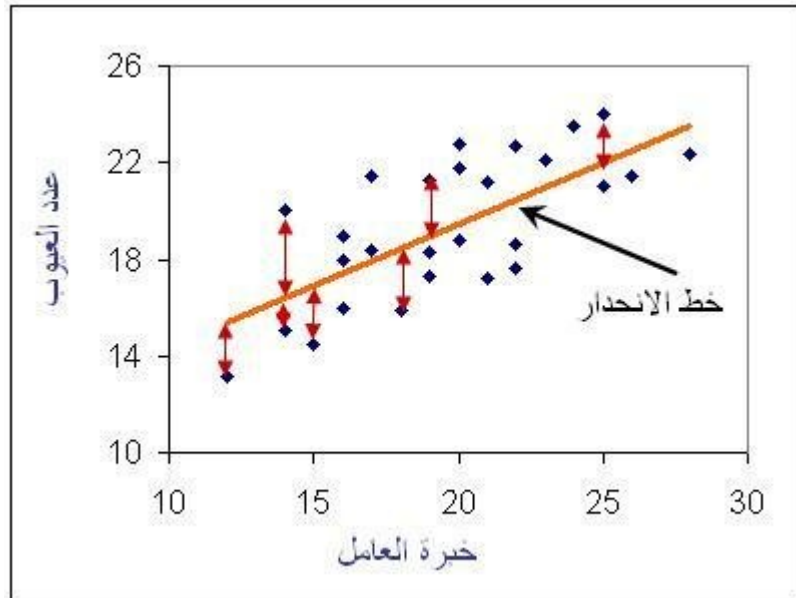


تحليل الانحدار

نستكمل في هذه المقالة موضوع تحليل الانحدار الخطي Linear Regression فنستعرض فرضيات تحليل الانحدار وكيفية التحقق منها. والأسلوب المستخدم للتأكد من تحقق هذه الفرضيات هو تحليل البواقي.

البواقي Residuals

ما هي البواقي؟ البواقي Residuals هي الفرق بين القيمة التي نحسبها من نموذج الانحدار والقيمة الحقيقية. فمثلا إذا قمنا بتحليل الانحدار لحجم المبيعات بناء على سعر البيع والجودة وعدد منافذ البيع فالباقي هو الخطأ في النموذج. فعند مقارنة حجم المبيعات لإحدى الحالات المعلومة لدينا بنتيجة النموذج فإننا نجد فارقا بينهما وهذا الفارق هو الخطأ في النموذج أو الباقي. يمكن أن نقول أن وجود هذا الفارق أو الخطأ هو من طبيعة تحليل الانحدار فمن النادر أن يكون تحليل الانحدار صحيحا بنسبة مائة بالمائة.



هذا الشكل يوضح البواقي عند القيام بتحديد خط الانحدار أي تحديد العلاقة الخطية بين خبرة العامل وعدد العيوب في المنتج. البيانات الأساسية التي سجلناها هي عبارة عن النقاط الزرقاء المبعثرة. وعندما رسمنا خط الانحدار وهو الخط المستقيم باللون البرتقالي فإنه لا ينطبق بطبيعة الحال على كل النقاط. ولذلك فهناك فارق بين البيانات المسجلة وعدد العيوب التي سننتجها من خط الانحدار أو معادلته. هذا الفارق هو الفارق بين كل نقطة من البيانات الأساسية وقيمة خط الانحدار. فعلى سبيل المثال فإنه إذا كانت خبرة العامل هي 25 عاما فإن البيانات المسجلة تبين أن عدد العيوب في المنتج كانت 24 تقريبا ولكننا لو استخدمنا خط الانحدار لوجدناه يعطينا قيمة مختلفة وهي 21 تقريبا. الفارق بين القيمة الحقيقية وتلك التي نحصل عليها من معادلة الانحدار أو خط الانحدار هو الباقي وهو بالنسبة لهذه النقطة يساوي $24 - 21 = 3$.

عندما قررنا أن نستخدم تحليل الانحدار الخطي فإننا افترضنا أن العلاقة بين عدد العيوب وخبرة العامل هي علاقة خطية أي أن عدد العيوب = ثابت + معامل * خبرة العامل بالسنين + خطأ. فنحن نفترض علاقة خط مستقيم. لذلك فإن أي شيء يبين أن العلاقة ليست خطية فهو ببساطة يهدم فرضنا الأساسي وبالتالي يجعلنا نبحث عن طريقة أخرى لدراسة العلاقة بين المتغيرين.

فرضيات تحليل الانحدار:

ينبغي تحليل الانحدار على عدة فرضيات لا بد أن نضمن صحتها عند إجراء هذا التحليل. هذه الفرضيات هي:

1- علاقة خطية Linearity بمعنى أن العلاقة هي علاقة خط كستقيم وليس خطا منحنيا

2- التجانس Homoscedasticity ومعناه ثبات التغير (التباين) في قيمة البواقي. عندما يكون هناك تجانس فإن البواقي ستكون متساوية إلى حد ما عند جميع القيم أو بمعنى آخر لن نلاحظ اتجاه لزيادة أو نقصان البواقي مع تغير قيمة المتغير المستقل. فمثلا لو حاولنا دراسة العلاقة بين حجم المبيعات وسعر البيع فإننا لن نلاحظ أن البواقي تتجه للزيادة مع زيادة سعر البيع.

3- استقلالية البواقي Independence of Residuals بمعنى أن الباقي لأي نقطة لا يعتمد على الباقي في النقطة أو النقاط السابقة. عندما تكون البواقي غير مستقلة فإننا نحتاج أن نستخدم نمودجا آخر يأخذ في الاعتبار هذه العلاقة.

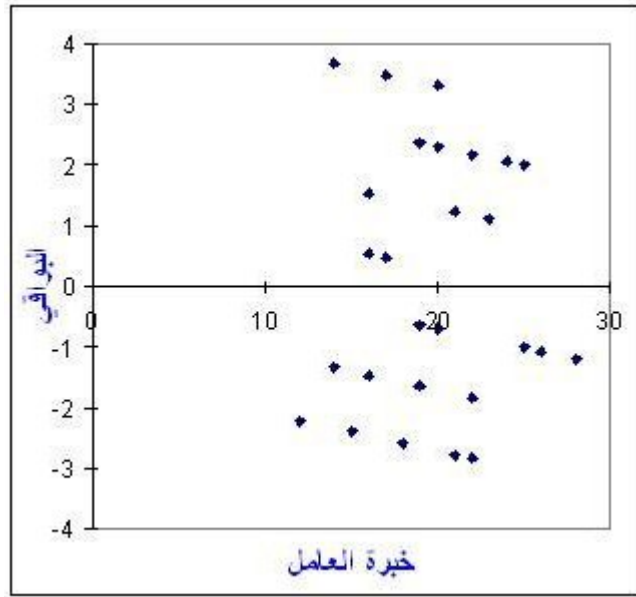
4- التوزيع الطبيعي للبواقي Normality of Residuals يبني تحليل الانحدار على أن البواقي موزعة توزيعاً طبيعياً عند كل النقاط للمتغير المستقل مثل سعر البيع. وهذا يعني أنها تتغير من سالب لموجب حول قيمة الصفر بشكل توزيع طبيعي وبحيث يكون مجموعها صفراً.

هذه هي الفرضيات باختصار وسوف نزيد الأمر وضوحاً في الأقسام التالية.

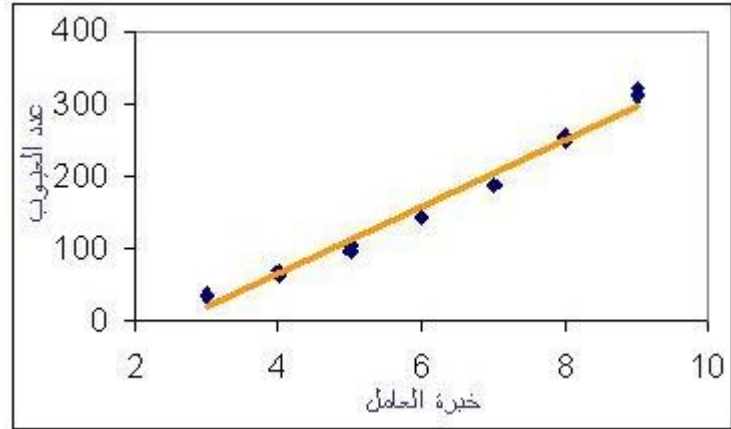
التأكد من تحقق فرضيات تحليل الانحدار:

لكي نتأكد من أن البيانات التي ندرسها تخضع للفرضيات التي نفترضها في تحليل الانحدار الخطي فإننا نلجأ لدراسة البواقي فنرسم مجموعة من الرسوم البيانية التي تبين تحقق هذه الفرضيات من عدمه.

العلاقة بين البواقي وكل متغير مستقل: لا بد أن تظهر هذه العلاقة كنقاط مبعثرة بشكل عشوائي في الاتجاهين السالب والموجب بدون وجود أي شكل أو منحنى. الشكل أدناه يبين هذه العلاقة للمثال السابق. لاحظ أن النقاط لا تأخذ شكلاً محدداً وهو ما يعني أن الخطأ هو خطأ عشوائي.



انظر إلى الشكل أدناه والذي يبين علاقة أخرى بين عدد العيوب وخبرة العامل. إن العلاقة هنا تأخذ شكلاً مختلفاً ولكن ربما لا تلاحظ شيئاً مزعجاً في خط الانحدار. وعندما ندرس نتائج تحليل الانحدار نجد أن النتائج مرضية فنسبة R square كبيرة وقيمة P صغيرة جداً ولكن دعنا ننظر لشكل البواقي.

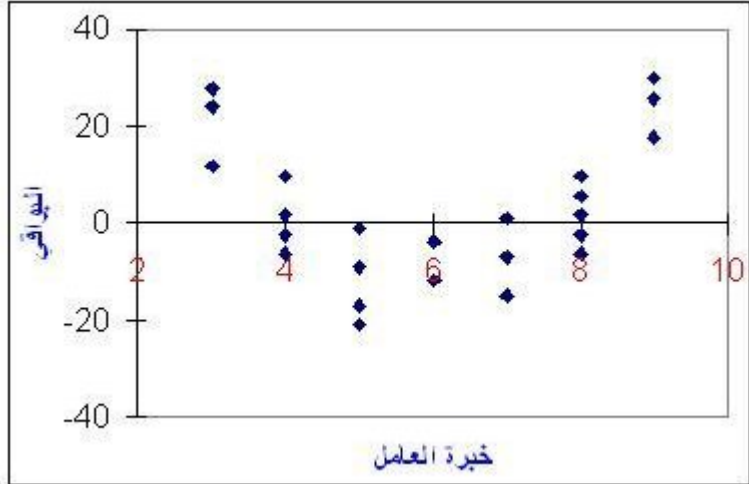


SUMMARY OUTPUT

Regression Statistics		
Multiple R	0.989311682	
R Square	0.978737604	
Adjusted R Square	0.977950108	
Standard Error	13.98105733	
Observations	29	

	Coefficients	P-value
Intercept	-113.938623	2.18E-14
خبرة العامل	45.33436406	4.07E-24

الشكل أدناه يبين العلاقة بين الفوائد وخبرة العامل. المحور س يبين خبرة العامل والمحور ص يبين الباقي المناظر لها. ماذا نلاحظ في هذا الشكل؟ إن هناك اتجاه واضحاً للفوائد فهي كانت موجبة ثم سالبة ثم موجبة مرة أخرى. هذا يعني أن العلاقة هي علاقة غير خطية. فلو كانت العلاقة خطية لما وجدنا هذا الاتجاه الواضح للبواقي ولو جدنا البواقي مبعثرة بشكل عشوائي. ففي مثالنا هذا يكون الأخذ بالنموذج الخطي هو أمر غير صحيح لأن العلاقة في الحقيقة غير خطية.



ماذا لو أخذنا بهذه العلاقة

الخطية وأهملنا مشكلة البواقي؟ في هذه الحالة نكون قد استخدمنا معادلة غير مناسبة وهذا يعني أننا لو استخدمنا هذا النموذج لتوقع عدد العيوب المناظرة لخبرة عامل ما فإن هناك خطأ في التقدير. ربما ستقول لي ولكن العلاقة البيانية بين خط الانحدار والنقاط الأصلية لا يبدو كبيرا في الشكل الأول؟ هذا صحيح ولكن مقياس الرسم يتدخل في هذا الأمر. لو نظرت إلى رسم البواقي لوجدت أن الخطأ في تقدير عدد العيوب المناظرة لخبرة عامل 3 سنوات يصل إلى أكثر من عشرين. هل هذا خطأ بسيط؟ لو نظرت إلى الرسم لاحظت أن عدد العيوب الحقيقي المناظر لخبرة عامل 3 سنوات يتراوح بين 35 و45 تقريبا. فنسبة الخطأ هنا هي حوالي $40 / 20 = 50\%$. هل هذه نسبة مقبولة.

دعك من هذا. لنرى كيف يمكننا استخدام النموذج لتوقع نسبة الخطأ المناظرة لعامل لديه خبرة قدرها سنة واحدة. إن النموذج الرياضي الذي استنتجناه هو

$$\text{عدد العيوب} = -113.9 + 45.33 * \text{خبرة العامل} + \text{الباقى (الخطأ)}$$

$$\text{عند التعويض بخبرة عامل قدرها سنة واحدة نحصل على عدد العيوب} = -68$$

بالطبع لا توجد عيوب أقل من الصفر. فالمعادلة هنا غير معبرة بالمرّة. ماذا لو قدرنا الخطأ لعامل لديه خبرة قدرها عشرين سنة؟ إن النتيجة تكون 792. هل هذه نتيجة صحيحة؟ باستخدام العلاقة الحقيقية التي أنشأت بها هذه البيانات لاستخدامها في هذا المثال وهي:

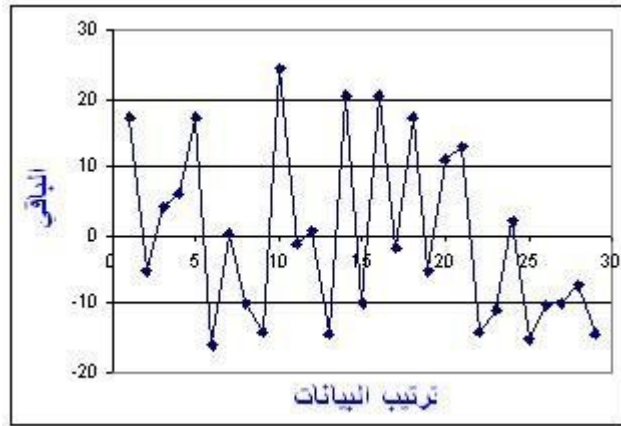
$$\text{عدد العيوب} = 2.5 + 3.85 * \text{عدد العيوب}^2$$

فإننا نجد أن نسبة العيوب المناظرة لخبرة عامل عشرين سنة هي 1542. فنسبة الخطأ هنا تقارب 50% .

ربما في بعض النقاط نجد نسبة الخطأ قليلة جدا وتقترب من الصفر ولكن هذا لا يعني صحة العلاقة بشكل عام.

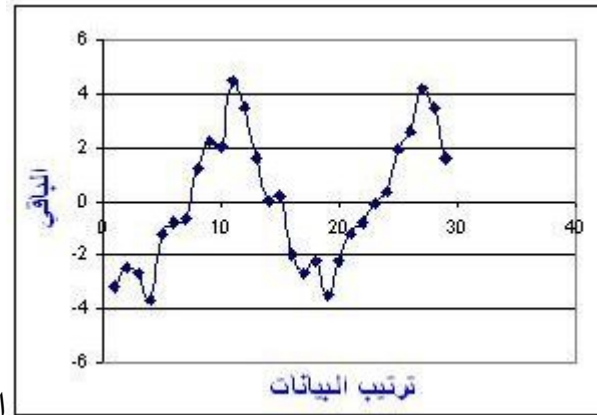
بالطبع هذا المثال هو مثال توضيحي ولا يقصد به العلاقة الحقيقية بين خبرة العامل وعدد العيوب.

العلاقة بين البواقي وترتيب البيانات: يمكننا كذلك رسم العلاقة بين البواقي وترتيب تسجيل البيانات والذي ينبغي ألا يظهر اتجاهها متزايدا أو شكل دوري متكرر. إن أحد فرضيات تحليل الانحدار هو استقلالية البواقي أي أن الباقي عند أي نقطة لا يعتمد على قيمة الباقي عند النقطة السابقة (أو النقاط السابقة) أي أن البواقي عشوائية.



هذا الشكل لا يظهر أي تزايد أو أي

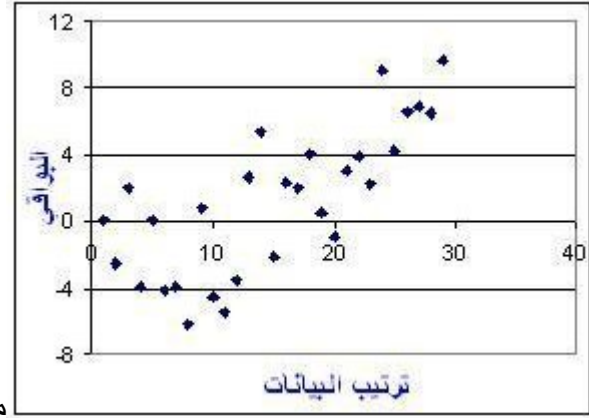
تغير دوري للبواقي. ولكن انظر إلى المثال التالي



إن هذا الشكل يظهر تغير دوري شبه

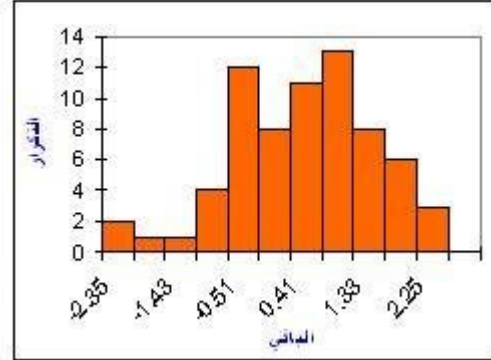
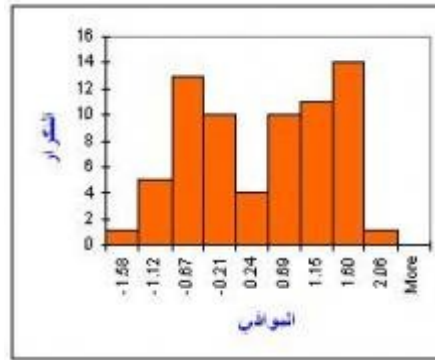
متكرر للبواقي. إذن فالبواقي غير مستقلة بل لها اتجاه محدد. في هذه الحالة مثلا يبدو أن هناك تغير موسمي seasonal في المتغير الذي نقيسه مع الزمن. فمثلا لو افترضنا أن هذا المتغير هو درجة حرارة سائل ما فإنه من الواضح تأثير الليل والنهار على درجة حرارة هذا السائل. في هذه الحالة فإن استخدامنا لنموذج تحليل الانحدار الخطي

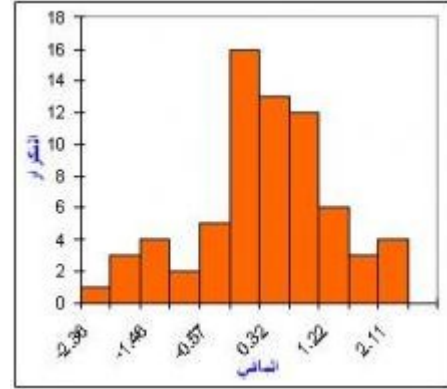
ليس هو الحل السليم بل يمكننا استخدام أسلوب التنبؤ باستخدام نموذج موسمي أي نموذج يأخذ في اعتباره هذا التغير الدوري في قيمة المتغير.



ماذا تلاحظ في هذا الشكل (أعلاه)؟ إن هناك تزايدا في قيمة البواقي مع الوقت وبالتالي فهي غير مستقلة. معنى ذلك أن الخطأ في النموذج يتزايد مع مرور الوقت فهو عند القراءات الأولى صغير وعند القراءات الأخيرة يأخذ قيما أكبر.

توزيع البواقي: هناك طريقتان تستخدمان للتأكد من توزيع البواقي توزيعا طبيعيا. الأول هو رسم التوزيع التكراري Histogram. فإذا كانت البواقي تتبع التوزيع الطبيعي فإن الفرض يكون قد تحقق. أما الثاني فهو منحنى الاحتمال الطبيعي Normal ProbabilityPlot وهو أسلوب يستخدم للتحقق من أن مجموعة بيانات تتبع التوزيع الطبيعي. فإن كانت البيانات تتبع توزيعا طبيعيا فإنها تأخذ شكل خط مستقيم تقريبا وإن كانت غير ذلك فإنها تأخذ اتجاهات مختلفة حول هذا الخط المستقيم. لا يهمنا الاستنتاج الرياضي لمنحنى الاحتمال الطبيعي ولكن يهمنا التعرف عليه واستخدامه.

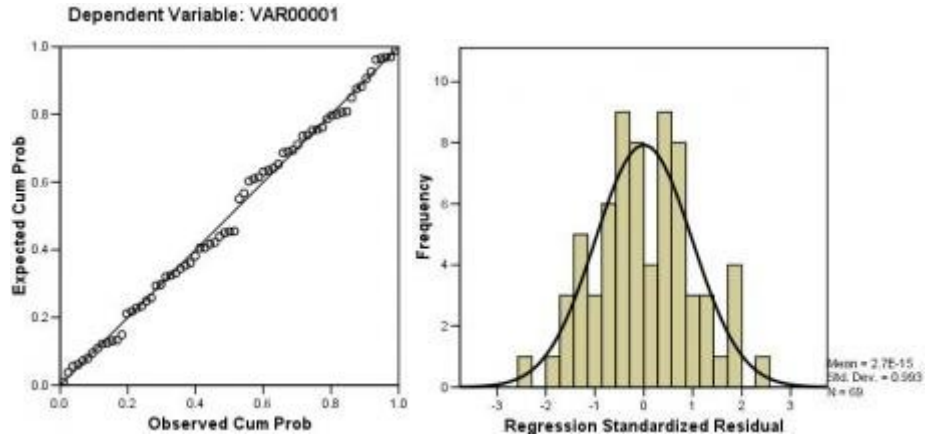




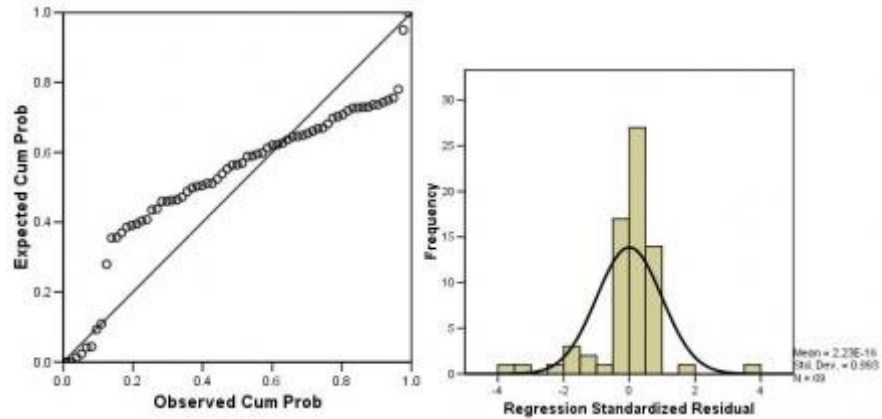
قد تجد أن التوزيع التكراري للأخطاء يشبه التوزيع الطبيعي مثل الأشكال الثلاثة أعلاه وقد تجده يختلف كثيرا. في حالة أن التوزيع لا يتبع التوزيع الطبيعي بالمرّة فإننا نبحث عن وسيلة أخرى غير تحليل الانحدار الخطي. الأمر المزعج هنا هو أن الحكم على الشكل يخضع للتقدير بشكل كبير وقد يختلف الرأي من شخص لآخر. بالإضافة لذلك فإن صغر حجم العينة قد يجعل الحكم على التوزيع الطبيعي أمرا غير دقيق.

التوزيع التكراري للبواقي لا يظهر مع النتيجة بشكل تلقائي في برنامج إكسل ولكن يمكننا رسمه باستخدام Tools...DataAnalysis...Histogram. وربما نشرح ذلك في مقالة أخرى. وأما في البرامج المتخصصة مثل SPSS, Minitab فإنك تحصل عليه مباشرة من النتيجة.

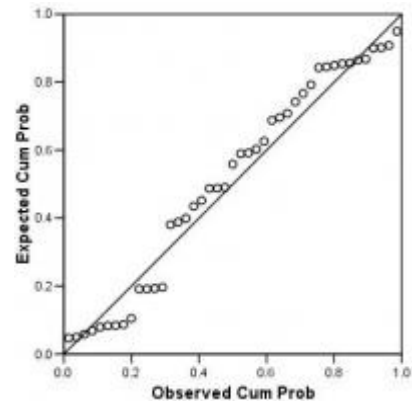
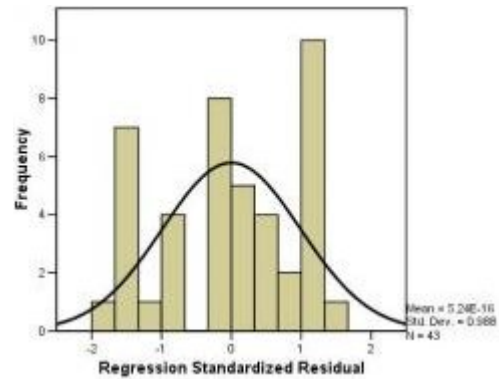
أما منحنى الاحتمال الطبيعي NormalProbability Plot فتحصل عليه من البرامج المتخصصة ولا تحصل عليه مباشرة من إكسل ويمكنك رسمه بإجراء بعض الحسابات. وهذا الرابط يقدم ملفا لرسم هذا المنحنى: [ملف يشرح كيفية رسم منحنى الاحتمال الطبيعي](#). وإن شاء الله أتناول هذه التفاصيل في المستقبل. والذي يهمنا الآن هو نتيجة هذا الاختبار.



لاحظ في الشكلين أعلاه كيف أن منحني التوزيع التكراري يشبه إلى حد كبير منحني التوزيع الطبيعي. والشكل على اليمين هو منحني الاحتمال الطبيعي وتلاحظ فيه أن النقاط تنطبق كثيرا على الخط المائل والذي يمثل منحني التوزيع الطبيعي. ولذلك فإن الشكل يبين أن البيانات والتي هي البواقي القياسية في هذه الحالة تتبع توزيعا طبيعيا.



أما الشكلان أعلاه فيظهرا انحرافا كبيرا عن التوزيع الطبيعي. انظر كيف تبعد النقاط -في الشكل على اليمين- عن الخط المائل الذي يمثل التوزيع الطبيعي.



والشكلان أعلاه يبينان بعدا عن التوزيع الطبيعي كذلك. بهذا تستطيع أن تحكم على نتائج منحنى الاحتمال الطبيعي.

موضوع دراسة البواقي ربما بدا معقدا بعض الشيء ولكن بالممارسة تعتاد عليه وتفهمه. وكما عرفنا فهو أمر مهم للتأكد من صحة استخدامنا لتحليل الانحدار.

من مراجع الموضوع:

1. Discovering Statistics using SPSS for Windows, A .Field, Sage, 2003.
2. Statistics for Managers, Levine et al., PrenticeHall, 1999.

مصدر : موقع الإدارة والهندسة الصناعية