2-طريقة الحذف العكسي The Back word elimination procedure

في هذه الأسلوب يتم وضع جميع المتغيرات المستقلة في معادلة انحدار ثم نبدأ بحذف المتغيرات ذات التأثير غير المعنوي واحدا بعد الآخر حتى نصل إلى الصيغة النهائية التي تحتوي على المتغيرات ذات التأثير المعنوي.

فالمتغير يستبعد على أساس مساهمته في تخفيض مجموع مربعات الخطأ، فالمتغير الأول الذي يحذف هو المتغير ذو المساهمة الأقل في تخفيض مجموع المربعات الخطأ أو بمعنى آخر حذف المتغير الذي يمتلك أقل نسبة t في معادلة فإذا كانت جميع نسب t معنوية فان كل المتغيرات تبقى في المعادلة

ويمكن توضيح طريقة الحذف العكسي من خلال الخطوات التالية

لنفرض أن لدينا أربعة متغيرات $x_1x_2x_3x_4$ مع متغير الاستجابة y

الخطوى الاولى: نجد معادلة الانحدار المقدرة لكل المتغيرات باي طريقة

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3} + \hat{\beta}_4 X_{i4}$$

ثم نجري العمليات التالية:

1. قيمة (F) الجزئية لكل متغير من المتغيرات بثبوت بقية المتغيرات وكما يلي:

$$F_1 = \frac{\mathrm{SSR}(x_1/x_2x_3x_4)}{\mathrm{MSe}(x_1x_2x_3x_4)} = \left(\frac{\hat{\beta}_1}{S(\hat{\beta}_1)}\right)^2 = t^2$$

$$F_2 = \frac{\mathrm{SSR}(x_2/x_1x_3x_4)}{\mathrm{MSe}(x_1x_2x_3x_4)} = \left(\frac{\hat{\beta}_2}{S(\hat{\beta}_2)}\right)^2 = t^2$$

$$F_3 = \frac{\mathrm{SSR}(x_3/x_1x_2x_4)}{\mathrm{MSe}(x_1x_2x_3x_4)} = \left(\frac{\hat{\beta}_3}{S(\hat{\beta}_3)}\right)^2 = t^2$$

$$F_4 = \frac{\mathrm{SSR}(x_4/x_1x_2x_3)}{\mathrm{MSe}(x_1x_2x_3x_4)} = \left(\frac{\hat{\beta}_4}{S(\hat{\beta}_4)}\right)^2 = t^2$$

2. نختار المتغير الذي له اقل قيمة (F) جزئية من الخطوات السابقة ولتكن F_2
3. نقارن F_2 مع قيمة (F) الجدولية فإذا كانت:

- $F_2 < F_{(1-\alpha,1,n-5)}$ نحذف x_2 وننتقل إلى الخطوة الثانية مع $x_1x_3x_4$ فقط
- $F_2 \geq F_{(1-\alpha,1,n-5)}$ نبقي x_2 مع $x_1x_3x_4$ ونتوقف عن الاختيار ويكون الأنموذج بكل المتغيرات ملائم للبيانات

الخطوة الثانية

عند حذف المتغير x_2 من الخطوة الأولى سيبقى لدينا $x_1x_3x_4$ ،وفي هذه الخطوة نقوم بنفس العمل السابق أي نستخرج معادلة الانحدار الكلية كما يلي:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_3 X_{i3} + \hat{\beta}_4 X_{i4}$$

ثم نجري العمليات التالية:

1. نبدأ بحساب قيمة (F) الجزئية لكل متغير مستقل بثبوت بقية المتغيرات، وكما يلي:

$$F_1 = \frac{\text{SSR}(x_1/x_3x_4)}{\text{MSe}(x_1x_3x_4)} = \left(\frac{\hat{\beta}_1}{S(\hat{\beta}_1)}\right)^2 = t_1^2$$

$$F_2 = \frac{\text{SSR}(x_3/x_1xx_4)}{\text{MSe}(x_1x_3x_4)} = \left(\frac{\hat{\beta}_3}{S(\hat{\beta}_3)}\right)^2 = t_2^2$$

$$F_3 = \frac{\text{SSR}(x_4/x_1x_3)}{\text{MSe}(x_1x_3x_4)} = \left(\frac{\hat{\beta}_4}{S(\hat{\beta}_4)}\right)^2 = t_3^2$$

وتجدر الملاحظة بأن قيم المعلمات المقدرة وإنحرافاتها المعيارية $\hat{\beta}_j$ و $S(\hat{\beta}_j)$ $j = 1,3,4$ تختلف عن القيم التي تم حسابها في الخطوة الاولى.

2. ونختار المتغير الذي له اقل قيمة (F) جزئية من الخطوات السابقة ولتكن F_2 ايضا
3. نقارن F_2 مع قيمة (F) الجدولية فإذا كانت:

- $F_2 < F_{(1-\alpha,1,n-4)}$ نحذف x_3 وننتقل إلى الخطوة الثانية مع المتغيرين x_1x_4 فقط

- $F_2 \geq F_{(1-\alpha,1,n-4)}$ نبقي (نثبت) x_3 في الانموذج مع المتغيرين x_1x_4 ونتوقف عن الاختيار ويكون الأنموذج بكل المتغيرات ملائم للبيانات

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_3 X_{i3} + \hat{\beta}_4 X_{i4}$$

وهكذا نقوم بخطوات اخرى تساوي او اقل من عدد المتغيرات 4 الاصلية للمعادلة إلى ان نصل إلى المعادلة المثلى، وبتطبيق هذه الطريقة على المثال (ص 273-274 خاشع الراوي):

S.O.V.	df	ss	MS	F
$R(x_1x_2x_3x_4)$	4			
$R(x_1/x_2x_3x_4)$	1			**4.34**
$R(x_2/x_1x_3x_4)$	1			**0.5**
$R(x_3/x_1x_2x_4)$	1			**0.02**
$R(x_4/x_1x_2x_3)$	1			**0.04**
$E(x_1x_2x_3x_4)$	8			
Total	**12**			

لدينا:

$$F_{(out)} = F_{(0.05,1,8)} = 5.32$$

$$Cal.F = 0.02 < F_{(out)} = 5.32$$

∴ نحذف x3 من النموذج

نقوم الان بعمل جدول لبقية المتغيرات $(x_1x_2x_4)$، وكما يلي:

s.o.v	df	ss	MS	F
$R(x_1x_2x_4)$	3			
$R(x_1/x_2x_4)$	1			**154.01**
$R(x_2/x_1x_4)$	1			**5,03**
$R(x_4/x_1x_2)$	1			**1**
$E(x_1x_2x_4)$	9			
Total	**12**			

لدينا:

$$F_{(out)} = F_{(0.05,1,9)} = 5.12$$

$$Cal.F = 1 < F_{(out)} = 5.12$$

∴ نحذف x_4 من النموذج

الآن نعمل جدول جديد لـ x_1x_2 وكما يلي:

S.O.V.	df	SS	MS	F
$R(x_1x_2)$	2			
$R(x_1/x_2)$	1			**146.52**
$R(x_2/x_1)$				**208.5**
$E(x_1x_2x_4)$				
	10			
Total	**12**			

لدينا:

$$F_{(out)} = F_{(0.05,1,10)} = 4.096$$

$$Cal.F = 208.5 > F_{(out)} = 4.096$$

∴ نتوقف عن الحل والأنموذج لدينا هو أفضل أنموذج، وكما يلي:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2}$$

3-طريقة الاختيار الأمامي أو المباشر: The Forward Selection Procedure

أن هذا الأسلوب يبدأ من معادلة لاتحتوي على متغير تفسيري وإنما حد ثابت، والمتغير الأول الذي يدخل المعادلة هو الذي له أعلى قيمة F محسوبة واكبر من الجدولية F_{in} أو له ارتباط بسيط مع y كقيمة مطلقة. أما بوجود المتغير الأول من الخطوة الأولى واكبر من F_{in} أو المتغير الذي له أعلى ارتباط مع y بثبوت المتغير الذي دخل أولا. وتتوقف عملية الإضافة إلى أن نصل إلى أعلى F جزئية تقل عن F_{in} فنتوقف عن الإضافة.

فعلى سبيل المثال إذا كان لدينا أربعة متغيرات فتتم عملية الاختيار كما هو موضح في الجدول التالي حيث أن علامة(*) تشير إلى المتغير المعنوي:

S.O.V	S.O.V	S.O.V	S.O.V
$R(x_1)$	$R(x_1/x_2)$	$R(x_1/x_2x_3)$	
$R(x_2)$ *			$R(x_4/x_1x_2x_3)$
$R(x_3)$	$R(x_3/x_2)$ *		
$R(x_4)$	$R(x_4/x_2)$	$R(x_4/x_2x_3)$ *	

مثال بتطبيق الطريقة في مثال ص (280- 285 خاشع الراوي)

نبدأ بكل متغير على حدى ، و F الجدولية هنا نسميها F_{in} بدل من F_{out} في الحذف العكسي

S.0.V.	df	SS	MS	F
$R(x_1)$ $E(x_1)$	1 11			$\frac{1450.0766}{115.0624} = 12.6$
Total	**12**			

S.0.V.	df	SS	MS	F
$R(x_2)$ $E(x_2)$	1 1			$\frac{1809.4268}{82.3942} = 21.96$
Total	**12**			

وهكذا الحال بالنسبة ل x_3x_4

$$Cal.F(X_3) = \frac{776.3626}{176.3091} = 4.4 \qquad Cal.F(X_4) = \frac{1831.8962}{80.3515} = 22.8^*$$

$$Max\ Cal.F(X_4) = 22.8^* > F(0.05,1,11) = 4.48$$

وهنا سنختار X_4 الذي يمتلك أعلى قيمة F محسوبة 22.8 اكبر من F_{in} الجدولية 4.48 لذا فهو المتغير الأول الذي يدخل إلى النموذج ،أما إذا كانت أكبر قيمة محسوبة غير معنوية نتوقف ونقول أن النموذج غير موجود

s.o.v	df	ss	MS	F
$R(x_1x_4)$ $R(x_1/x_4)$ $E(x_1x_4)$	2 1 10			**108.22**
Total	**12**			

s.o.v	df	ss	MS	F
$R(x_2x_4)$ $R(x_2/x_4)$ $E(x_2x_4)$	2 1 10			**0.1725**
Total	**12**			

وهكذا بالنسبة ل $F(X_3/X_4)$=40.29

F_{IN}=4.96

إذا سيدخل X1 وx4 إلى النموذج لان له أعلى F محسوبة واكبر من F_{IN}

الآن نختبر X1 وx3 بثبوت x1 وx4 وكما يلي

s.o.v	df	ss	MS	F
$R(x_1x_2x_4)$ $R(x_2/x_1x_4)$ $E(x_1x_2x_4)$	3 1 9			**5.02**
Total	**12**			

s.o.v	df	ss	MS	F
$R(x_1x_3x_4)$ $R(x_3/x_1x_4)$ $E(x_1x_2x_4)$	3 1 9			**0.1725**
Total	**12**			