Weekly Quiz 4

The **due date** for this quiz is **Sun 17 Feb 2013 8:30 PM PST**.

Question 1

Load the movies data available from the course website here:

https://spark-public.s3.amazonaws.com/dataanalysis/movies.txt

Fit a linear regression model by least squares where the Rotten Tomatoes score is the outcome and the box office gross is the only covariate. What is the regression coefficient for the slope and it's interpretation?

○ The regression coefficient is 50.07559. The interpretation is that an increase of one dollar in box office gross is associated with a 50.07559 increase in Rotten Tomatoes Score.

● The regression coefficient is 0.09676. The interpretation is that an increase of one million dollars in box office gross is associated with a 0.09676 increase in Rotten Tomatoes Score.

○ The regression coefficient is 0.09676. The interpretation is that an increase of one dollar in box office gross is associated with a 0.09676 increase in Rotten Tomatoes Score.

○ The regression coefficient is 50.07559. The interpretation is that a movie that made zero dollars at the box office would have an average score of 50.07559.

Question 2

Load the movies data available from the course website here:

https://spark-public.s3.amazonaws.com/dataanalysis/movies.txt

Fit a linear regression model by least squares where the Rotten Tomatoes score is the outcome and the box office gross is the only covariate. What is the 90% confidence interval for the intercept term and what can you deduce from the 90% confidence interval?

○ The 90% confidence interval for the intercept is (47.52, 52.63). If we repeated this study 100 times, we would expect our calculated interval to cover the true value on average 90% of the time.

○ The 90% confidence interval for the intercept is (0.04, 0.16). If we repeated this study 100 times, we would expect our calculated interval to cover the true value on average 90% of the time.

○ The 90% confidence interval for the intercept is (-75.25, -10.61). If we repeated this study 100 times, we would expect our calculated interval to cover the true value on average 90% of the time.

○ The 90% confidence interval for the intercept is (47.03, 53.12). If we repeated this study 100 times, we would expect our calculated interval to cover the true value on average 95% of the time.

Question 3

Load the movies data available from the course website here:

https://spark-public.s3.amazonaws.com/dataanalysis/movies.txt

Fit a linear regression model by least squares where the Rotten Tomatoes score is the outcome and box office gross and running time are the covariates. What is the value for the regression coefficient for running time? How is it interpreted?

○ The coefficient is 0.12752. That means that an increase of one hour in running time is associated with an average increase of 0.12752 in score.

○
The coefficient is 0.08239. That means that if two movies have the same box office gross, increase of one minute in running time is associated with an average increase of 0.08239 in score.

◉
The coefficient is 0.12752. That means that if two movies have the same box office gross, an increase of one minute in running time is associated with an average increase of 0.12752 in score.

○
The coefficient is 0.12752. That means that an increase of one minute in running time is associated with an average increase of 0.12752 in score.

Question 4

Load the movies data available from the course website here:

https://spark-public.s3.amazonaws.com/dataanalysis/movies.txt

Fit a linear regression model by least squares where the Rotten Tomatoes score is the outcome and box office gross and running time are the covariates. Is running time a confounder for the relationship between Rotten Tomatoes score and box office gross? Why or why not?

○
Yes running time is a confounder. It is correlated with the box office gross.

○
Yes running time is a confounder. It is correlated with the Rotten Tomatoes score.

○
No running time is not a confounder. It is correlated with both box office gross and Rotten Tomatoes score.

◉
Yes running time is a confounder. It is correlated both with the Rotten Tomatoes score and the box office gross.

Question 5

Load the movies data available from the course website here:

https://spark-public.s3.amazonaws.com/dataanalysis/movies.txt

Fit a linear regression model by least squares where the Rotten Tomatoes
score is the outcome and box office gross and running time are the
covariates. Make a plot of the movie running times versus movie score. Do
you see any outliers? If you do, remove those data points and refit the same
regression (Rotten Tomatoes score is the outcome and box office gross and
running time are the covariates). What do you observe?

○ Yes there are two outliers. After removing them and refitting the
regression line, the running time coefficient has a larger magnitude and is
more statistically significant.

○ Yes there are two outliers. After removing them and refitting the
regression line, the running time coefficient does not change.

○ Yes there are two outliers. After removing them and refitting the
regression line, there is no difference in the box office coefficient.

○ No there are no outliers.

Question 6

Load the movies data available from the course website here:

https://spark-public.s3.amazonaws.com/dataanalysis/movies.txt

Fit a linear regression model by least squares where the Rotten Tomatoes
score is the outcome and running time and box office gross are covariates.
What is the P-value for running time and how is it interpreted?

○

The P-value is 2.379. It is the probability there is no relationship between Rotten Tomatoes score and running time for a fixed box office gross.

◉

The P-value is 0.0187. It is the probability of observing a t-statistic as big as, or larger than, the one we saw, if there was no relationship between Rotten Tomatoes score and running time for a fixed box office gross.

○

The P-value is 0.0187. It is the probability there is a relationship between Rotten Tomatoes score and running time for a fixed box office gross.

○

The P-value is 0.0187. It is the probability there is no relationship between Rotten Tomatoes score and running time for a fixed box office gross.

Question 7

Load the movies data available from the course website here:

https://spark-public.s3.amazonaws.com/dataanalysis/movies.txt

Fit a linear model by least squares where Rotten Tomatoes score is the outcome and the covariates are movie rating, running time, and an interaction between running time and rating are the covariates. What is the coefficient for the interaction between running time and the indicator/dummy variable for PG rating?

○ The coefficient is -0.9387

○ The coefficient is 38.2850

◉ The coefficient is -0.6901.

○ The coefficient is 58.9816

Question 8

Load the movies data available from the course website here:

https://spark-public.s3.amazonaws.com/dataanalysis/movies.txt

Fit a linear model by least squares where Rotten Tomatoes score is the outcome and the covariates are movie rating, running time, and an interaction between running time and rating are the covariates. What is the estimated average change in score for a PG movie for a one minute increase in running time?

- ◯ -0.6901
- ◯ 1.8753
- ◉ 0.4951
- ◯ 1.1852

Question 9

Load the data on number of breaks during weaving into R with the command:

```
data(warpbreaks)
```

Fit a linear model where the outcome is the number of breaks and the covariate is tension. What is a 95% confidence interval for the average difference in number of breaks between medium and high tension?

- ◯ (-10.00, -14.72)
- ◉ (-3.23, 12.67)
- ◯ (6.77, 22.67)
- ◯ (-17.95, -2.04)
- ◯ (16.04, 27.29)

Question 10

Based on this graph is there a significant association between organic food sales and autism rates? Would you believe this association could be used to reduce autism rates? Why or why not?

The real cause of increasing autism prevalence?
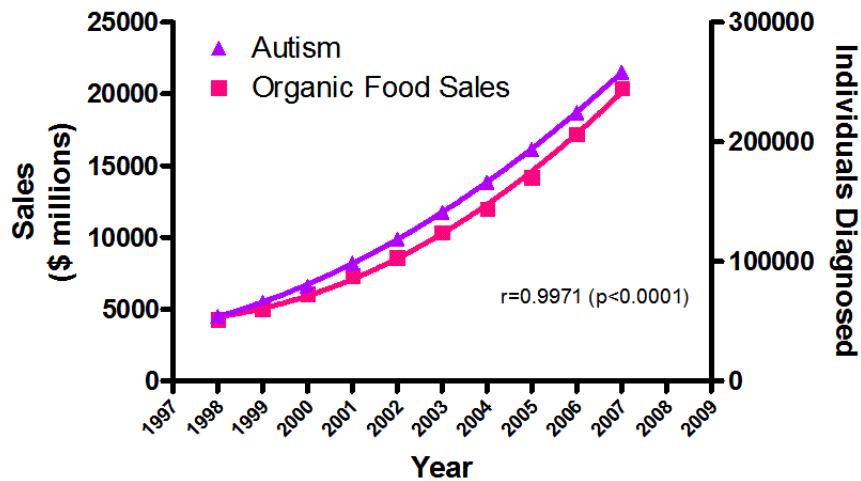
Sources: Organic Trade Association, 2011 Organic Industry Survey; U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS), OMB# 1820-0043: "Children with Disabilities Receiving Special Education Under Part B of the Individuals with Disabilities Education Act

○ There is a statistically significant association. The association may be useful for reducing autism rates by reducing organic food consumption.

◉ There is a statistically significant association. We may be skeptical this association could be used to reduce autism rates, since there are many possible explanations for the association that do not involve a direct relationship between organic foods and autism.

○ There is not a statistically significant association. We may be skeptical this association could be used to reduce autism rates, since there are many possible explanations for the association that do not involve a direct relationship between organic foods and autism.

○ There is not a statistically significant association. The association may be useful for reducing autism rates by reducing organic food consumption.

☑ In accordance with the Honor Code, I certify that my answers here are my own work.

[Submit Answers]   [Save Answers]