Lesson 25: Practice Problems
Answer Key

1.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.987479
R Square	0.975114
Adjusted R Square	0.974574
Standard Error	35.45634
Observations	48

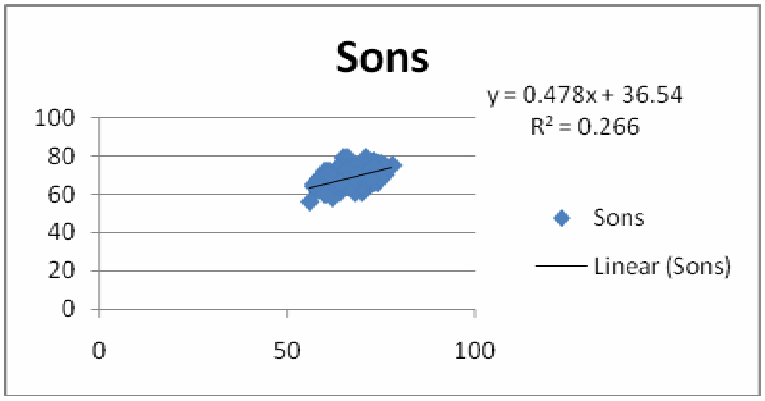
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2265973	2265973	1802.465	1.51E-38
Residual	46	57829	1257.152		
Total	47	2323802			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-260.046	19.28559	-13.484	1.36E-17	-298.866	-221.226	-298.866	-221.226
Weight	3866.567	91.07351	42.45545	1.51E-38	3683.246	4049.888	3683.246	4049.888

➔Price = -260.05 + 3,866.57*weight

2.



➔According to the scatter plot trend line, **Sons = 36.54 + 0.478*fathers**

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.516277
R Square	0.266542
Adjusted R Square	0.264699
Standard Error	3.224615
Observations	400

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1503.936	1503.936	144.635	1.25E-28
Residual	398	4138.462	10.39814		
Total	399	5642.398			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	36.54138	2.678644	13.64174	4.94E-35	31.27532	41.80744	31.27532	41.80744
Fathers	0.478923	0.039823	12.02643	1.25E-28	0.400635	0.557212	0.400635	0.557212

➔According to 'Regression', **Sons= 36.54 + 0.479*fathers**

➔There is only a very small difference of 0.001 in the slope of the regression.

3.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.224823
R Square	0.050545
Adjusted R Square	0.046717
Standard Error	8.282066
Observations	250

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	905.5965	905.5965	13.20254	0.00034
Residual	248	17010.97	68.59261		
Total	249	17916.56			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	17.93333	11.47593	1.56269	0.119401	-4.66939	40.53604	-4.66939	40.53604
Height	0.604111	0.16626	3.63353	0.00034	0.276649	0.931573	0.276649	0.931573

a) Income = 17.39 + 0.604*Height
b) Intercept Interpretation: CEOs with a height of 0 would have an income of 17.39
   Slope Interpretation: An increase of 1 inch of height implies an increase of $0.604

c) $1.812, on average.

4.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.537831
R Square	0.289262
Adjusted R Square	0.277008
Standard Error	5.88848
Observations	60

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	818.4965	818.4965	23.60535	9.35E-06
Residual	58	2011.103	34.6742		
Total	59	2829.6			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.635664	2.225896	1.633349	0.107813	-0.81995	8.091279	-0.81995	8.091279
Length	0.267483	0.055054	4.858533	9.35E-06	0.15728	0.377685	0.15728	0.377685

➔ Test_score = 3.636 + 0.267*Length
➔ An increase of one second implies an increase of around 1/3 of a point in the score.

5.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.291909
R Square	0.085211
Adjusted R Square	0.047094
Standard Error	132.9601
Observations	26

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	39520.86	39520.86	2.235548	0.147903
Residual	24	424281.2	17678.38		
Total	25	463802			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	296.9197	64.30518	4.617352	0.00011	164.2003	429.639	164.2003	429.639
Ads	21.35597	14.28326	1.495175	0.147903	-8.12323	50.83517	-8.12323	50.83517

Due to the output of the regression, one can observed that there is a fixed number of 296 customers that will enter the store per week. Moreover, as ads take place every week, the number of customers entering the store will increase in 21 per each weekly ad. However, based on the 'R-Squared' it can be noticed that the number of weekly ads only explain an 8.5% of the variation in the number of customer. Furthermore, the number of ads per week is only significant variable at a 15% level.

6.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.43402
R Square	0.188373
Adjusted R Square	0.177968
Standard Error	1.813176
Observations	80

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	59.51651	59.51651	18.10328	5.77E-05
Residual	78	256.4335	3.287609		
Total	79	315.95			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	30.63307	1.044142	29.33805	6.85E-44	28.55435	32.7118	28.55435	32.7118
Age	-0.11692	0.027479	-4.2548	5.77E-05	-0.17162	-0.06221	-0.17162	-0.06221

Results show that the older the employee is or get, the smaller the employments spell will be. More precisely, the fact of being or getting one year older implies that the employee will work 0.11 months less.

➔ This will not help too much for a HR policy. Even though the age of the employee is significant at a 5% level for explaining the length of employment spells, it only explains 18.83% percent of its variance. Hence, other relevant factors would have been ignored if decision were made based on this result.

7.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.307947
R Square	0.094831
Adjusted R Square	0.082921
Standard Error	67.30373

Observations	78

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	36067.27	36067.27	7.962237	0.006092
Residual	76	344264.2	4529.792		
Total	77	380331.4			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	114.0854	12.06375	9.456875	1.79E-14	90.05836	138.1125	90.05836	138.1125
Revenues	0.000725	0.000257	2.821744	0.006092	0.000213	0.001237	0.000213	0.001237

Since we want to explore the number of desk copies given the revenues made, 'Revenues' will be considered as independent variable and 'Copies' as dependent.

The results show that when revenues are zero, each representative has to give on average 114 desk copies. Moreover, for every increase of $10,000 in revenues, on average each sales representative should give in addition 7 desk copies.

If we plug in the formula the average revenue, $ 36,385.7692, we obtain that the number of copies should be 140.47. This value is very similar to the average number of copies, obtained from the data. Hence, it is not too many.