مشكلة التعدد الخطي في الانحدار

Multicollinearity

1. تعريف التعدد الخطي.
2. عواقبه.
3. اكتشاف.
4. علاجه.
5. تطبيق

1. تعريف التعدد الخطي

يعرف التعدد الخطي بوجود علاقة خطية بين المتغيرات المستقلة $(X_1, X_2,..., X_p)$ ، وهذا ينافي أحد الافتراضات التي يستند عليها نموذج الانحدار الخطي المتعدد، وهو افتراض أن المتغيرات المستقلة يجب أن تكون مستقلة خطيا، أي لا توجد بينها علاقة.

2. عواقب التعدد الخطي

وجود علاقة بين المتغيرات المستقلة يجعل الأخطاء المعيارية لمعاملا الانحدار $S.E_{\hat{\beta}}$ كبيرة حسب درجة الارتباط، ويترتب على ذلك احتمالات كبيرة في قبول الفرض العدم $H_0 : \beta = 0$ (المتغير المستقل X ليس له أثر معنوي) وقد يكون الفرض غير صحيح ومن ثم يقع الباحث في خطأ من النوع الثاني $Error\ II$.

3. اكتشاف التعدد الخطي

يمكن للباحث اكتشاف مشكلة التعدد الخطي من خلال عدة طرق منها ما يلي:

- عندما يكون معامل التحديد كبير $(R^2 > 0.7)$، و من ثم كبر قيمة إحصائية الاختبار $(F^* = MSR / MSE)$ مشيرة بصلاحية النموذج الخطي للتنبؤ، وفي الوقت نفسه يوجد متغير أو أكثر من المتغيرات المستقلة غير معنوي وفقا لاختبار t .
- البحث في معاملات الارتباط بين المتغيرات المستقلة r_{x_i, x_j} إذا كانت قوية،

يشك في وجود ارتباط خطي بين المتغيرات المستقلة.

- من خلال حساب قيمة محدد مصفوفة معاملات الارتباط بين المتغرات المفسرة،

$$|\mathbf{R}_{x_i,x_j}| = \begin{vmatrix} 1 & r_{x_1x_2} & r_{x_1x_3} & \dots & r_{x_1,x_p} \\ r_{x_1x_2} & 1 & r_{x_2x_3} & \dots & r_{x_2,x_p} \\ & & & & \\ r_{x_px_1} & r_{x_px_2} & r_{x_px_3} & \dots & 1 \end{vmatrix}$$

إذا كان قيمة المحدد $|\mathbf{R}_{x_i,x_j}| = 0$ دل ذلك على وجود ارتباط خطي تام بين

4. علاج مشكلة التعدد الخطي

يوجد عدد من طرق علاج التعدد الخطي بين المتغيرات المستقلة، منها ما يلي:

- **حذف بعض المتغيرات المستقلة من النموذج.**

من خلال استخدام معامل تضخم التباين $[VIF = (1/(1-R_X^2)]$ لكل متغير مستقل ، حيث أن R_X^2 هو معامل التحديد في نموذج انحدار المتغير المستقل X على باقي المتغيرات المستقلة، ويحسب هذا المعيار لكل متغير مستقل. وعلى سبيل المثال إذا كان عدد المتغيرات المستقلة أربعة هي: (X_1, X_2, X_3, X_4) يكون لدينا أربع معادلات انحدار، هي:

$$\begin{aligned} &equation\quad 1: \; X_1 = \alpha_0 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \xi_1 \\ &equation\quad 2: \; X_2 = \alpha_0 + \alpha_1 X_1 + \alpha_3 X_3 + \alpha_4 X_4 + \xi_2 \\ &equation\quad 3: \; X_3 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_4 X_4 + \xi_3 \\ &equation\quad 4: \; X_4 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \xi_4 \end{aligned}$$

ويحسب لكل معادلة معامل التحديد R_X^2 ثم يحسب المعامل $[VIF = (1/(1-R_X^2)]$ ، إذا كان VIF يزيد عن 5 يتم استبعاد المتغير المستقل المناظر لهذا المعيار.

- **معلومات اضافيه للعينة أو استعمال معلومات من خارج العينة.**

كزيادة حجم العينة، أو توافر معلومات سابقة عن بعض معاملات الانحدار، يتم الاستناد إليها في التحليل.

- **وضع المتغيرات في شكل نسب يعني قسمة النموذج على أحد المتغيرات المستقلة.**

عندما يكون تباين الخطأ العشوائي متناسب مع أحد هذه المتغيرات ، يتم قسمة المتغيرات المستقلة على هذا المتغير.

- **إضافة معلمة الـ Ridge وتطبيق طريقة المربعات الصغرى**

حيث يتم إضافة ثابت لمصفوفة التباين والتغاير ويتم الحصول على تقدير المربعات الصغرى العادي

$$\hat{\beta}(K) = (X'X + KI)^{-1} X'Y \;, K \geq 0$$

$$0 < k < 1$$

$0 < k < 1$ معلمة التحيز، ويمكن التطبيق بسهولة في برنامج SPSS.

5. تطبيق

فيما يلي بيانات عن الإنفاق العائلي خلال الشهر كمتغير تابع ، والدخل وعدد أفراد الأسرة كمتغيرين مستقلين.

التطبيق في ملف EXCEL