] كون العينات لا تحتوى على قيم شاذة تم تغيير البيانات الخاصة بالولايتين من اجل التوضيح فقط.

[1] البيانات المرصودة في الجدول هي بيانات التضخم، الأسعار التي يدفعها المستهلكون (% سنويا) مع تقريب القيم من أجل عرضها في الجدول.

بعد تعبئة البيانات في البرنامج Stata، يجب أولا التعريف بمتغير الزمن باستخدام الأمر[2] tsset حتى يتسنى للبرنامج معرفة ان الامر يتعلق ببيانات سلاسل زمنية وليس مقطعية، فنكتب الأمر: tsset année وبتطبيقه يعر البرنامج الجدول التالي:

```
. tsset année
        time variable:  année, 1989 to 2019
                delta:  1 unit
```

وهو دليل على اعتبار المتغير année متغير الزمن.

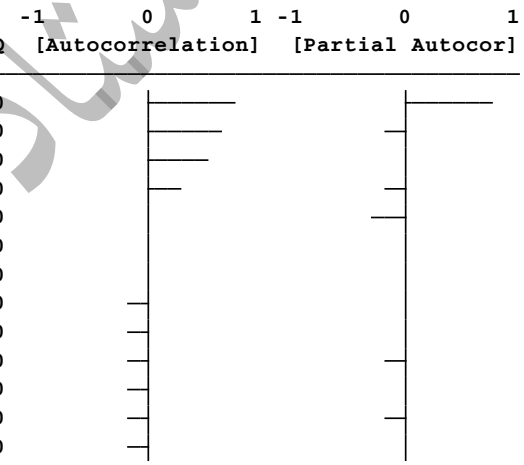
وبعد ذلك نكتب الأمر الخاص باستخراج التمثيل البياني لدالة الارتباط الذاتي والارتباط الذاتي الجزئي على النحو التالي:

corrgram inf, lags(13)

بحيث تم تحديد عدد فترات الابطاء بـ 13 فقط نظرا لكون السلسلة قصيرة نوعا ما (31 مشاهدة فقط)، بينما في حالة السلاسل الزمنية الطويلة يمكن تحديد عدد التأخيرات الى غاية 60 حتى تتضح الصورة أكثر حول سلوك السلسلة عبر الزمن.

وبتطبيق الأمر يعرض البرنامج الشكل التالي:

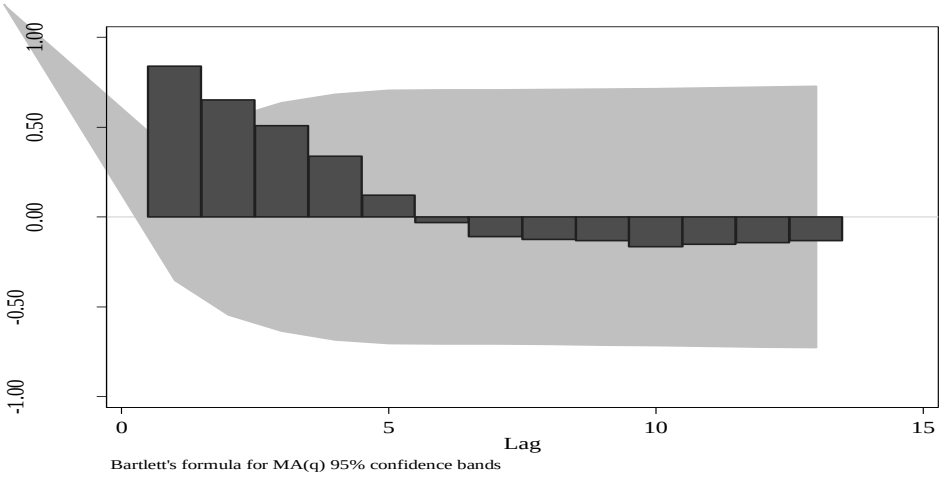
LAG	AC	PAC	Q	Prob>Q	[Autocorrelation]	[Partial Autocor]
1	0.8385	0.8548	23.976	0.0000		
2	0.6518	-0.1867	38.963	0.0000		
3	0.5095	0.0423	48.448	0.0000		
4	0.3382	-0.2356	52.781	0.0000		
5	0.1219	-0.2812	53.365	0.0000		
6	-0.0320	-0.0212	53.407	0.0000		
7	-0.1098	-0.0649	53.921	0.0000		
8	-0.1254	-0.0234	54.62	0.0000		
9	-0.1318	-0.0359	55.428	0.0000		
10	-0.1640	-0.2227	56.738	0.0000		
11	-0.1543	-0.1141	57.956	0.0000		
12	-0.1445	-0.2163	59.081	0.0000		
13	-0.1341	-0.1242	60.104	0.0000		



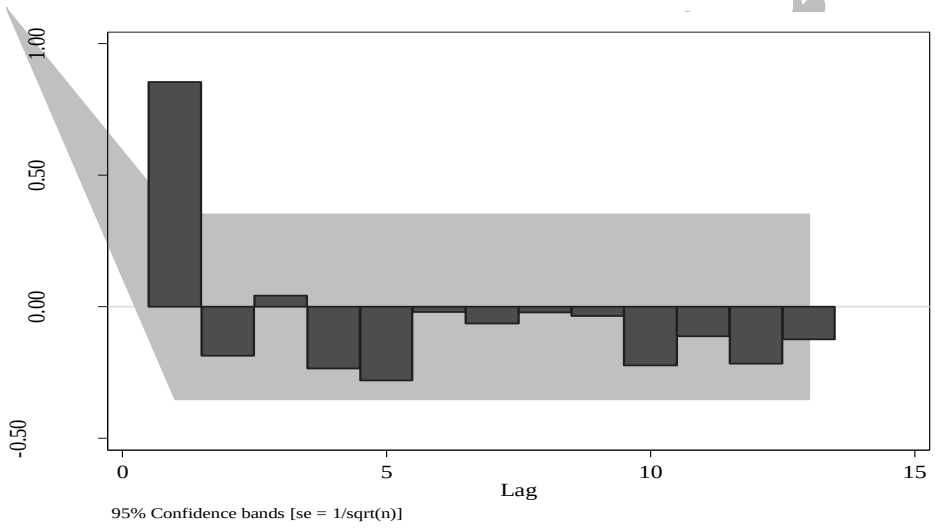
يتضح من خلال الشكل ان البرنامج يعرض الى جانب شكل دالة الارتباط الذاتي Autocorrelation ودالة الارتباط الذاتي الجزئي Partial Autocorrelation معاملات الارتباط المقدرة وقيمة احصائية Q (احصائية Ljung-box) لاختبار المنوية الاحصائية لمعاملات الارتباط.

اذا اراد المستخدم استخراج التمثيل البياني لدالة الارتباط الذاتي بوضوح دون اللجوء الى استخراج الـ correlogram يمكن الاستعانة بالأمر التالي: ac inf, recast(bar) الذي يقوم باعداد رسم لمعالات الارتباط الذتي للسلسلة inf في شكل أعمدة الى جانب مجال الثقة 95% كما هو موضح في الصورة التالية:

[2] لان البرنامج Stata يشترط التعريف بالبعد الزمني في حالة بيانات السلاسل الزمنية، كما هو الحال في بيانات البانل.



وكذلك الأمر بالنسبة لدالة الارتباط الذاتي الجزئي بكتابة الأمر: pac inf, recast(bar)، الذي يقوم باعداد الشكل التالي:



ملاحظة:

للتوسع أكثر في اعداد الرسومات البيانية الخاصة بالسلاسل الزمنية يمكن للمستخدم الاستعانة بالايقونة الخاصة بالمخططات البيانية Graphics من شريط الادوات، واختيار الخيار Times –series graphs الذي يحتوي على مختلف الأشكال البيانية الخاصة بهذا المجال حسب الحاجة.

الفصل الخامس: تحليل التباين والمقارنات المتعددة باستخدام برنامج Stata

يتناول هذا الفصل احد اهم انواع التحليل الاحصائي للعينات، وهو يتضمن مجموعة من الطرق لاختبار الفرضيات حول الاختلاف بين المتوسطات، ويمكن تطبيق هذا التحليل على نطاق واسع بدأ من التحليلات البسيطة التي يمكن من خلالها المقارنة بين متوسط متغير كمي ضمن فئات متغير نوعي، وصولا الى التحليلات المعقدة مع المتغيرات التصنيفية المتعددة والمتواصلة للمتغيرات. وبناء على ذلك يتمحور مضمون هذا الفصل في العناصر التالية:

- اختبار متوسط العينة الواحدة؛
- اختبار متوسطي عينتين مستقلتين؛
- اختبار متوسطي عينة مزدوجة؛
- اختبار متوسط اكثر من عينتين.

1. اختبار متوسط العينة الواحدة:

يفيد هذا التحليل في اختبار فرضية وجود فروق معنوية بين متوسط العينة وقيمة ثابتة معلومة. مع تحديد مجال ثقة لمتوسط المجتمع. بحيث يتم اختبار فرضية العدم التي تقر بتساوي متوسط العينة لمقدار معين ضد الفرضية البديلة عند مستوى معنوية 5%.

$$H_0 : \overline{X} = \mu$$
$$H_1 : -\overline{X} \neq \mu$$
$$-\overline{X} < \mu$$
$$-\overline{X} > \mu$$

بحيث تكون قيمة t المحسوبة وفق العلاقة التالية:

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \neg t_{(n-1)}^{(\alpha/2)}$$

ونقارنها بالقيمة المجدول بدرجة حرية n-1 من جدول ستودنت. بحيث يتم الاقرار بتساوي المتوسطات اذا كانت الاحصائية المحسوبة اقل من القيمة المجدول يتم الاقرار بان متوسط العينة يساوي الى المقدار المفترض (معدل 11)، أو باستخدام القيمة الاحتمالية التي يجب ان تكون أكبر من مستوى المعنوية الافتراضي (5%).

فمثلا لو فرضنا اننا نريد اختبار مساواة معدلات الطلبة للمعدل اللافتراضي 20/11 لعينة تتكون من 25 طالب من قسم السنة الثالثة تخصص: اقتصاد كمي وفق البيانات المبينة في الجدول التالي:

الجدول رقم (5-1): معدلات عينة طلبة السنة الثالثة تخصص اقتصاد كمي

no	moyenne	no	moyenne
1	9,13	14	12,25
2	11,68	15	9,88
3	12,33	16	10,92
4	10,33	17	11,01
5	10,01	18	10,82
6	11,33	19	11,18
7	10,38	20	11,82
8	9,64	21	11,91
9	10,15	22	10,67
10	10,5	23	12,9
11	11,98	24	10,3
12	14,22	25	12,18
13	11,92		

لاجراء الاختبار المتمثل في اختبار ستودنت باستعمال Stata نقوم بكتابة الأمر التالي:

ttest moyenne=11

فيعرض البرنامج جدول الاختبار كما يلي:

```
One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
moyenne |      25     11.1776    .2303955    1.151978    10.70209    11.65311
------------------------------------------------------------------------------
    mean = mean(moyenne)                                      t =   0.7708
Ho: mean = 11                                 degrees of freedom =       24

    Ha: mean < 11                Ha: mean != 11                 Ha: mean > 11
 Pr(T < t) = 0.7758         Pr(|T| > |t|) = 0.4483          Pr(T > t) = 0.2242
```

القيمة المحسوبة للاختبار

من خلال نتائج الاختبار المبينة في الجدول اعلاه، يتبين ان قيمة t ستودنت المحسوبة تقدر بـ 0.7708 وهي اقل من القيمة المجدولة عندة مستوى 5% بدرجة حرية 24 مما يقودنا الى قبول فرضية العدم التي تقر بمساوات المعدل المحصل من طرف الطلبة بالقيمة الافتراضية 11.

2. اختبار متوسطي عينتين مستقلتين

يقوم هذا الإجراء على المقارنة بين متوسطي عينتين مستقلتين تحت الفرضية:

$$H_0 : \overline{X}_1 = \overline{X}_2$$
$$H_1 : \overline{X}_1 \neq \overline{X}_2$$
$$\overline{X}_1 < \overline{X}_2$$
$$\overline{X}_1 > \overline{X}_2$$

باستعمال اختبار t ستودنت وفق العلاقة التالية:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sum(X_{i1} - \overline{X}_1)^2 + \sum(X_{i2} - \overline{X}_2)^2}{N_1 + N_2 - 1}} \sqrt{\frac{1}{N_1} + \frac{1}{N2}}}$$

بحيث N_1 وN_2 عدد المشاهدات في العينة الأولى والثانية على التوالي. وX_i تمثل القيم المشاهدة للعينة في الحالة i.

مثال:

لدينا البيانات الخاصة بأداء الاطارات الوسيطية في شركتين تعمل الاولى في قطاع تجهيزات النقل (32 اطار) والثانية تعمل في انتاج المنتجات الالكترونية (34 اطار)، والناتجة عن اختبار تقييمي لمختلف الخصائص الادارية. وتمثلت هذه البيانات في مجموع النقاط المحصلة في الامتحان والموضحة في الجدول التالي:

الجدول رقم (5-2): مؤشر اداء الاطارات الوسيطية في الشركة حسب قطاع النشاط:

cadre	Transport	Electronique
1	200	210
2	198	208
3	177	183
4	210	223
5	214	227
6	189	197
7	168	172
8	168	172
9	178	184
10	191	199
11	213	225
12	195	204
13	195	204
14	184	191
15	201	211
16	170	174
17	206	217

cadre	Transport	Electronique
18	159	162
19	188	196
20	204	215
21	183	190
22	129	155
23	221	235
24	178	184
25	206	218
26	147	147
27	212	225
28	178	183
29	182	189
30	150	150
31	192	201
32	174	179
33	/	199
34	/	186

المصدر: Gérald Baillargeon, Fernando ouellet : «*Analyse des données avec SPSS pour Windows* » ; Les Edition SMG, Canada, 2008, p 72.

قبل تعبئة هذه البيانات في برنامج Stata، يجب الاشارة إلى أنه يجب تعريف متغيرين هما متغير النتائج (Resultat) ومتغير القطاع (Secteur)، ومن ثم تعبئة البيانات الخاصة بنتائج التقييم في المتغير الاول، بينما المتغير الثاني فيأخذ القيمة 1 لقطاع النقل و2 لقطاع الالكترونيك. وهي الطريقة الصحيحة لتعبئة البيانات التي تمكننا من اجراء اختبار وجود فروقات معنوية بين اداء الاطارات الوسيطية في القطاعين.

بتعبئة البيانات في البرنامج Stata وكتابة الأمر edit تظهر لنا شاشة متصفح البيانات على الشكل التالي:

الجدول رقم (5-3): بيانات الجدول (5-2) بعد التعبئة الى البرنامج Stata

	secteur	resultat
1	1	200
2	1	198
3	1	177
4	1	210
5	1	214
6	1	189
7	1	168
8	1	168
9	1	178
10	1	191
11	1	213
12	1	195
13	1	195
14	1	184
15	1	201
16	1	170
17	1	206
18	1	159
19	1	188
20	1	204
21	1	183
22	1	129
23	1	221
24	1	178
25	1	206
26	1	147

	secteur	resultat
27	1	212
28	1	178
29	1	182
30	1	150
31	1	192
32	1	174
33	2	210
34	2	208
35	2	183
36	2	223
37	2	227
38	2	197
39	2	172
40	2	172
41	2	184
42	2	199
43	2	225
44	2	204
45	2	204
46	2	191
47	2	211
48	2	174
49	2	217
50	2	162
51	2	196
52	2	215

	secteur	resultat
53	2	190
54	2	155
55	2	235
56	2	184
57	2	218
58	2	147
59	2	225
60	2	183
61	2	189
62	2	150
63	2	201
64	2	179
65	2	199
66	2	186

لاحظ أنه تم انشاء متغير جديد يعبر عن القطاع يأخذ القيمة واحد لقطاع النقل والقيمة 2 لقطاع الالكترونيك

بهذه الطريقة نكون قد عرفنا المتغير الكمي resultat والمتغير النوعي secteur الذي من خلاله يتم تجزئة المتغير الكمي الى عينتين مختلفتين الأولى تتعلق بقطاع النقل والثاني يتعلق بقطاع الالكترونيك.

ان اختبار فرضية تساوي متوسطي الكفاءة الانتاجية في القطاعين (العينتين)، انما يشترط أولا معرفة ما اذا كانت تباينات العينتين متساوي أم لا، وعليه يجب اختبار تساوي تباين العينتين قبل اختبار تساوي المتوسطات[*].

[*] بعض البرامج الاحصائية الأخرى كبرنامج SPSS يقوم باجراء اختبار تجانس تباين العينتين بطريقة تلقائية الى جانب اختبار تساوي المتوسطات، بينما باستخدام البرنامج Stata يجب على الباحث اجراء اختبار تجانس التباين باستخدام الاوامر الخاصة به مسبقا.

ولاجراء الاختبار على عينتي العمال في شركة انتاج المركبات وشركة المنتجات الالكترونية نتبع الخطوات التالية:

1- اختبار تساوي التباين باستعمال الأمر **Sdtest** متبوع باسم المتغير، مع المحدد **by** لتبيين متغير التجزئة أي كتابة الأمر كما يلي: **Sdtest resultet, by(secteur)**

بحيث يتم اختبار تساوي التباين باستعمال اختبار فيشر، ويكون القرار بمقارنة قيمة F المحسوبة بالقيمة المجدولة بدرجة حرية (n1-1,n2-1) عند مستوى معنوية 5% بحيث n1 هي حجم العينة الأولى وn2 هي حجم العينة الثانية، أو باستعمال قيمة الاحتمال. فاذا كانت القيمة المحسوبة أقل من القيمة المجدولة نقبل فرضية تساوي التباين والعكس. أو إذا كانت قيمة الاحتمال أكبر من 0.05 فنقر بان تباين العينتين متساوي.

بتنفيذ الأمر يظهر البرنامج النتائج كما هي مبينة في الجدول التالي:

```
. sdtest resultat ,by( secteur)

Variance ratio test
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
transpor	32	186.25	3.764488	21.29516	178.5723	193.9277
electron	34	194.5588	3.887956	22.67049	186.6487	202.4689
combined	66	190.5303	2.737823	22.24218	185.0625	195.9981

قيمة اختبار فيشر المحسوبة

```
    ratio = sd(transpor) / sd(electron)                    f =   0.8823
Ho: ratio = 1                     degrees of freedom =  31, 33

    Ha: ratio < 1               Ha: ratio != 1                 Ha: ratio > 1
  Pr(F < f) = 0.3643         2*Pr(F < f) = 0.7285          Pr(F > f) = 0.6357
```

من خلال النتائج المبينة في الجدول، نلاحظ ان القيمة المحسوبة لفيشر تساوي 0.8823 وهي أقل من القيمة المجدولة المقدرة بـ 1.84، مما يقودنا الى قبول فرضية العدم التي تقر بتجانس تباين العينتين، وهذا الأمر طبيعي كون اننا لو تمعننا في قيم الانحراف المعياري للعينتين نجد انها متقاربة وتقترب من المساوات فهي تقدر بـ 3.76 بالنسبة للعينة الأولى و 3.88 بالنسبة للعنة الثانية. وعليه يمكن الانتقال الى الخطوة الثانية وهي اختبار تساوي متوسطي العينتين.

2- اختبار تساوي المتوسطين: بعد معرفة ما إذا كان تباين العينتين متساوي من عدمه، يتم اجراء اختبار تساوي متوسطين وفق كل حالة، بكتابة الأمر: **ttest** متبوع باسم المتغيرة، **by** متغير التجزئة متبوعة بكلمة **unequal** في حالة عدم التساوي، بينما نستغني عن هذه العبارة في حال تجانس تباين العينتين.

وفي مثالنا، بما ان تباين العينتين نكتفي بكتابة الأمر: **ttest resultat, by(secteur)**.

فيقوم البرنامج باختبار الفرق بين متوسطي العينتين بحساب قيمة t بدرجة حرية (n-2)، كما هو مبين في الجدول التالي:

```
Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
transpor	32	186.25	3.764488	21.29516	178.5723	193.9277
electron	34	194.5588	3.887956	22.67049	186.6487	202.4689
combined	66	190.5303	2.737823	22.24218	185.0625	195.9981
diff		-8.308824	5.42222		-19.14095	2.523305

قيمة t المحسوبة

```
    diff = mean(transpor) - mean(electron)                        t =  -1.5324
Ho: diff = 0                                     degrees of freedom =       64

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0652         Pr(|T| > |t|) = 0.1304          Pr(T > t) = 0.9348
```

يكون القرار بمقارنة قيمة t المحسوبة بتلك المجدولة عند مستوى معنوية 5%، فاذا كانت ttab>tcal يتم الاقرار بتساوي المتوسطين. أو باستخدام قيمة الاحتمال، فإذا كانت p-value المحسوبة أكبر من 0.05 يتم الاقرار بتساوي متوسطي العينتين.

وفي حالتنا نلاحظ أن قيمة الاحتمال من أجل التساوي هي 0.13 مما يدل على تساوي متوسطي العينتين. فمن خلال الجدول نلاحظ أن قيمة احصائية t تقدر بـ 1.532- ، وهي أقل من القيمة المجدولة عند مستوى معنوية 5% ($t_{64}^{0.025} \approx 2.0003$)، وهذا يقودنا إلى قبول فرضية العدم القائلة بأنه لا توجد فروقات معنوية بين متوسطي العينتين أي أن $\overline{X}_1 = \overline{X}_2$ بالرغم من وجود فرق بين المتوسطين بمقدار 8.309.

3. اختبار تساوي المتوسط للعينات المزدوجة

في بعض الاحيان يتم اختبار ادخال تغيير معين على عينة مما يؤدي الى وجوب اختبار اثر ذلك على العينة، ويتم ذلك باستعمال اختبار تساوي المتوسط، ومن أمثلة ذلك اختبار الشفاء من المرض نتيجة تلقي علاج معين، أو اختبار تحسن القدرة الانتاجية نتيجة تلقي تكوين معين وغيرها من الامثلة.

مثال: قامت مديرية الموارد البشرية بتوفير دورة تكيونية لعمال وحدة التركيب في مؤسسة لتركيب أجهزة كهرومنزلية، ونريد اختبار ما إذا كانت الدورة قد حسنت من انتاجية العامل أم لا. و هنا تم اختيار عينة عشوائية تتكون من 15 عامل، حيث تم حساب عدد الوحدات المركبة خلال فترة زمنية محددة قبل وبعد الدورة التكوينية فكانت النتائج كما يبينها الجدول التالي:

الجدول رقم (5-4): انتاجية العمال في مؤسسة تركيب الأجهزة الكهرومنزلية قبل وبعد الدورة

individus	avant le programme	après le programme
1	15	17
2	13	16
3	8	10
4	9	9
5	7	9
6	12	13
7	11	14
8	12	15
9	11	14
10	9	11
11	10	14
12	12	11
13	11	13
14	7	10
15	12	13

من اجل اختبار فرضية ان الدورة التكوينية ساهمت في تحسين اداء العمال من خلال مؤشر الانتاجية، نقوم باجراء اختبار تساوي متوسط الانتاجية قبل الدورة بنظيره بعد الدورة. وفي هذه الحالة يتم تعبئة البيانات الى البرنامج Stata كما هي مبينة في الجدول وليس بالطريقة المعمول بها في حالة العينتين المستقلتين. وعليه ففي هذه الحالة يكون بكتابة الأمر ttest متبوع بعبارة تَساوي المتغيرين أي كتابة الأمر التالي:

ttest avant=aprés

فتكون النتيجة في جدول Paired t.test التالي:

```
Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   avant |      15        10.6    .5839113    2.261479    9.347635    11.85237
   après |      15        12.6     .645866    2.501428    11.21476    13.98524
---------+--------------------------------------------------------------------
    diff |      15          -2    .3380617    1.309307   -2.72507    -1.27493
------------------------------------------------------------------------------
    mean(diff) = mean(avant - après)                              t =  -5.9161
Ho: mean(diff) = 0                               degrees of freedom =       14

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

بحيث يتم اختبار فرضية أن يكون متوسط الفرق بين العينتين يساوي إلى 0 باستعمال اختبار t ستودنت، فإذا كانت القيمة المحسوبة للاختبار أكبر من القيمة المجدولة بدرجة حرية n-1 يتم الاقرار

برفض فرضية العدم أي ان المتوسطين غير متساويين وأن الفرق بينهما أقل من الصفر بمعنى ان الدورة التكوينية حسنت من انتاجية العمال، وهذا ما يظهر من خلال Pr(T < t) = 0.0000 ويمكن الاقرار بذلك كذلك باستعمال قيمة الاحتمال إذا كانت القيمة المحسوبة أقل من مستوى المعنوية 0.05. ومن خلال النتائج المبينة في الجدول نلاحظ أن قيمة t المحسوبة تقدر بـ -5.91 وهي أكبر من القيمة المجدولة (بالقيمة المطلقة)، مما يقودنا الى رفض فرضية العدم القائلة ان متوسطي العينتين متساويين ونقر بوجود اختلاف معنوي بين المتوسطين وهذا الفرق سالب دليل على أن متوسط انتاجية العمل بعد الدورة أكبر منه قبل الدورة، أي ان الدورة التكوينية قد حسنت من انتاجية العمال في وحدة التركيب.

4. اختبار تساوي المتوسطات لأكثر من عينتين مستقلتين

مثلا اختبار تساوي متوسطات 3 عينات مستقلة، ففي هذه الحالة يكون الاختبار عبارة عن تحليل التباين بعامل واحد، بحيث يتم اختبار الفرضية القائلة أن متوسطات العينات متساوية.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = ... = \mu_k$$
$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq ... \neq \mu_k$$

نفرض انه لدينا ثلاث عينات من ثلاث شركات، حيث نضيف شركة لانتاج التجهيزات الكهرومنزلية للعينتين المدروستين في المثال السابق (العينتين المستقلتين).

ان اجراء الاختبار باستعمال البرنامج stata يكون بالأمر oneway متبوعة باسم المتغير ومتغير التجزئة (الشركات)، وفي هذا المثال نكتب:

oneway resultat secteur

فيعرض البرنامج نتائج الاختبار في جدول تحليل التباين كما يلي:

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	2834.65723	2	1417.32862	2.61	0.0787
Within groups	50459.5824	93	542.576154		
Total	53294.2396	95	560.991996		

Bartlett's test for equal variances: chi2(2) = 1.1976 Prob>chi2 = 0.549

بحيث يقوم البرنامج بتجزئة التباين الكلي للمتغير التابع الى جزئين تباين بين المجموعات والذي يرجع مصدره الى الفروقات في المتوسطات، وتباين ضمن المجموعات والذي لا يعرف مصدره الى البواقي، ويقارن بين تباين الجزئين باستعمال اختبار فيشر، فإذا كانت قيمة F المحسوبة أقل من القيمة

المجدول فعندئذ يمكن الاقرار بتساوي المتوسطات والعكس اذا كانت القيمة المحسوبة أكبر من تلك المجدولة عند مستوى 5%.

وكما تشير اليه نتائج الاختبار فإن القيمة المحسوبة لـ F هي 2.61 والقيمة المجدولة هي F(2,93)=3.11، وبالتالي يمكن الاقرار بتساوي متوسطات العينات. كما يمكن مقارنة قيمة الاحتمال بمستوى المعنية 0.05 نجد أن 0.05<0.07 وبالتالي نقبل H_0.

كما يعطي البرنامج القيمة الاحصائية قيمة اختبار تجانس التباين لـ Bartlett الذي يتبع توزيع chi2 بدرجة حرية 2 (k-1). وكلما كانت قيمة هذا الاختبار صغيرة كلما تحققت فرضية تساوي التباينات. وكانت القيمة المحسوبة لـ (2) Khi هي 1.19 وان القيمة المجدولة لـ (2) Khi بدرجة حرية (2) تساوي 5.99 عند مستوى معنوية 5% فيمكن الاقرار بتجانس تباين العينات.

- يمكن اظهار قيم مختلف المتوسطات والانحراف المعياري لكل عينة وإجمالي العينات باستخراج جدول خاص، وذلك بإضافة العبارة t الى الأمر **oneway** كما يلي:

oneway resultat secteur, t

بحيث يقدم البرنامج اختبار وجود فروقات في المتوسط بالإضافة الى جدول خاص بالمتوسطات والانحراف المعياري.

كما يقدم البرنامج اختبار وجود فروقات في المتوسط بين كل زوج من العينات في شكل مصفوفة مقارنة لبونفيروني Bonferroni والذي يعتمد على اختبار ستودنت للفرق بين متوسطي عينتين. بحيث نظيف bon الى الأمر السابق فنحصل على الجدول التالي:

Summary of resultat

secteur	Mean	Std. Dev.	Freq.
1	186.25	21.295161	32
2	194.55882	22.670487	34
3	199.6	25.891817	30
Total	193.36458	23.68527	96

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	2834.65723	2	1417.32862	2.61	0.0787
Within groups	50459.5824	93	542.576154		
Total	53294.2396	95	560.991996		

Bartlett's test for equal variances: chi2(2) = 1.1976 Prob>chi2 = 0.549

في هذه الحالة وبالرغم من عدم وجود اخلاف معنوي في متوسطات العينات فإننا قمنا بعرض جدول المقارنة من أجل التوضيح فقط، بحيث يظهر الجدول اختبار تساوي المتوسط لكل زوج من العينات، بعرض الفرق بين كل متوسطين وقيمة الاحتمال الخاص به. فيظهر من خلال الجدول أن قيمة الاحتمال في كل الحالات كان أكبر من 0.05 مما يدل على عدم وجود فرق في المتوسط بين كل زوج من العينات عند مستوى معنوية 5%.

مثال 2:

لدينا المعطيات الخاصة بنتائج امتحان مادة الاحصاء في ثلاث تخصصات مختلفة (علوم اقتصادية، تقنيات رياضة، تكنولوجيا)، كما يبينها الجدول التالي:

الجدول رقم (5-5): علامات عينة من الطلبة في مادة الاحصاء حسب الشعب

Seco	**10**	**12**	**15**	**20**	**18**	**17**	**19**	**14**	**10**
Sport	**15**	**10**	**8**	**14**	**3**	**18**	**10**	**14**	**7**
Tec	**18**	**12**	**8**	**5**	**10**	**16**	**11**	**4**	**10**

ونريد اختبار ما إذا كان متوسط النقاط المحصلة في المادة متساوي بين مختلف الاقسام.

يجب أولا اعادة بناء الجدول في شكل طولي بحيث يشمل متغيرين متغير خاص بالنقاط المحصلة وآخر خاص بالكلية التي سحبت منها العينة بحيث يتم ترقيم الكليات من 1 إلى 3، ثم اعطاء الاسم الخاص بكل كلية في عناوين القيم في شاشة Data Editor من قائمة خصائص المتغيرات لتظهر في المخرجات وتسهيل عملية التحليل، كما فعلنا سابقا في حالة العينتين المستقلتين.

لاجراء الاختبار نكتب الأمر:

oneway note fac

فتظهر نتائج الاختبار على النحو التالي:

```
. oneway note fac

                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      111.185185      2   55.5925926       2.92     0.0730
 Within groups      456.222222     24   19.0092593
------------------------------------------------------------------------
    Total           567.407407     26   21.8233618

Bartlett's test for equal variances:  chi2(2) =   0.3984  Prob>chi2 = 0.819
```

نتائج اختبار فيشر

يظهر من خلال النتائج أنه يمكن الاقرار بتساوي المتوسطات، لأن القيمة المحسوبة لإحصائية فيشر أقل من القيمة المجدولة عند مستوى معنوية 5% (3.40>2.92)، كما أن احصائية كاي 2

المحسوبة 0.39 أقل من القيمة المجدولة (5.99) مما يدل على قبول الفرضية H_0 أي قبول فرضية تساوي تباين العينات.

ان الفرضية البديلة تنص على وجود أحد المتوسطات يختلف عن باقي المتوسطات وليس بالضرورة اختلاف كل المتوسطات فيما بينها.

إلى جانب الاختبار الاحصائي المذكور، يمكن استخدام المقارنات البعدية لمقارنة متوسطات المتغير التابع لكل زوجين من الفئات على حدة، فإذا كان لدينا ثلاث فئات سيتم المقارنة بين كل زوج منها كما يلي: الأولى-الثانية؛ الثانية-الثالثة والأولى-الثالثة. وفي هذه الحالة يتم استخدام اختبار t للمقارنة بين كل زوج من المتوسطات وفق الطريقة السابقة باضافة العبارة t متبوعة باسم طريقة من بين الطرق الثلاثة التالية:

- bon: طريقة بونفيروني Bonferroni ونكتب الأمر : oneway note fac, t bon
- sch: طريقة شيفيه Scheffe ونكتب الأمر : oneway note fac, t sch
- sid: طريقة سيداك Sidak ونكتب الأمر : oneway note fac, t sid

فتطبيق احدى الطرق (طريقة بونفيروني مثل) يعرض البرنامج نتائج اختيار تساوي المتوسط بين كل زوج من العينات، وهنا تتم مقارنة متوسط علامات مقياس الاحصاء بين كل زوج من عينات الطلبة والمعبر عنها بالكلية، على النحو التالي:

```
. oneway note fac, t bon

                Summary of note
      fac |      Mean   Std. Dev.       Freq.
----------+--------------------------------------
Seconomiq |        15   3.7749172          9
    Sport |        11   4.6636895          9
Technolog |  10.444444   4.5856055          9
----------+--------------------------------------
    Total |  12.148148   4.6715481         27

                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      111.185185      2   55.5925926      2.92     0.0730
 Within groups      456.222222     24   19.0092593
------------------------------------------------------------------------
    Total           567.407407     26   21.8233618

Bartlett's test for equal variances:  chi2(2) =   0.3984  Prob>chi2 = 0.819

                Comparison of note by fac
                        (Bonferroni)
Row Mean-
Col Mean |   Seconomi      Sport
---------+--------------------
   Sport |         -4
         |      0.190
         |
Technolo |   -4.55556   -.555556
         |      0.109      1.000
```

جدول الاحصائيات الوصفية

جدول تحليل التباين

جدول المقارنات المتعددة

من خلال هذه النتائج لاحظ أن جدول المقارنات المتعددة يحتوي على الفرق بين المتوسط لكل زوج من العينات (الشعب)، بالاضافة الى القيمة الاحتمالية لاختبار ستودنت التي من خلالها يتم الاقرار

بوجود فروقات معنوية بين المتوسطات. فبما ان القيمة الاحتمالية للاختبار بين كل زوج أكبر من 0.05 نقبل فرضية العدم التي تقر بأن متوسط كل زوج من العينات متساوي، ولا توجد فروقات معنوية بين متوسطات علامات الطلبة في مقياس الاحصاء بين الشعب (العلوم الاقتصادية، التربية البدنية والتكنولوجيا).

5. تحليل التباين الثنائي

في كثير من الحالات يستدعي الأمر تحليل البيانات الكمية بدلالة متغيرات اسمية أو نوعية، فبالإضافة الي تحليل المتوسطات ما بين العينات، يمكن اجراء تحليل المتوسطات باستعمال متغيرات أخرى اسمية تتوزع الى فئتين أو أكثر. ويكون التحليل هنا بمثابة البحث عن أثر متغيرين مستقلين أو أكثر على متغير تابع واحد مع اشتراط ان يكون المتغير التابع متغير كمي والمتغيرات المستقلة اسمية. مثلا تأثير حرفة الأب وجنس الطالب على نجاح الطالب في البكالوريا، تأثير الجنس والشهادة المحصل عليها على الأجر المحصل وغيرها من الحالات.

- ان تحليل التباين الثنائي انما يستعمل لمعرفة ان كان هناك فروق معنوية بين المتوسطات. ومن خلال هذا التحليل يتم اختبار ثلاث فرضيات أساسية:

1. الأثر الرئيسي للمتغير العاملي الأول على المتغير التابع: $H_0 : \mu_{a1} = \mu_{a2} = \mu_{a3}$
2. الأثر الرئيسي للمتغير العاملي الثاني على المتغير التابع: $H_0 : \mu_{b1} = \mu_{b2} = \mu_{b3}$
3. أثر التفاعل بين المتغيرين العاملين $a*b$ على المتغير التابع: $H_0 : \sum(\alpha.\beta)^2 = 0$

ويشترط اجراء هذا التحليل أن تكون العينات مأخوذة من مجتمع يتبع التوزيع الطبيعي، وفي غير ذلك يمكن زيادة حجم العينات لتشمل أكثر من 15 حالة بحيث تكون نتائج التحليل أكثر دقة. وكذلك شرط تساوي تباين المتغير التابع لكل مجتمع من مجتمعات المعرفة في كل خلية من خلايا التجربة. بالإضافة الى شرط العشوائية أي ان العينات مختارة بطريقة عشوائية. لكن يمكن الاستغناء عن الشرطين الأول والثاني في الكثير من الحالات وعندها يتم اتباع طرق مختلفة في التحليل.

ويكتب الأمر هنا كما يلي:

anova depvar var1 var2 var1#var2

مثال:

لدينا البيانات الخاصة بنتائج امتحان البكالوريا لسنة 2006 لعينة من المترشحين المبينة في الجدول التالي:

الجدول رقم (5-6): نتائج امتحان البكالوريا لعينة من المترشحين

no	sexe	Sbac	moyen	no	sexe	Sbac	moyen	no	sexe	Sbac	moyen
1	Femme	LLE	13,875	31	Femme	SNV	12,375	61	homme	GE	10,125
2	homme	LLE	11,125	32	homme	SNV	10,375	62	homme	GE	14
3	Femme	LLE	12,375	33	homme	SNV	11	63	Femme	GE	12,25
4	Femme	LLE	12,25	34	homme	SNV	15,75	64	Femme	GE	9,125
5	Femme	LLE	11	35	Femme	SNV	13	65	Femme	GE	11,5
6	Femme	LLE	10	36	homme	SNV	11,375	66	homme	GE	9,125
7	Femme	LLE	9,875	37	homme	SNV	11,25	67	Femme	GE	13,25
8	Femme	LLE	12	38	Femme	SNV	10,375	68	Femme	GE	12,125
9	Femme	LLE	9,75	39	homme	SNV	9,375	69	homme	GE	11,375
10	homme	LLE	9,75	40	Femme	SNV	10,375	70	homme	GE	12,375
11	Femme	LLE	10	41	homme	SNV	16,375	71	Femme	GE	10,125
12	homme	LLE	13,625	42	Femme	SNV	15,375	72	homme	GE	10,25
13	homme	LLE	10,125	43	homme	SNV	10,75	73	homme	GE	10,875
14	homme	LLE	11,25	44	homme	SNV	11,375	74	homme	GE	9,625
15	Femme	LLE	8,75	45	homme	SNV	11,75	75	homme	GE	11,125
16	homme	LLE	13,75	46	homme	SNV	10,5	76	Femme	GE	12,25
17	Femme	LLE	13	47	Femme	SNV	13,25	77	Femme	GE	10,125
18	Femme	LLE	10,75	48	homme	SNV	12,625	78	homme	GE	10
19	Femme	LLE	9,875	49	Femme	SNV	10,75	79	Femme	GE	12,125
20	Femme	LLE	13,875	50	homme	SNV	14,875	80	Femme	GE	10,5
21	Femme	LLE	9,75	51	Femme	SNV	9,5	81	Femme	GE	8,875
22	Femme	LLE	10	52	homme	SNV	9,625	82	homme	GE	11,75
23	Femme	LLE	11,75	53	homme	SNV	11,125	83	Femme	GE	10,375
24	Femme	LLE	11,125	54	homme	SNV	11,25	84	Femme	GE	10,875
25	Femme	LLE	10	55	homme	SNV	10,75	85	homme	GE	10
26	Femme	LLE	9,875	56	homme	SNV	11,5	86	Femme	GE	9,25
27	Femme	LLE	14,375	57	Femme	SNV	12,125	87	homme	GE	10,125
28	Femme	LLE	11,25	58	Femme	SNV	13,375	88	Femme	GE	10
29	homme	LLE	12,5	59	homme	SNV	12,125	89	Femme	GE	11,125
30	Femme	LLE	10,25	60	homme	SNV	13,375	90	homme	GE	10,125

يحتوي الجدول على المتغيرات التالية: الجنس (Sexe)، شعبة البكالوريا (Sbac) ويحتوي على 3 شعب: تسيير واقتصاد ، العلوم الطبيعية والحياة ، اداب ولغات اجنبية، ومعدل البكالوريا (Moyen). ويتم تعبئة هذه البيانات باعادة ترميزها باعطاء ارقام لفئات المتغيرات الاسمية عوض عناوينها.

المطلوب:

اختبار وجود فروقات معنوية بين معدلات البكالوريا حسب متغير الجنس وشعبة البكالوريا.

لاجراء الاختبار نكتب الأمر التالي:**anova moyen sexe1 sbac sexe1#sbac**

فيظهر البرنامج نتائج التحليل في الجدول التالي:

```
      Number of obs =        90     R-squared     =  0.0886
      Root MSE      =   1.58783     Adj R-squared =  0.0343
```

Source	Partial SS	df	MS	F	Prob>F
Model	20.581728	5	4.1163456	1.63	0.1603
sexe1	.22936807	1	.22936807	0.09	0.7637
sbac	18.102213	2	9.0511065	3.59	0.0319
sexe1#sbac	2.2356639	2	1.117832	0.44	0.6434
Residual	211.78216	84	2.5212162		
Total	232.36389	89	2.6108302		

يتضح من خلال نتائج تحليل التباين الثنائي، انه يوجد تأثر معنوي لمتغير شعبة البكالوريا لأن القيمة الاحتمالية لاحصائية فيشر أقل من 0.05 وهذا يدل على وجود فروقات معنوية في معدلات البكالوريا حسب متغير الشعبة، بينما لايظهر التأثير المعنوي لمتغيرة الجنس وكذلك التفاعل بينه ومتغير الشعبة وانما تأثير هذين الأخيرين ضعيف لأن القيمة الاحتمالية لاحصائية فيشر اكبر بكثير من 0.05.

ويمكن الاستعانة بنماذج الانحدار المتعدد لتفسير تأثير المتغيرت المدرجة على معدل البكالوريا بتقدير معاملات النموذج لكن الاشكال يبقى مطروح في تفسير معاملات المتغيرات الاساسية في حضور تأثير التفاعلات. ويتم تقدير النموذج باستعمال الأمر:

reg moyen i.sexe1 i.sbac i.sexe1#i.sbac

بحيث يستخدم الرمز i. لاستعمال المتغيرات الثنائية (0.1) لمتغير الجنس والشعبة دون تكوين مصفوفة الوحدة، وتكون نتائج تقدير النموذج على الشكل التالي:

```
. reg moyen i.sexe1 i.sbac i.sexe1#i.sbac
```

Source	SS	df	MS		
Model	20.5817278	5	4.11634557	Number of obs	= 90
Residual	211.782161	84	2.5212162	F(5, 84)	= 1.63
				Prob > F	= 0.1603
Total	232.363889	89	2.61083021	R-squared	= 0.0886
				Adj R-squared	= 0.0343
				Root MSE	= 1.5878

moyen	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sexe1						
Home	-.0904018	.5810877	-0.16	0.877	-1.245958	1.065155
sbac						
SNV	.2523777	.5169082	0.49	0.627	-.775551	1.280306
LLE	1.182813	.6400763	1.85	0.068	-.0900494	2.455674
sexe1#sbac						
Home#SNV	.7029794	.8985848	0.78	0.436	-1.083955	2.489914
Home#LLE	-.1033482	.8460764	-0.12	0.903	-1.785864	1.579168
_cons	10.86719	.3969585	27.38	0.000	10.07779	11.65658

ويتم تفسير المعاملات المقدرة للنموذج بدرجة تأثر فئات المتغير بالنسبة لفئة الأساس[1] التي يعتمدها البرنامج بالترتيب الابجدي في حالة الرميز الاسمي أو الترتيب التصاعدي في حالة الترميز الرقمي.

[1] للمزيد من التفاصيل أنظر في ذلك:

MAXWELLE Scott E., Harold D. DELANEY. (2004), «Designing Experiments and analysis Data, a model comparison perspective », second édition, TLFeBook, édition Mahwah, New jersey, London.

الفصل السادس: تحليل الارتباط والانحدار باستخدام برنامج Stata

في غالب الاحيان يلجأ الباحث الى دراسة وتحليل مجموعة من المتغيرات الاحصائية في آن واحد، كما انه في بعض الحالات نجد ان بعض المتغيرات لها علاقة بمتغيرات اخرى بحيث لا يمكن تحليل المتغير الواحد دون ربطه بتحليل متغير آخر. ويمكن التمييز بين نوعين من التحليل:

1. **تحليل الارتباط**: وهو يقيس درجة الارتباط الخطي بين متغيرين ودرجتها حيث يأخذ معامل الارتباط قيمه في المجال (1 و -1) للتعبير عن الدرجة والاتجاه، ويميز بين عدة معاملات للارتباط ومن أهمها:
   - معامل الارتباط لبيرسون (Pearson): يقيس العلاقة بين متغيرين أحدهما تابع والآخر مستقل بحيث يقيس التغير الذي يطرأ على المتغير التابع نتيجة تغير المتغير المستقل، و يستخدم في حالة العينات الكبيرة والبيانات الكمية.
   - معامل الارتباط لسبيرمان (Spearman): وهو يقيس مقدار قوة الارتباط بين متغيرين، على شكل بيانات وصفية (اسمية) وعينات صغيرة يمكن وضعها في شكل ترتيبي. مثل درجة الأهمية: 1. جيدة، 2. متوسطة، 3. ضعيفة، وغيرها.
2. **تحليل الانحدار:** ويقيس العلاقة الخطية بين متغيرين أو أكثر في شكل علاقة خطية يكون فيها أحد المتغيرات تابع والمتغيرات الأخرى مستقلة، بالإضافة الى قياس درجة المعنوية الاحصائية للمعلمات المقدرة، وكذا قياس القدرة التفسيرية للنموذج باستخدام معامل التحديد وهو عبارة عن مربع معامل الارتباط.

1. تحليل الارتباط باستخدام STATA

يتم تحليل الارتباط باستخدام Stata عن طريق الأوامر التالية:

-correlate أو pwcorr: لاستخراج مصفوفة الارتباط التي تحتوي على معاملات الارتباط الخطي البسيط بين كل زوج من المتغيرات المختارة.

-pcorr لاستخراج مصفوفة الارتباط الجزئي بين مختلف المتغيرات وهو يستخدم في حالة التحليل لاكثر من متغيرين.

- sig star(.05): لتحديد مستوى معنوية اختبار الفرضية: H0: معامل الارتباط=0 ضد الفرضية البديلة H1: معامل الارتباط يختلف عن الصفر.

لتحليل الارتباط والانحدار باستخدام متغيرين كميين، نستعين بالمثال التالي:

لدينا المعطيات الخاصة بمتوسط المبيعات الشهرية لأحد الجرائد اليومية في 12 مدينة وعدد السكان في كل مدينة المبينة في الجدول التالي:

ville	1	2	3	4	5	6	7	8	9	10	11	12
population	119564	52900	100702	62339	167264	53775	77042	38981	59714	74240	76786	38714
ventes	1772	674	2572	1184	3492	814	1772	660	864	1884	2026	452

ولتحليل الارتباط والانحدار للعلاقة بين المبيعات وعدد السكان نتبع الخطوات التالية:

كتابة الأمر "correlate" متبوع بأسماء المتغيرات لاستخراج مصفوفة الارتباط بين المتغيرين فيظهر البرنامج النتائج كما يلي:

	ventes	popula~n
ventes	1.0000	
population	0.9134	1.0000

بحيث يعرض البرنامج قيمة معامل الارتباط لبيرسون للعلاقة بين المتغيرين: المبيعات وعدد السكان والذي يقدر بـ 0.9134 وهو يدل على قوة الارتباط بين المتغيرين، وأي تغير في عدد السكان سوف يكون له تأثر ايجابي على المبيعات.

ولتحليل الارتباط بين ثلاث متغيرات (ventes، population وpvente) بحيث pvente تدل على عدد نقاط البيع في كل مدينة، نستعين بالأمرين pwcorr و pcorr لاستخراج مصفوفة الارتباط والارتباط الجزئي وشبه الجزئي semipartial. وبتطبيق الأمرين تظهر النتائج التالية:

	ventes	popula~n	pvente
ventes	1.0000		
population	0.9134	1.0000	
pvente	0.6165	0.8016	1.0000

```
. pcorr ventes population pvente
(obs=12)

Partial and semipartial correlations of ventes with
```

Variable	Partial Corr.	Semipartial Corr.	Partial Corr.^2	Semipartial Corr.^2	Significance Value
population	0.8906	0.7012	0.7931	0.4917	0.0002
pvente	-0.4754	-0.1935	0.2260	0.0374	0.1395

ملاحظة: تعتبر معاملات الارتباط شبه الجزئية مهمة في تقييم الاهمية النسبية للمتغيرات في التنبؤ بالمتغير التابع، والاختلاف الاساسي بين معاملات الارتباط الجزئية والارتباط شبه الجزئية، هو أنه في الحالة الاولى نستبعد أثر المتغير المستقل عن المتغيرين التابع والمستقل، بينما في حالة معامل الارتباط شبه الجزئي يتم استبعاد أثر احد المتغيرين المستقلين عن متغير آخر فقط.

كما يمكن استخراج قيمة الاحتمال أو الدلالة الاحصائية للمعامل المحسوب، وذلك بإضافة العبارة: star sig (مستوى الدلالة المرغوب).

مثلا: كتابة الأمر: **pwcorr ventes population pvente, sig star(5)** في بيانات هذا المثال فينتج الجدول التالي:

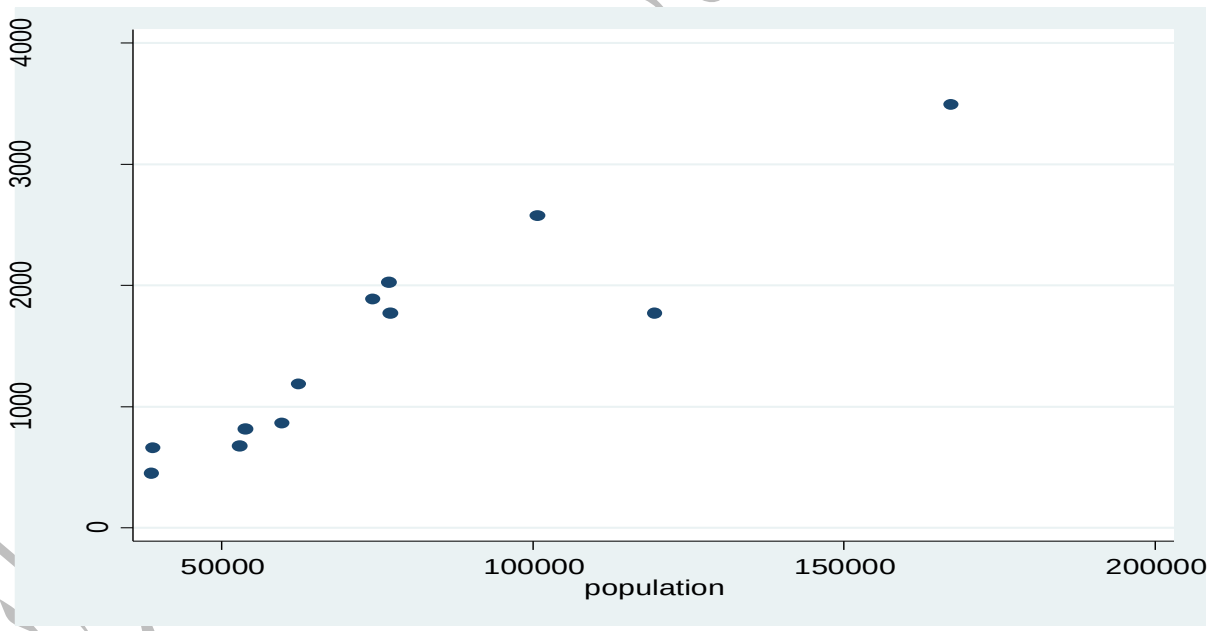
```
. pwcorr ventes population pvente, sig star(5)
```

	ventes	popula~n	pventes
ventes	1.0000		
population	0.9134*	1.0000	
	0.0000		
pventes	0.6165*	0.8016*	1.0000
	0.0328	0.0017	

كما يمكن تحليل الارتباط بين المتغيرات باستعمال التمثيل البياني في شكل انتشار وهو أحسن تمثيل في مثل هكذا حالات، بحيث يتم التأكد من وجود علاقة خطية من عدمه من خلال الشكل البياني، ففي كثير من الأحيان وبالرغم من ظهور معامل الارتباط كبير نسبيا ومعنوي الا ان الشكل البياني يظهر علاقة غير خطية بين المتغيرات. ويتم استخراج شكل الانتشار باستعمال الأمر:

scatter ventes population

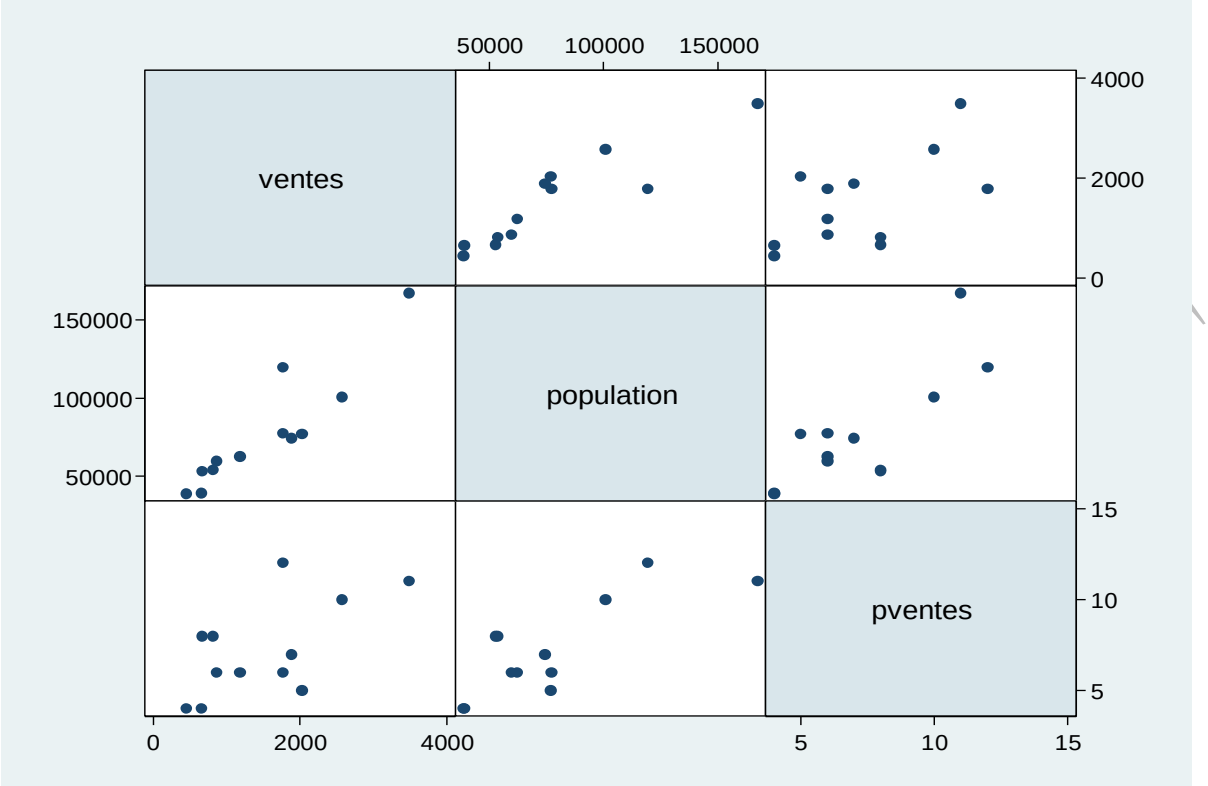
حيث يظهر التمثيل البياني للعلاقة بين المبيعات وعدد السكان. كما هو مبين في الشكل التالي:



كما يمكن استعمال الأمر graph matrix لاستخراج تمثيل بياني للعلاقة بين مختلف المتغيرات بكتابة الأمر:

graph matrix population ventes pvente

فيظهر البرنامج المخطط البياني لمصفوفة الانتشار على الشكل التالي:



معامل الارتباط لسبيرمان:

يستخدم هذا المعامل في حالة المتغيرات الترتيبية، يمكن حساب معامل الارتباط لسبيرمان حسب حاجة الباحث باستعمال الأمر التالي:

spearman ventes population, stats(rho p) star(.05)

بحيث يتم تحديد القيم الاحصائية المراد استخراجها (معامل الارتباط، وقيمة الاحتمال) ومستوى المعنوية للاختبار. فتكون المخرجات على الشكل التالي:

```
. spearman ventes population

 Number of obs =      12
Spearman's rho =      0.9177

Test of Ho: ventes and population are independent
    Prob > |t| =      0.0000
```

معامل الارتباط لكندال:

كما يمكن استخراج معامل كاندال من النوع (أ) و (ب) في حالة المتغيرات الترتيبية، وذلك باستخدام الامر ktau متبوع باسماء المتغيرات، بحيث يتم حساب قيم تو (أ) وتو (ب) وهي عبارة عن نسبة الارتباط بين المتغيرين. وفي مثالنا بكتابة الأمر:

ktau ventes population

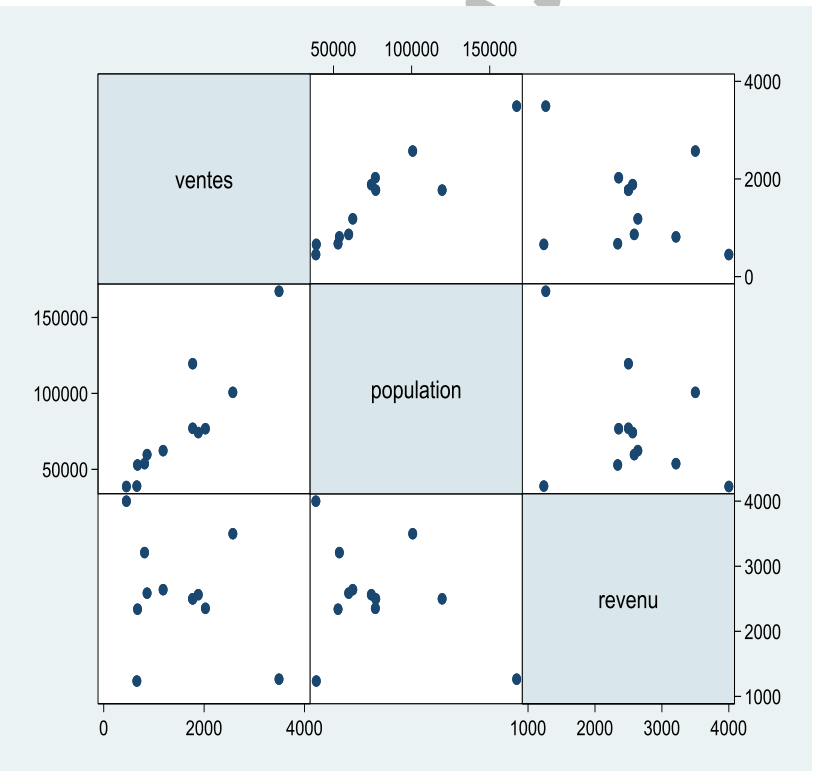
ينتج لدينا:

```
. ktau ventes population

  Number of obs =      12
Kendall's tau-a =      0.8333
Kendall's tau-b =      0.8397
Kendall's score =      55
    SE of score =      14.549    (corrected for ties)

Test of Ho: ventes and population are independent
     Prob > |z| =      0.0002   (continuity corrected)
```

بالمقارنة بين مختلف معاملات الارتباط المحسوبة، نلاحظ أنه بالرغم من وجود اختلاف طفيف في القيم المحسوبة، إلا أنها مرتفعة قريبة من الواحد الصحيح وتدل على وجود ارتباط موجب وقوي بين حجم المبيعات وعدد السكان، الأمر الذي يقود الى امكانية استخدام النموذج الخطي لتحليل العلاقة بين المتغيرين. بينما نشير الى عدم وجود علاقة خطية (علاقة ضعيفة) بين حجم المبيعات ومتوسط الدخل الفردي بحيث أن معاملات الارتباط كانت ضعيفة وغير معنوية، وهذا ما يؤكده عدم ظهور علاقة خطية من خلال مصفوفة الانتشار المبينة أدناه:

```
. ktau ventes revenu

  Number of obs =      12
Kendall's tau-a =      -0.1364
Kendall's tau-b =      -0.1385
Kendall's score =      -9
    SE of score =      14.515   (corrected for ties)

Test of Ho: ventes and revenu are independent
     Prob > |z| =      0.5815   (continuity corrected)
```

```
. pwcorr ventes revenu, sig star(.05)

             |   ventes   revenu
-------------+------------------
      ventes |   1.0000
             |
             |
      revenu |  -0.3036   1.0000
             |   0.3373
```




```
. spearman ventes revenu

 Number of obs =      12
Spearman's rho =      -0.1368

Test of Ho: ventes and revenu are independent
    Prob > |t| =      0.6715
```

2. تحليل الانحدار

2-1. تحليل الانحدار البسيط

لتقدير العلاقة بين حجم المبيعات بدلالة عدد السكان يكفي كتابة الأمر reg متبوع باسم المتغير التابع والمتغير المستقل على الترتيب أي:

reg ventes population

والذي يعبر عن انحدار المبيعات على عدد السكان، فيقوم البرنامج بتقدير العلاقة ويظهر النتائج كما يبينها الجدول التالي:

```
. reg ventes population
```

Source	SS	df	MS	Number of obs	=	12
				F(1, 10)	=	50.35
Model	7632238.49	1	7632238.49	Prob > F	=	0.0000
Residual	1515941.18	10	151594.118	R-squared	=	0.8343
				Adj R-squared	=	0.8177
Total	9148179.67	11	831652.697	Root MSE	=	389.35

ventes	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
population	.0225018	.0031713	7.10	0.000	.0154358	.0295678
_cons	-215.0934	268.3377	-0.80	0.441	-812.987	382.8002

ومن خلال هذه النتائج يمكن كتابة النموذج المقدر على الشكل التالي:

$$ventes = -215.09 + 0.0225 * population$$
$$(-0.80)\quad (7.1)$$
$$R^2 = 0.8343, \overline{R}^2 = 0.8177, F_{(1,10)} = 50.35$$

ويمكن اجراء اختبارات فرضيات النموذج بالاستعانة بالأوامر المتوفرة في البرنامج، نذكر منها مايلي:

1. اختبار وجود ارتباط ذاتي للاخطاء:

يتم اجراء هذا الاختبار بعد عملية التقدير مباشرة، مع ضرورة تعريف المتغير الخاص بالزمن أو بالافراد، وهذا بكتابة الأمر tsset ville بالنسبة لمثالنا وغالبا ما يكون متغير خاص بالزمن في حالة استخدام بيانات السلاسل الزمنية (t). وبعد ذلك نكتب الامر estat dwatson لحساب قيمة داربين واتسون للارتباط الذاتي للاخطاء من الدرجة الأولى، وتكون النتائج كما يلي:

```
. tsset ville
        time variable:  ville, 1 to 12
                delta:  1 unit

. estat dwatson

Durbin-Watson d-statistic(  2,    12) =  1.620586
```

2. اختبار تجانس التباين arch-test:

بالنسبة لهذا الأختبار، نلجأ الى اختبار مضاعف لاغراج ARCHLM-test، وذلك بكتابة الأمر:

estat archlm, lags(p) بحيث p هي عدد التأخيرات (Lag) المختارة من طرف المستخدم.

وفي مثالنا نختار الدرجة 2، فنكتب: estat archlm, lags(2)

فنتحصل على النتائج التالية:

```
. estat archlm, lags(2)
LM test for autoregressive conditional heteroskedasticity (ARCH)
```

lags(p)	chi2	df	Prob > chi2
2	1.263	2	0.5317

H0: no ARCH effects *vs.* H1: ARCH(p) disturbance

القيمة المحسوبة للإختبار هي قيمة R^2 مضروبة في عدد المشاهدات تقارن بقيمة كاي تربيع بدرجتي حرية.

2-2. الانحدار المتعدد:

لا تختلف كثيرا طريقة التعامل مع النموذج المتعدد عن طريقة تقدير النموذج البسيط الا في عدد المتغيرات المدرجة في النموذج، وبعض الاختبارات الاحصائية الخاصة بالتشخيص. ولتوضيح طريقة تقدير نماذج الانحدار المتعدد والاختبارات الاحصائية الملحقة بها نستعين بالمثال التالي:

لتكن لدينا البيانات التالية:

année	Y	X2	X3	X4	X5	X6
2000	10227	112	121,3	776,8	4,89	79367
2001	10872	111	125,3	839,6	4,55	82153
2002	11350	111,1	133,1	949,8	7,38	85064
2003	8775	117,5	147,7	1038,4	8,61	86794
2004	8539	127,6	161,2	1142,8	6,16	85846
2005	9994	135,7	170,5	1252,6	5,22	88752
2006	11046	142,9	181,5	1379,3	5,5	92017
2007	11164	153,8	195,3	1551,2	7,78	96048
2008	10559	166	217,7	1729,3	10,25	98824
2009	8979	179,3	247	1918	11,28	99303
2010	8535	190,2	272,3	2127,6	13,73	100397
2011	7980	197,6	286,6	2261,4	11,2	99526
2012	9179	202,6	297,4	2428,1	8,69	100834
2013	10394	208,5	307,6	2670,6	9,65	105005
2014	11039	215,2	318,5	2841,1	7,75	107150
2015	11450	224,4	323,4	3022,1	6,31	109597

المصدر: Damodar N. Gujarati , Op Cit, p 384

والنموذج المراد تقديره هو:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \beta_5 x_{5t} + \beta_6 x_{6t} + \varepsilon_t$$

بحيث تشير رموز المتغيرات في النموذج الى اللوغاريتم النيبير للمتغيرات الموجودة في الجدول، بما ان البرنامج stata يمسز بين الرموز الصغيرة والرممور الكبيرة للمتغيرات، أي اننا نقوم بتحويل المتغيرات باستخدام العلاقة: var=log(VAR) بالنسبة لكل المتغيرات.

ويكون تقدير النموذج بكتابة الامر: reg y x2 x3 x4 x5 x6 فنتحصل على النتائج التالية:

```
. reg y x2 x3 x4 x5 x6
```

Source	SS	df	MS		
				Number of obs	= 16
				F(5, 10)	= 11.77
Model	.183346079	5	.036669216	Prob > F	= 0.0006
Residual	.031143304	10	.00311433	R-squared	= 0.8548
				Adj R-squared	= 0.7822
Total	.214489383	15	.014299292	Root MSE	= .05581

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x2	1.790182	.8732524	2.05	0.068	-.1555459	3.735909
x3	-4.108581	1.599697	-2.57	0.028	-7.672927	-.5442344
x4	2.127249	1.257848	1.69	0.122	-.675411	4.929909
x5	-.030443	.1218488	-0.25	0.808	-.301939	.241053
x6	.2777009	2.036988	0.14	0.894	-4.260991	4.816393
_cons	3.255715	19.11667	0.17	0.868	-39.33889	45.85032

بالاضافة الى الأوامر الخاصة بالاختبارات في نموذج الانحدار البسيط، يمكن اجراء بعض الاختبارات الخاصة بنموذج الانحدارالمتعدد من خلال الأوامر التالية:

الأمر estat hettest: لاختبار فرضية تجانس التباين، وهنا يمكن اجراء الأمر في حالتين:

1- كتابة الأمر: estat hettest x2 x3 x4 x5 X6, iid، لحساب قيمة $N*R^2$ التي تقارن بقيمة كاي تربيع بدرجة حرية (5). بحيث نختبر فرضية العدم القائلة أنه لا يوجد اختلاف في التباين. والتي يتم قبولها في حال القيم المحسوبة اقل من القيمة المجدولة أو اذا كانت p-value>0.05. وفي مثالنا نحصل على النتائج التالية:

```
. estat hettest x2 x3 x4 x5 x6, iid

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: x2 x3 x4 x5 x6

         chi2(5)      =    10.09
         Prob > chi2  =   0.0726
```

لاحظ أن القيمة الاحتمالية للاختبار أكبر من 0.05 وبالتالي يمكن القبول بالفرضية H0 التي تقر بعدم وجود اختلاف في تباينات الاخطاء.

2- كتابة الأمر estat hettest x2 x3 x4 x5 X6, fstat، لحساب قيم احصائية فيشر:

$$F = \frac{(TSS - RSS)/(K-1)}{RSS/(N-K-)} \rightarrow F_{(K-1)(N-K-1)}$$

لانحدار تربيع بواقي التقدير على المتغيرات المفسرة، بحيث نختبر فرضية تساوي معاملات النموذج الى الصفر، بنفس طريقة اختبار معنوية المعلمات المقدرة في نموذج الانحدار.

ونتائج الاختبار تكون على الشكل التالي:

```
. estat hettest x2 x3 x4 x5 x6, fstat

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: x2 x3 x4 x5 x6

         F(5 , 10)     =     3.42
         Prob > F      =   0.0464
```

لاحظ أن القيمة الاحتمالية للاختبار أقل من 0.05 مما يقودنا الى قبول الفرضية H1 التي تقر باختلاف المعلمات المقدرة في النموذج عن الصفر، وبالتالي لا يوجد اختلاف في تباينات الاخطاء.

- الامر estat vif: لحساب معامل تضخ التباين، لاختبار وجود تعدد خطي بين المتغيرات المفسرة.

```
. estat vif
```

Variable	VIF	1/VIF
x4	1525.70	0.000655
x3	1524.37	0.000656
x2	243.46	0.004108
x6	194.98	0.005129
x5	7.72	0.129574
Mean VIF	699.25	

- الامر estat vce, correlation: لاستخراج هذه المصفوفة يكفي كتابة الأمر التالي بعد عملية التقدير.

```
. estat vce, correlation

Correlation matrix of coefficients of regress model
```

e(V)	x2	x3	x4	x5	x6	_cons
x2	1.0000					
x3	-0.8149	1.0000				
x4	0.6051	-0.9454	1.0000			
x5	0.6989	-0.8916	0.8601	1.0000		
x6	-0.6225	0.8791	-0.9467	-0.8257	1.0000	
_cons	0.5897	-0.8607	0.9413	0.8139	-0.9990	1.0000

- الأمر estat vce: لاستخراج مصفوفة التباين والتباين المشترك للمعلمات المقدرة.

```
. estat vce

Covariance matrix of coefficients of regress model
```

e(V)	x2	x3	x4	x5	x6	_cons
x2	.76256977					
x3	-1.1383026	2.5590297				
x4	.66462522	-1.9022366	1.5821816			
x5	.07436376	-.17378874	.13182221	.01484712		
x6	-1.107226	2.8645984	-2.4256742	-.20493911	4.1493202	
_cons	9.8442707	-26.322466	22.633345	1.8959482	-38.902265	365.44719

- الأمر stepwise, pr(.05):reg y x2 x3 x4 x5 x6: لاجراء الانحدار المتدرج واختيار احسن نموذج تكون فيه كل المعلمات المقدرة معنوية عند مستوى 5%.

```
. stepwise, pr(.05):reg y x2 x3 x4 x5 x6
                    begin with full model
p = 0.8943 >= 0.0500  removing x6
p = 0.8034 >= 0.0500  removing x5
```

Source	SS	df	MS	Number of obs	=	16
				F(3, 12)	=	23.34
Model	.183103778	3	.061034593	Prob > F	=	0.0000
Residual	.031385604	12	.002615467	R-squared	=	0.8537
				Adj R-squared	=	0.8171
Total	.214489383	15	.014299292	Root MSE	=	.05114

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x2	1.933943	.568746	3.40	0.005	.6947523	3.173134
x3	-4.414658	.5502504	-8.02	0.000	-5.61355	-3.215765
x4	2.332129	.3349456	6.96	0.000	1.602345	3.061913
_cons	5.769139	.697152	8.28	0.000	4.250175	7.288103

لاحظ ان النموذج المقدر في هذه المرحلة يحتوي على ثلاث متغيرات مفسرة، بحيث تم الاستغناء عن المتغيرة x5 و x6 نظرا لعدم معنويتها في النموذج. وهذه الطريقة تفيدنا في ايجاد احسن نموذج ممكن وبمعلمات معنوية احصائيا.

- الأمر estat ic: يسمح بحساب معيار أكايك الخاص بالنموذج المقدر والذي من خلاله يمكن المفاضلة بين النماذج. فمثلا بالنسبة للنموذج الاول يكون معيار اكايك كما يلي:

```
. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	16	11.79365	27.23094	6	-42.46188	-37.82635

Note: N=Obs used in calculating BIC; see [R] BIC note.

بينما في النموذج الثاني يكون:

```
. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	16	11.79365	27.16894	4	-46.33788	-43.24753

Note: N=Obs used in calculating BIC; see [R] BIC note.

لاحظ أن معيار Aic في هذا النموذج أقل من المعيار المحسوب النموذج الأول (-46.34<-42.46)، مما يقودنا الى تفضيل النموذج الثاني.

المتغيرات الصورية:

في بعض الأحيان يلجأ الباحث في اطار تقدير نماذج القياس الاقتصادي الى استعمال المتغيرات الصورية وهي عبارة عن متغيرات وهمية تأخذ القيمة 1 للتعبير عن حدوث الحدث و 0 لعدمه. وهذا النوع من المتغيرات كثيرة الاستعمال في شتى المجالات فمثلا لقياس الأجر يتم اشراك متغير deplome للتعبير الشهادة مثلا: الليسانس والماستر. أو استعماله لتحديد فترة زمنية معينة مثلا: قبل الأزمة وبعد الأزمة وغيرها.

ويسمح برنامج Stata بانشاء هذا النوع من المتغيرات بطريقة تلقائية باستعمال عدة أوامر حسب الحالة:

- لانشاء متغير يعبر عن قيمتين مختلفين أو عدة قيم مثلا فترات زمنية معينة

gen dummy=(x==value1 | x==value2)

مثلا فترات الأزمة معبر عنها بأرقام (1،2....) أو فترات الجفاف، أو فترات زمنية مثلا فصول السنة.

- لانشاء متغير صوري للتعبير عن قيم معينة للمتغير x: نكتب احدى العبارات التالية حسب الحاجة:

— **gen dummy1=(x=>value1)**

— **gen dummy2=(x<value1 & x=>value2)**

— **gen dummy3=(x<value2 & x. =!)**

بحيث يتم انشاء 3 متغيرات صورية تسمح بتقسيم متغير x الى ثلاث اجزاء حسب رغبة الباحث وتحديده لقيم المتغيرx، والعبارة الأخيرة دليل على احتساب القيم المفقودة ضمن المتغير d3.

- لانشاء مجموعة من المتغيرات الصورية للتعبير عن الأفراد، المؤسسات، القطاعات، البلدان وغيرها والتي تكون في شكل أرقام (code)، مثلما هو الحال في معطيات البانل، نستعين بالأمر :

tab countrycode ,gen(country)

وفي هذه الحالة يقوم البرنامج بانشاء عدة متغيرات في آن واحد حسب عدد الأفراد.

في حالة ما إذا كنا بصدد تقدير علاقة انحدارية لمتغير تابع ومتغيرات مستقلة بالإضافة الى متغيرات صورية خاصة بالأفراد (بشرط ان يكون معبر عنهم بأرقام code) يسمح الأمر xi بهذا الإجراء دون اللجوء الى انشاء متغيرات صورية بحيث يكفي كتابة الأمر التالي:

xi: reg y x1 x2 i.codp

بحيث يكون المتغير codp هو المتغير الخاص بترميز الأفراد .

الفصل السابع: مدخل لتحليل بيانات البانل باستخدام برنامج Stata

يقصد بمعطيات البانل (الطولية) تلك البيانات التي تمزج بين البيانات المقطعية وبيانات السلاسل الزمنية، أي مشاهدة ظاهرة معينة لمجموعة من الأفراد عبر الزمن، وهي بذلك تأخذ في الاعتبار اثر التغير في الزمن وكذلك اثر التغير في المشاهدات المقطعية (الأفراد).

ويمكن التمييز بين نوعين من المعطيات الطولية حسب المعطيات المتوفرة حول عينة الدراسة أي الافراد والفترة الزمنية: المعطيات الطولية المتوازنة التي يكون فيها عدد المشاهدات متساوي لجميع الأفراد (القطاعات) أي نفس عدد المشاهدات لكل فرد، و المعطيات الطولية غير المتوازنة، بحيث يكون فيها عدد المشاهدات يختلف من فرد لآخر.

1. الشكل العام للنموذج:

يأخذ النموذج العام لمعطيات البانل في شكله الخطي الشكل التالي:

$$Y_{it} = X_{it}.B + u_{it}$$

بحيث:

i=1,2,....,N : يشير الى الأفراد (الوحدات، أو القطاعات)، و t=1,2,…,T تشير الى الفترة الزمنية.

Y_{it}: تمثل المتغير التابع ذو البعد الفردي والزمني (i,t)، وX_{it} مصفوفة المتغيرات المفسرة عددها k لكل فرد في الفترة t. u_{it} حد الخطأ العشوائي، B شعاع معلمات مجهولة للتقدير تتعلق بكل فرد i في الفترة الزمنية t وفي الحالة العامة يتم اعتبار المعاملات مختلفة من فرد لآخر ومن فترة زمنية لأخرى. ويمكن التعبير على النموذج بالصيغة التالية:

$$y_{it} = \sum_{k=1}^{K} \beta_{kit} x_{kit} + \varepsilon_{it}$$

- ان النموذج وفق هذه العلاقة يعتبر نموذجا شاملا لكل السلوكات الممكنة للأفراد عبر الزمن، وعليه يتم وضع قيود على النموذج، ويمكن التمييز بين 5 عائلات من النماذج نذكرها فيما يلي:

أ. كل معاملات النموذج مختلفة من أجل كل فرد عبر الزمن، ويكتب على الشكل:

$$y_{it} = \sum_{k=1}^{K} \beta_{kit} x_{kit} + \varepsilon_{it}$$

ب. معاملات المتغيرات المفسرة ثابتة والحد الثابت يختلف من فرد الى آخر، وفيه لا يوجد تأثير الزمن لكن يوجد تأثير الافراد يظهر في الحد الثابت بحيث يكون لكل فرد حد ثابت خاص به. ويكتب على الشكل:

$$y_{it} = \beta_{1i} + \sum_{k=2}^{K} \beta_k x_{kit} + \varepsilon_{it}$$

ج. معاملات المتغيرات المفسرة ثابتة والحد الثابت يختلف من فرد الى آخر ومن لحظة زمنية الى أخرى (يتغير حسب الافراد وحسب الزمن). ويكتب على الشكل:

$$y_{it} = \beta_{1it} + \sum_{k=2}^{K} \beta_k x_{kit} + \varepsilon_{it}$$

د. الحد الثابت لا يختلف من فرد الى آخر لكن معاملات المتغيرات المستقلة تختلف حسب الأفراد:

$$y_{it} = \beta_1 + \sum_{k=2}^{K} \beta_{ki} x_{kit} + \varepsilon_{it}$$

هـ. كل معاملات النموذج تتغير حسب الافراد وحسب الزمن أي أن كل المعلمات تختلف من فرد الى آخر ومن لحظة زمنية الى اخرى. ويكتب على الشكل:

$$y_{it} = \beta_{1it} + \sum_{k=2}^{K} \beta_{kit} x_{kit} + \varepsilon_{it}$$

و. يكون فيه لكل الأفراد نفس معلمات النموذج (لايوجد اختلاف في النموذج بين الأفراد)

$$y_{it} = \beta_1 + \sum_{k=2}^{K} \beta_k x_{kit} + \varepsilon_{it}$$

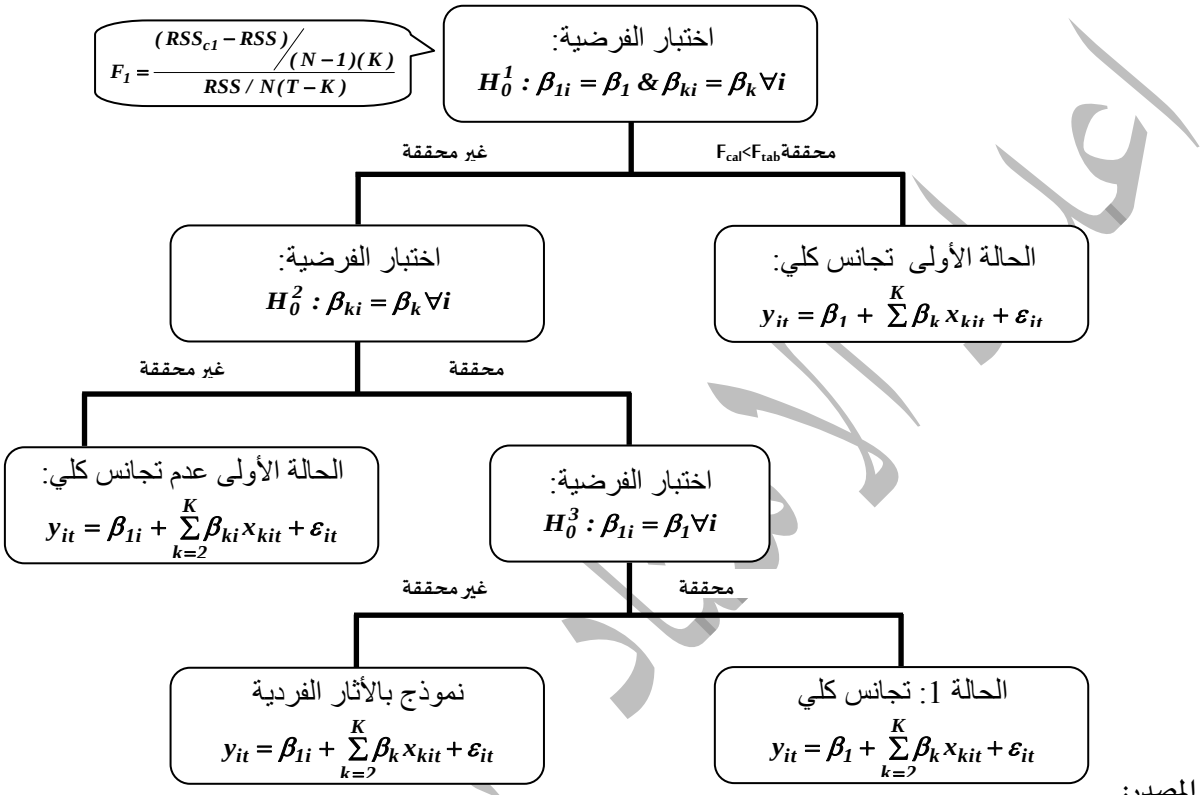
ان استعمال البيانات في شكلها المقطعي والزمني، انما يقودنا الى طرح سؤال رئيس مهم، آل وهو هل البيانات محل الدراسة متجانسة حسب الأفراد والزمن؟ لكي يتم استخدامها في بناء نموذج قياسي يتناولها في آن واحد. ومن هنا نشير الى انه قبل الوصول الى تقدير اي من العلاقات سابقة الذكر، يجب اختبار التجانس عبر الافراد وعبر الزمن للبيانات، وهذا ما يعرف باختبار التجانس وفق منهجية هيساو، ومن أجل تناول الموضوع ببساطة نفترض ان البانات متجانسة عبر الزمن ونكتفي فقط باختبار التجانس حسب الافراد[*]، والتي نشرح طريقة اجرائه باستخدام برنامج Stata.

[*] ان اختبار التجانس حسب الزمن يقوم على نفس الخطوات المتعلقة بالافراد، بحيث يتم اعتبار متغيرات صورية لمتغير الزمن بنفس الطريقة المعتمدة في حالة الافراد واختبار تساوي معاملاتها بحد ثابت واحد متعلق بالزمن الى جانب حد ثابت أو عشوائي بالنسبة للافراد. وكون التطرق الى ذلك يزيد المشكلة تعقيدا وصعوبة الفهم ارتأينا الاكتفاء بتجانس الافراد فقط.

2. اختبار التجانس لنماذج المعطيات الطولية باستخدام البرنامج Stata

يعمد هذا الاختبار على تحليل التباين والتباين المشترك وفق منهجية نلخصها في المخطط التالي:

المخطط رقم (7-1): مخطط توضيحي لاختبار التجانس لهيساو Hisao



المصدر:

بتصرف Régis Bourbonnais : « Économétrie , Cours et exercices corrigés », 9e Edition, DANOD, 2015, p349

- **اختبار الفرضيات:**

نفرض أنه لدينا نموذج من الشكل:

$$y_{it} = \beta_{1i} + \beta_{2i} x_{2it} + \beta_{3i} x_{3it} + \varepsilon_{it} \qquad (1)$$
$$i = 1,2,....,9; k = 1,2,3; t = 1,2,...,25$$

- **الخطوة الأولى:**

في هذه الحالة يتم تقدير النموذج العام (الشكل (1)) الذي يفترض أن لكل فرد نموذج خاص به وهي حالة عدم التجانس الكلي للبيانات بحيث نقدر (N*K) معلمة، ونحتفظ مجموع مربعات البواقي الناتجة عن التقدير (RSS)، ثم نقدر النموذج المقيد بحيث نفرض أن معلمات النموذج متساوية بين الأفراد أي انه يوجد نموذج واحد يعبر عن كل الافراد دون تمييز وبالتالي يكون لدينا K معلمة فقط (بما فيها الحد الثابت)، ونحتفظ مجموع مربعات البواقي الناتجة عن التقدير. وهنا نشير الى انه تم وضع (N-1)*K قيد على النموذج العام.

ومن ثم نقوم بحساب احصائية الاختبار وفق العلاقة التالية:

$$F_1 = \frac{(RSS_{c1} - RSS) / (N-1)(K)}{RSS / N(T-K)}$$

مثال:

لدينا البيانات الخاصة بنفقات التعليم و اجمالي النفقات والناتج الداخلي الخام في 9 بلدان لمدة 25 سنة، المبينة في الجدول التالي:

الجدول (7-1): تطور نفقات التعليم واجمالي النفقات والناتج الداخلي الخام لعينة من البلدان

ind	année	y	x1	x2	ind	année	y	x1	x2
1	1	512	3,15	25,92	6	1	1337	17,67	10,34
1	2	687	5,32	22,51	6	2	1301	22,02	12,04
..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..
1	24	2166	6,29	19,76	6	24	1246	12,83	22,25
1	25	1214	11,62	19,76	6	25	675	17,67	27,36
2	1	74	8,47	35,99	7	1	759	2,66	21,6
2	2	67	11,37	17,8	7	2	1836	2,9	6,54
..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..
2	24	5	2,66	12,43	7	24	460	3,39	30,37
2	25	3	4,11	11,13	7	25	381	4,6	15,45
3	1	419	5,81	13,61	8	1	64	1,94	21,07
3	2	10238	7,02	3,27	8	2	220	2,42	11,52
..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..
3	24	1939	12,83	14,27	8	24	393	3,63	12,3
3	25	2254	16,7	14,14	8	25	16	5,57	15,31
4	1	498	0,24	21,34	9	1	375	4,36	20,68
4	2	81	0,24	1,31	9	2	55	5,81	10,08
..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..
4	24	413	4,11	21,86	9	24	28	4,84	12,96
4	25	268	5,32	23,3	9	25	170	3,87	12,83
5	1	194	15,49	10,08					
5	2	50	14,76	2,75					
..	..	..	..	..					
..	..	..	..	..					
5	24	83	5,08	9,16					
5	25	5	9,68	7,72					

المصدر: Régis Bourbonnais , Op cit, p 352،

متوفرة على الموقع:

https://regisbourbonnais.dauphine.fr/

اسم الملف: **fichier C13EX1**

باستخدام البرنامج Stata، نتبع الخطوات التالية:

1. تقدير النموذج العام بكتابة الأمر:

```
reg y x2 x3 i.ind i.ind#c.x1 i.ind#c.x2
```

نتحصل على التقدير التالي:

```
. reg y x2 x3 i.ind i.ind#c.x2 i.ind#c.x3
```

Source	SS	df	MS	Number of obs	=	225
				F(26, 198)	=	2.42
Model	58444651	26	2247871.19	Prob > F	=	0.0003
Residual	183583378	198	927188.779	R-squared	=	0.2415
				Adj R-squared	=	0.1419
Total	242028029	224	1080482.27	Root MSE	=	962.91

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x2	37.04003	119.3801	0.31	0.757	-198.3797	272.4598
x3	33.05892	27.98951	1.18	0.239	-22.13688	88.25472
ind						
2	-113.2177	796.433	-0.14	0.887	-1683.797	1457.362
3	2525.067	977.0358	2.58	0.010	598.3351	4451.798
4	655.6455	748.4505	0.88	0.382	-820.312	2131.603
5	-155.0801	763.9943	-0.20	0.839	-1661.69	1351.53
6	1874.224	1124.656	1.67	0.097	-343.6177	4092.065
7	233.6722	1370.501	0.17	0.865	-2468.979	2936.324
8	100.7176	901.3993	0.11	0.911	-1676.858	1878.293
9	-371.6792	1192.135	-0.31	0.756	-2722.589	1979.231
ind#c.x2						
2	-37.2206	134.997	-0.28	0.783	-303.437	228.9958
3	-178.5407	133.0561	-1.34	0.181	-440.9297	83.84827
4	-202.4471	142.759	-1.42	0.158	-483.9702	79.0761
5	-29.8576	126.7171	-0.24	0.814	-279.7459	220.0307
6	-123.86	130.6701	-0.95	0.344	-381.5438	133.8237
7	-42.33938	372.534	-0.11	0.910	-776.983	692.3042
8	-50.35638	255.0189	-0.20	0.844	-553.2581	452.5453
9	1.581652	218.6536	0.01	0.994	-429.6071	432.7704
ind#c.x3						
2	-32.55252	38.72346	-0.84	0.402	-108.9159	43.81082
3	32.5826	54.32379	0.60	0.549	-74.54487	139.7101
4	29.24078	39.60947	0.74	0.461	-48.86979	107.3514
5	-29.13225	72.50468	-0.40	0.688	-172.1127	113.8482
6	-11.46923	41.58537	-0.28	0.783	-93.4763	70.53784
7	-31.14316	40.3535	-0.77	0.441	-110.721	48.43464
8	-41.82091	50.6966	-0.82	0.410	-141.7955	58.15367
9	-18.95136	53.43159	-0.35	0.723	-124.3194	86.41668
_cons	171.5795	641.9581	0.27	0.790	-1094.373	1437.532

ونحتفظ بقيمة مجموع مربعات البواقي (RSS) المقدرة بـ RSS=183583378.

- **الخطوة الثانية:**

➢ **اختبار الفرضية الأولى:**

ويقوم هذا الاختبار على اختبار فرضية أن معلمات النموذج (1) تتساوى فيما بين الأفراد بكتابة الأمر:

testparm i.ind i.ind#c.x2 i.ind#c.x3

فنتحصل على قيمة الاختبار F_1=2.26 الموضحة في الصورة التالية:

```
F( 24,   198) =    2.26
     Prob > F =    0.0012
```

ويمكن التحقق من ذلك بتقدير النموذج المقيد باعتبار ان المعطيات عبارة عن سلاسل لا تأخذ بعين الاعتبار الأفراد ونقدر النموذج في شكله التجميعي من الشكل (2) التالي:

$$y_{it} = \beta_1 + \beta_2 x_{2it} + \beta_3 x_{3it} + \varepsilon_{it} \qquad (2)$$
$$i = 1,2,....,9; k = 1,2,3; t = 1,2,...,25$$

بكتابة الأمر:

reg y x2 x3

فمتحصل على النتائج التالية:

```
. reg y x2 x3
```

Source	SS	df	MS	Number of obs	=	225
				F(2, 222)	=	3.91
Model	8245474.1	2	4122737.05	Prob > F	=	0.0213
Residual	233782555	222	1053074.57	R-squared	=	0.0341
				Adj R-squared	=	0.0254
Total	242028029	224	1080482.27	Root MSE	=	1026.2

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x2	29.32698	13.82629	2.12	0.035	2.079408	56.57455
x3	21.51258	11.47806	1.87	0.062	-1.107322	44.13248
_cons	146.2167	138.0735	1.06	0.291	-125.8859	418.3192

ونستخرج قيمة مجموع مربعات البواقي المقدرة هي: RSS_{c1}=233782555

ونحسب قيمة الاختبار، بحيث درجة حرية البسط 24=3*3=8*3=(9-1)*(K(N-1))، ودرجة حرية المقام (N*(T-K)=9*(25-3)=9*22=198 فينتج لدينا:

$$F_1 = \frac{(233782555 - 183583378)/(9-1)(3)}{183583378/9(25-3)} \approx 2.26$$

وهي أكبر من القيمة المجدولة عند مستوى 5% المقدرة بـ 1.52 مما يقودنا الى رفض فرضية تساوي كل معاملات النموذج فيما بين الأفراد، وعليه ننتقل الى اختبار الفرضية الثانية.

- **اختبار الفرضية الثانية:**

ويقوم على اختبار فرضية أن معاملات المتغيرات المستقلة متساوية بين الأفراد بينما الحد يتغير حسب الأفراد (أي حد ثابت لكل فرد). وفي هذه الحالة بعد تقدير النموذج العام من جديد اذا كنا قد قدرنا النموذج الثاني، نكتب الأمر التالي:

testparm i.ind#c.x2 i.ind#c.x3

فنتحصل على النتائج التالية:

```
F( 16,   198) =    0.80
     Prob > F =    0.6801
```

ويمكن التحقق من النتائج بتقدير نموذج بـ 9 معلمات للحد الثابت ومعلمتين فقط للمتغيرات المستقلة وذلك بطريقتين: إما كتابة الأمر: reg y x2 x3 i.ind أو:

كتابة الأمر: ta ind, gen(d)، بحيث يقوم البرنامج بانشاء مصفوفة المتغيرات الصورية الخاصة بالأفراد، ومن ثم تقدير النموذج من الشكل:

$$y_{it} = \beta_1 + d_2 + d_3 + d_4 + d_5 + d_6 + d_7 + d_8 + d_9 + \beta_2 x_{2it} + \beta_3 x_{3it} + \varepsilon_{it}$$

$$i = 1,2,....,9; k = 1,2,3; t = 1,2,...,25$$

بحيث يتم الاستغناء عن الفرد الأول وقيمة الحد الثابت الخاص به هو معلمة الحد الثابت المقدرة من طرف البرنامج.

فنتحصل على التقدير التالي:

```
. reg y x2 x3 i.ind
```

Source	SS	df	MS		
Model	46516654.8	10	4651665.48	Number of obs	= 225
Residual	195511374	214	913604.553	F(10, 214)	= 5.09
				Prob > F	= 0.0000
Total	242028029	224	1080482.27	R-squared	= 0.1922
				Adj R-squared	= 0.1544
				Root MSE	= 955.83

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x2	-49.50377	23.64297	-2.09	0.037	-96.10669	-2.900844
x3	22.17142	11.02105	2.01	0.046	.447697	43.89514
ind						
2	-487.5155	271.2143	-1.80	0.074	-1022.109	47.07813
3	1107.967	312.4725	3.55	0.000	492.0487	1723.885
4	53.41096	273.9096	0.19	0.846	-486.4953	593.3172
5	-458.6714	272.6048	-1.68	0.094	-996.0058	78.66306
6	745.7494	368.5273	2.02	0.044	19.34111	1472.158
7	-322.4856	274.8602	-1.17	0.242	-864.2657	219.2945
8	-550.5831	277.0173	-1.99	0.048	-1096.615	-4.551224
9	-530.7585	270.5637	-1.96	0.051	-1064.07	2.55263
_cons	704.1205	240.8149	2.92	0.004	229.4475	1178.794

بحيث قيمة مجموع مربعات البواقي المقدرة هي: RSSc2= 195511374

اختبار الفرضية يكون وفق العلاقة التالية:

$$F_1 = \frac{(RSS_{c1} - RSS) / (N-1)(K-1)}{RSS / N(T-K)}$$

فتكون قيمة الاختبار:

$$F_2 = \frac{(195511374 - 183583378) / (9-1)*(3-1)}{183583378 / 9(25-3)} \approx 0.68$$

وبالمقارنة مع القيمة المجدولة لفيشر ($F_{(16,198)}$=1.72) نلاحظ أن القيمة المحسوبة أقل من المجدولة وبالتالي نقبل فرضية العدم H_2. اذن ننتقل الى اختبار الفرضية الثالثة H_3 .

- **اختبار الفرضية الثالثة:**

في هذه المرحلة نختبر الفرضية القائلة أن معلمات الحد الثابت تتساوى بالنسبة لكل الافراد انطلاقا من تقدير النموذج الثالث، وهنا نكتب الأمر testparm i.ind مباشرة بعد التقدير بواسطة الأمر:

```
reg y x2 x3 i.ind
```

فينتج لدينا قيمة الاختبار كما يلي:

```
F(  8,   214) =    5.24
     Prob > F =    0.0000
```

ونشير هنا الى أن قيمة درجات الحرية الخاصة بالمقام تتغير بحيث تم تغيير نموذج الأساس من النموذج العام الى النموذج الثالث الذي يفترض أن معلمات الحد الثابت تختلف حسب الافراد بينما معلمات المتغيرات المستقلة هي نفسها بالنسبة لكل الأفراد، فتصبح (N*T-N-K-1) بدل (N*T-N*K)، أما بالنسبة للبسط فإن درجة الحرية الخاصة به تساوي الى عدد القيود المفروضة على النموذج الثالث للحصول على النموذج الثاني وتساوي (N-1) قيد.

فيصبح الاختبار على الشكل التالي:

$$F_1 = \frac{(RSS_{c3} - RSS_{c2})/(N-1)}{RSS_{c2}/[N(T-1)-(K-1)]}$$

أي:

$$F_1 = \frac{(233782555 - 195511374)/(9-1)}{195511374/9(25-1)-(3-1)} \approx 5.24$$

بالمقارنة مع القيمة المجدولة لفيشر المقدرة بـ 1.98 تقريبا، وهي اقل من القيمة المحسوبة نرفض الفرضية H_3 ونقر بأن النموذج الأفضل للمعطيات المتوفرة هو نموذج الآثار الفردية.

$$F_2 = \frac{(195511374 - 183583378)/(9-1)*(3-1)}{183583378/9(25-3)} \approx 0.68$$

وبعد التحقق من صلاحية المعطيات المتوفرة للنمذجة بمعطيات البانل يتم اختبار نوعين من نماذج البانل والمعرفة كما يلي:

3. نموذج الأثر الثابت ونموذج الاثر العشوائي

3-1. نموذج الأثر الثابت

ينشأ هذا الشكل من النماذج، انطلاقا من فرضية أن العلاقة بين المتغير التابع والمتغيرات المفسرة متطابقة بالنسبة لجميع الأفراد في عينة الدراسة عبر الزمن، ويهدف الى تحديد سلوك كل مجموعة بيانات مقطعية(بالنسبة لكل فرد) على حدة من خلال السماح للاعتراض (معلمة القطع) أن تتغير بتغير الأفراد مع بقاء معالم الانحدار (الميل) ثابتة لكل الأفراد، وبالتالي فان صياغة الاثار الثابتة بين المجموعات (الأفراد) يمكن التعبير عنها من خلال الحد الثابت للنموذج والذي يتم اعتباره كمعلمة مجهولة للتقدير.

وبالتالي يتم التعامل مع حالة عدم التجانس في التباين بين الأفراد، بحيث يأخذ النموذج شكل النموذج B المذكور سابقا وفق العلاقة التالية:

$$Y_{it} = \beta_{1i} + \sum_{k=2}^{K} \beta_k X_{kit} + \varepsilon_{it}$$
$$i = 1,2,3,....,N\,;\, t = 1,2,3,...,T$$

بحيث يعتبر الحد الثابت مختلف من فرد لآخر، بينما يشترك كل الأفراد في معامل الانحدار لكل متغير مفسر. وحد الخطأ العشوائي تتوفر فيه الفرضيات الكلاسيكية.

ولتقدير النموذج وفق صياغة الاثار الثابتة بطريقة المربعات الصغرى العادية يتم الاستعانة بمتغيرات صورية فيصبح النموذج وفق العلاقة:

$$Y_{it} = X'\beta_k + D\beta_1 + \varepsilon_{it}$$
$$i = 1,2,3,....,N\,;\, t = 1,2,3,...,T\,;\, k = 2,3....,K$$

بحيث تمثل:

X' مصفوفة المتغيرات المفسِّرة، و β_k شعاع المعلمات المتعلقة بالمتغيرات المفسرة (معلمات الميل)، β_1 معلمات الحد الثابت المتعلقة بكل فرد (مجموعة)، D مصفوفة المتغيرات الصورية. وتعطي طريقة المربعات الصغرى تقدير لـ K معلمة متعلقة بالمتغيرات المفسرة وN معلمة متعلقة بالحد الثابت، وتسمى الطريقة عندئذ بطريقة المربعات الصغرى بالمتغيرات الصورية variable dummy least squares LSDV.

3-2-. نموذج الأثر العشوائي

يفترض نموذج الأثر العشوائي على خلاف نموذج الأثر الثابت -الذي يفترض ان لكل فرد أو مجموعة تأخذ قاطعا خاص بها-، أن العلاقة بين المتغير التابع والمتغيرات المفسرة عشوائية وليست ثابتة، بحيث يعتبر أن الاثار الفردية ليست ثابتة بل هي متغيرات عشوائية مستقلة بمتوسط معدوم

وتباين معلوم تدخل ضمن حد الخطأ العشوائي للنموذج بحيث يكون الحد الثابت في النموذج عبارة عن حد ثابت مضافا اليه حد عشوائي أي أن النموذج يكتب على الشكل:

$$Y_{it} = (\beta_1 + u_i) + \sum_{k=2}^{K} \beta_k X_{kit} + \varepsilon_{it}$$

$$i = 1,2,3,....,N ; t = 1,2,3,...,T$$

بحيث: ui عبارة عن حد عشوائي تحت الفرضيات الكلاسيكية:

ويقوم هذا النموذج على فرضية اساسية وهي أن الاختلاف بين الأفراد (المجموعات) يكون في الحد العشوائي.

في الحالة العامة لنموذج الأثر العشوائي الذي يشمل الأثر الفردي (المقطعي) والزمني يسمى النموذج بنموذج الخطأ المركب *Modele à Erreur Composée*. وتكون فيه عبارة معلمة الحد الثابت على شكل:

$$\beta_{1i} = \beta_1 + u_i$$

بحيث: β_1 حد ثابت بالنسبة لكل الأفراد، و μ_i حد عشوائي يعبر عن الأثار الفردية (المقطعية)، وبالتالي يتم تضمينه ضمن حد الخطأ العشوائي في النموذج، بالاضافة الى الاثار الزمنية فيصبح عندئذ الخطأ العشوائي للنموذج على الشكل:

$$\varepsilon_{it} = u_i + \lambda_t + \upsilon_{it}$$

و يكون بذلك الخطأ العشوائي مركب من الآثار الفردية والزمنية بالإضافة الى الخطأ العشوائي المرتبط بهذه الآثار. بينما في الحالة التي يتم فيها استبعاد الآثار الزمنية أي اعتبار $\lambda_t = 0$ في حد الخطأ فإن نموذج الأثر العشوائي يكتب على الشكل:

$$Y_{it} = \beta_1 + \sum_{k=2}^{K} \beta_k X_{kit} + \omega_{it}; \quad \omega_{it} = \varepsilon_{it} + \mu_i$$

4. تقدير نماذج البانل باستخدام Stata

4-1. تقدير النموذج التجميعي

في هذه المرحلة يتم تقدير النموذج في شكل تجميعي دون الاخذ بعين الاعتبار العنصر المكاني، فيكون النموذج في شكل نموذج انحدار متعدد، واعتبار ان البيانات الخاصة بكل متغير وكأنها سلسلة زمنية ذات البعد N*T.

ويتم ذلك من خلال كتابة الأمر:

```
reg y x1 x2
```

بحيث يتم تقدير النموذج بطريقة المربعات الصغرى العادية مع احتساب الحد الثابت دون الاشارة اليه في المعادلة. أي تقدير النموذج التالي:

$$y_{it} = \beta_1 + \beta_2 x_{1it} + \beta_3 x_{2it} + \varepsilon_{it}$$

وبتطبيق الأمر تظهر لنا النتائج على النحو التالي:

```
. reg y x1 x2
```

Source	SS	df	MS	Number of obs	=	40
				F(2, 37)	=	4.89
Model	4.82416154	2	2.41208077	Prob > F	=	0.0130
Residual	18.2468816	37	.493158963	R-squared	=	0.2091
				Adj R-squared	=	0.1663
Total	23.0710432	39	.59156521	Root MSE	=	.70225

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x1	.1674908	.0537162	3.12	0.004	.0586515	.2763302
x2	.0568985	.1670706	0.34	0.735	-.2816186	.3954156
_cons	22.34798	2.846089	7.85	0.000	16.58125	28.1147

2-4. التعريف بمعطيات البانل (البعد الفردي والبعد الزمني)

في هذه المرحلة نقوم بتعريف البيانات على أنها بيانات سلاسل زمنية ومقطعية، بحيث نقوم بانشاء متغير جديد يعبر عن رموز البلدان اذا كان متغير الفرد معبر عنه بأسماء البلدان مثلا، بكتابة الأمر:

egen codp=group(pays)

يتم تحديد البعد الزمني والمقطعي للبرنامج، وذلك بكتابة الأمر:

tsset pays année

للتعريف بمتغير الفرد والزمن وهما pays و année على التوالي.

فيظهر البرنامج الجدول التالي:

```
. tsset codp année
      panel variable:  codp (strongly balanced)
       time variable:  année, 2001 to 2011, but with gaps
               delta:  1 unit
```

مخرجات الأمر tsset التي تبين ان البرنامج يعتبر البيانات ذات البعد الزمني والمكاني في شكل متوازن

3-4. تقدير نموذج المربعات الصغرى للمتغيرات الصورية LSDV

وهو نموذج أثار ثابتة، يظهر فيه الاختلاف بين الأفراد في الحد الثابت، بحيث يكون لكل فرد حد ثابت خاص به، بينما يشترك الأفراد في معاملات المتغيرات المستقلة، وفي حالتنا يكتب النموذج على الشكل التالي:

$$Y_{it} = D_1 + D_2 + D_3 + D_4 + \beta_2 x_{1it} + \beta_3 x_{2it} + \varepsilon_{it}$$

بحيث تمثل Di متغيرات صورية تدل عن الفرد (i) أي: 1 للفرد المعني و 0 غير ذلك ... وهكذا.

وفي هذه الحالة يتم الاستغناء عن الحد الثابت في التقدير بطريقة المربعات الصغرى، أو الاستغناء عن أحد المتغيرات الصورية. وفي حالة عدم التدخل والاشارة الى ذلك يقوم البرنامج تلقائيا بازالة احد المتغيرات الصورية.

ولتقدير النموذج LSDV يجب اضافة مصفوفة المتغيرات الصورية الى مصفوفة المتغيرات، ويمكن انشاء هذه المصفوفة تلقائيا عن طريق كتابة الأمر:

tabulate codp, generate (D)

فيظهر في مخرجات البرنامج جدول موضح عدد المتغيرات التي تم انشاؤها، بينما في متصفح البيانات تظهر لدينا متغيرات جديدة الى جانب المتغيرات الموجودة من قبل تتمثل في المتغيرات الصورية.

```
. tabulate codp, generate(D)
```

codp	Freq.	Percent	Cum.
Btp	10	25.00	25.00
agr	10	25.00	50.00
ind	10	25.00	75.00
serv	10	25.00	100.00
Total	40	100.00	

يتم تقدير نموذج LSDV بطريقة المربعات الصغرى العادية بكتابة الأمر:

reg y D2 D3 D4 x1 x2

بحيث قمنا بازالة المتغير الصوري المتعلق بالبلد الأول من أجل امكانية التقدير، لكن يمكن اضافة العبارة noconstant لتقدير النموذج بدون حد ثابت مع استخدام كل المتغيرات الصورية.

```
. reg y D2 D3 D4 x1 x2
```

Source	SS	df	MS		
				Number of obs	= 40
				F(5, 34)	= 498.07
Model	22.760301	5	4.5520602	Prob > F	= 0.0000
Residual	.310742187	34	.009139476	R-squared	= 0.9865
				Adj R-squared	= 0.9846
Total	23.0710432	39	.59156521	Root MSE	= .0956

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
D2	1.310136	.1465328	8.94	0.000	1.012346	1.607927
D3	1.916025	.050566	37.89	0.000	1.813263	2.018788
D4	1.267579	.1757824	7.21	0.000	.9103462	1.624812
x1	.2606395	.0278623	9.35	0.000	.2040166	.3172625
x2	.113568	.0993567	1.14	0.261	-.088349	.315485
_cons	17.85898	.9953567	17.94	0.000	15.83617	19.88179

كما يتم الحصول على نفس النتائج بكتابة الأمر :

i.codp x2 reg y x1

فنحصل على نفس النتائج بدون انشاء المتغيرات الصورية

ويتم تفسير معاملات المتغيرات الصورية بأنها الحد الثابت الخاص بكل بلد. بحيث: D1=cons، بينما الحدود الأخرى فهي عبارة عن الحد المقدر بهذه الطريقة مضاف اليه الحد الثابت cons المقدر:

Di=Di+cons

مع العلم أن: D1=cons

وبالضغط على Entrer ينتج لدينا نتائج التقدير الخاصة بالنموذج.

4-4. تقدير نموذج الآثار الثابتة FEM

في هذه المرحلة نقوم بتقدير نموذج الآثار الثابتة أو ما يعرف بنموذج الانحدار داخل المجموعات، والذي يعتبر أن الاختلاف بين الأفراد يكون في الحد الثابت، بحيث يكون الحد الثابت في في هذا النموذج عبارة عن متوسط معلمات المتغيرات الصورية المقدرة في نموذج LSDV.

يتم تقدير نموذج الآثار الثابتة ببرنامج Stata بكتابة الأمر:

xtreg y x1 x2 ,fe

بحيث تعبر العبارة fe عن الأثر الثابت

وبالضغط على Entrer نتحصل على نتائج التقدير على الشكل التالي:

```
Fixed-effects (within) regression               Number of obs      =        40
Group variable: codp                            Number of groups   =         4

R-sq:                                           Obs per group:
     within  = 0.8615                                         min =        10
     between = 0.1530                                         avg =      10.0
     overall = 0.2089                                         max =        10

                                                F(2,34)            =    105.71
corr(u_i, Xb)  = -0.2794                        Prob > F           =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .2606395   .0278623     9.35   0.000     .2040166    .3172625
          x2 |    .113568   .0993567     1.14   0.261     -.088349     .315485
       _cons |   18.98242   1.02628     18.50   0.000     16.89677    21.06807
-------------+----------------------------------------------------------------
     sigma_u |  .80538611
     sigma_e |  .09560061
         rho |  .98610571   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(3, 34) = 654.16                     Prob > F = 0.0000
```

من خلال نتائج التقدير، تظهر لنا نتائج اختبار معنوية المعلمات المقدرة للنموذج (اختبار t و F)، لكن يظهر ان معامل الارتباط بين حد الخطأ والمتغيرات المستقلة سالب وكبير نوعا ما وهذا مخالف لفرضية العدم التي تقر بعدم وجود ارتباط.

بينما تظهر القيمة المحسوبة لاختبار تجانس الأفراد، بمعنى اختبار تساوي معاملات الحد الثابت بحد ثابت واحد أكبر من تلك المجدولة مما يقودنا الى قبول فرضية عدم تجانس الأفراد، وبالتالي نقر بان نموذج الأثر الثابت هو الأفضل عن النموذج التجميعي.

ملاحظة: لغرض اجراء اختبار المفاضلة بين نموذج الآثار الثابتة ونموذج الآثار العشوائية يتوجب علينا تخزين نتائج التقدير لنموذج FEM بكتابة الأمر:

estimates store fixe

5-4. تقدير نموذج الآثار العشوائية REM

يعتبر هذا النموذج أن الاختلاف بين الأفراد لا يعود الى الاختلاف في الحد الثابت للأفراد، وانما الى حد عشوائي يتم تدقيره ضمن الخطأ العشوائي للنموذج ولهذا يعرف بنموذج الخطأ المركب.
ويتم تقدير هذا النموذج بطريقة المربعات الصغرى المعممة بدل طريقة المربعات الصغرى العادية.
وباستعمال البرنامج Stata يمكن تقدير هذا الأخير بكتابة الأمر:

xtreg y x1 x2,re

وبتطبيق الأمر نتحصل على النتائج التالية:

```
Random-effects GLS regression                   Number of obs      =        40
Group variable: codp                            Number of groups   =         4

R-sq:                                           Obs per group:
     within  = 0.8615                                         min =        10
     between = 0.1529                                         avg =      10.0
     overall = 0.2089                                         max =        10

                                                Wald chi2(2)       =    222.92
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	.2600425	.0269725	9.64	0.000	.2071774	.3129076
x2	.1148455	.0961278	1.19	0.232	-.0735614	.3032525
_cons	18.98055	1.187204	15.99	0.000	16.65367	21.30742
sigma_u	1.3263584					
sigma_e	.09560061					
rho	.99483168	(fraction of variance due to u_i)				

6-4. المفاضلة بين النماذج

تتعدد الاختبارات المستعملة في المفاضلة بين مختلف نماذج البانل، ونخص بالذكر:

- **اختبار والد (Wald test):**

ويستخدم للمفاضلة بين نموذج الاثار الثابتة والنموذج التجميعي باختبار فرضية تساوي معاملات الحد الثابت في نموذج FEM وبالتالي تحقق الفرضية يعني النموذج التجميعي هو الأفضل.
وهذا الاختبار يقوم به Stata تلقائيا كما تمت الاشارة اليه عند تقدير نموذج الآثار الثابتة من خلال اختبار التجانس بين الأفراد.

- **اختبار هوسمان (Hausman test):**

يستخدم هذا الاختبار للمفاضلة بين نموذج الآثار الثابتة ونموذج الآثار العشوائية، بحيث يختبر فرضية العدم القائلة أن النموذج الأفضل هو نموذج الآثار العشوائية.
ويتم اجراء الاختبار ببرنامج Stata عن طريق كتابة الأمر:

fe hausman

لكن هذا يكون عبر اتباع الخطوات المذكورة سابقا أي:

1. تقدير نموذج الاثر الثابت:

xtreg y x1 x2, fe

2. تخزين نتائج التقدير: **estimate store fixe**

3. تقدير نموذج الاثار العشوائية: **xtreg y x1 x2, re**

4. اختبار هوسمان: **hausman fixe**

بتنفيذ هذه الاوامر تظهر لنا تائج اختبار هوسمان على الشكل التالي:

```
                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       fixe           .       Difference          S.E.
-------------+------------------------------------------------------------
          x1 |    .2606395     .2600425         .000597       .0069851
          x2 |     .113568     .1148455        -.0012775       .0251237
------------------------------------------------------------------------------
                    b = consistent under Ho and Ha; obtained from xtreg
     B = inconsistent under Ha, efficient under Ho; obtained from xtreg

  Test:  Ho:  difference in coefficients not systematic

                chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                        =          0.01
              Prob>chi2 =          0.9961
```

يعرض البرنامج هنا القيمة المحسوبة للاختبار 0.01 وهي أقل من القيم المجدولة لكاي تربيع المقدرة بـ 5.99 عند مستوى 5%، مما يقودنا الى قبول فرضية العدم أي ان نموذج الآثار العشوائية هو الأفضل.

مثال تطبيقي:

نريد تقدير نموذج دالة انتاج من شكل كوب دوغلاص لستة دول من دول حوض المتوسط من الشكل:

$$Yit = \beta 1i.K^{\beta 2i}.L^{\beta 3i}.e^{\varepsilon it}$$

ومن أجل امكانية تقدير النموذج بطريقة المربعات الصغرى يجب تحويله الى شكل خطي، بحيث يتم تحويله الى الشكل الخطي بادخال اللوغاريتم النيبيري على أطراف النموذج ليصبح على الشكل:

$$Y_{it} = \beta_{1i} + \beta_{2i}\, log(K_{it}) + \beta_{3i}\, log(L_{it}) + \varepsilon_{it}$$

بحيث تمثل:

Y: الناتج الداخلي الخام؛ K: مخزون رأس المال؛ L: العمل؛

βik: معلمات للتقدير.

i=1,2,...,N. يعبر عن الأفراد (الدول)

t=1,2,...,T: يعبر عن الزمن.

Log: يعبر عن اللوغاريتم النيبيري.

والبيانات الخاصة بمختلف المتغيرات مبينة في الجدول التالي:

الجدول رقم (7-1): بيانات الانتاج وعوامل الانتاج لعينة من بلدان حوض المتوسط

pays	année	log(Y)	log(K)	log(L)	pays	année	log(Y)	log(K)	log(L)	pays	année	log(Y)	log(K)	log(L)
DZA	1998	25,26	23,78	15,87	EGY	1998	25,39	23,45	16,68	ESP	1998	27,57	26,06	16,63
DZA	1999	25,30	23,82	15,91	EGY	1999	25,43	23,55	16,70	ESP	1999	27,60	26,08	16,64
DZA	2000	25,31	23,82	15,93	EGY	2000	25,49	23,68	16,72	ESP	2000	27,63	26,13	16,65
DZA	2001	25,36	23,86	15,96	EGY	2001	25,53	23,88	16,73	ESP	2001	27,67	26,23	16,67
DZA	2002	25,39	23,88	15,98	EGY	2002	25,59	23,92	16,79	ESP	2002	27,72	26,33	16,69
DZA	2003	25,43	23,91	16,01	EGY	2003	25,64	23,90	16,81	ESP	2003	27,77	26,40	16,72
DZA	2004	25,46	23,93	16,03	EGY	2004	25,67	23,88	16,83	ESP	2004	27,81	26,45	16,72
DZA	2005	25,51	24,00	16,06	EGY	2005	25,70	23,93	16,85	ESP	2005	27,84	26,49	16,76
DZA	2006	25,58	24,05	16,08	EGY	2006	25,73	23,84	16,90	ESP	2006	27,87	26,56	16,80
DZA	2007	25,62	24,13	16,11	EGY	2007	25,77	23,89	16,95	ESP	2007	27,90	26,61	16,83
DZA	2008	25,68	24,21	16,13	EGY	2008	25,81	24,01	16,99	ESP	2008	27,94	26,68	16,87
DZA	2009	25,70	24,28	16,15	EGY	2009	25,88	24,14	16,99	ESP	2009	27,98	26,75	16,90
DZA	2010	25,73	24,37	16,17	EGY	2010	25,95	24,35	17,04	ESP	2010	28,02	26,80	16,93
DZA	2011	25,75	24,46	16,19	EGY	2011	26,02	24,49	17,07	ESP	2011	28,03	26,76	16,96
DZA	2012	25,77	24,72	16,22	EGY	2012	26,06	24,39	17,09	ESP	2012	27,99	26,57	16,97
DZA	2013	25,81	24,79	16,24	EGY	2013	26,11	24,46	17,12	ESP	2013	27,99	26,52	16,97
DZA	2014	25,83	24,82	16,26	EGY	2014	26,13	24,44	17,14	ESP	2014	27,98	26,45	16,98
DZA	2015	25,87	24,89	16,29	EGY	2015	26,15	24,42	17,16	ESP	2015	27,95	26,38	16,98
DZA	2016	25,90	24,97	16,31	EGY	2016	26,17	24,33	17,18	ESP	2016	27,94	26,35	16,97
DZA	2017	25,93	25,03	16,33	EGY	2017	26,19	24,35	17,20	ESP	2017	27,95	26,39	16,97

pays	année	log(Y)	log(K)	log(L)	pays	année	log(Y)	log(K)	log(L)	pays	année	log(Y)	log(K)	log(L)	pays	année	log(Y)	log(K)	log(L)
FRA	1998	28,34	26,76	17,09	ITA	1998	28,25	26,58	16,94	TUN	1998	23,82	22,22	14,87	TUR	1998	26,74	24,97	16,85
FRA	1999	28,35	26,76	17,10	ITA	1999	28,27	26,60	16,94	TUN	1999	23,89	22,24	14,90	TUR	1999	26,81	25,10	16,87
FRA	2000	28,38	26,77	17,10	ITA	2000	28,29	26,62	16,95	TUN	2000	23,94	22,32	14,93	TUR	2000	26,88	25,24	16,87
FRA	2001	28,41	26,83	17,11	ITA	2001	28,30	26,66	16,95	TUN	2001	23,99	22,38	14,95	TUR	2001	26,91	25,20	16,89
FRA	2002	28,45	26,91	17,12	ITA	2002	28,32	26,70	16,96	TUN	2002	24,05	22,46	14,97	TUR	2002	26,87	25,02	16,91
FRA	2003	28,48	26,97	17,12	ITA	2003	28,35	26,76	16,96	TUN	2003	24,10	22,50	14,98	TUR	2003	26,94	25,18	16,88
FRA	2004	28,50	27,00	17,13	ITA	2004	28,37	26,79	16,97	TUN	2004	24,13	22,58	15,00	TUR	2004	26,88	24,83	16,89
FRA	2005	28,51	26,99	17,14	ITA	2005	28,37	26,83	16,98	TUN	2005	24,15	22,56	15,01	TUR	2005	26,94	24,96	16,91
FRA	2006	28,52	27,01	17,16	ITA	2006	28,38	26,83	16,99	TUN	2006	24,19	22,55	15,03	TUR	2006	26,99	25,10	16,90
FRA	2007	28,55	27,04	17,16	ITA	2007	28,39	26,85	17,01	TUN	2007	24,25	22,56	15,04	TUR	2007	27,08	25,35	16,90
FRA	2008	28,57	27,07	17,18	ITA	2008	28,40	26,87	17,01	TUN	2008	24,29	22,58	15,05	TUR	2008	27,16	25,51	16,92
FRA	2009	28,59	27,11	17,18	ITA	2009	28,42	26,90	17,01	TUN	2009	24,34	22,69	15,07	TUR	2009	27,23	25,63	16,94
FRA	2010	28,61	27,16	17,19	ITA	2010	28,44	26,91	17,01	TUN	2010	24,40	22,80	15,10	TUR	2010	27,27	25,66	16,95
FRA	2011	28,61	27,17	17,20	ITA	2011	28,42	26,88	17,03	TUN	2011	24,44	22,90	15,12	TUR	2011	27,28	25,60	16,98
FRA	2012	28,58	27,07	17,21	ITA	2012	28,37	26,78	17,02	TUN	2012	24,47	23,01	15,14	TUR	2012	27,23	25,39	17,02
FRA	2013	28,60	27,09	17,21	ITA	2013	28,38	26,77	17,02	TUN	2013	24,51	23,11	15,15	TUR	2013	27,32	25,65	17,06
FRA	2014	28,63	27,11	17,21	ITA	2014	28,39	26,75	17,02	TUN	2014	24,49	23,01	15,17	TUR	2014	27,40	25,82	17,10
FRA	2015	28,63	27,12	17,22	ITA	2015	28,36	26,66	17,04	TUN	2015	24,53	23,10	15,18	TUR	2015	27,42	25,79	17,12
FRA	2016	28,63	27,11	17,22	ITA	2016	28,34	26,59	17,05	TUN	2016	24,56	23,15	15,20	TUR	2016	27,46	25,83	17,14
FRA	2017	28,64	27,10	17,22	ITA	2017	28,35	26,56	17,07	TUN	2017	24,59	23,09	15,21	TUR	2017	27,49	25,82	17,16

تجدر الاشارة الى انه تم تقريب القيم بعد القاصلة وحذف المتغيرات الاصلية قبل ادخال اللوغاريتم النيبيري حتى يتسنى لنا وضع الجدول في الملف. بالاضافة الى انه تم تحويل الجدول من شكله العرضي الى الشكل الطولي بحيث يشمل خمسة اعمدة فقط.

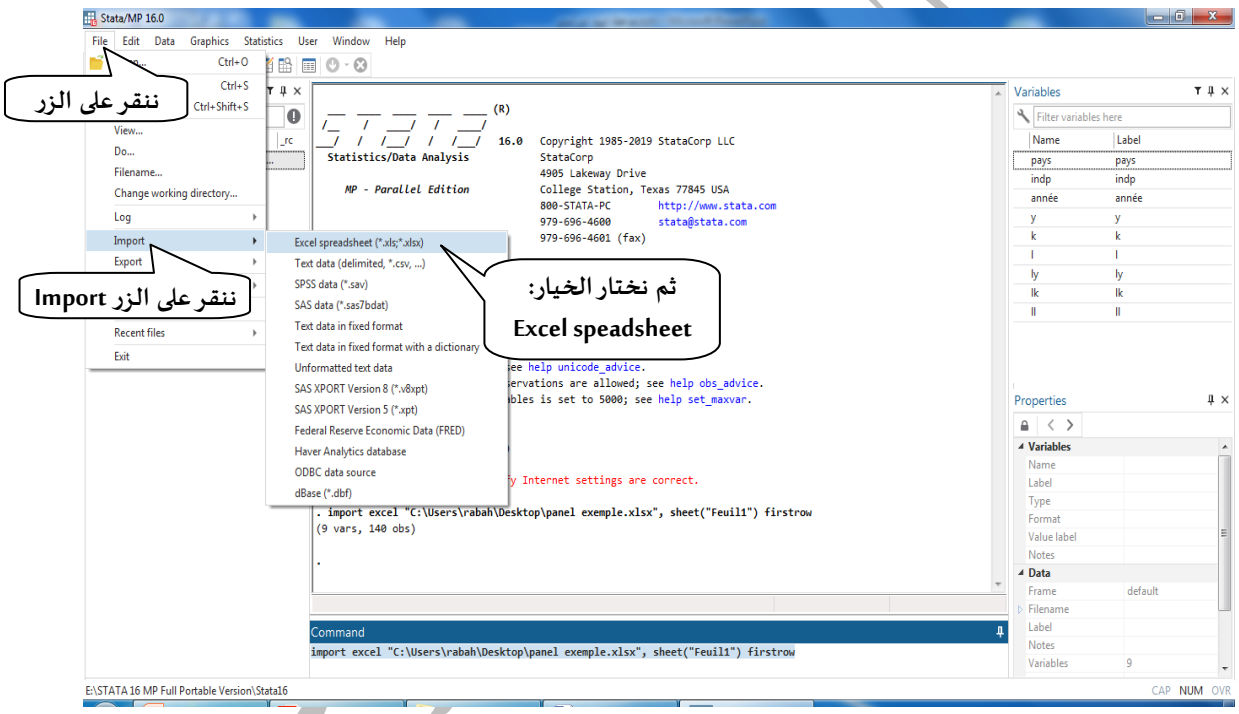
تقدير نماذج البانل باستخدام برنامج Stata:

- الخطوة الأولى:

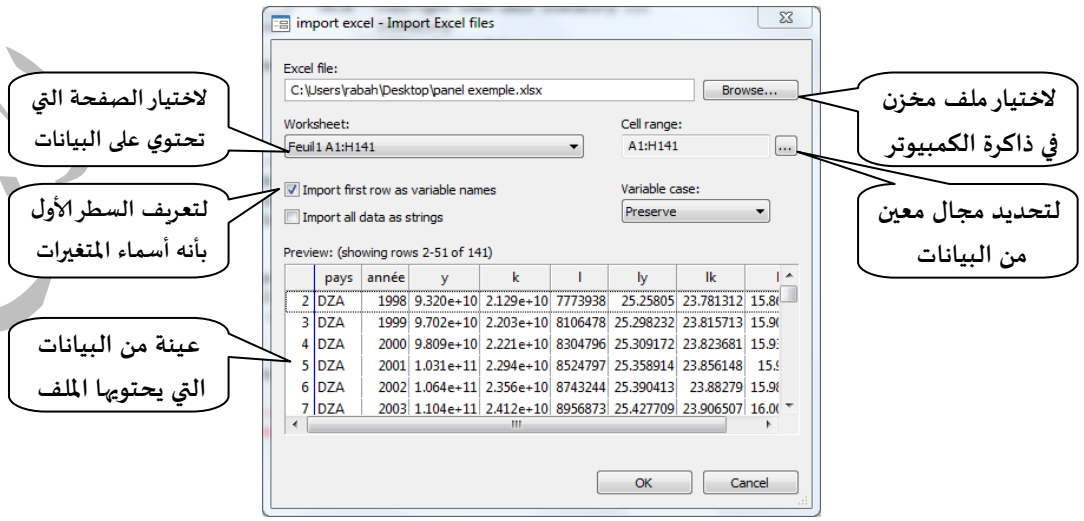
ادخال البيانات الى الرنامج: في هذه المرحلة يتم نقل البيانات المتوفرة في برنامج Excel الي Stata باستخدام الأمر :

import excel "C:\Users\rabah\Desktop\panel exemple.xlsx", sheet("Feuil1") firstrow

- **استيراد البيانات من ملف Excel**



فيظهر لدينا صندوق حوار على النحو التالي:



- الخطوة الثانية:

1- تقدير النموذج التجميعي:

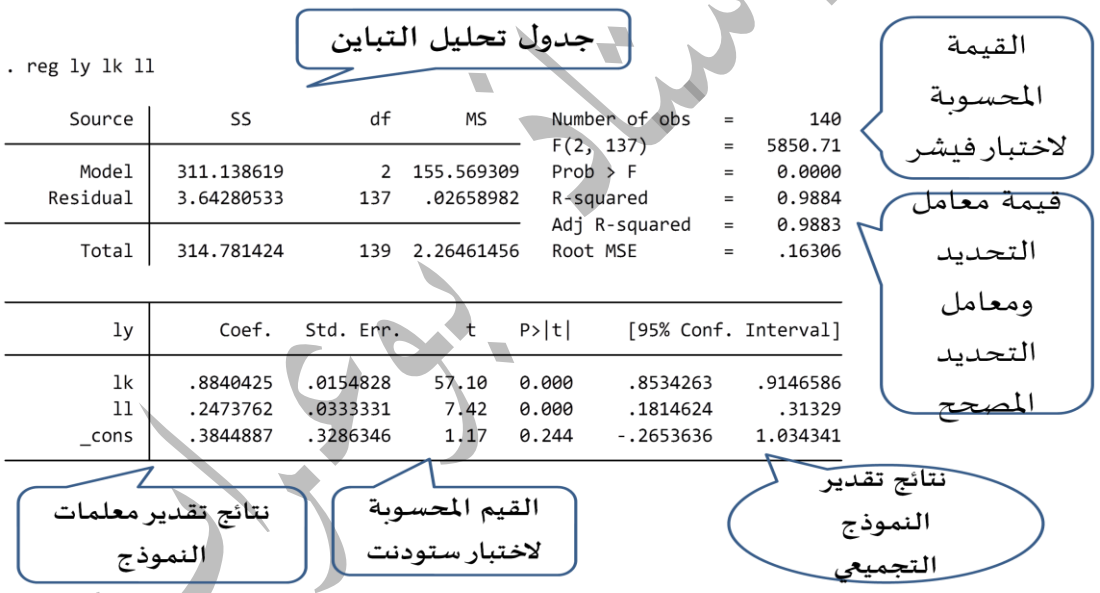
في هذه المرحلة يتم تقدير النموذج في شكل تجميعي دون الاخذ بعين الاعتبار العنصر المكاني، فيكون النموذج في شكل نموذج انحدار متعدد[4](*)، واعتبار ان البيانات الخاصة بكل متغير وكأنها سلسلة زمنية ذات البعد N*T.

ويتم ذلك من خلال كتابة الأمر :

regress ly lk ll

أو اختصارا reg ly lk ll. بحيث يتم تقدير النموذج بطريقة المربعات الصغرى العادية مع احتساب الحد الثابت دون الاشارة اليه في المعادلة.

وبتطبيق الأمر تظهر لنا النتائج على النحو التالي:




```
. reg ly lk ll
```

Source	SS	df	MS		
				Number of obs	= 140
				F(2, 137)	= 5850.71
Model	311.138619	2	155.569309	Prob > F	= 0.0000
Residual	3.64280533	137	.02658982	R-squared	= 0.9884
				Adj R-squared	= 0.9883
Total	314.781424	139	2.26461456	Root MSE	= .16306

ly	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lk	.8840425	.0154828	57.10	0.000	.8534263	.9146586
ll	.2473762	.0333331	7.42	0.000	.1814624	.31329
_cons	.3844887	.3286346	1.17	0.244	-.2653636	1.034341

في هذه المرحلة لم يتم بعد التعامل مع البيانات بأنها ذات بعد زماني ومكاني في آن واحد. لكن يجب تقدير هذا النموذج قبل التعريف ببيانات البانل لبرنامج Stata من أجل مقارنته لاحقا بباقي النماذج.

- الخطوة الثالثة:

في هذه المرحلة نقوم بتعريف البيانات على أنها بيانات سلاسل زمنية ومقطعية، بحيث نقوم بانشاء متغير جديد يعبر عن رموز البلدان من 1 إلى 7 بكتابة الأمر: egen codp=group(pays)

(*) يجب تقدير هذا النموذج قبل التعريف بالبعد الزمني والمقطعي لبيانات البانل بحيث يتم اعتبار ان المتغيرات عبارة عن سلاسل مشاهدات ذات البعد N*T.

وبعدها نقوم تحديد البعد الزمني والمقطعي للبرنامج[*]، وذلك من خلال كتابة الأمر:

tsset codp année

للتعريف بمتغير الفرد والزمن وهما pays وannée على التوالي.

فيظهر البرنامج الجدول التالي:

```
tsset codp année
     panel variable:  codp (strongly balanced)
      time variable:  année, 1998 to 2017
              delta:  1 unit
```

مخرجات الأمر tsset التي تبين ان البرنامج يعتبر البيانات ذات البعد الزمني والمكاني في شكل متوازن

2. تقدير نموذج المربعات الصغرى للمتغيرات الصورية LSDV:

وهو نموذج أثار ثابتة، يظهر فيه الاختلاف بين الأفراد في الحد الثابت، بحيث يكون لك فرد حد ثابت خاص به، بينما يشترك الأفراد في معاملات المتغيرات المستقلة في النموذج، وفي حالتنا يكتب النموذج على الشكل التالي:

بحيث تمثل Di متغيرات صورية تدل عن الفرد (i) أي:

1 للجزائر و 0 غير ذلك، 1 لمصر و 0 غير ذلك... وهكذا.

وفي هذه الحالة يتم الاستغناء عن الحد الثابت في التقدير بطريقة المربعات الصغرى، أو الاستغناء عن أحد المتغيرات الصورية. وفي حالة عدم ذلك يقوم البرنامج تلقائيا بازالة احد المتغيرات الصورية.

ولتقدير النموذج LSDV يجب اضافة مصفوفة المتغيرات الصورية الى مصفوفة المتغيرات، ويمكن انشاء هذه المصفوفة تلقائيا عن طريق كتابة الأمر: tabulate codp, generate(D) فيظهر في مخرجات البرنامج الجدول التالي:

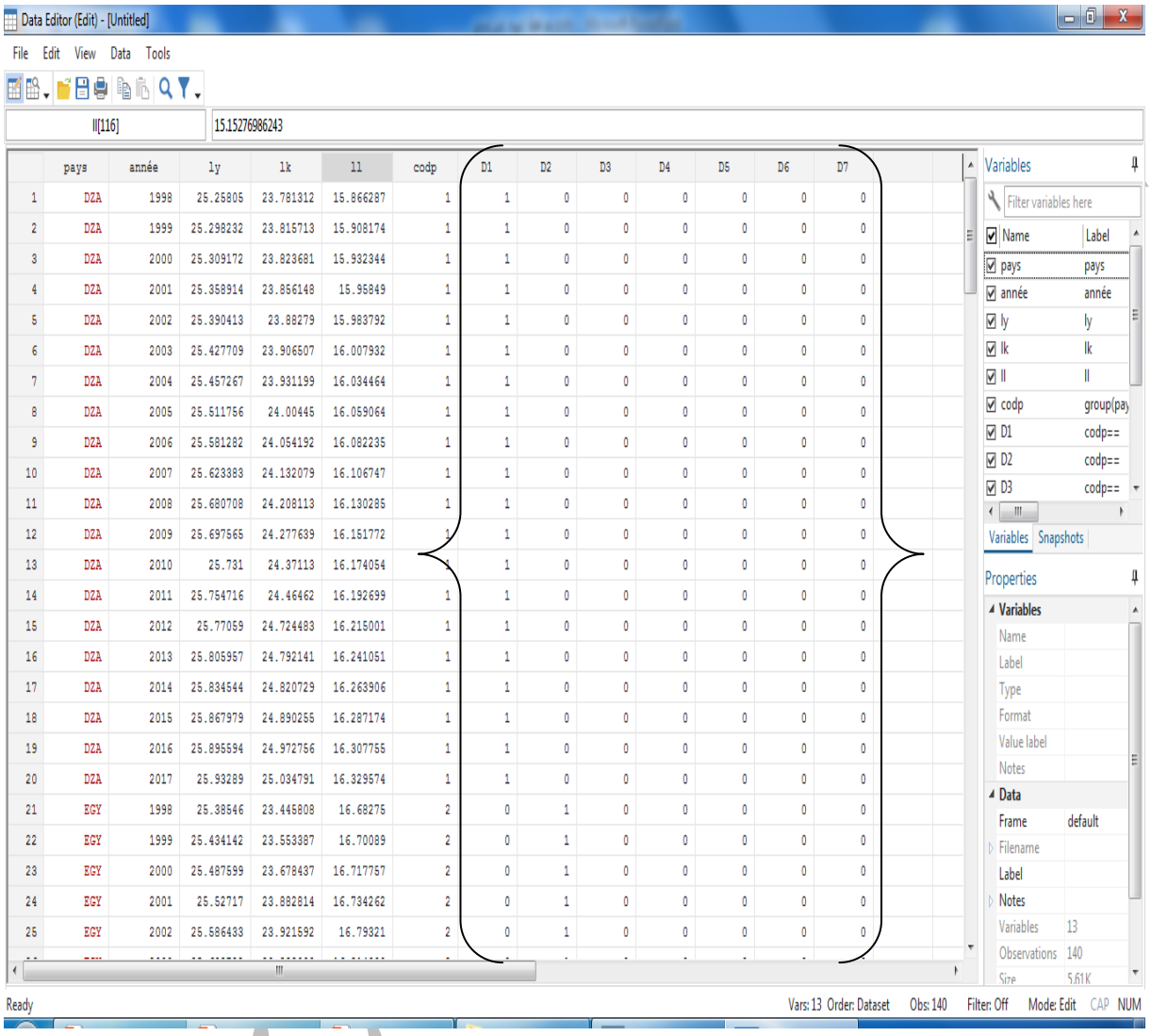
```
. tabulate codp, gen(D)
```

group(pays)	Freq.	Percent	Cum.
1	20	14.29	14.29
2	20	14.29	28.57
3	20	14.29	42.86
4	20	14.29	57.14
5	20	14.29	71.43
6	20	14.29	85.71
7	20	14.29	100.00
Total	140	100.00	

جدول يوضح انه تم انشاء 7 متغيرات صورية خاصة بالأفراد (البلدان لك منها 20 مشاهدة

[*] تجدر الاشارة هنا الى ان البرنامج يشترط في تحديد البعد المقطعي ان يكون معبر عنه بارقام وليس باسماء المجموعات أو الأفراد، ولهذا قمنا بترميز البلدان بالارقام في المتغير الجديد codp.

بينما في متصفح البيانات تظهر لدينا متغيرات جديدة الى جانب المتغيرات الموجودة من قبل تتمثل في المتغيرات الصورية، كما هي موضحة في ما يلي:



	pays	année	ly	lk	ll	codp	D1	D2	D3	D4	D5	D6	D7
1	DZA	1998	25.25805	23.781312	15.866287	1	1	0	0	0	0	0	0
2	DZA	1999	25.298232	23.815713	15.908174	1	1	0	0	0	0	0	0
3	DZA	2000	25.309172	23.823681	15.932344	1	1	0	0	0	0	0	0
4	DZA	2001	25.358914	23.856148	15.95849	1	1	0	0	0	0	0	0
5	DZA	2002	25.390413	23.88279	15.983792	1	1	0	0	0	0	0	0
6	DZA	2003	25.427709	23.906507	16.007932	1	1	0	0	0	0	0	0
7	DZA	2004	25.457267	23.931199	16.034464	1	1	0	0	0	0	0	0
8	DZA	2005	25.511756	24.00445	16.059064	1	1	0	0	0	0	0	0
9	DZA	2006	25.581282	24.054192	16.082235	1	1	0	0	0	0	0	0
10	DZA	2007	25.623383	24.132079	16.106747	1	1	0	0	0	0	0	0
11	DZA	2008	25.680708	24.208113	16.130285	1	1	0	0	0	0	0	0
12	DZA	2009	25.697565	24.277639	16.151772	1	1	0	0	0	0	0	0
13	DZA	2010	25.731	24.37113	16.174054	1	1	0	0	0	0	0	0
14	DZA	2011	25.754716	24.46462	16.192699	1	1	0	0	0	0	0	0
15	DZA	2012	25.77059	24.724483	16.215001	1	1	0	0	0	0	0	0
16	DZA	2013	25.805957	24.792141	16.241051	1	1	0	0	0	0	0	0
17	DZA	2014	25.834544	24.820729	16.263906	1	1	0	0	0	0	0	0
18	DZA	2015	25.867979	24.890255	16.287174	1	1	0	0	0	0	0	0
19	DZA	2016	25.895594	24.972756	16.307755	1	1	0	0	0	0	0	0
20	DZA	2017	25.93289	25.034791	16.329574	1	1	0	0	0	0	0	0
21	EGY	1998	25.38546	23.445808	16.68275	2	0	1	0	0	0	0	0
22	EGY	1999	25.434142	23.553387	16.70089	2	0	1	0	0	0	0	0
23	EGY	2000	25.487599	23.678437	16.717757	2	0	1	0	0	0	0	0
24	EGY	2001	25.52717	23.882814	16.734262	2	0	1	0	0	0	0	0
25	EGY	2002	25.586433	23.921592	16.79321	2	0	1	0	0	0	0	0

يتم تقدير نموذج LSDV بطريقة المربعات الصغرى العادية بكتابة الأمر:

reg ly D2 D3 D4 D5 D6 D7 lk ll

بحيث قمنا بازالة المتغير الصوري المتعلق بالبلد الأول من أجل امكانية التقدير، لكن يمكن اضافة العبارة noconstant لتقدير النموذج بدون حد ثابت مع استخدام كل المتغيرات الصورية.

كما يتم الحصول على نفس النتائج بكتابة الأمر :

reg ly lk ll i.codp

فنحصل على نفس النتائج بدون انشاء المتغيرات الصورية

ويتم تفسير معاملات المتغيرات الصورية بأنها الحد الثابت الخاص بكل بلد. بحيث: D1=cons، بينما الحدود الأخرى فهي عبارة عن الحد المقدر بهذه الطريقة مضاف اليه الحد الثابت cons المقدر:

Di=Di+cons، مثلا: D2=2.963828-0.4920529=2.4717751

وبالضغط على Entrer ينتج لدينا نتائج التقدير على الشكل التالي:

```
. reg ly D2 D3 D4 D5 D6 D7 lk ll
```

Source	SS	df	MS	Number of obs	=	140
				F(8, 131)	=	18028.10
Model	314.495766	8	39.3119708	Prob > F	=	0.0000
Residual	.285657767	131	.002180594	R-squared	=	0.9991
				Adj R-squared	=	0.9990
Total	314.781424	139	2.26461456	Root MSE	=	.0467

ly	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
D2	-.4920529	.0616548	-7.98	0.000	-.6140208	-.3700851
D3	.8949307	.0338796	26.42	0.000	.8279088	.9619526
D4	1.081171	.0420589	25.71	0.000	.997968	1.164373
D5	1.14723	.0378546	30.31	0.000	1.072344	1.222115
D6	.1360435	.0435457	3.12	0.002	.0498998	.2221873
D7	.3693601	.0382523	9.66	0.000	.2936878	.4450323
lk	.3184143	.0262087	12.15	0.000	.2665673	.3702613
ll	.9255516	.065636	14.10	0.000	.795708	1.055395
_cons	2.963828	.6234848	4.75	0.000	1.730427	4.19723

الحد الثابت الخاص بكل بلد

3. تقدير نموذج الآثار الثابتة FEM:

في هذه المرحلة نقوم بتقدير نموذج الآثار الثابتة أو ما يعرف بنموذج الانحدار داخل المجموعات، والذي يعتبر أن الاختلاف بين الأفراد يكون في الحد الثابت، بحيث يكون الحد الثابت في في هذا النموذج عبارة عن متوسط معلمات المتغيرات الصورية المقدرة في نموذج LSDV.

يتم تقدير نموذج الآثار الثابتة ببرنامج Stata بكتابة الأمر:

xtreg ly lk ll, fe

بحيث تعبر العبارة fe عن الأثر الثابت

وبالضغط على Entrer نتحصل على النتائج التالية:

```
. xtreg ly lk ll, fe

Fixed-effects (within) re                       Number of obs      =       140
Group variable: codp                            Number of groups   =         7

R-sq:                                           Obs per group:
     within  = 0.9437                                         min =        20
     between = 0.9076                                         avg =      20.0
     overall = 0.9066                                         max =        20

                                                F(2,131)           =   1097.56
corr(u_i, Xb)  = 0.6009                         Prob > F           =    0.0000

------------------------------------------------------------------------------
          ly |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          lk |   .3184143   .0262087    12.15   0.000     .2665673    .3702613
          ll |   .9255516   .0       
       _cons |   3.411926   .6       
-------------+----------------------------------------------------------------
     sigma_u |  .61610049
     sigma_e |  .04669683
         rho |  .99428806   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(6, 131) = 256.59                     Prob > F = 0.0000
```

معامل التحديد بين المجموعات، داخل المجموعات والكلي

معامل الارتباط بين المتغيرات المستقلة والخطأ العشوائي

نسبة تباين الخطأ العشوائي بالنسبة للتباين الكلي

اختبار تجانس الأفراد

من خلال نتائج التقدير تظهر لنا معنوية المعلمات المقدرة للنموذج، لكن يظهر ان معامل الارتباط بين حد الخطأ والمتغيرات المستقلة موجب وكبير نوعا ما وهذا مخالف لفرضية العدم التي تقر بعدم وجود ارتباط.

بينما تظهر القيمة المحسوبة لاختبار تجانس الأفراد، بمعنى اختبار تساوي معاملات الحد الثابت بحد ثابت واحد أكبر من تلك المجدولة وبالتالي نقبل فرضية عدم تجانس الأفراد، وبالتالي نقر بان نموذج الأثر الثابت هو الأفضل عن النموذج التجميعي.

ملاحظة: لغرض اجراء اختبار المفاضلة بين نموذج الآثار الثابتة ونموذج الآثار العشوائية يتوجب علينا تخزين نتائج التقدير لنموذج FEM بكتابة الأمر:

estimates store fixe

4. تقدير نموذج الآثار العشوائية REM:

يعتبر هذا النموذج أن الاختلاف بين الأفراد لا يعود الى الاختلاف في الحد الثابت للأفراد، وانما الى حد عشوائي يتم تدقيره ضمن الخطأ العشوائي للنموذج ولهذا يعرف بنموذج الخطأ المركب.

ويتم تقدير هذا النموذج بطريقة المربعات الصغرى المعممة بدل طريقة المربعات الصغرى العادية.

وباستعمال البرنامج Stata يمكن تقدير هذا الأخير بكتابة الأمر:

Xtreg ly lk ll,re

وبتطبيق الأمر نتحصل على النتائج التالية:

```
Random-effects GLS regression                   Number of obs      =       140
Group variable: codp                            Number of groups   =         7

R-sq:                                           Obs per group:
     within  = 0.9417                                         min =        20
     between = 0.9339                                         avg =      20.0
     overall = 0.9320                                         max =        20

                                                Wald chi2(2)       =   1737.07
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

ly	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lk	.3904879	.0306317	12.75	0.000	.3304509	.4505249
ll	.827602	.07508	11.02	0.000	.680448	.974756
_cons	3.21742	.7355977	4.37	0.000	1.775675	4.659165
sigma_u	.15053973					
sigma_e	.04669683					
rho	.91222433	(fraction of variance due to u_i)				

5. المفاضلة بين النماذج:

تتعدد الاختبارات المستعملة في المفاضلة بين مختلف نماذج البانل، ونخص بالذكر:

- **اختبار والد wald test:** والذي يستخدم للمفاضلة بين نموذج الاثار الثابتة والنموذج التجميعي باختبار فرضية تساوي معاملات الحد الثابت في نموذج FEM، وبالتالي تحقق الفرضية يعني النموذج التجميعي هو الأفضل، وهذا الاختبار يقوم به Stata تلقائيا كما تمت الاشارة اليه سابقا من خلال تقدير نموذج الآثار الثابتة في آخر جدول التقدير معبر عنه باختبار فيشر.
- **اختبار هوسمان Hausman:**

يستخدم هذا الاختبار للمفاضلة بين نموذج الآثار الثابتة ونموذج الآثار العشوائية، ويتم اجراء الاختبار ببرنامج Stata عن طريق كتابة الأمر:

Hausman fe

لكن هذا يكون عبر اتباع الخطوات المذكورة سابقا أي:

1. تقدير نموذج الاثر الثابت: xtreg ly lk ll, fe
2. تخزين نتائج التقدير: estimate store fe
3. تقدير نموذج الاثار العشوائية: xtreg ly lk ll, re
4. اختبار هوسمان: hausman fe

وبتطبيق الأمر الأخير يظهر لنا نتائج الاختبار كما يوضحه الجدول التالي:

```
. hausman fixe

                ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       fixe          .       Difference          S.E.
-------------+----------------------------------------------------------------
          lk |    .3184143     .3904879      -.0720736               .
          ll |    .9255516      .827602       .0979496               .
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =   -24.09    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test
```

الا أنه من خلال هذه النتائج، تظهر قيمة الاختبار سالبة وتقدر بـ -24.09 وهذا غير منطقي وراجع الى خلل في البيانات المستعملة كونها اختيرت من أجل توضيح الخطوات الاساسية التي يتم اتباعها لتقدير نماذج البانل باستخدام البرنامج Stata.

وبالعودة الى الاسس النظرية للاختبار يشترط ان تكون مصفوفة التباين والتباين المشترك للمعلمات المقدرة محددة موجبة، وان المعلمات المقدرة في نموذج الاثر الثابت أكبر من مثيلتها في نموذج الاثار العشوائية، وعدم توفر ذلك نتج عنه قيمة سالبة للاختبار.

5. بعض الاختبارات حول بيانات البانل:

ان تقدير النماذج القياسية يستدعي اجراء بعض الاختبارات الاحصائية الى جانب اختبارات الفرضيات الخاصة بالمعلمات المقدرة والمفاضلة بين النماذج القياسية المقدرة. ونجد من بين هذا الاختبارات اختبارات الفرضيات الاساسية للنموذج.

5-1. اختبار التوزيع الطبيعي للبواقي (الأخطاء العشوائية):

ان اغلب النماذج القياسية تنطلق اساسا من فرضية التوزيع الطبيعي لبواقي التقدير أو ما يعرف بالأخطاء العشوائية ، وهي الفرضية المعبر عنها بالمتوسط المعدوم والتباين الثابت للبواقي، ويتم اختبار هذه الفرضية بالاختبار المشهور لجارك بيرا. ويقوم هذا الاختبار على اختبار الفرضية التالية:

H_0: البواقي تتبع التوزيع الطبيعي.

H_1: البواقي لا تتبع التوزيع الطبيعي.

ويعتمد هذا الاختبار اساسا على اختبار تساوي معامل الالتواء بالقيمة 0 ومعامل التفلطح بالقيمة 3، وهي القيم التي تميز التوزيع الطبيعي المعتدل.

ولاختبار هذه الفرضية باستخدام البرنامج Stata يجب أولا تقدير قيم البواقي بكتابة الأمر:

predict residu

ليقوم البرنامج بانشاء سلسلة البواقي، ومن ثم اجراء الاختبار بكتابة الأمر:

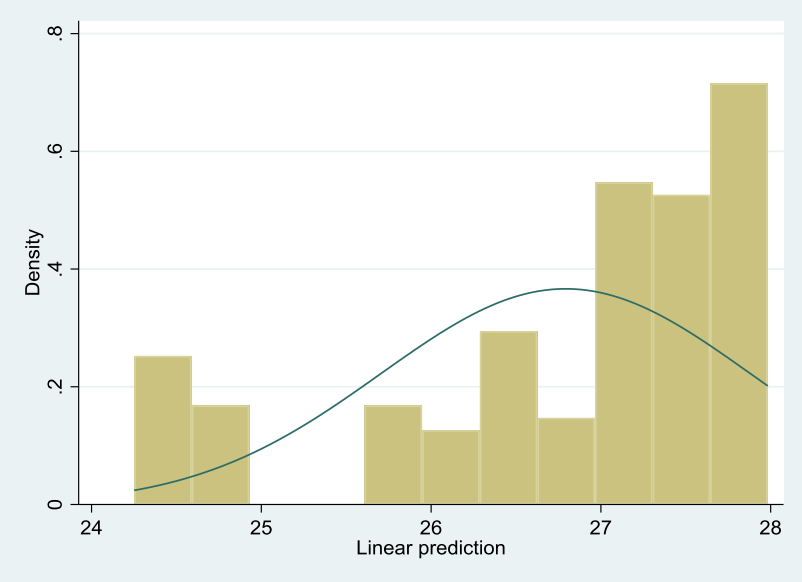
sktest residu

وبافتراض ان النموذج الافضل هو نموذج الاثر الثابت بالعودة الى المثال السابق، تكون نتائج الاختبار على النحو التالي:

```
. sktest residu

                    Skewness/Kurtosis tests for Normality
                                                      ------- joint ------
    Variable |   Obs  Pr(Skewness)  Pr(Kurtosis) adj chi2(2)   Prob>chi2
-------------+-------------------------------------------------------------
      residu |   140      0.0000        0.9294        16.40        0.0003
```

من خلال نتائج الاختبار، لاحظ أن القيمة الاحتمالية لنسب الالتواء والتفلطح معا تساوي الى 0.0003 وهي أقل من 0.05 مما يقودنا الى قبول الفرضية البديلة للاختبار مفادها ان البواقي لا تتوزع طبيعيا، بالرغم من القيمة الاحتمالية لنسبة الالتواء تقريبا معدومة.

ويمكن الاستدلال بالمدرج التكراري للبواقي باستخدم الأمر histogram residu لعرض المدرج التكراري على النحو التالي:



المدرج التكراري يوضح البواقي لا تتوزع طبيعيا.

2-5. اختبار الاثر العشوائي للافراد:

يقوم هذا الاختبار باختبار فرضية الخطأ المركب في نموذج الأثر العشوائي، ويُعتمد في ذلك على اختبار Breusch-Pagan أو مضاعف لاغرانج، ويقوم على اختبار الفرضية التالية:

$$H_0 : \sigma_u^2 = 0$$

$$H_1 : \sigma_u^2 \neq 0$$

بحيث: σ_u^2 تعبر عن تباين الخطأ المرتبط بالافراد.

ويتم اجراء الاختبار في حالة تحقق فرضية نموذج الأثر العشوائي هو الافضل، وبالتالي يتم كتابة الأمر xttest0 بعد عملية تقدير نموذج الاثر العشوائيعن طريق الأمر: xtreg ly lk ll, re.

وبافتراض نموذج الأثر العشوائي في مثالنا هو الافضل نقوم باجراء الاختبار بكتابة الامر xttest0 فينتج لدينا النتائج التالية:

```
Breusch and Pagan Lagrangian multiplier test for random effects

        ly[codp,t] = Xb + u[codp] + e[codp,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                      ly |   2.264615       1.504864
                       e |   .0021806       .0466968
                       u |   .0226622       .1505397

        Test:   Var(u) = 0
                                  chibar2(01) =   318.62
                              Prob > chibar2 =   0.0000
```

القيمة المحسوبة لمضاعف لاغراج LM-test

القيمة المحسوبة لمضاعف لاغرانج من خلال اختبار Breusch-pgen تتبع توزيع كاي تربيع بدرجة حرية واحدة، والملاحظ هنا ان القيمة الاحتمالية للاختبار أقل من 0.05 مما يقودنا الى قبول الفرضية البديلة التي تقر بوجود اختلاف في تباين الاخطاء حسب الافراد.

ملاحظة:

حاولنا من خلال هذا المثال توضيح طريقة استخدام البرنامج Stata في تحليل بيانات البانل بالرغم من عدم توفر بعض الشروط في بعض الحالات، بحيث كان الهدف هو تناول مختلف الأومر الاساسية التي من شانها تقدير النماذج واجراء بعض الاختبارات من خلال معطيات واقعية دون مراعاة اطار نظري محدد. ويمكن للطالب الاستعانة بمختلف المراحل والاوامر وتطبيقها على بيانات واقعية والتوسع اكثر في هذا المجال الواسع والذي يعتبر هذا الجزء نقطة بداية له.

اعداد الاستاذ بوعزار رابح

خاتمة

حاولنا من خلال هذه المطبوعة توضيح بعض المفاهيم والطرق الاحصائية الأكثر استخداما من قبل الطلبة والباحثين في ميدان العلوم الاقتصادية، وبالاستناد إلى البرنامج المقرر من طرف الوزارة الوصية لطلبة السنة الثانية ماستر تخصص اقتصاد اقتصاد كمي في مقياس برمجية Stata، تناولنا بعض تقنيات التحليل الاحصائي للبيانات والتي تعتبر جزء بسيط من الطرق والأساليب الكثيرة التي يمكن تطبيقها باستخدام البرنامج Stata، والتي تستدعي الالمام الواسع بالجوانب النظرية لأساليب التحليل الاحصائي للبيانات وطرق القياس الاقتصادي. بحيث اعتمدنا على أمثلة بسيطة تساعد الطالب على الفهم وإعادة التدريب عليها بطريقة سهلة، وبالتالي فتح المجال لتعميمها على مختلف البيانات الممكن توفرها في البحوث والدراسات التطبيقية.

وقد اعتمدنا استخدام الاصدار 14 من البرنامج Stata، كونه يتميز بنوع من التطور في معالجة البيانات مقارنة بالإصدارات السابقة له، لكن نشير الى أنه كل الإصدارت صالحة وتقوم بنفس الوظائف، ويبقى الفرق في بعض الاضافات يغلبها الجانب الشكلي وإضافة بعض التقنيات في تنفيذ الأوامر Command.

يبقى هذا العمل عبارة عن بداية لأعمال أخرى للتوسع أكثر في استعمال البرنامج Stata وبرامج أخرى في معالجة البيانات وتحليلها بطرق متقدمة. ونرجو من زملائنا الاساتذة وطلبتنا الاعزاء اشعارنا بأي ملاحظات او أخطاء لاحظوها في هذه المطبوعة من أجل العمل على تصحيحها.

نسال الله لنا ولكم التوفيق والسداد

قائمة المراجع:

1. جيلالي جلاطو: "**الاحصاء مع تمارين ومسائل محلولة**"، ديوان المطبوعات الجامعية، الجزائر، طبعة 2001.

2. رائد ادريس محمود الخفاجي وعبد الله مجيد حميد العتابي: " **الوسائل الاحصائية في البحوث التربوية والنفسية** (مفهومها، أهميتها وتطبيقاتها باستخدام الحقيبة الاحصائية SPSS)"، دار دجلة للنشر، الأردن، الطبعة الأولى 2015.

3. سعد زغلول بشير: "**دليلك الى البرنامج SPSS الاصدار العاشر**"، **المعهد العربي للتدريب والبحوث** الإحصائية، بغداد، 2003.

4. غيث البحر ، معن التنجي: " **التحليل الإحصائي للاستبيانات باستخدام برنامج IBM SPSS Statistics**"، مركز سبر للدرسات الإحصاييية رالسياسات العامة، تركيا، 2014.

5. محمد شامل بهاء الدين فهمي: **"الاحصاء بلا معاناة المفاهيم مع التطبيقات باستخدام البرنامج SPSS"** مركز البحوث بمعهد الادارة العامة، المملكة العربية السعودية، الجزء الأول والثاني، 2005.

6. Andrew Rutherford : «***Introducing Anova And Ancova A Glm Approach***», SAGE Publications, First published, London, 2001.

7. CHENG HSIAO : «Analysis of Panel Data», Second Edition, Cambridge University Press, New York, 2003.

8. Damodar N. Gujarati: « ***Basic Econometrics*** », Fourt Edition, The McGraw Hill Companies, 2004.

9. George G. Judge, William E. Griffiths, R. Carter Hill Helmut Liitkepohl Tsoung-Chao Lee : "The Theory and practice of econometrics", John Wiley & Sons 1985.

10. Gérald Baillargeon, Fernando ouellet : «*Analyse des données avec SPSS pour Windows* » ; Les Edition SMG, Canada, 2008.

11. Lawrence C. Hamilton : « ***Statistics With Stata*** », Version 12,BROOKS/COLE, 2013.

12. Lucien LEBOUCHER ; Marie-José VOISIN : « ***Introduction à la statistique descriptive Cours et exercices avec tableur*** » ; CÉPADUÈS-ÉDITIONS, France , 2011.

13. MAXWELLE Scott E., Harold D. DELANEY. (2004), «***Designing Experiments and analysis Data, a model comparison perspective*** », second édition, TLFeBook, édition Mahwah, New jersey, London.

14. **Régis Bourbonnais : « *Économétrie , Cours et exercices corrigés* », 9e Edition, DANOD, 2015.**

15. Sophia Rabe-Hesketh, Anders Skrondal: «***Multilevel and Longitudinal Modeling Using Stata***» ; 2 Vols. stata Press (2012)

16. Sophia Rabe-Hesketh, Brian Everitt: «***A Handbook of Statistical Analyses using Stata***»; *Second Edition,* Chapman & Hall/CRC; 1999.