أسلوب الانحدار التدريجي

Stepwise

٩-١ مقدمة:

في كثير من المجالات التطبيقية لنموذج الانحدار الخطي المتعدد، يرغب الباحث في تحديد أهم المتغيرات المستقلة، والتي تحسن القدرة التنبؤية للنموذج، وفي المقابل إزالة المتغيرات المستقلة الأخرى التي يكون إضافة أحدها أو بعضها للنموذج غير مؤثر معنويا في تحسين القدرة التنبؤية للنموذج، ومن ثم يمكن تطبيق أسلوب الانحدار التدريجي، ويستند على إدخال المتغيرات واحد تلو الآخر حسب نسبة الجزء الذي يعزى إليه في تفسير الاختلافات الكلية في المتغير التابع.

وتظهر أهمية تطبيق أسلوب الانحدار التدريجي في الآتي:

١- أحد طرق علاج مشكلة الارتباط الخطي Multicollinearity بين المتغيرات المستقلة.

٢- ترتيب المتغيرات حسب أهميتها في تفسير المتغير التابع.

٩-٢ خطوات تطبيق أسلوب الانحدار التدريجي

بفرض أن $(x_1, x_2, ..., x_k)$ هي مجموعة المتغيرات المفسرة، يمكن تطبيق أسلوب الانحدار التدريجي، بإدخال المتغيرات واحد تلو الآخر، وفي المرحلة رقم r يصاغ الفرض العدم H_0 على الصورة التالية:

$$\mathrm{H}_0 : \beta_{(x_r^*|x_1^*, x_2^*, ..., x_{r-1}^*)} = 0 \qquad \textbf{(9-1)}$$

ويعني الفرض العدم أعلاه أن إضافة المتغير x_r^* للنموذج الذي يشمل المتغيرات $(x_1^*, x_2^*, ..., x_{r-1}^*)$ لا يحسن قدرته التنبؤية، وتستخدم إحصائية الاختبار " $Partial\ F$ "، وتحسب بتطبيق المعادلة التالية:

$$F^*_{x^*_r|x^*_1,x^*_2,...,x^*_r} = \frac{\dfrac{\left(SSR(x^*_1,x^*_2,...,x^*_r) - SSR(x^*_1,x^*_2,...,x^*_{r-1})\right)}{(r-(r-1))}}{\dfrac{SSE(x^*_1,x^*_2,...,x^*_r)}{(n-r-1)}} = \frac{\dfrac{\left(R^2(x^*_1,x^*_2,...,x^*_r) - R^2(x^*_1,x^*_2,...,x^*_{r-1})\right)}{(r-(r-1))}}{\dfrac{\left(1-R^2(x^*_1,x^*_2,...,x^*_r)\right)}{(n-r-1)}} \qquad \textbf{(9-2)}$$

حيث أن:

$SSR(x^*_1,x^*_2,...,x^*_r)$ **: هو مجموع مربعات الانحدار في نموذج الانحدار:** $Y = \beta_0 + \beta_1 x^*_1 + ... + \beta_r x^*_r$

$SSR(x^*_1,x^*_2,...,x^*_{r-1})$ **: هو مجموع مربعات الانحدار في نموذج الانحدار:** $Y = \beta_0 + \beta_1 x^*_1 + ... + \beta_r x^*_{r-1}$

$SSE(x^*_1,x^*_2,...,x^*_r)$ **: هو مجموع مربعات الأخطاء في نموذج الانحدار:** $Y = \beta_0 + \beta_1 x^*_1 + ... + \beta_r x^*_r$

وإحصائية الاختبار أعلاه تتبع توزيع F بدرجات حرية بسط = $(r-(r-1)=1)$ ، ودرجات حرية مقام = $(n-(r+1)=n-r+1)$.

كما يلاحظ عند تطبيق هذا الأسلوب مراعاة الآتي:

١- استخدام معيار $R^2(x^*_1,x^*_2,...,x^*_r)$ في تحديد من المتغير الأفضل (الذي ينتج عن إضافته أعلى معامل تحديد)

٢- بعد تحديد المتغير الأفضل يجب اختبار معنوية إضافته، فإذا أثبتت الاختبارات الإحصائية معنوية إضافته في تحسين القدرة التنبئية للنموذج يجب إبقائه كأحد المتغيرات الهامة في تفسير سلوك المتغير التابع.

تطبيق

أخذت عينة عشوائية من الأطفال ذات المرحلة العمرية (7-6) حجمها 20 طفل ، وتم تسجيل إجمالي ما يتناول يومياً من الغذاء بالجرام لمدة شهر لثالثة أنواع من الغذاء p, v, sf، تم تسجيل وزن الطفل w_i بعد ثلاثة أشهر من التغذية، والجدول التالي يبين بيانات هذه التجربة خلال هذه الفترة.

w_i	w	p	v	sf
6.135	3.80	2.42	6.80	4.564
5.341	3.00	3.23	7.40	5.262
6.893	3.60	3.33	7.50	5.361
8.394	4.10	3.85	7.60	5.668
7.663	4.00	4.05	7.80	5.866
7.406	3.30	4.31	8.00	6.093
7.217	3.50	4.33	8.20	6.202
7.296	3.10	4.44	8.30	6.306
8.721	3.70	4.50	8.30	6.336
8.561	3.80	4.54	8.50	6.455
6.987	3.60	4.61	6.70	5.598
8.159	3.90	4.63	6.80	5.658
6.577	3.40	4.79	6.90	5.787
6.404	3.00	5.24	7.00	6.059
7.348	3.50	5.50	7.20	6.287
8.310	3.50	5.37	7.40	6.321
8.038	3.40	5.58	7.50	6.475
7.599	2.90	5.59	7.60	6.529
7.882	3.30	6.12	7.90	6.940
9.174	3.40	5.79	8.40	7.024

حيث: يمثل المتغير w وزن الطفل عند بداية التجربة.

والمطلوب استخدام أسلوب الانحدار التدريجي في تحديد أهم هذه المتغيرات المؤثرة على زيادة الوزن $(y = w\_i - w)$ من المتغيرات المفسرة.