أولا: مشكل التوزيع غير الطبيعي للأخطاء (Normality)

تمهيد:

تُعد "مشكلة التوزيع غير الطبيعي للأخطاء" من افتراضات الشروط الأساسية في نماذج الانحدار الخطي، وتظهر عندما **لا يتبع مُربعات الخطأ (البواقي) التوزيع الطبيعي**، بل يميل أو ينحرف أو يكون له ذيول ثقيلة (زيادة القيم الشاذة).

1. ماذا تعني هذه المشكلة؟

- في الانحدار الخطي نفترض أن الأخطاء (البواقي) تتوزع توزيعًا طبيعيًّا بمتوسط 0، وهذا يسمح باستخدام اختبارات مثل (t) و(F) وبناء فترات الثقة بشكل سليم.
- عندما يكون التوزيع غير طبيعي تصبح النتائج الاحتمالية غير موثوقة: قد تُبالغ في معنوية بعض المتغيرات أو تُقلل منها.

2. مصادر التوزيع غير الطبيعي للأخطاء

- **وجود قيم شاذة (Outliers)** أو قطع بيانات من عمليات مختلفة تُخلّ بالتوزيع المنتظم.
- **حدوث قيود طبيعية** في البيانات (كأن تكون القيم قريبة من الصفر أو الحد الأقصى)، مما يسبب انحرافًا أو توزيعًا غير متماثل.
- **اختيار نموذج خاطئ** (علاقة غير خطية، أو حذف متغيرات مهمة) فيؤدي إلى باقٍ غير عشوائي وغير طبيعي.

3. آثار هذه المشكلة

- تقل دقة اختبارات الدلالة الإحصائية (مثل اختبار (t) لمعامل الانحدار) وفترات الثقة.
- قد تزداد احتمالات **الخطأ من النوع الأول أو الثاني** في نتائجك، أي أنك ترى معنوية غير حقيقية أو تفوّت معنوية حقيقية.

4. كيفية التحقق من المشكلة:

- رسم أو استخدام **مخطط التوزيع Histogram**.
- استخدام اختبارات إحصائية مثل **Shapiro-Wilk** أو **Kolmogorov–Smirnov** لفحص التوزيع الطبيعي للبواقي.
- استخدام اختبار **(Jarque-Bera)** وهو اختبار للعينات الكبيرة وقد لا يكون مناسبا في العينات الصغيرة.

5. أختبار (Jarque-Bera): يتم حساب إحصائية (JB) كما يلي:

$$JB = n[\frac{s^2}{6} + \frac{(k-3)^2}{24}] \sim X_2^2$$

حيث: n حجم العينة، S معامل الالتواء، k معامل التفرطح.

بالنسبة للمتغير الذي يتبع التوزيع الطبيعي $X \sim N(\mu, \sigma_X^2)$ يكون لدينا (S=0, k=3) ؛ على هذا الأساس يتضح من إحصائية JB أنه إذا كانت S=0 و k=3 فإن قيمتها تكون صفرا، فكلما اقتربت قيمة JB من الصفر كلما كان افتراض الحالة الطبيعية أفضل.

يمكننا استخدام توزيع مربع كاي لإيجاد المعنوية الإحصائية لإحصائية JB.

يمكن مقارنة **إحصائية جارك–بيرا** JB و**توزيع كاي-تربيع** χ^2 من خلال فهم دور كل منهما:

- إحصائية جارك–بيرا هي **إحصاء لاختبار طبيعيّة التوزيع** يُستمد من الانحراف والتفرطح،
- بينما توزيع كاي-تربيع هو **توزيع احتمالي** يُستعمل في اختبارات عدة (جودة التوفيق، الاستقلال، وغيرها).

كيف ترتبط جارك–بيرا بتوزيع كاي-تربيع؟

تحت الفرض العدمي بأن البيانات موزعة توزيعًا طبيعيًّا، فإن إحصائية جارك–بيرا تَسلك **توزيعًا تقريبيًا كاي-تربيع بدرجتي حرية 2**، أي:

$$JB \sim \chi^2(2)$$

مثال تطبيقي بسيط

- تستخرج JB من بياناتك (مثلاً $JB = 6.5$) (ثم تقارنها بقيمة حرجة لـ $\chi^2(2)$ عند $\alpha = 0.05$) تقريبًا 5.99.
- إذا كانت $JB > 5.99$ ترفض فرض العدمي للطبيعيّة (أي تقول إن التوزيع غير طبيعي)

5. حلول وطرق المعالجة:

- **تحويل المتغير التابع أو المتغيرات المستقلة**: مثل التحويل اللوغاريتمي $\log(y)$، أو الجذر التربيعي، أو التحويل العكسي، لجعل التوزيع أقرب إلى الطبيعي.
- **استخدام نماذج غير خطية أو نماذج عامة خطية GLM** تُناسب نوع التوزيع الفعلي (مثل بواسون، ثنائي، إلخ) بدلًا من الاعتماد على الافتراض الطبيعي.
- في بعض الحالات، الاعتماد على **اختبارات لا بارامترية** أو **تقديرات حصينة (Robust regression)** إذا كانت القيم الشاذة هي السبب الأساسي.

ثانيا: أخطاء في القياس:

تمهيد:

مشاكل الانحدار الناتجة عن **أخطاء القياس** تحدث عندما تكون **القيم المسجلة للمتغيرات، خاصة المتغيرات المستقلة (Xi) غير دقيقة أو تحتوي على أخطاء**، مما يشوّه نتائج التقدير ويخالف أحد افتراضات نموذج الانحدار الخطي.

1. ماهية أخطاء القياس في الانحدار:

- خطأ القياس يعني أن **القيمة التي تُستخدم في التقدير ليست القيمة الحقيقية** للمتغير، بل مقيّمة مع وجود خطأ عشوائي أو منظم (مثلاً بيانات رسمية غير دقيقة، إجابات غير صادقة، أو خطأ في جمع البيانات).
- إذا كان الخطأ في المتغير المستقل (Xi) ، يُصبح هذا النوع من المشاكل خاصًا لأنه يُسبب **تحيزًا في معاملات الانحدار** وغالبًا يُقلل من حجم المعامل المُقدَّر.

2. آثار أخطاء القياس في الانحدار:

- **تحيز في التقديرات** :خصوصًا عند وجود خطأ في المتغير المستقل، يصبح مُقدَّر الميل (معامل الانحدار ,b1) أقل من قيمته الحقيقية (منتقصًا نحو الصفر)، أي تقدير غير دقيق لقوة العلاقة.
- **زيادة تباين التقديرات** :حتى لو بقيت التقديرات غير متحيزة في بعض الحالات (خاصة عند خطأ في المتغير التابع فقط)، يزداد تباين المعاملات، فتنخفض دقة الفرضيات وفترات الثقة.
- **ارتباط المشاكل بمشاكل أخرى** :أخطاء القياس قد تساهم في ظهور **عدم ثبات تباين حد الخطأ (Heteroscedasticity)**.

3. طرق تقليل أو معالجة أخطاء القياس

- **استخدام أدوات قياس أكثر دقة** :مثل بيانات رسمية موثوقة، أو مصادر ثانوية أدق، أو طرق جمع بيانات أكثر احترافية) مسوحات، لوحة بيانات.(panel data)
- **استخدام نماذج خاصة** :مثل **نماذج المتغيرات الصورية (Instrumental Variables)** أو النماذج شبه المعلمية التي تُصمَّم خصيصًا لمعالجة تأثير خطأ القياس في المتغيرات.
- **التحويلات والتدقيق** :تطبيق تحويلات (مثل اللوغاريتم) أو تنظيف البيانات من القيم الشاذة التي يمكن أن تُنتج خطأ قياس كبير.