

---

**SOLUTIONS TO**

**TEST YOUR KNOWLEDGE: Regression**

A company wants to determine the relationship between years of education and how long it takes to a certain key task (in hours). The data is below:

<u>X (years of education)</u>	<u>Y (time to complete task in hours)</u>
9	13.0
10	12.0
10	10.5
11	3.5
11	6.5
11	8.0
12	9.0
12	9.5
12	8.5
13	10.0
14	9.0
15	10.0
15	7.0
15	7.5
16	6.5
16	7.0
17	10.0
18	4.0
19	6.0
20	7.0
20	5.5
20	6.0

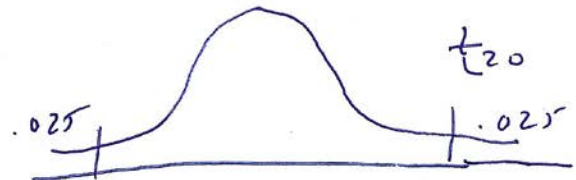
$$\sum X = 316; \quad \sum Y = 176; \quad \sum XY = 2429.5; \quad \sum X^2 = 4802; \quad \sum Y^2 = 1530 \quad n = 22$$

(a) Calculate the correlation coefficient and test for significance at the .05 level.

$$\begin{aligned} r &= \frac{22(2429.5) - 316(176)}{\sqrt{[22(4802) - (316)^2][22(1530) - (176)^2]}} = \frac{-2167}{\sqrt{[5788][2684]}} \\ &= -2167/3941.4 = -.55 \end{aligned}$$

$$(a) \quad r = -0.55; \quad r^2 = (-0.55)^2 = 0.3025 = 30.25\%$$

$$t_{20} = \frac{-0.55\sqrt{20}}{\sqrt{1-0.3025}} = \frac{-2.46}{0.835} = -2.94$$



(b) Do the regression and explain the meaning of the slope term and the Y-intercept.

$$(b) \quad b_1 = \frac{-2167}{5788} = -0.374$$

$$b_0 = \frac{176}{22} - (-0.374)\left(\frac{316}{22}\right) = 8 + 5.372 = 13.37$$

$$\hat{y}_{\text{Time}} = 13.37 - 0.374(X) \quad \leftarrow \begin{array}{l} \text{year of} \\ \text{education} \end{array}$$

(c) What percentage of the variation in time to complete task is explained by years of education?

(c) 30.25% explained; Every year of education reduces by 0.374 hours the time it takes to complete task

(d) According to the regression model, how long should the task take for someone who has 14.5 years of education?

$$(d) \quad \hat{y} = 13.37 - 0.374(14.5) = 13.37 - 5.42 = 7.95$$

7.95 hours

(e) Using Excel show the scatterplot. Also, do the regression in MS Excel and see if your calculations above are correct.

Years of Education	Time to Complete Task (hours)				
9	13	117	81	169	
10	12	120	100	144	
10	10.5	105	100	110.25	
11	3.5	38.5	121	12.25	
11	6.5	71.5	121	42.25	
11	8	88	121	64	
12	9	108	144	81	
12	9.5	114	144	90.25	
12	8.5	102	144	72.25	
13	10	130	169	100	
14	9	126	196	81	
15	10	150	225	100	
15	7	105	225	49	
15	7.5	112.5	225	56.25	
16	6.5	104	256	42.25	
16	7	112	256	49	
17	10	170	289	100	
18	4	72	324	16	
19	6	114	361	36	
20	7	140	400	49	
20	5.5	110	400	30.25	
20	6	120	400	36	
316	176	2429.5	4802	1530	

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.549798302
R Square	0.302278173
Adjusted R	0.267392082
Standard E	2.063032512
Observatio	22

ANOVA

	df	SS	MS	F	nificance F
Regression	1	36.87793711	36.87794	8.664718843	0.008
Residual	20	85.12206289	4.256103		
Total	21	122			

	Coefficients	Standard Error	t Stat	P-value	ower 95	Upper 95%
Intercept	13.37767795	1.879112964	7.119145	6.72863E-07	9.458	17.29744
X Variable :	-0.374395301	0.127190054	-2.94359	0.008031609	-0.64	-0.109081