

AE Additional Exercises

1. Drew is undecided about whether to go back to school and get his master's degree. He is trying to perform a cost-benefit analysis to determine whether the cost of attending the school of his choice will be outweighed by the increase in salary he will receive after he attains his degree. He does research and compiles data on annual salaries in the industry he currently works in (he has been working for 10 years), along with the years of experience for each employee and whether or not the employee has a master's degree. Earning his master's degree will require him to take out approximately \$20,000 worth of student loans. He has decided that if the multiple regression model shows, with 95% confidence, that earning a master's degree is significant in predicting annual salary, and the estimated increase in salary is at least \$10,000, he will enroll in a degree program.

Data

This data set can be found at stat.hawkeslearning.com by navigating to **Discovering Business Statistics, Second Edition > Data Sets > Industry Salaries**.

Industry Salaries		
Salary (\$)	Years of Experience	Master's Degree
37,620	22	No
67,080	27	Yes
31,280	15	No
21,500	2	No
75,120	28	Yes
59,820	25	Yes
40,180	15	Yes
81,360	32	Yes
35,080	19	No
36,080	12	Yes
36,680	22	No
29,200	11	Yes
33,040	18	No
30,060	14	No
53,300	21	Yes
22,820	7	No
72,900	31	Yes
55,920	22	Yes
19,280	0	No
26,000	7	No

- a. Create an indicator variable, degree, that is equal to 1 if the employee has a master's degree and equal to 0 if the employee does not have a master's degree.
- b. Using statistical software, estimate the following multiple regression model.

$$\text{Salary} = \beta_0 + \beta_1 (\text{Experience}) + \beta_2 (\text{Degree}) + \varepsilon_i$$

Write the estimated multiple regression equation.

- c. According to the model, how much does salary increase on average with each additional year of experience?
- d. According to this model, will Drew decide to enroll in a master's program? Explain your answer.
- e. Why should Drew be cautious when using this model to make his decision?

Data

This data set can be found at stat.hawkeslearning.com by navigating to **Discovering Business Statistics, Second Edition > Data Sets > Sports Club Membership**.

2. A chain of sports clubs wishes to use regression analysis to help determine which features should be included in their new location. They believe that median income in the area is a significant factor in determining the number of people who join a neighborhood sports club. The CEO of the chain gathered data from existing sports clubs regarding the number of members each club had, the median income in the area in which they were located, and whether or not the clubs had a pool, racquetball courts, or group fitness classes. If management can determine with 90% confidence that a pool, racquetball courts, or group fitness classes produces significantly more memberships than sports clubs without those features, they will include them in the new location.

Sports Club Membership				
Number of Members	Median Income (\$)	Pool?	Racquetball Courts?	Fitness Classes?
1258	32,223	No	No	No
1479	34,975	No	No	No
1480	43,187	No	Yes	No
1701	44,337	No	No	No
2014	52,167	No	No	Yes
2271	57,521	No	No	Yes
2615	58,347	No	Yes	No
2632	60,960	Yes	No	No
2737	62,201	Yes	No	Yes
2810	67,993	No	No	Yes
3563	68,770	No	No	Yes
3765	81,289	Yes	Yes	Yes
3792	83,902	No	No	Yes
4069	84,594	Yes	No	Yes
4393	86,855	Yes	Yes	Yes
4787	88,381	Yes	Yes	Yes

- a. What sign do you expect the coefficient of median income to have? Explain why.
- b. Create three dummy variables, pool, courts, and classes, that are equal to 1 if the observation contains this feature and equal to 0 if the observation does not contain this feature.
- c. Use statistical software to estimate the following regression models. In each case, write the estimated regression equation and state whether the coefficient of the independent variable is significant at the 0.10 level.
 - i. $Members = \beta_0 + \beta_1 (Pool) + \varepsilon_i$
 - ii. $Members = \beta_0 + \beta_1 (Courts) + \varepsilon_i$
 - iii. $Members = \beta_0 + \beta_1 (Classes) + \varepsilon_i$
- d. Estimate the following multiple regression model.
 $Members = \beta_0 + \beta_1 (Income) + \beta_2 (Pool) + \beta_3 (Courts) + \beta_4 (Classes) + \varepsilon$
 Write the estimated regression equation.
- e. Are any of the coefficients of the indicator variables significant at the 0.10 level?
- f. Explain why it is important to include the income variable in the regression model.
- g. After studying these regression results, how would you suggest the management of the sports club chain go about building their new location? Should they use any of the regression models you have estimated? Explain why or why not.

3. The amount of a certain additive injected into a chemical process has a direct effect on the yield. The following table contains data on the amount of additive and yield.

Amount of Additive and Yield											
Additive	12.0	6.7	5.6	13.2	8.9	7.8	12.9	16.4	4.5	9.6	5.8
Yield	96	50	42	82	76	70	89	94	15	75	32

- Assuming that yield is the dependent variable, plot yield against additive. Does the relationship appear to be linear?
 - Using statistical software, estimate the simple linear regression model. Identify R^2 and s_e^2 .
 - In instances such as this where linearity does not hold, polynomial regression can be used to provide a better fit to the data. Polynomial regression is a special case of multiple regression where new predictor variables are formed by raising other predictor variables to integral powers. In this exercise, a new predictor will be formed by squaring the values of additive (Add_sq). Yield will then be fitted to the predictors Additive and Add_sq. The prediction equation based upon the polynomial regression is Estimated Yield = $-67.53 + 23.04(\text{Additive}) - 0.82(\text{Add_sq})$. R^2 and s_e^2 are 0.95 and 47.53, respectively. Predict the yield when Additive = 16. Make this prediction using both the linear and polynomial fits. Compare your results.
 - Compare the linear and polynomial fits to the data by the values for R^2 and s_e^2 .
 - Which model do you believe is best to use for estimation and prediction? Explain your answer.
4. Suppose that an association of real estate professionals has reported home sales for 2011 in a data set titled Home Sales. The table contains the current sales by region and the inventory for existing-home sales (single-family and condos/co-ops). An excerpt of the full table is given below.

Home Sales							
Sale Price	Region	Home Type	Inventory	Sale Price	Region	Home Type	Inventory
\$237,000	NE	Condo/Co-op	185,000	\$239,600	NE	Single-Family	550,000
\$225,400	NE	Condo/Co-op	188,000	\$242,400	NE	Single-Family	550,000
\$235,200	NE	Condo/Co-op	205,000	\$244,600	NE	Single-Family	560,000
\$144,900	MW	Condo/Co-op	80,000	\$138,800	MW	Single-Family	850,000
\$145,000	MW	Condo/Co-op	79,000	\$138,900	MW	Single-Family	840,000
\$139,200	MW	Condo/Co-op	82,000	\$138,600	MW	Single-Family	900,000
\$110,400	S	Condo/Co-op	194,000	\$153,400	S	Single-Family	1,520,000
\$108,100	S	Condo/Co-op	176,000	\$153,100	S	Single-Family	1,520,000
\$112,100	S	Condo/Co-op	200,000	\$150,800	S	Single-Family	1,570,000
\$154,600	W	Condo/Co-op	90,000	\$223,100	W	Single-Family	940,000
\$152,900	W	Condo/Co-op	91,000	\$216,300	W	Single-Family	940,000
\$146,800	W	Condo/Co-op	74,000	\$216,900	W	Single-Family	1,050,000

Data

The full table can be found at stat.hawkeslearning.com by navigating to **Discovering Business Statistics, Second Edition > Data Sets > Home Sales**.

- Suggest a regression model that would allow you to predict sale price as a function of inventory, region, and whether you have a condo/co-op or a single-family home.
- Estimate the model that you suggested in part a.
- What is the estimated equation for predicting sale price by region?
- What is the estimated equation for predicting sale price by type of home?

Data

This data set can be found at stat.hawkeslearning.com by navigating to **Discovering Business Statistics, Second Edition > Data Sets > Tablet Survey**.

e. Is your model estimated in part b. statistically useful for predicting sale price at a 1% significance level? Explain your answer.

5. Given the digital revolution in the United States and the trend towards textbooks and lecture materials being made available through electronic means, one wonders if having these technologies will help improve students' grades. To help answer this question, a survey was taken on a university campus inquiring if students used a tablet for their classes, which tablet was used, and their grade point average, household income, and the highest level of education attained by their parents. Using the data from the survey presented in the following table, answer the following questions.

Key for Student Survey		
Tablet:	Household Income:	Parent's Highest Level of Education:
1 = Motorola Xoom	1 = < \$30,000	1 = Some High School
2 = Samsung Galaxy	2 = \$30,000–\$49,999	2 = High School Diploma
3 = Apple iPad	3 = \$50,000–\$74,999	3 = Some College
4 = No Tablet	4 = > \$75,000	4 = College Graduate

Student Data							
GPA	Tablet	Income	Education	GPA	Tablet	Income	Education
3.9622	4	3	1	3.0530	2	3	4
2.9555	1	2	3	3.8034	2	3	3
3.2058	4	2	2	2.3986	2	2	4
3.6487	1	4	2	3.1191	1	3	2
3.4459	1	4	4	2.9556	1	3	3
3.5222	4	1	1	3.3100	2	3	1
3.9964	2	1	3	3.5477	1	3	2
2.4374	1	3	4	3.7710	3	3	1
2.9262	4	3	3	3.2706	1	3	1
3.8684	2	2	3	3.8039	2	1	3
3.2102	1	3	3	3.5149	1	1	2
3.4394	2	3	4	2.0290	2	3	1
3.8169	4	1	3	3.2985	3	1	4
3.6107	2	2	1	2.2877	2	2	2
2.7475	3	2	3	2.4569	3	2	1

- Suggest a multiple regression model to predict GPA from tablet use and which type of tablet is being used.
- What is the estimated regression equation for the model proposed in part a.?
- Is the model useful in predicting GPA from whether a student uses a tablet or not and which type of tablet is being used at the 0.05 level? Justify your answer.
- Discuss whether there is a difference between tablet use and household income regarding how the variables affect one's GPA.
- Discuss whether there is a difference in the extent to which each type of tablet affects one's GPA.

6. The following table contains a list of high-dividend exchange-traded funds (ETFs). Exchange-traded funds are investment funds traded on stock exchanges, much like stocks. ETFs are traditionally index funds, but ETFs can hold assets such as stocks, commodities, or bonds, and trade at approximately the same price as the net asset value of their underlying assets over the course of the trading day. ETFs may be attractive as investments because of their low costs, tax efficiency, and stock-like features.

Data

This data set can be found at stat.hawkeslearning.com by navigating to **Discovering Business Statistics, Second Edition > Data Sets > Exchange-Traded Funds.**

Exchange-Traded Funds							
ETF	Share Price (\$)	Dividend Per Share (\$)	Dividend Yield (%)	ETF	Share Price (\$)	Dividend Per Share (\$)	Dividend Yield (%)
1	4.32	0.28	6.49	26	23.82	0.78	3.28
2	15.38	0.92	6.02	27	19.58	0.64	3.26
3	25.25	1.43	5.66	28	136.94	4.41	3.22
4	21.28	1.13	5.31	29	83.57	2.65	3.17
5	698.75	36.88	5.28	30	6.44	0.20	3.12
6	120.55	6.22	5.16	31	13.50	0.42	3.07
7	23.00	1.09	4.74	32	47.47	1.45	3.06
8	13.34	0.62	4.66	33	30.28	0.92	3.03
9	24.82	1.15	4.63	34	14.56	0.43	2.94
10	22.96	1.04	4.53	35	36.06	1.06	2.93
11	78.51	3.21	4.09	36	14.26	0.41	2.90
12	24.22	0.99	4.08	37	369.35	10.56	2.86
13	8.70	0.35	4.02	38	7.07	0.20	2.83
14	15.08	0.58	3.86	39	20.12	0.57	2.82
15	23.44	0.90	3.84	40	1.00	0.03	2.81
16	22.81	0.87	3.80	41	104.15	2.91	2.79
17	22.80	0.84	3.70	42	32.4	0.89	2.75
18	113.84	4.21	3.69	43	47.58	1.30	2.73
19	13.62	0.49	3.60	44	18.43	0.49	2.64
20	22.86	0.82	3.58	45	27.78	0.73	2.63
21	15.58	0.56	3.57	46	30.10	0.76	2.52
22	121.33	4.16	3.43	47	26.55	0.66	2.49
23	15.27	0.51	3.34	48	19.53	0.46	2.35
24	8.70	0.29	3.30	49	27.43	0.62	2.26
25	45.40	1.50	3.30	50	30.33	0.62	2.04

- Using the data in the table, can dividend yield be predicted by share price and dividend per share? Is it a useful model? Justify your answers.
- Which variable explains the greatest amount of variability in dividend yield? Explain your answer.
- Can you detect any multicollinearity in the model containing share price and dividend per share? Explain your answer.

Data

This data set can be found at stat.hawkeslearning.com by navigating to **Discovering Business Statistics, Second Edition > Data Sets > SNAP Benefits**.

7. The Supplemental Nutrition Assistance Program (SNAP) provides monthly benefits that help eligible low-income households buy the food they need for good health. For most households, SNAP funds account for only a portion of their food budgets, so they must also use their own funds to buy enough food to last throughout the month. Eligible households can receive food assistance through regular SNAP or through the Louisiana Combined Application Project (LaCAP). Using the data in the table, answer the following questions to help predict monthly benefits to eligible households.

SNAP Benefits					
Monthly Benefit (\$)	Family Size	Gross Monthly Income (\$)	Monthly Benefit (\$)	Family Size	Gross Monthly Income (\$)
603.41	5	3753	556.42	1	3098
560.69	3	3778	569.05	8	3707
623.24	6	3609	365.80	8	2071
416.12	5	2262	489.08	5	3166
323.90	1	1966	495.86	4	3126
418.78	4	2736	642.77	4	3933
506.46	2	3274	364.81	8	1925
552.53	2	3480	619.30	6	3736
586.46	7	3741	238.71	1	1453
637.18	8	3684	378.94	4	2538
244.49	2	1476	302.58	1	1798
507.19	5	2835	231.74	8	1189
512.56	5	2873	428.67	6	2247
312.89	4	1618	286.99	5	1460
329.05	4	1565	268.81	1	1567
243.49	6	1582	329.81	6	1622
560.37	8	3380	627.25	3	3828
599.90	3	3922	421.52	6	2782
657.09	5	3845	656.38	2	3978
394.82	5	2233	400.64	3	2493

- Suggest a regression model that will assist SNAP administrators in providing a monthly benefit to eligible households.
- Fit the model that you suggested in part **a**. Is this model useful in predicting monthly benefits? Justify your answer.
- Are all independent variables in the model helpful in explaining the variation in monthly benefits? Explain your answer.
- Give a 95% confidence interval for average monthly benefits for a four-member household with a gross monthly income of \$2500. Interpret this interval.
- Provide a 99% prediction interval for a four-member household with a gross monthly income of \$2500. Interpret this interval.
- What is the difference between the intervals found in parts **d**. and **e**.?