المحاضرة الخامسة:

استخراج التوزيع الطبيعي و القيم المتطرفة باستخدام برنامج SPSS

تمهيد :

يعد التوزيع الطبيعي للبيانات أحد الخطوات المهمة لاختيار الوسيلة الإحصائية المناسبة في البيانات الكمية ،مثلما اشرنا سابقا و هناك عدة طرق للتأكد من التوزيع الطبيعي للبيانات حيث يجب أولا ادراك بعض الأمور قبل التطرق في لكيفية استخراج التوزيع الطبيعي و علاقته باختيار الوسيلة الإحصائية تبعا لكل إحصاء (معلمي و غير معلمي) .

كما ان برنامج SPSS يعطي ميزة اخرة اثناء استخراج التوزيع الطبيعي من الخيار Explorer هو استخراج القيم المتطرفة ،و هي قيم تكون بعيدة جدا عن متوسط البيانات و تؤثر في قيمة المتوسط و في نتائج التحليل الاحصائي و هناك العديد من الباحثين يغفلون هاذين الخطوتين اثناء التحليل في المستوى الكمي وهو ما يؤثر أيضا على النتائج .

1-في حالة البيانات تخضع للتوزيع الطبيعي مع عدم وجود قيم متطرفة :

في عديد من الحالات قد تؤثر القيم المتطرفة على التوزيع الطبيعي لذا سوف نعالج عديد الحالات للتعرف على أهمها و كيفية التعامل مع كل حالة باستخدام برنامج SPSS

- **مثال عملي :**

في احدى الدراسات تحصل أحد الباحثين على مجموعة من البيانات و أراد التأكد من التوزيع الطبيعي وخلو البينات من القيم المتطرفة باستخدام برنامج SPSS :

الرياضة	ضغوط	الرياضة	ضغوط	الرياضة	ضغوط	الرياضة	ضغوط	الرياضة	ضغوط
كرة القدم	210	كرة القدم	180	كرة اليد	110	الكرة الطائرة	136	الكرة الطائرة	125
كرة القدم	135	كرة القدم	195	كرة اليد	120	الكرة الطائرة	140	الكرة الطائرة	95
كرة القدم	130	كرة القدم	170	كرة اليد	150	الكرة الطائرة	174	الكرة الطائرة	115
كرة القدم	205	كرة القدم	100	كرة اليد	150	الكرة الطائرة	140	الكرة الطائرة	160
كرة القدم	150	كرة القدم	110	كرة اليد	165	الكرة الطائرة	119	الكرة الطائرة	155
كرة القدم	155	كرة القدم	220	كرة اليد	175	الكرة الطائرة	160	الكرة الطائرة	120
كرة القدم	110	كرة القدم	150	كرة اليد	150	الكرة الطائرة	125	الكرة الطائرة	150
كرة القدم	147	كرة القدم	200	كرة اليد	185	الكرة الطائرة	105	الكرة الطائرة	90
كرة القدم	180	كرة القدم	160	كرة اليد	145	الكرة الطائرة	90	الكرة الطائرة	185
كرة القدم	160	كرة القدم	165	كرة اليد	135	الكرة الطائرة	100	الكرة الطائرة	155
كرة القدم	155	كرة القدم	135	كرة اليد	140	الكرة الطائرة	135	الكرة الطائرة	145
كرة القدم	180	كرة القدم	140	كرة اليد	110	الكرة الطائرة	145	الكرة الطائرة	175
كرة القدم	195	كرة اليد	120	كرة اليد	160	الكرة الطائرة	145	الكرة الطائرة	140
كرة القدم	155	كرة اليد	185	كرة اليد	150	الكرة الطائرة	125	الكرة الطائرة	150
كرة القدم	166	كرة اليد	175	كرة اليد	155	الكرة الطائرة	115	الكرة الطائرة	180
كرة القدم	150	كرة اليد	180	الكرة الطائرة	170	الكرة الطائرة	165	الكرة الطائرة	135
كرة القدم	180	كرة اليد	100	الكرة الطائرة	154	الكرة الطائرة	175	الكرة الطائرة	130
كرة القدم	150	كرة اليد	145	الكرة الطائرة	104	الكرة الطائرة	150	الكرة الطائرة	115

المطلوب :التأكد من التوزيع الطبيعي وخلو البينات من القيم المتطرفة :

المرحلة الأولى: ادخال البيانات :

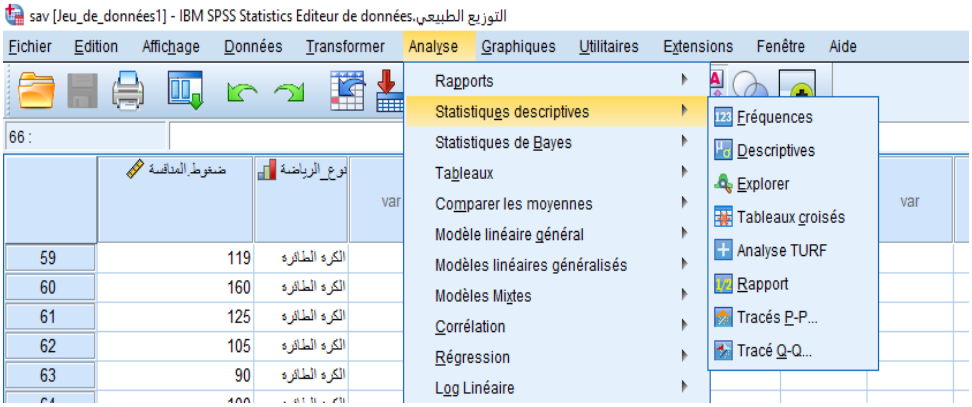
– و تتم وفق ما تناولناه سابقا .

المرحلة الثانية : استخراج النتائج

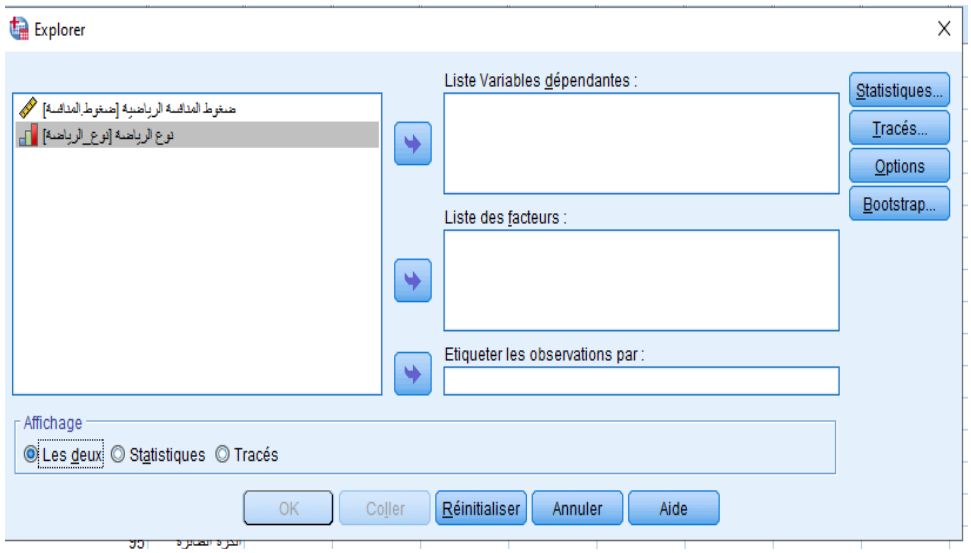
1– اضغط على Analyse

2– ثم Statistique descriptives

3– ثم أضغط على Explorer



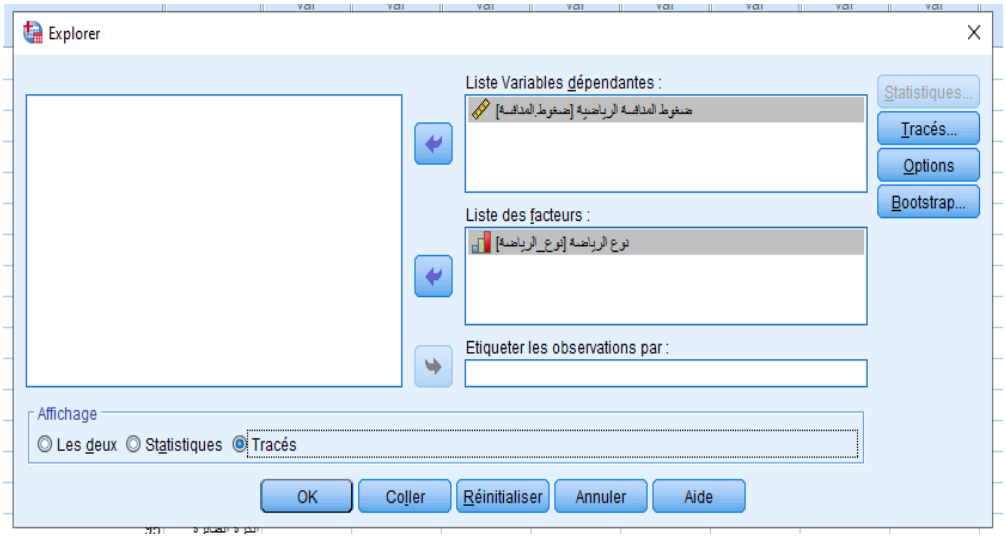
4– بعد الضغط على Explorer تظهر لنا قائمة:



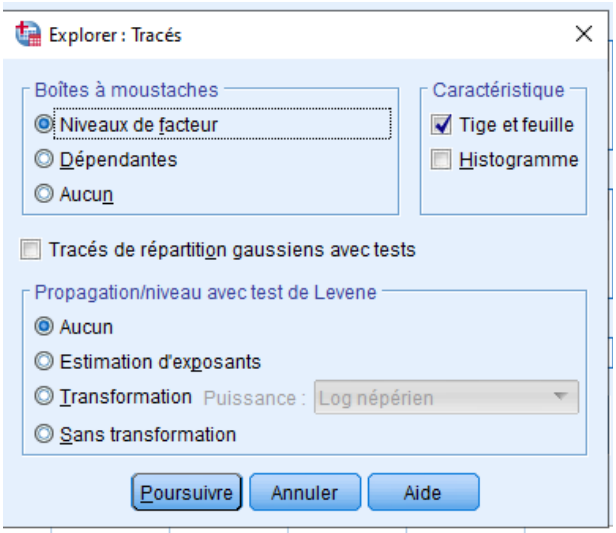
5– أقوم بنقل متغير ضغوط المنافسة الى مربع: Liste variables dépendantes

6– أقوم بنقل متغير نوع الرياضة الى مربع :Liste des facteurs

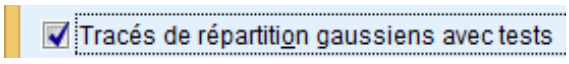
7– من Affichage اسفل الشاشة أقوم باختيار Tracés



8- ثم اضغط على الزر Tracé يمين الشاشة فتظهر لنا نافذة اخرى:



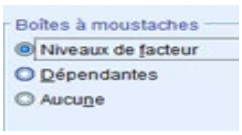
9- اذهب و افعل اختبار التوزيع الطبيعي في المربع الصغير اما العبارة



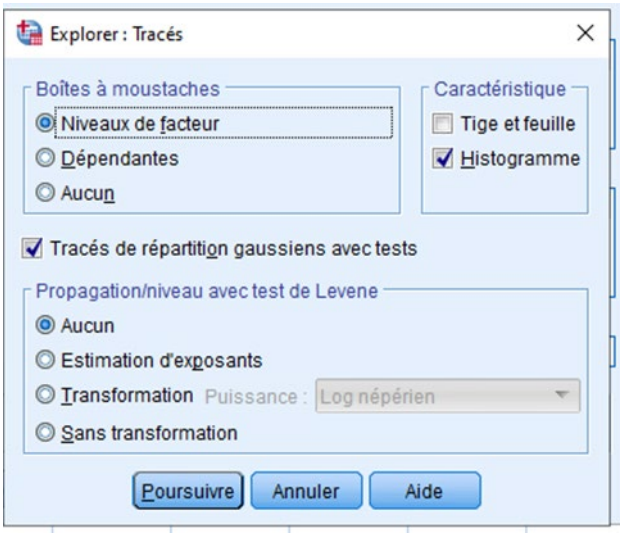
10- من قائمة Caractéristique اختار Histogramme

ان اردنا استخراج التوزيع التكراري (في حالة عدم رغبتنا في استخراج الاشكال فاني انزع الإشارة)

11- نذهب الى القائمة Boites à moustaches وهذا لاستخراج شكل الصندوق الخاص بمعرفة القيم المتطرفة:



- في حالة رغبتنا في معرفة القيم المتطرفة نختار Niveaux de facteur
- في حالة عدم رغبتنا في معرفة القيم المتطرفة اختار aucune



12- ثم اضغط على poursuivre اسفل الشاشة للعودة الى القائمة السابقة

13- اعود الى الشاشة الرئيسية أقوم بالضغط على OK

14- تظهر لنا النتائج في صفحة المخرجات (Sortie) فاقوم بتخزينها في مجلد معين وفق إجراءات التخزين المناسبة.

15-

المرحلة الثانية: نقل النتائج والتعليق عليها

- الجدول الأول يبين عدد افراد العينة حسب كل رياضة، و عدد القيم المفقودة.

Récapitulatif de traitement des observations

		Observations					
		Valide		Manquant		Total	
	نوع الرياضة	N	Pourcentage	N	Pourcentage	N	Pourcentage
ضغوط المنافسة الرياضية	كرة القدم	30	100,0%	0	0,0%	30	100,0%
	كرة اليد	21	100,0%	0	0,0%	21	100,0%
	الكرة الطائرة	39	100,0%	0	0,0%	39	100,0%

الجدول الثاني خاص باختبار التوزيع الطبيعي و هو اهم جدول هنا

Tests de normalité

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	نوع الرياضة	Statistiques	ddl	Sig.	Statistiques	ddl	Sig.
ضغوط المنافسة الرياضية	كرة القدم	,086	30	,200*	,978	30	,775
	كرة اليد	,122	21	,200*	,950	21	,343
	الكرة الطائرة	,065	39	,200*	,975	39	,528

*. Il s'agit de la borne inférieure de la vraie signification.

a. Correction de signification de Lilliefors

و يمكن كتابته باللغة العربية في متن الدراسة او البحث كما يلي :

جدول يوضح اختبار التوزيع الطبيعي لـ (kolmodrove–Smirnove و Shapiro–Wilk) لمتغير ضغوط المنافسة لدى عينة الدراسة تبعا لمتغير أو عامل نوع الرياضة الممارسة

نوع الرياضة	Kolmogorov–Smirnov		القرار	Shapiro–Wilk		القرار
	الاحصائية	Sig		الإحصائية	Sig	
كرة القدم	0,086	0,200*	يخضع	0,978	0,775	يخضع
كرة اليد	0,122	0,200*	يخضع	0,950	0,343	يخضع
الكرة الطائرة	0,065	0,200*	يخضع	0,975	0,528	يخضع

في حال :

- في حالة كانت الاختبارات غير دالة احصائيا أي ان القيمة الاحتمالية Sig أكبر من 0.05 فهذا يدل على ان البيانات تخضع للتوزيع الطبيعي .
- في حالة كانت الاختبارات دالة احصائيا أي ان القيمة الاحتمالية Sig أقل من 0.05 فهذا يدل على ان البيانات تخضع للتوزيع الطبيعي .

في حالتنا هذه فان جميع القيم الاحتمالية للدلالة Sig أكبر من 0.05 و بالتالي فان البيات الضغوط عند لاعبي الرياضات الثلاثة (كرة القدم ،كرة اليد ،الكرة الطائرة) تخضع للتوزيع الطبيعي ،و باتالي يمكن استخدام الاختبارات المعلمية (البارامترية)

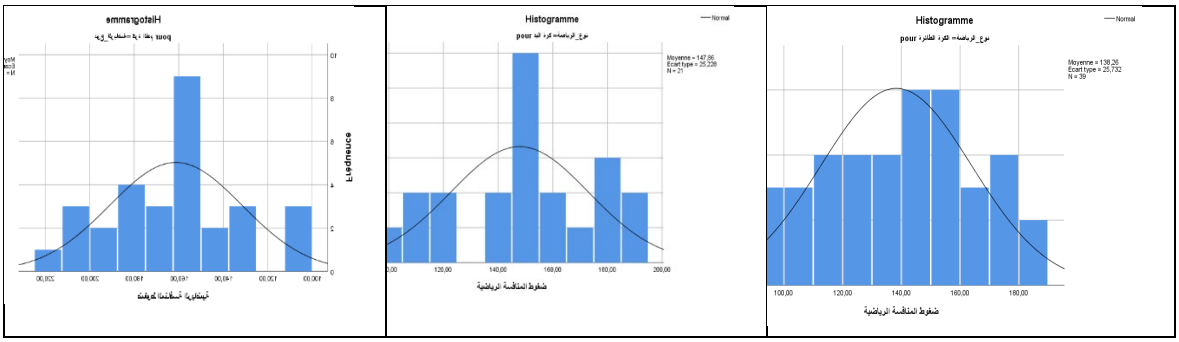
و يمكننا التعليق على النتائج كما يل :

لتعليق: من خلال الجدول رقم (...) يتبين مايلي :

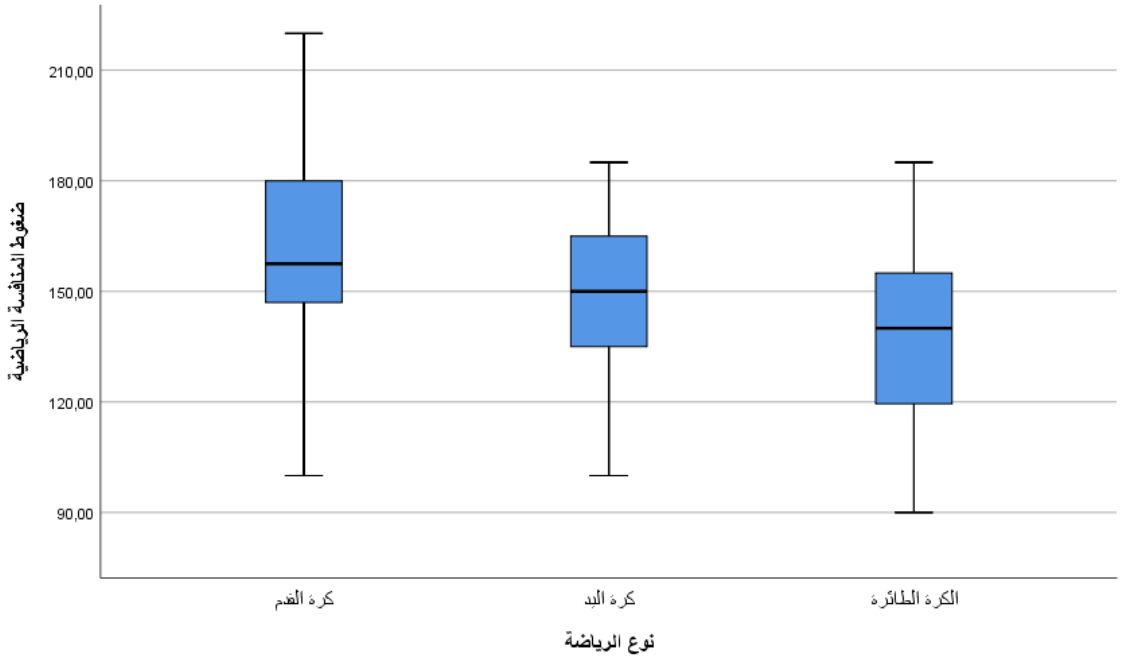
- البيانات الخاصة بمتغير ضغوط المنافسة لدى لاعبي كرة القدم تخضع للتوزيع الطبيعي و هذا لعدم وجود دلالة إحصائية ، لان القيمة الاحتمالية المرموز لها بالرمز Sig اكبر من مستوى الدلالة α =0,05 لكل من اختبار Kolmogorov–Smirnov واختبار Shapiro–Wilk

- البيانات الخاصة بمتغير ضغوط المنافسة لدى لاعبي كرة اليد تخضع للتوزيع الطبيعي و هذا لعدم وجود دلالة إحصائية ، لان القيمة الاحتمالية المرموز لها بالرمز Sig اكبر من مستوى الدلالة α =0,05 لكل من اختبار Kolmogorov-Smirnov واختبار Shapiro-Wilk
- البيانات الخاصة بمتغير ضغوط المنافسة لدى لاعبي الكرة الطائرة تخضع للتوزيع الطبيعي وهذا لعدم وجود دلالة إحصائية، لان القيمة الاحتمالية المرموز لها بالرمز Sig أكبر من مستوى الدلالة α =0,05 لكل من اختبار Kolmogorov-Smirnov واختبار Shapiro-Wilk
- ومنه نستنتج ان بيانات العينات الثلاث لمتغير ضغوط المنافسة تخضع للتوزيع الطبيعي ويمكن اجراء عليها مختلف الوسائل الإحصائية الخاصة بالاحصاء المعلمي (البارمتري)

اشكال تمثل منحنيات التوزيع الطبيعي للمتغير ضغوط المنافسة بالنسبة للعينات الثلاثة



شكل يمثل رسم الصندوق للكشف عن القيم المتطرفة :



التعليق: من خلال الشكل يتبين أنه لا توجد أي قيم متطرفة حيث لا توجد أي نجمة ولا دائرة صغيرة (* أو○) تحمل رقم اعلى او أسفل الصندوق ،وهذا الرقم هو رقم المعلمة للقيمة المتطرفة.

و منه نستنتج ان البيانات الخاصة بالعينات الثلاث لا تحوي على أي قيم متطرفة.

ملاحظة هامة :احيانا تظهر لنا قيم متطرفة في متغير او متغيرين ،او في مجموعة دون اخرى هنا يجب معالجة القيم المتطرفة (بالحذف مثلا للمعلمة) ثم اعادة التاكد من التوزيع الطبيعي لان عملية الحذف قد تؤثر على النتائج السابقة وهذا يذكر اثناء البحث

2-في حالة وجود قيم متطرفة لا تؤثر على التوزيع الطبيعي :

لنفترض ان باحث اراد معرفة العلاقة بين العمر التدريبي و مستوى قلق المنافسة لدى الرياضيين فاختار عينة عشوائية تكونت من 40 رياضي و قام باختبار للتوزيع الطبيعي و كذلك قام بالكشف عن القيم المتطرفة قبل القيام بخطوات الارتباط :

فتحصل على ما يلي من مخرجات برنامج SPSS :

في هذا المثال نلاحظ ان هناك متغيرين كميين ،و كما اشرنا سابقا يجب التأكد من التوزيع الطبيعي و القيم المتطرفة للمتغيرين و هنا عند القيام بالعملية ينقل متغير العمر التدريبي ،و قلق المنافسة الى مربع (Liste variables dépendantes) و هذا في الخطوة الخامسة (5) من التسلسل الذي تطرقنا له :

فتحصلنا على النتائج التالية بعد عملية التحليل :

جدول رقم () يبين اختبار التوزيع الطبيعي لمتغيري العمر التدريبي و قلق المنافسة الرياضية

Tests de normalité

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistiques	ddl	Sig.	Statistiques	ddl	Sig.
قلق المنافسة	,103	40	,200*	,961	40	,177
العمر التدريبي	,104	40	,200*	,980	40	,694

*. Il s'agit de la borne inférieure de la vraie signification.

a. Correction de signification de Lilliefors

من خلال جدول المخرجات يتبين أن :

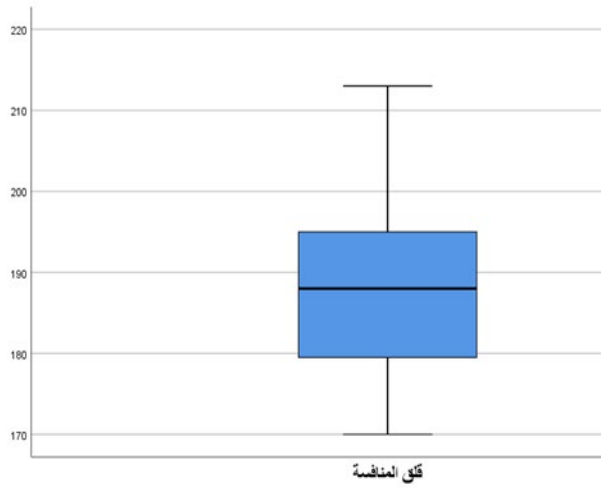
- القيم الخاصة بمتغير قلق المنافسة تخضع للتوزيع الطبيعي حيث اختبار Kolmogorov–Smirnov غير دال احصائيا وهذا لان القيمة الاحتمالية Sig تساوي 0.200 وهي اكبر من مستوى

الدلالة 0.05 ،وهو ما يؤكده كذلك اختبار Shapiro–Wilk الذي بدوره غير دال احصائيا حيث ان قيمة Sig للاختبار تساوي 0.177 وهي اكبر من مستوى الدلالة 0.05.

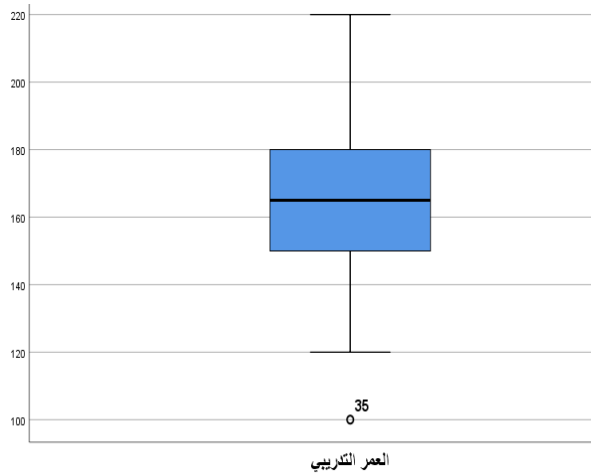
- القيم الخاصة بمتغير العمر التدريبي تخضع للتوزيع الطبيعي حيث اختبار Kolmogorov–Smirnov غير دال احصائيا وهذا لان القيمة الاحتمالية Sig تساوي 0.200 وهي اكبر من مستوى الدلالة 0.05 ،وهو ما يؤكده كذلك اختبار Shapiro–Wilk الذي بدوره غير دال احصائيا حيث ان قيمة Sig للاختبار تساوي 0.694 وهي اكبر من مستوى الدلالة 0.05.

- بعد التأكد من التوزيع الطبيعي ننتقل الى الخطوة الثانية و هي التأكد من عدم وجود القيم المتطرفة و هو ما يبينه الشكلين التاليين:

شكل رقم(01) اختباررسم الصندوق للمعرفة وجود القيم المتطرفة لمتغير قلق المنافسة



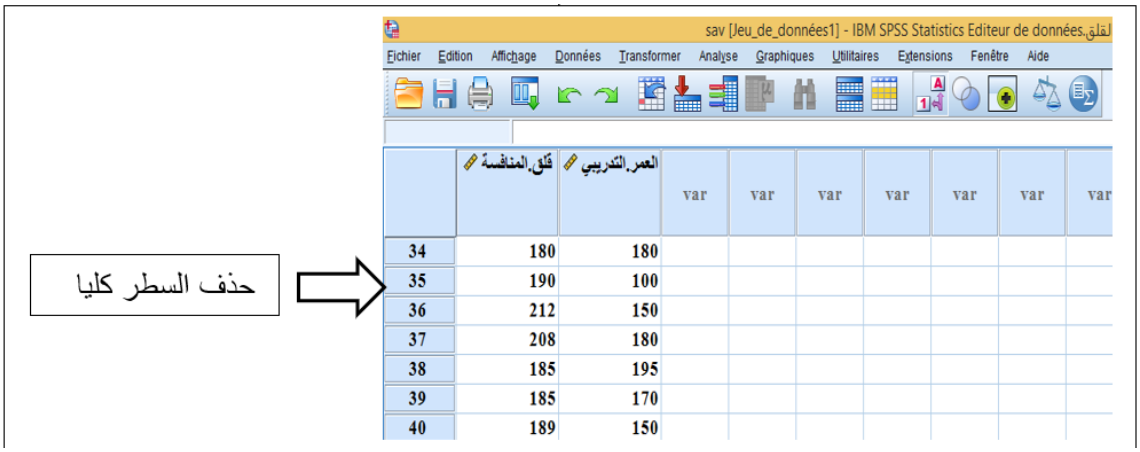
شكل رقم(01) اختباررسم الصندوق للمعرفة وجود القيم المتطرفة لمتغير العمر التدريبي



- من خلال الشكل رقم 01 الذي يمثل اختبار رسم الصندوق للكشف عن القيم المتطرفة لمتغير قلق المنافسة يتبين لنا عدم وجود أي قيمة متطرفة حيث لا نلاحظ وجود أي دائرة او نجمة (* أو○) اسفل او اعلى الصندوق .

- من خلال الشكل رقم 02 الذي يمثل اختبار رسم الصندوق للكشف عن القيم المتطرفة لمتغير العمر التدريبي يتبين لنا ان هناك قيمة متطرفة دنيا حيث نلاحظ وجود دائرة صغيرة بجنبها الرقم 35 و وهو رقم المعلمة (او الصف/السطر) الذي يحوي القيمة المتطرفة في العمر التدريبي ،و هي قيمة متطرفة دنيا لانها اسفل الصندوق هنا و بعد العودة للصف تبين ان هذه القيمة هي 100 شهر وهو يبعد عن المتوسط الحسابي المقدر (163.75) بـ (63.5 شهر) أي 5 سنوات تقريبا .

- و هناك عدة اجراءات لمعالجة القيم المتطرفة الا اننا سوف نتطرق فقط الى طريقة الحذف ثم اعادة الاختبارات



	قلق.المنافسة	العمر.التدريبي
34	180	180
35	190	100
36	212	150
37	208	180
38	185	195
39	185	170
40	189	150

- بعد حذف السطر نعيد الاختبارات السابقة للتأكد من التوزيع الطبيعي و عدم ظهور قيم متطرفة أخرى لانه في بعض الأحيان قد يخلق الحذف بعض المشكلات .

Tests de normalité

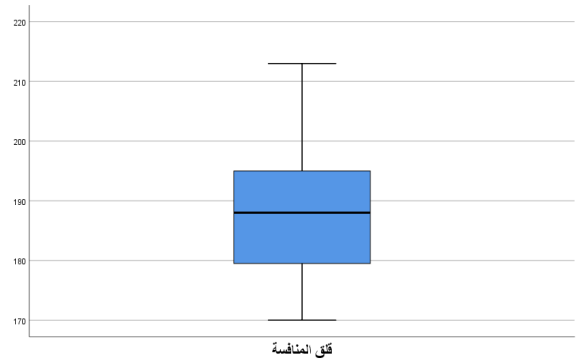
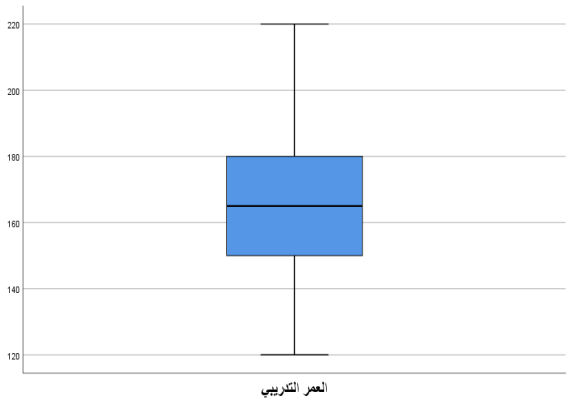
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistiques	ddl	Sig.	Statistiques	ddl	Sig.
قلق المنافسة	,100	39	,200*	,960	39	,177
العمر التدريبي	,104	39	,200*	,973	39	,450

*. Il s'agit de la borne inférieure de la vraie signification.

a. Correction de signification de Lilliefors

- يتبين من خلال الجدول ان القيم الخاصة بمتغير قلق المنافسة بعد حذف القيم المتطرفة تخضع للتوزيع الطبيعي حيث اختبار Kolmogorov–Smirnov غير دال احصائيا وهذا لان القيمة الاحتمالية Sig تساوي 0.200 وهي اكبر من مستوى الدلالة 0.05 ،وهو ما يؤكده كذلك اختبار Shapiro–Wilk الذي بدوره غير دال احصائيا حيث ان قيمة Sig للاختبار تساوي 0.177 وهي اكبر من مستوى الدلالة 0.05.

- كما يتبين ان القيم الخاصة بمتغير العمر التدريبي بعد حذف القيم المتطرفة تخضع للتوزيع الطبيعي حيث اختبار Kolmogorov–Smirnov غير دال احصائيا وهذا لان القيمة الاحتمالية Sigتساوي 0.200 وهي اكبر من مستوى الدلالة 0.05 ،وهو ما يؤكده كذلك اختبار Shapiro–Wilk الذي بدوره غير دال احصائيا حيث ان قيمة Sig للاختبار تساوي 0.450 وهي اكبر من مستوى الدلالة 0.05.

شكل رقم(01) اختبارسم الصندوق للمعرفة وجود القيم المتطرفة لمتغير قلق المنافسة بعد حذف المعلمة 35	شكل رقم(01) اختبارسم الصندوق للمعرفة وجود القيم المتطرفة لمتغير العمر التدريبي بعد حذف المعلمة 35
	

من خلال الشكلين رقم (01) و (02) يتبين أنه لا يوجد قيم متطرفة أخرى بعد الحذف الأول للمعلمة رقم (35) و هذا نظرا لعدم وجود أي علامة سواء الدائرة او النجمة (* أو○)

<u>ملاحظة هامة:</u> بما اننا حذفنا المعلمة رقم 35 هذا يعني انني قد حذفنا الفرد المقابل لهذه المعلمة من العينة ككل حيث ان العينة كان عددها 40 لاعبا ،و بالتالي ستصبح بعد الحذف 39 لاعبا ،و هنا من الضروري العودة الى عنصر (عينة ومجتمع الدراسة) و التعديل في حجم العينة مع ذكر السبب ،و المتمثل في الحذف بسبب القيم المتطرفة التي تؤثر على نتائج الدراسة

3-في حالة وجود قيم متطرفة مؤثرة في التوزيع الطبيعي.

قد يقع الباحث في بعض التناقضات عند قراءة الجداول و من بينها ما سوف نراه في المثال الموالي حول قيام أحد الباحثين باستعمال اختباري (Kolmogorov-Smirnov وShapiro-Wilk) للتاكد من التوزيع الطبيعي للبيانات الخاصة بضغوط مهنة التدريس لدى أساتذة التربية البدنية و الرياضية مقسمين بحسب المراحل التعليمية ابتدائي ،متوسط (الاعدادي) ،و الثانوي ،فتحصل على ما يلي : مثلا قد نتحصل على جدول التالي لاختبار اختبار التوزيع الطبيعي لـ (Kolmogorov-Smirnov وShapiro-Wilk):

Tests de normalité

	المرحلة التعليمية	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistiques	ddl	Sig.	Statistiques	ddl	Sig.
ضغوط مهنة التدريس	مرحلة الابتدائي	,129	27	,200*	,963	27	,422
	مرحلة المتوسط	,175	31	,016	,951	31	,166
	مرحلة الثانوي	,108	32	,200*	,983	32	,880

نلاحظ من خلال الجدول ما يلي:

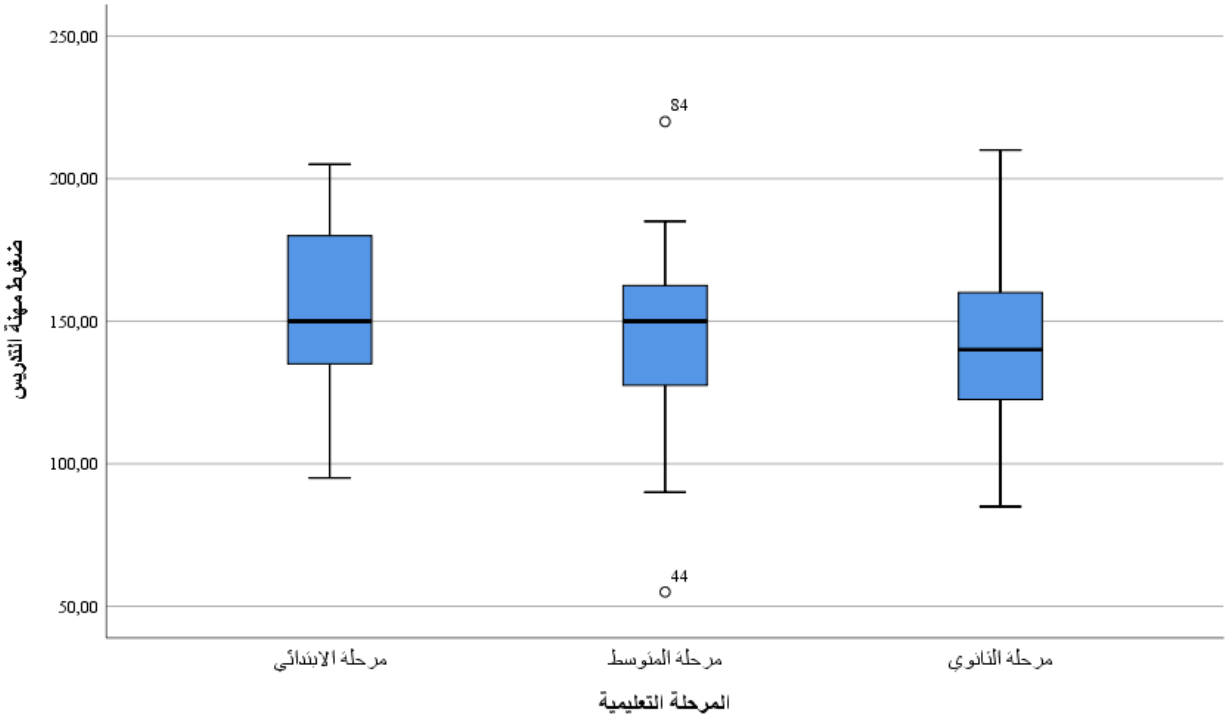
1- كل من العينات الخاصة بالمرحلة الابتدائية و العينات و العينات الخاصة بالمرحلة المتوسطة تخضع للتوزيع، و هذا لان اختبار **Kolmogorov-Smirnov** ،و اختبار **Shapiro-Wilk** غير دالان احصائيا ،و هذا لأن قيمة Sig للاختبارين اكبر من مستوى الدلالة α =0,05

2- فيما يخص العينة الخاصة بالمرحلة المتوسطة نلاحظ من خلال الجدول ان اختبار **Kolmogorov-Smirnov** دال احصائيا وهذا لان قيمة Sig= **0,016** وهي أقل من مستوى الدلالة α =0,05 وهو ما يدل على عدم اعتدالية توزيع البيانات الخاصة بهذه العينة، ولكن في المقابل نجد ان اختبار **Shapiro-Wilk** يبين ان البيانات تخضع للتوزيع الطبيعي وهذا لان قيمة Sig=0,166 وهي أكبر من مستوى الدلالة α =0,05 وهنا باي الاختبارين نأخذ.

- **اكتشاف القيم المتطرفة او الشاذة:**

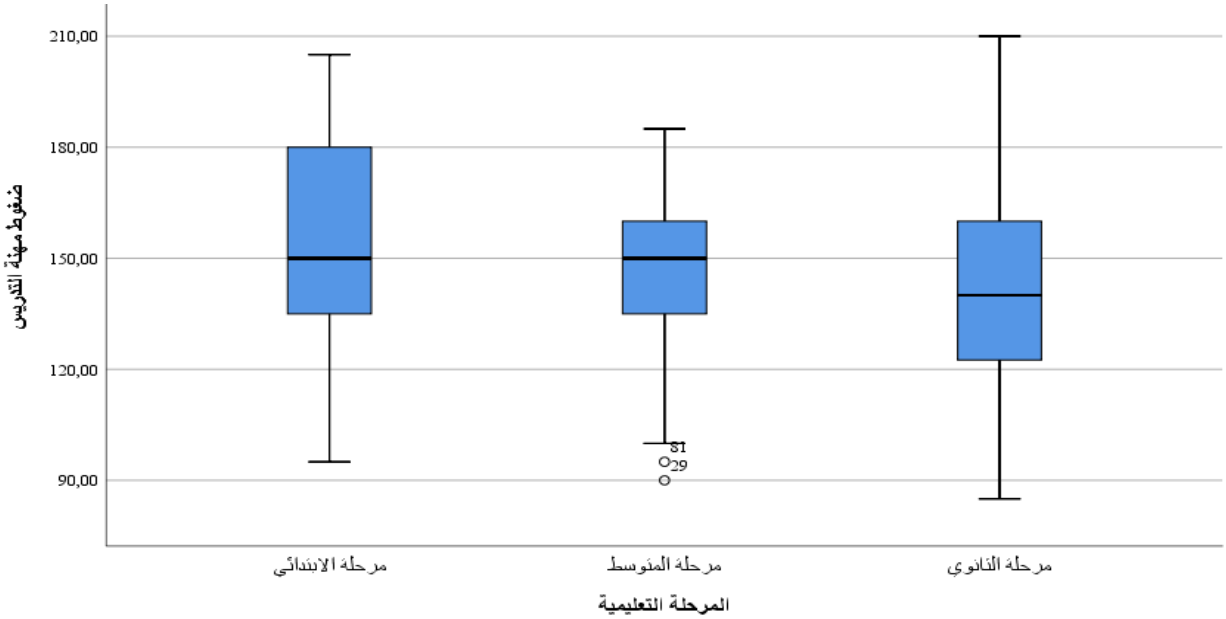
قد نتحصل في بعض المعالجات على قيم متطرفة او شاذة يتوجب علينا معالجتها والا سوف تتأثر النتائج بها والمثال التالي يبين ان أحد الباحثين قام باختبار رسم الصندوق وتحصل على بعض القيم الشاذة:

شكل يوضح رسم الصندوق الخاصة بضغوط مهنة التدريس بحسب الخبرة المهنية والتي قسمناها الى ثلاث فئات أو عينات:



نلاحظ من خلال الشكل وجود معلمتين شاذتين و هو الفرد الذي يحمل الترتيب 44 وهو الذي لديه قيمة متطرفة دنيا ،و الشخص الذي يحمل الترتيب 84 الذي لديه قيمة متطرفة قصوى او عليا ،و هناك عدة طرق لمعالجة القيم الشاذة او المتطرفة منها الحذف ،و لكن مع الحذر و هذا لإمكانية بروز قيم شاذة اخرى ،كما قد تؤثر على اعتدالية التوزيع و سوف أقوم بهذه العملية أي الحذف :

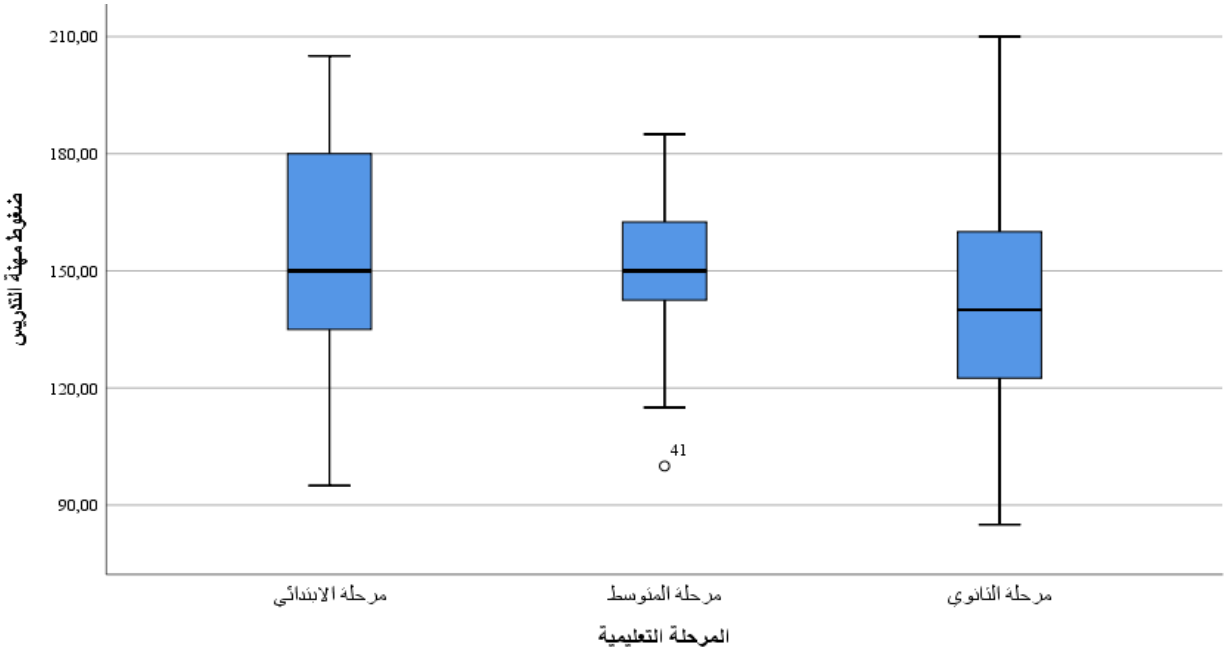
شكل يوضح رسم الصندوق الخاصة بضغوط مهنة التدريس بحسب الخبرة المهنية والتي قسمناها الى ثلاث فئات أو عينات بعد الحذف الاول:



نلاحظ بعد الحذف الاول بروز قيم شاذة اخرى و هو ما حذرنا منه سابقا و نلاحظ هنا بروز الفرد الذي يحمل الترتيب 81 و الفرد الذي يحمل الترتيب 29

و سوف نقوم بعملية حذف للمرة الثانية:

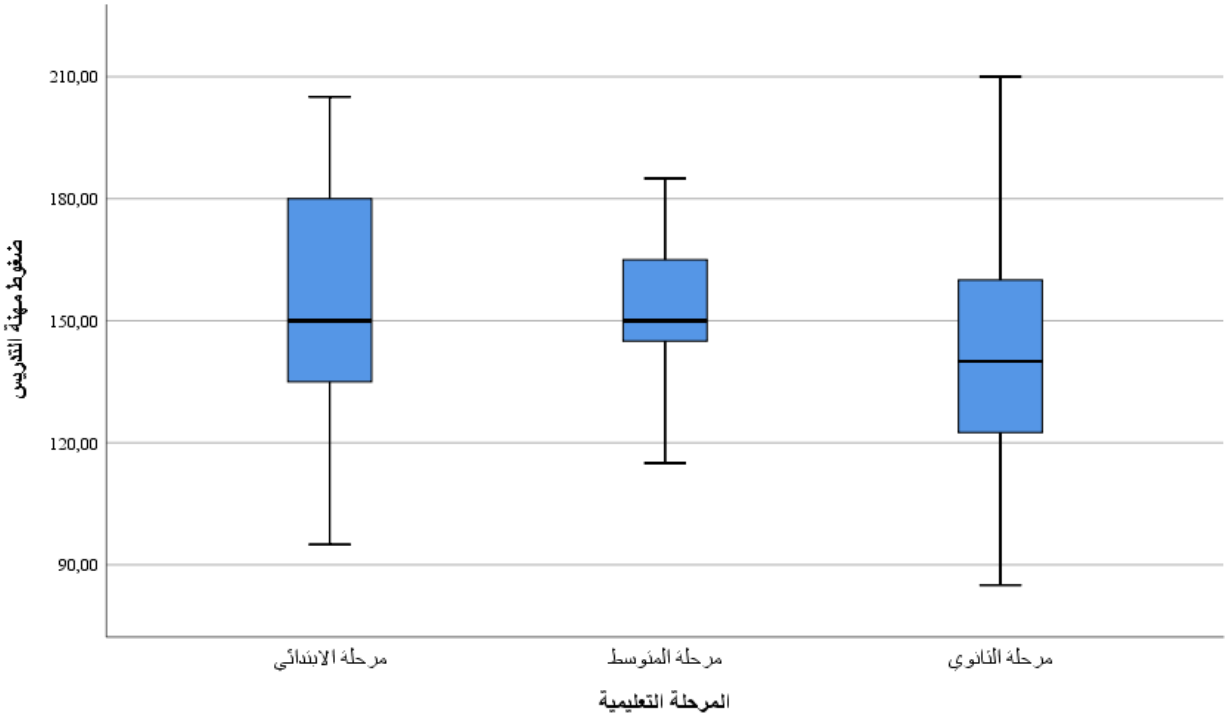
شكل رقم (..) يوضح رسم الصندوق Boxplot الخاصة بضغوط مهنة التدريس بحسب الخبرة المهنية و التي قسمناها الى ثلاث فئات أو عينات بعد الحذف الثاني:



نلاحظ بعد الحذف الاول بروز قيم شاذة اخرى و هو ما حذرنا منه سابقا و نلاحظ هنا بروز الفرد الذي يحمل الترتيب 41 بعد الحذف الثاني

و سوف نقوم بعملية حذف للمرة الثالثة :

شكل رقم يوضح رسم الصندوق الخاصة بضغوط مهنة التدريس بحسب الخبرة المهنية و التي قسمناها الى ثلاث فئات أو عينات بعد الحذف الثالث :



بعد الحذف الثالث نلاحظ انه لم يعد لدينا أي قيم شاذة أو متطرفة ويمكننا القيام بعملية المعالجة مع ملاحظة ان عدد افراد العينة الخاصة بمن لديهم خبرة (من 6 الى 10 سنوات) قد نقص عددهم من 31 الى 26

وهنا لا يجب ان يؤثر الحذف من اجل القيم المتطرفة على عدد العينة المناسب.

ملاحظة :

في حالة عدم وجود قيم متطرفة مع عدم وجود توزيع طبيعي فإننا نلجأ للإحصاء لا معلمي .

خلاصة :

حاولنا في هذه المحاضرة اظهار خطوات استخراج التوزيع الطبيعي للبيانات باستخدام برنامج SPSS و كيفية قراءة المخرجات ،من جداول ورسومات بيانية ،كما حاولنا تبيين مختلف الحالات التي تؤثر على التوزيع الطبيعي و المتعلقة بالقيم المتطرفة و بعض طرق معالجتها و كيف نغير من توزيع البيانات .

❖ **أسئلة التقويم :**

- لما نلجأ لاختبار التوزيع الطبيعي ؟
- ماهي خطوات استخراج التوزيع الطبيعي باستخدام برنامج SPSS ؟
- هل هناك طرق أخرى تمكننا من استخراج التوزيع الطبيعي غير الاختبارات التي ذكرناها في المحاضرة ؟
- كيف تؤثر القيم المتطرفة (الشاذة) على التوزيع الطبيعي ؟
- ماهي أهم عيوب طريقة الحذف (القيم المتطرفة) ؟