Statistics – Linear Regression Example – Prof. Richard B. Goldstein

Sample Data:

x	y
2	3
5	5
7	12
7	15
9	10
12	18
14	23

$\Sigma x = 56$ =sum(array)
$\Sigma y = 86$
$\Sigma x^2 = 548$ =sumsq(array)
$\Sigma xy = 848$ =sumproduct(array1,array2)
$\Sigma y^2 = 1{,}356$

$S_{xx} = 548 - (56)^2/7 = 100$
$S_{xy} = 848 - (56)(86)/7 = 160$
$S_{yy} = 1{,}356 - (86)^2/7 = 299.4286$

Line of Regression: $\hat{y} = a + bx$

$b = S_{xy}/S_{xx} = 160/100 = 1.6 \qquad a = \bar{y} - b\bar{x} = 86/7 - 1.6(56/7) = -0.5143$

$$s_e^2 = \frac{SSE}{n-2} = \frac{S_{yy} - bS_{xy}}{n-2} = \frac{299.4286 - 1.6(160)}{7-2} = 8.6857,\ s_e = 2.9472$$

$t_{0.025,\,5} = 2.5706 = \text{tinv}(0.05, 5)$

standard errors:

$$s(b) = \frac{s_e}{\sqrt{S_{xx}}} = \frac{2.9472}{\sqrt{100}} = 0.29472 \quad s(a) = \frac{s_e\sqrt{\Sigma x^2}}{\sqrt{nS_{xx}}} = \frac{2.9472\sqrt{548}}{\sqrt{7(100)}} = 2.6077$$

confidence intervals:

$b \pm t_{\alpha/2}s(b) = 1.6 \pm 2.5706(0.29472) = [0.8424, 2.3576]$
$a \pm t_{\alpha/2}s(a) = -0.5143 \pm 2.5706(2.6077) = [-7.2177, 6.1891]$

$$\text{mean response: } -0.5143 + 1.6x_0 \pm 2.5706(2.9472)\sqrt{\frac{1}{7} + \frac{(x_0 - 8)^2}{100}}$$

$$\text{single response: } -0.5143 + 1.6x_0 \pm 2.5706(2.9472)\sqrt{1 + \frac{1}{7} + \frac{(x_0 - 8)^2}{100}}$$

analysis of variance

$SSE = S_{yy} - bS_{xy} = 43.4286 \qquad SST = S_{yy} = 299.4286 \qquad SSR = SST - SSE = 256$

$R^2 = SSR/SST = 256/299.4286 = 0.8550 \qquad r = \sqrt{R^2} = 0.9246$

$$\text{Adjusted } R^2 = 1 - \frac{n-2}{n-1}(1 - R^2) = 1 - \frac{6}{5}(1 - 0.8550) = 0.8260$$

Excel's Output:

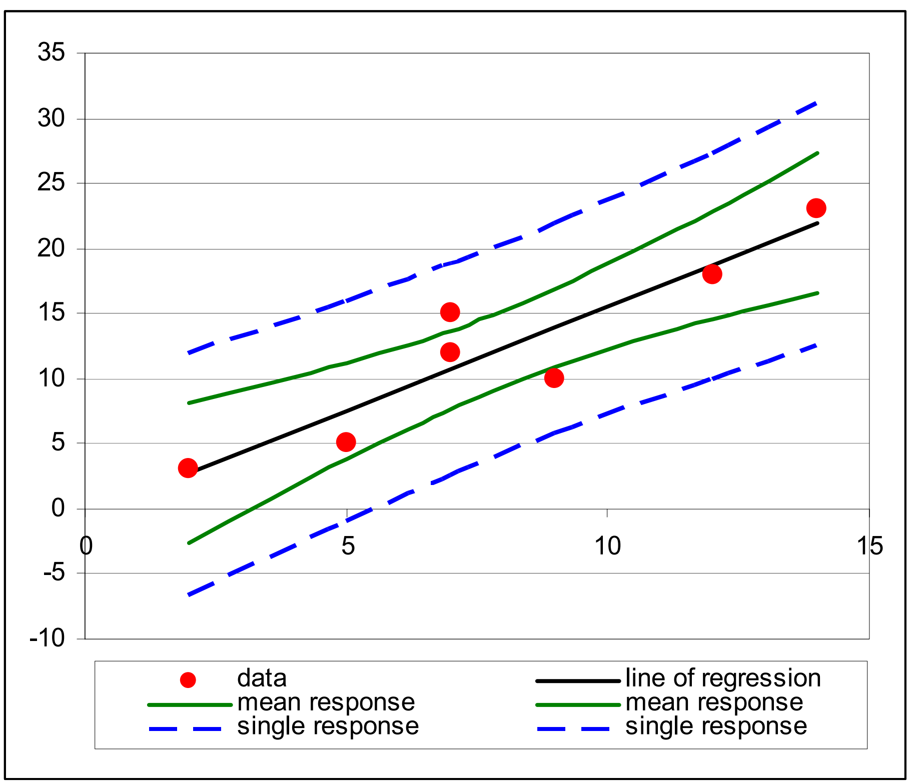
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.924641
R Square	0.854962
Adjusted R Square	0.825954
Standard Error	2.947154
Observations	7

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	256	256	29.4737	0.00287424
Residual	5	43.42857143	8.685714		
Total	6	299.4285714			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.51429	2.607618352	-0.19722	0.85142	-7.21738208	6.1888107
X	1.6	0.294715359	5.428967	0.00287	0.842410052	2.3575899



Walpole – page 378 example

x	y	y fit	group m	(y-ym)^2	(yfit-ym)^2	(group m-fit)^2	(y-group m)^2
150	77.4	77.743333	77.43333	82.50694	76.3876	0.0961	0.00111111
150	76.7	77.743333	77.43333	95.71361	76.3876	0.0961	0.53777778
150	78.2	77.743333	77.43333	68.61361	76.3876	0.0961	0.58777778
200	84.1	83.57	84.1	5.680278	8.487511111	0.2809	0.00000000
200	84.5	83.57	84.1	3.933611	8.487511111	0.2809	0.16000000
200	83.7	83.57	84.1	7.746944	8.487511111	0.2809	0.16000000
250	88.9	89.396667	89.26667	5.840278	8.487511111	0.0169	0.13444444
250	89.2	89.396667	89.26667	7.380278	8.487511111	0.0169	0.00444444
250	89.7	89.396667	89.26667	10.34694	8.487511111	0.0169	0.18777778
300	94.8	95.223333	95.13333	69.16694	76.3876	0.0081	0.11111111
300	94.7	95.223333	95.13333	67.51361	76.3876	0.0081	0.18777778
300	95.9	95.223333	95.13333	88.67361	76.3876	0.0081	0.58777778
				513.1167	509.2506667	1.2060	2.6600

mean y (ym)	86.48333
a	60.26333
b	0.116533

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.996226
R Square	0.992466
Adjusted R Square	0.991712
Standard Error	0.621772
Observations	12

ANOVA

	df	*SS*	*MS*	*Computed f*	*P-values*
Regression	1	509.2507	509.2507	1531.58	5.99373E-12
Error	10	3.8660			
Lack of fit	2	1.2060	0.6030	1.81	0.2241
Pure error	8	2.6600	0.3325		
Total	11	513.11667			

	Coeffs	*St Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	60.26333	0.7443968	80.95592	2.02E-15	58.6047138	61.92195286
X Variable 1	0.116533	0.0032108	36.294	5.99E-12	0.109379192	0.123687475

Conclusion: "The partitioning of the total variation in this manner reveals a significant variation accounted for by the linear model and an insignificant amount of variation due to lack-of-fit."