AE Additional Exercises

1. A pharmacist is interested in studying the relationship between the amount of a particular drug in the bloodstream (in mg) and reaction time (in seconds) of subjects taking the drug. Ten subjects are randomly selected and administered various doses of the drug. The reaction times (in seconds) are measured 15 minutes after the drug is administered with the following results.

Reaction Times

Amount of Drug (mg)	Reaction Time (Seconds)	Amount of Drug (mg)	Reaction Time (Seconds)
1	0.5	6	0.8
2	0.7	7	0.9
3	0.6	8	0.6
4	0.7	9	0.9
5	0.8	10	1.0

A regression analysis has been performed to estimate the model, and the following output was produced.

$$\text{Reaction Time} = \beta_0 + \beta_1\left(\text{Amount of Drug}\right) + \varepsilon_i$$

Regression Analysis: Reaction Time (Seconds) versus Amount of Drug (mg)

The regression equation is
Reaction Time (Seconds) = 0.533 + 0.0394 Amount of Drug (mg)

Predictor	Coef	SE Coef	T	P
Constant	0.53333	0.07521	7.09	0.000
Amount of Drug (mg)	0.03939	0.01212	3.25	0.012

S = 0.110096 R-Sq = 56.9% R-Sq(adj) = 51.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.12803	0.12803	10.56	0.012
Residual Error	8	0.09697	0.01212		
Total	9	0.22500			

Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	0.6909	0.0393	(0.6003, 0.7815)	(0.4214, 0.9605)

Values of Predictors for New Observations

New Obs	Amount of Drug (mg)
1	4.00

a. Draw a scatterplot of the data. Describe the relationship you observe between the reaction time and the amount of drug in the bloodstream. Are there any unusual observations?

b. Find and interpret the standard deviation of the error terms in the output.

c. Interpret the slope coefficient.

d. What fraction of the variation in reaction time is explained by the amount of drug in the bloodstream? What other factors might affect reaction time?

e. Is there evidence of a linear relationship between the amount of drug in the bloodstream and reaction time? Test at the 0.05 significance level and the 0.01 significance level.

f. Construct and interpret a 95% confidence interval for β_1, the slope of the line.

g. Find the predicted value of the reaction time of an individual who has 4 mg of the drug in the bloodstream.

h. Find and interpret a 95% confidence interval for the average reaction time of all individuals who have 4 mg of the drug in their bloodstreams.

i. Suppose a particular individual has 4 mg of the drug in the bloodstream. What would be the 95% prediction interval for the reaction time?

2. A sample of 11 lonely hearts advertisements, all placed by males, was selected from the local newspaper. In each of the selected ads, the males gave their heights, along with other physical characteristics and preferences. Some of the males obviously felt that being taller than average might result in more responses to the ad. Suppose that y, the number of responses to the ad over the next 30 days, was determined for each male. The following table contains the data.

Height and Response

Height (Inches)	70	62	67	75	78	69	70	64	66	69	75
y	14	7	10	18	17	12	15	9	12	14	17

a. Draw a scatterplot of the data. Does the relationship appear to be linear?

b. Estimate the slope and intercept of the regression equation using statistical software.

c. Is there evidence of a linear relationship between the number of responses and height? Test at the $\alpha = 0.01$ significance level.

d. Interpret the regression coefficient corresponding to height.

e. Construct a 95% confidence interval for the slope.

f. Compute R^2 and interpret this value.

g. Estimate the number of responses for a male 6 feet tall. Round your answer to the nearest whole number.

h. Construct and interpret a 95% prediction interval for the number of responses for a male who is 6 feet tall.

i. Construct and interpret a 95% confidence interval for the average number of responses for a male who is 6 feet tall.

3. It is believed that when one is in the process of buying a home, the interest rate that is given on the loan is a function of his or her credit score. The Fair Isaac Corporation (FICO) is a major producer of credit scores. They have collected data from major lenders about buyers' history of borrowing and paying back credit. The following table contains 20 randomly selected loan applicants along with their FICO scores and the interest rate that they were given when financing their homes. With the data given, answer the following questions.

Credit Scores and Interest Rates

Observation	FICO Score	Interest Rate (%)
1	756	6.32
2	679	7.85
3	527	10.20
4	839	5.52
5	677	7.30
6	686	7.37
7	512	9.67
8	590	8.40
9	765	5.82
10	502	10.01
11	819	5.86
12	630	8.51
13	704	6.83
14	679	7.72
15	663	7.68
16	542	9.53
17	575	6.86
18	508	9.65
19	689	7.75
20	750	6.89

a. Draw a scatterplot of the data. Does there appear to be a linear relationship between FICO score and interest rate?

b. Estimate the simple linear regression equation using statistical software.

c. What is the estimate of the mean square error? Interpret this value.

d. Test at the 5% significance level if a linear relationship exists between FICO scores and interest rates.

e. Interpret the regression coefficient corresponding to FICO score.

f. Construct a 95% confidence interval for the slope. Interpret the interval.

g. Compute the coefficient of determination. Interpret this value.

h. Calculate the correlation coefficient. Interpret this value.

i. What is the average interest rate for a credit score of 725?

j. Construct a 90% confidence interval for the average interest rate for people who have FICO scores of 725. Interpret this interval.

k. Construct a 90% prediction interval for the interest rate for a person with a FICO score of 725. Interpret this interval.

4. It appears that many cellular phone service providers are making huge profits from customers using their messaging services such as text and multimedia messaging services (MMS). To that end, the cellular phone companies are using their marketing campaigns to target kids rather than adults. The belief is that kids tend to utilize their messaging services much more than adults. In fact, it is the belief that the younger one is, the more texts and MMS sent via his or her cell phone. Using the data given which reports the number of monthly messages sent by age, formulate a simple linear regression model to answer the following questions.

Age and Message Use

Age	Number of Messages	Age	Number of Messages
78	7	37	1541
36	1607	69	6
11	3037	69	25
69	26	55	517
56	491	39	1439
74	0	20	2505
22	2373	14	2845
74	5	10	3048
10	3059	80	0
26	2155	59	295
18	2619	40	1374
68	17	67	35
10	3067		

a. Draw a scatterplot of the data. Does there appear to be a linear relationship between age and the number of messages that one sends?

b. What is the estimated simple linear regression equation?

c. What is the estimate of the coefficient of determination? Interpret this value.

d. Test at the 5% significance level if a linear relationship exists between age and the number of messages sent via a cellular phone.

e. Interpret the regression coefficient corresponding to age.

f. Construct a 95% confidence interval for the slope. Interpret this interval.

g. Calculate the correlation coefficient. Interpret this value.

h. What is the average number of messages sent by a 15-year-old? Round your answer to the nearest whole number.

i. Construct a 95% confidence interval for the average number of messages sent by a 15-year-old. Interpret this interval.

j. Suppose Jacob's parents are contemplating giving him a cell phone but with a limited messaging plan at 500 per month. Eager to get the cell phone, Jacob, at 15 years old, promises that he won't send more than 500 messages per month and he'll also limit the number of friends that will have his phone number. In spite of Jacob's honesty and loyalty, should his parents believe that he won't send more than 500 messages per month? Explain your answer.

5. For the last 10 years, the Virginia Department of Mines, Minerals, and Energy (VDMME) has been promoting Energy Star, a resource for energy-efficient products and solutions. VDMME wants all energy consumers to take responsibility and exercise leadership by practicing conservation and efficiency on a daily basis. The average annual energy usage for a 1800 square foot home is 18,000 kilowatt hours. VDMME believes that this number can be significantly reduced if consumers started using Energy Star appliances. Answer the following questions based on data of 25 randomly selected homes with Energy Star appliances built within the last five years.

Home Size and Energy Usage

Home Size (Square Feet)	Annual Energy Usage (kWh)	Home Size (Square Feet)	Annual Energy Usage (kWh)
2895	15,200	2180	13,227
3650	17,333	4492	19,492
2927	15,050	6450	25,353
6289	24,763	1583	11,075
7252	27,098	4170	18,557
4147	18,291	4189	18,636
6505	25,028	3920	18,210
1413	11,099	6833	26,075
2279	13,110	4469	19,232
3251	15,844	6141	24,225
2992	14,904	5084	21,530
6912	26,329	6746	26,333
2503	13,765		

a. Draw a scatterplot of the data. Does there appear to be a linear relationship between home size and the amount of annual kWh used?

b. What is the estimated simple linear regression equation?

c. What is the estimate of the coefficient of determination? Interpret this value.

d. Test at the 5% significance level if a linear relationship exists between home size and the annual amount of kWh used.

e. Interpret the regression coefficient corresponding to home size.

f. Construct a 99% confidence interval for the slope. Interpret the interval.

g. Calculate the correlation coefficient. Interpret this value.

h. Suppose the James family constructed a 3200 square foot home using all Energy Star appliances. How many kilowatt hours should they expect to use in their first year in the home?

i. Construct a 95% confidence interval for the average number of kWh that will be used by the James family. Interpret this interval.

6. With grade inflation being a major problem in many U.S. high schools, college admissions offices are beginning to look at other performance measures when evaluating student applications. It is believed that many students with high grade point averages in high school will not necessarily score high on the SAT. Using the data of 30 randomly selected students that took the SAT, answer the following questions to determine if there is a linear relationship between high school GPA and SAT score.

High School GPA and SAT Score

High School GPA	SAT Score	High School GPA	SAT Score
3.21	1448	4.95	1960
2.23	1435	4.69	1717
2.89	1411	2.49	1365
1.84	1291	2.45	1561
3.34	1462	2.57	1474
2.42	1357	1.28	1328
2.75	1396	1.94	1302
2.35	1549	4.75	1622
4.80	1829	1.91	1499
1.98	1508	4.25	1566
2.92	1514	1.15	1413
4.18	1658	2.17	1428
4.50	1694	4.73	1720
4.42	1686	4.39	1783
4.78	1840	2.92	1614

a. Draw a scatterplot of the data. Does there appear to be a linear relationship between high school GPA and SAT score?

b. What is the estimated simple linear regression equation?

c. What is the coefficient of determination? Interpret this value.

d. Test at the 5% significance level if a linear relationship exists between high school GPA and SAT score.

e. Interpret the regression coefficient corresponding to high school GPA.

f. Construct a 95% confidence interval for the slope. Interpret the interval.

g. Calculate the correlation coefficient. Interpret this value.

h. What SAT score would you expect for students with a GPA of 3.5? Round your answer to the nearest whole number.