

Standardized residuals and leverage points - example

The rain/wheat data:

	rain	wheat
1	12	310
2	14	320
3	13	323
4	16	330
5	18	334
6	20	348
7	19	352
8	22	360
9	22	370
10	20	344
11	23	370
12	24	380
13	26	385
14	27	393
15	28	395
16	29	400
17	30	403
18	31	406
19	26	383
20	27	388
21	28	392
22	29	398
23	30	400
24	31	403
25	20	270
26	50	260

For this data the variance of *rain* is: 59.85385, therefore $\sum_{i=1}^{26} (x_i - \bar{x})^2 = (26 - 1)(59.85385) = 1496.346$. The mean of *rain* is computed $\bar{x} = 24.42308$.

We now run the regression of *wheat* on *rain* and obtain:

Call:

```
lm(formula = wheat ~ rain)
```

Residuals:

Min	1Q	Median	3Q	Max
-131.57	-18.89	14.44	28.49	36.25

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	334.137	26.938	12.404	6.28e-12 ***
rain	1.149	1.053	1.091	0.286

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 40.74 on 24 degrees of freedom

Multiple R-Squared: 0.04722, Adjusted R-squared: 0.007518

F-statistic: 1.189 on 1 and 24 DF, p-value: 0.2863

Therefore the estimate of σ^2 is $s_e = 40.74$.

The residuals, leverage values, and standardized residuals from this regression are listed below (from R):

Residuals:

1	2	3	4	5	6	7
-37.9216553	-30.2190978	-26.0703766	-22.5165403	-20.8139828	-9.1114253	-3.9627040
8	9	10	11	12	13	14
0.5911322	10.5911322	-13.1114253	9.4424110	18.2936898	20.9962473	27.8475260
15	16	17	18	19	20	21
28.6988048	32.5500835	34.4013623	36.2526410	18.9962473	22.8475260	25.6988048
22	23	24	25	26		
30.5500835	31.4013623	33.2526410	-87.1114253	-131.5730626		

Leverage values:

1	2	3	4	5	6	7	8
0.14160134	0.11106542	0.12566508	0.08587585	0.06603264	0.05153579	0.05811592	0.04238530
9	10	11	12	13	14	15	16
0.04238530	0.05153579	0.03981493	0.03858116	0.04012338	0.04289937	0.04701195	0.05246112
17	18	19	20	21	22	23	24
0.05924688	0.06736923	0.04012338	0.04289937	0.04701195	0.05246112	0.05924688	0.06736923
25	26						
0.05153579	0.47564580						

Standardized residuals:

1	2	3	4	5	6	7
-1.00454535	-0.78663519	-0.68428208	-0.57799735	-0.52858656	-0.22961625	-0.10021199
8	9	10	11	12	13	14
0.01482573	0.26562795	-0.33041991	0.23650058	0.45790119	0.52596973	0.69860967
15	16	17	18	19	20	21
0.72151748	0.82069231	0.87049160	0.92132225	0.47586842	0.57317489	0.64609440
22	23	24	25	26		
0.77026588	0.79457964	0.84508044	-2.19528759	-4.45945089		

Let's see how R computes these values. The leverage value for point i is equal to:

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Therefore, the leverage value of point 1 is:

$$h_{11} = \frac{1}{26} + \frac{(12 - 24.42308)^2}{1496.346} = 0.14160.$$

The standardized residual for point i is computed as follows:

$$e_i^* = \frac{e_i}{sd(e_i)} = \frac{e_i}{s_e \sqrt{1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} = \frac{e_i}{s_e \sqrt{1 - h_{ii}}}.$$

Therefore the standardized residual for point 1 is equal to:

$$e_1^* = \frac{e_1}{s_e \sqrt{1 - h_{11}}} = \frac{-37.9216553}{40.74 \sqrt{1 - 0.14160134}} = -1.004666.$$

Any differences are due to rounding.