

## (2) تحديد مشاهدات قاصية في Y (المتغير التابع) – بواقي الحذف المعيرة :-

لتحديد المشاهدات القاصية في Y يستخدم عادة رواسب الحذف المعيرة (بواقي ستيودنت المحذوفة) (Studentized Deleted Residual) حيث نقدم الآن تحسين يجعل تحليل الرواسب أكثر فعالية في الكشف عن مشاهدات قاصية في Y .

التحسين هو أن نقيس الراسب :

$$e_i = Y_i - \hat{Y}_i$$

عندما يبني الانحدار التوفيقي على المشاهدات بعد أن نستثني المشاهدة i.

وسبب هذا التحسين هو أنه إذا كانت  $Y_i$  قاصية جدا فقد يؤثر هذا في دالة الانحدار التوفيقية المستندة إلى جميع المشاهدات لجعلها أقرب إلى Y . مما ينتج قيمة توفيقية  $\hat{Y}_i$  قريبة من  $Y_i$  وفي تلك الحالة ، سيكون الراسب  $e_i$  صغيرا وسوف لا يكشف عن كون  $Y_i$  قاصية . وإذا حذفت المشاهدة i قبل توفيق دالة الانحدار ، فإن قيمة المربعات الدنيا التوفيقية لا تتأثر بالمشاهدة القاصية، وسيميل الراسب عندئذ إلى أن يكون أكبر وبالتالي تزداد إمكانية كشفه للمشاهدة القاصية في قيمة Y .

### - حساب رواسب الحذف :

- نحذف المشاهدة (i) .
- توفيق دالة الانحدار للمشاهدات الـ n-1 الباقية.
- نحسب التقدير النقطي للقيمة المتوقعة الخاصة بالمشاهدة i المحذوفة ونرمز لها بالرمز  $\hat{Y}_{i(i)}$  .

ويتم حساب الراسب المحذوف (Deleted Residual) من العلاقة:-

$$d_i = Y_i - \hat{Y}_{i(i)}$$

حيث  $d_i \leftarrow$  الراسب المحذوف للمشاهدة رقم (i).

$Y_i \leftarrow$  القيمة الفعلية للمتغير التابع للمشاهدة رقم i.

$\hat{Y}_{i(i)} \leftarrow$  القيمة المقدرة للقيمة المتوقعة للمتغير التابع للمشاهدة رقم i بعد حذف المشاهدة i .

والعلاقة المكافئة جبريا لـ  $d_i$  والتي لا تستدعي إعادة حساب دالة الانحدار التوفيقية بعد حذف المشاهدة i ، هي:

$$d_i = \frac{e_i}{1 - h_{ii}}$$

حيث  $e_i \leftarrow$  البواقي لدالة الانحدار باستخدام جميع المشاهدات .

$h_{ii} \leftarrow$  قيم العزم ويتم استخراجها كما سبق من العنصر القطري في مصفوفة القبة.

وهكذا فإن رواسب الحذف ستحدد أحيانا المشاهدات القاصية في قيم  $\gamma$  حيث تفشل الرواسب العادية في القيام بذلك ،وفي أحيان أخرى تفقد رواسب الحذف إلى التحديدات نفسها التي تقود إليها الرواسب العادية.لذلك نستخدم رواسب الحذف المعيرة.

- رواسب الحذف المعيرة (بواقي ستودنت المحذوفة) (Studentized Deleted Residual) :-  
نلاحظ من علاقة الراسب المحذوف أنه كلما كانت قيمة العزم  $h_{ii}$  أكبر كلما كان الراسب المحذوف أكبر .

#### • حساب رواسب الحذف المعيرة :-

يتم حساب رواسب الحذف المعيرة للملاحظة رقم  $i$  من العلاقة :-

$$d_i^* = \frac{d_i}{s.e(di)}$$

حيث  $d_i^* \leftarrow$  راسب الحذف المعير للملاحظة رقم  $i$  .

$d_i \leftarrow$  راسب الحذف للملاحظة رقم  $i$  .

$$s.e(di) \leftarrow \text{الانحراف المعياري للراسب المحذوف للملاحظة رقم } i \text{ يساوي } \sqrt{\frac{MSE(i)}{1-h_{ii}}}$$

$MSE(i) \leftarrow$  متوسط مجموع مربعات الخطأ لنموذج الانحدار الخطي المقدر بعد حذف الملاحظة رقم  $i$  .

حيث أن رواسب الحذف المعيرة تتبع توزيع  $t$  بدرجة حرية  $n-p-1$  ولذلك جاءت التسمية (بواقي ستودنت المحذوفة) .

يمكن حساب رواسب الحذف المعيرة من العلاقة:

$$d_i^* = e_i \left[ \frac{n-p-1}{SSE(1-h_{ii}) - e_i^2} \right]^{\frac{1}{2}}$$

حيث  $e_i \leftarrow$  الرواسب العادية لنموذج الانحدار باستخدام جميع المشاهدات .

$n \leftarrow$  عدد المشاهدات .

$P \leftarrow$  عدد معالم النموذج  $B_0, B_1, \dots, B_{p-1}$  .

$SSE \leftarrow$  مجموع مربعات الخطأ لنموذج الانحدار باستخدام جميع المشاهدات .

- طريقة كشف المشاهدات القاصية في المتغير التابع  $Y$  :-

عند مستوى المعنوية  $\alpha$  ، نقارن القيمة المطلقة لرواسب الحذف المعيرة (بواقي ستودنت المحذوفة)  $|d_i^*|$  بتوزيع  $t$  عند درجة حرية  $n-p-1$  ومستوى معنوية  $(\alpha/2)$  فإذا كان:

$$|d_i^*| > t_{\alpha/2, n-p-1}$$

تعتبر المشاهدة رقم  $(i)$  قاصية في بيانات المتغير التابع  $Y$  وتستدعي دراستها وتحديد مدى تأثيرها على النموذج.

- تطبيق (2-1) :-

بالعودة إلى تطبيق (1-1) في تجربة لعميل من قسم الميكانيكية ، نستخرج في هذا التطبيق المشاهدات القاصية في الجهد (بالفولت)  $Y$  (المتغير التابع) وذلك باستخدام بواقي ستودنت المحذوفة بطريقتين :

• الطريقة الأولى :

باستخدام العلاقة التالية :

$$d_i^* = e_i \left[ \frac{n-p-1}{SSE(1-h_{ii}) - e_i^2} \right]^{\frac{1}{2}}$$

نقوم بحساب رواسب الحذف المعيرة تقديراً بأخذ المشاهدة الأولى (كمثال) :

$$d_1^* = e_1 \left[ \frac{n-p-1}{SSE(1-h_{11}) - e_1^2} \right]^{\frac{1}{2}} = 0.0703 \left[ \frac{18-3-1}{0.232(1-0.168134) - (0.0703)^2} \right]^{\frac{1}{2}} \\ = 0.60578$$

وهكذا لبقية المشاهدات الأخرى .

## • الطريقة الثانية :

### - التحليل الإحصائي :-

باستخدام برنامج الحزم الإحصائية Minitab سوف نقوم باستخراج بواقي ستودنت المحذوفة واكتشاف المشاهدات القاصية في الجهد (بالفولت) باستخدام مستوى المعنوية  $\alpha = 0.05$ .

```
MTB > Name c4 "TRES1"
MTB > Regress 'Y' 2 'X1' 'X2';
SUBC>   Tresiduals 'TRES1';
SUBC>   Constant;
SUBC>   Brief 2.
```

### Regression Analysis: Y versus X1; X2

The regression equation is  
 $Y = -1.64 + 0.000556 X1 - 67.4 X2$

Predictor	Coef	SE Coef	T	P
Constant	-1.6413	0.2470	-6.64	0.000
X1	0.00055571	0.00003448	16.11	0.000
X2	-67.396	4.481	-15.04	0.000

S = 0.124548    R-Sq = 96.1%    R-Sq(adj) = 95.6%

### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	5.6942	2.8471	183.54	0.000
Residual Error	15	0.2327	0.0155		
Total	17	5.9269			

Source	DF	Seq SS
X1	1	2.1849
X2	1	3.5094

### Inverse Cumulative Distribution Function

Student's t distribution with 14 DF

P( X <= x )	x
0.975	2.14479

- قيمة ( $t_{0.025,14} = 2.14479$ ) ، وبواقي ستودنت كما في الجدول التالي .

- جدول (2-1) بواقي ستودنت والمشاهدات القاصية في المتغير التابع الجهد.

i	Y	X1	X2	e(Y/X1,X2)	Hii	$d_i^*$	$ d_i^*  > t$
1	1.95	6336	0	0.070315	0.168134	0.60578	
2	2.5	7099	0	0.196309	0.129856	1.81414	
3	2.93	8026	0	0.111166	0.203451	1.00007	
4	1.69	6230	0	-0.13078	0.180514	-1.17452	
5	1.23	5369	0	-0.112315	0.344897	-1.12388	
6	3.13	8343	0	0.135006	0.25885	1.28629	
7	1.55	6522	0.006	-0.028672	0.10638	-0.23573	
8	1.94	7310	0.006	-0.076571	0.060196	-0.62105	
9	2.18	7974	0.006	-0.205562	0.095191	-1.87496	
10	2.7	8501	0.006	0.02158	0.171084	0.18408	
11	1.32	6646	0.012	0.076795	0.136976	0.65084	
12	1.6	7384	0.012	-0.053318	0.076233	-0.43318	
13	1.89	8000	0.012	-0.105635	0.089473	-0.88225	
14	2.15	8545	0.012	-0.148496	0.149695	-1.32516	
15	1.09	6755	0.018	0.190598	0.255876	1.92802	
16	1.26	7362	0.018	0.023283	0.186366	0.20051	
17	1.57	7934	0.018	0.015417	0.172565	0.13155	
18	1.92	8554	0.018	0.020878	0.214262	0.18291	

- الاستنتاج :-

نلاحظ من الجدول السابق أنه لا يوجد نقاط قاصية في الجهد (المتغير التابع) لأن جميع القيم المطلقة لبواقي ستودنت المحذوفة أقل من قيمة  $(t_{0.025,14} = 2.14479)$ .

مع ملاحظة أن المشاهدة رقم 5 كانت قاصية بالنسبة للمتغيرات المستقلة كما مر معنا سابقا .

- تطبيق (2-2) :-

بالعودة إلى تطبيق (2-1) في بيانات لمشروع تنفيذي ، نستخرج في هذا التطبيق المشاهدات القاصية في التكلفة (التعويضات) Y وذلك باستخدام بواقي ستودنت المحذوفة.

- التحليل الإحصائي :-

باستخدام برنامج الحزم الإحصائية Minitab سوف نقوم باستخراج بواقي ستودنت المحذوفة واكتشاف المشاهدات القاصية في التكلفة (التعويضات) (المتغير التابع).

## Regression Analysis: Y versus X3

The regression equation is

$$Y = 357 + 0.00124 X3$$

Predictor	Coef	SE Coef	T	P
Constant	356.98	22.22	16.07	0.000
X3	0.0012418	0.0003070	4.05	0.000

$$S = 102.534 \quad R\text{-Sq} = 34.5\% \quad R\text{-Sq}(\text{adj}) = 32.4\%$$

## Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	172041	172041	16.36	0.000
Residual Error	31	325911	10513		
Total	32	497952			

## Unusual Observations

Obs	X3	Y	Fit	SE Fit	Residual	St Resid
5	34000	676.0	399.2	18.1	276.8	2.74R
25	337119	698.0	775.6	92.0	-77.6	-1.72 X

R denotes an observation with a large standardized residual.

X denotes an observation whose X value gives it large influence.

## Inverse Cumulative Distribution Function

Student's t distribution with 30 DF

P( X ≤ x )	x
0.975	2.04227

- قيمة  $(t_{0.025,30} = 2.04227)$  ، وبواقي ستودنت كما في الجدول التالي .

- جدول (2-2) بواقي ستودنت والمشاهدات القاصية في المتغير التابع التكلفة.

i	Y	X3	e(Y/X3)	h <sub>ii</sub>	$d_i^*$	$ d_i^*  > t$
1	450	48000	33.41	0.03052	0.32613	
2	387	55900	-39.4	0.031775	-0.38512	
3	368	13783	-6.098	0.037999	-0.05965	
4	277	27765	-114.461	0.032407	-1.14035	
5	676	34000	276.796	0.031043	3.09999	3.09999
6	454	26500	64.11	0.032768	0.62953	
7	507	30800	111.77	0.031656	1.11196	
8	496	41000	88.103	0.030342	0.86914	
9	487	25900	97.855	0.03295	0.96954	
10	383	8600	15.338	0.040962	0.15033	
11	311	21075	-72.154	0.034645	-0.71047	
12	271	6874	-94.518	0.042056	-0.94007	
13	524	39000	118.587	0.030453	1.18208	
14	498	34300	98.423	0.030995	0.97434	
15	343	19405	-38.08	0.035329	-0.37284	
16	354	3586	-7.435	0.044287	-0.07297	
17	324	3905	-37.831	0.044062	-0.37209	
18	225	4139	-137.122	0.043899	-1.38797	
19	254	6255	-110.75	0.042461	-1.10785	
20	208	10605	-162.152	0.039759	-1.65882	
21	518	65392	79.812	0.034764	0.78742	
22	406	89400	-62.002	0.049532	-0.61399	
23	332	55200	-93.531	0.031619	-0.9248	
24	340	7800	-26.668	0.041463	-0.26163	
25	698	337119	-77.628	0.80529	-1.77421	
26	306	52000	-115.557	0.031016	-1.15089	
27	613	50500	193.305	0.030796	2.00626	
28	302	18625	-78.111	0.035666	-0.77067	
29	540	97937	61.396	0.057274	0.61044	
30	293	12300	-79.256	0.038798	-0.7835	
31	528	71800	81.854	0.037695	0.80926	
32	456	87700	-9.891	0.048147	-0.09728	
33	417	14600	41.887	0.037576	0.4108	

- الاستنتاج :-

نلاحظ من الجدول السابق وجود نقطة واحدة شاذة في متغير التكلفة (المتغير التابع) وهي المشاهدة رقم 5. وذلك

لأن قيمة بواقي الحذف المعيرة المطلقة أكبر من قيمة  $(t_{0.025,30} = 2.04227)$  . ويعتبر ذلك مؤشر جيد لاكتشاف المشاهدات القاصية في المتغير التابع. حيث أنها تساوي :

$$d_5^* = 3.09999$$

وتستدعي هذه المشاهدة القاصية دراستها وقياس أثرها على النموذج.

ويلاحظ أن المشاهدة رقم 25 كانت قاصية بالنسبة للمتغير المستقل X3 كما مر معنا سابقا .

### - تطبيق (2-3) :-

بالعودة إلى تطبيق (1-3) في بيانات لوحدة جراحة ، نستخرج في هذا التطبيق المشاهدات القاصية في الفترة التي يعيشها المريض بعد الجراحة Y (المتغير التابع) وذلك باستخدام بواقي ستودنت المحذوفة.

### - التحليل الإحصائي :-

باستخدام برنامج الحزم الإحصائية Minitab سوف نقوم باستخراج بواقي ستودنت المحذوفة واكتشاف المشاهدات القاصية في الفترة التي يعيشها المريض بعد الجراحة Y .

### Regression Analysis: Y versus X1; X2; X3

The regression equation is

$$Y = -659 + 38.3 X1 + 4.57 X2 + 4.49 X3$$

Predictor	Coef	SE Coef	T	P
Constant	-659.18	55.67	-11.84	0.000
X1	38.323	5.326	7.20	0.000
X2	4.5677	0.4996	9.14	0.000
X3	4.4850	0.4002	11.21	0.000

$$S = 61.2185 \quad R\text{-Sq} = 83.3\% \quad R\text{-Sq(adj)} = 82.2\%$$

### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	931546	310515	82.85	0.000
Residual Error	50	187385	3748		
Total	53	1118931			

Source	DF	Seq SS
X1	1	155274
X2	1	305511
X3	1	470760



## Unusual Observations

Obs	X1	Y	Fit	SE Fit	Residual	St Resid
13	5.8	830.00	512.89	23.67	317.11	5.62R
18	3.7	34.00	-100.54	21.82	134.54	2.35R
28	11.2	574.00	520.84	31.33	53.16	1.01 X
38	4.3	120.00	75.87	32.98	44.13	0.86 X

R denotes an observation with a large standardized residual.

X denotes an observation whose X value gives it large influence.

## Inverse Cumulative Distribution Function

Student's t distribution with 49 DF

P( X <= x )	x
0.975	2.00958

- قيمة ( $t_{0.025,49} = 2.00958$ ) ، وبواقي ستودنت كما في الجدول التالي .

- جدول (2-3) بواقي ستودنت والمشاهدات القاصية في الفترة التي يعيشها المريض بعد الجراحة Y.

I	Y	X1	X2	X3	e(Y/X1,X2,X3)	Hii	$d_i^*$	$ d_i^*  > t$
1	200	6.7	62	81	-44.07	0.026364	-0.72611	
2	101	5.1	59	66	-0.775	0.029496	-0.01272	
3	204	7.4	57	83	-53.028	0.045194	-0.88453	
4	101	6.5	73	41	-6.249	0.078851	-0.1053	
5	509	7.8	65	115	56.58	0.123355	0.98686	
6	80	5.8	38	72	20.411	0.06224	0.34124	
7	80	5.7	46	63	28.067	0.047159	0.46599	
8	127	3.7	68	81	-29.508	0.053411	-0.49166	
9	202	6	67	93	-91.903	0.031147	-1.54624	
10	203	3.7	76	94	-48.356	0.072011	-0.81723	
11	329	6.3	84	83	-9.201	0.05003	-0.1527	
12	65	6.7	51	43	41.606	0.08116	0.70544	
13	830	5.8	96	114	317.111	0.14949	9.15314	9.15314
14	330	5.8	83	88	-6.897	0.049811	-0.11444	
15	168	7.7	62	67	-51.603	0.047838	-0.86161	
16	217	7.4	74	68	-50.404	0.044803	-0.83994	

I	Y	X1	X2	X3	e(Y/X1,X2,X3)	Hii	$d_i^*$	$ d_i^*  > t$
17	87	6	85	28	2.405	0.149911	0.04218	
18	34	3.7	51	41	134.544	0.127056	2.46928	2.46928
19	215	7.3	68	74	-48.075	0.036141	-0.79697	
20	172	5.6	57	87	-33.987	0.025029	-0.55837	
21	109	5.2	52	76	-9.484	0.028868	-0.15566	
22	136	3.4	83	53	48.053	0.127368	0.83777	
23	70	6.7	26	68	48.674	0.123988	0.84708	
24	220	5.8	67	86	-34.844	0.022892	-0.57191	
25	276	6.3	59	100	-24.254	0.046282	-0.40227	
26	144	5.8	61	73	-25.132	0.01958	-0.41115	
27	181	5.2	52	86	17.666	0.031085	0.29047	
28	574	11.2	76	90	53.164	0.261873	1.01103	
29	72	5.2	54	56	34.082	0.047443	0.56653	
30	178	5.8	76	59	3.143	0.043047	0.05195	
31	71	3.2	64	65	23.684	0.080841	0.40013	
32	58	8.7	45	23	75.068	0.21128	1.39368	
33	116	5	59	73	-13.338	0.025231	-0.21857	
34	295	5.8	72	93	-14.078	0.034596	-0.23181	
35	115	5.4	58	70	-11.644	0.023871	-0.19066	
36	184	5.3	51	99	-36.904	0.047925	-0.61395	
37	118	2.6	74	86	-46.185	0.105898	-0.79491	
38	120	4.3	8	119	44.13	0.290152	0.85327	
39	151	4.8	61	76	6.736	0.026137	0.11039	
40	148	5.4	52	88	-31.969	0.031738	-0.52685	
41	95	5.2	49	72	8.16	0.035264	0.13436	
42	75	3.6	28	99	24.302	0.139301	0.42438	
43	483	8.8	86	88	17.431	0.124222	0.30148	
44	153	6.5	56	77	-38.059	0.02653	-0.62627	
45	191	3.4	77	93	-48.942	0.082842	-0.83221	
46	123	6.5	40	84	-26.371	0.063129	-0.44144	
47	311	4.5	73	106	-11.131	0.068607	-0.18658	
48	398	4.8	86	101	27.417	0.084136	0.46428	
49	158	5.1	67	77	-29.652	0.023326	-0.48636	
50	310	3.9	82	103	-16.792	0.094148	-0.28554	
51	124	6.6	77	46	-27.778	0.071383	-0.46717	
52	125	6.4	85	40	-28.745	0.105363	-0.49265	
53	198	6.4	59	85	-38.81	0.026362	-0.63867	
54	313	8.8	78	72	-44.266	0.094799	-0.75675	

## - الاستنتاج :-

نلاحظ من الجدول السابق أن هناك مشاهدتان قاصيتان في قيم المتغير التابع  $Y$  وهي المشاهدتان 13,18 ذلك لأن القيم المطلقة لبواقي ستودنت المعيرة لهذه المشاهدات أكبر من قيمة  $(t_{0.025,49} = 2.00958)$ .

$$d_{13}^* = 9.15314, \quad d_{18}^* = 2.46928$$

وتستدعي هذه المشاهدات دراستها وقياس أثرها على النموذج والتحليل الإحصائي . ونلاحظ أن المشاهدات رقم 13, 17, 28 ,32,38 كانت مشاهدات قاصية في المتغيرات المستقلة كما مر معنا سابقا.

## - تطبيق (2-4) :-

بالعودة إلى تطبيق (1-4) في بيانات لمبيعات في 15 منطقة بيع، نستخرج في هذا التطبيق المشاهدات القاصية في المبيعات  $Y$  (المتغير التابع) و ذلك باستخدام بواقي ستودنت المحذوفة.

## - التحليل الإحصائي :-

باستخدام برنامج الحزم الإحصائية Minitab سوف نقوم باستخراج بواقي ستودنت المحذوفة واكتشاف المشاهدات القاصية في المبيعات  $Y$  .

### Regression Analysis: Y versus X1; X2; X3

The regression equation is  
 $Y = 179 + 2.11 X1 + 3.56 X2 - 22.2 X3$

Predictor	Coef	SE Coef	T	P
Constant	178.521	8.318	21.46	0.000
X1	2.1055	0.6479	3.25	0.008
X2	3.56240	0.09945	35.82	0.000
X3	-22.1880	0.5286	-41.98	0.000

$S = 4.97952 \quad R-Sq = 99.7\% \quad R-Sq(adj) = 99.6\%$

### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	89274	29758	1200.14	0.000
Residual Error	11	273	25		
Total	14	89547			

Source	DF	Seq SS
X1	1	1074
X2	1	44505
X3	1	43695

```
MTB > InvCDF 0.975;
SUBC> T 10.
```

### Inverse Cumulative Distribution Function

Student's t distribution with 10 DF

```
P ( X <= x )      x
0.975      2.22814
```

- قيمة  $(t_{0.025,10} = 2.22814)$  ، وبواقي ستودنت كما في الجدول التالي .

- جدول (2-4) بواقي ستودنت والمشاهدات القاصية في المبيعات Y.

I	X1	X2	X3	Y	e(Y/X1,X2,X3)	hii	$d_i^*$	$ d_i^*  > t$
1	5.5	31	10	79.3	0.64432	0.223605	0.14015	
2	2.5	55	8	200.1	-2.11269	0.20889	-0.45959	
3	8	67	12	163.2	-4.59006	0.441628	-1.26708	
4	3	50	7	200.1	-7.54144	0.200942	-1.87908	
5	3	38	8	146	3.29538	0.231713	0.73929	
6	2.9	71	12	177.7	6.39862	0.446299	1.92855	
7	8	30	12	30.9	-5.08116	0.452146	-1.44523	
8	9	56	5	291.9	5.87488	0.476566	1.78558	
9	4	42	8	160	0.94022	0.139043	0.19439	
10	6.5	73	5	339.4	-1.9221	0.414587	-0.48668	
11	5.5	60	11	159.6	-0.17736	0.156596	-0.03698	
12	5	44	12	86.3	6.76184	0.169198	1.5899	
13	6	50	6	237.5	1.35393	0.182656	0.28794	
14	5	39	10	107.2	1.09787	0.124677	0.22526	
15	3.5	55	10	155	-4.94225	0.131457	-1.0722	

- الاستنتاج :-

نلاحظ من الجدول السابق أنه لا يوجد قيم قاصية في قيم المتغير التابع Y.

وكذلك أيضا نلاحظ أنه لا يوجد مشاهدات قاصية في قيم المتغيرات المستقلة كما مر معنا سابقا .

