BTRY 6020: Module 1

Your Name

Question 1 (3 pts)

We discussed the conceptual framework of a population process of interest, the data that is generated from that process, and a statistic which is calculated from the data. Briefly describe what that might look like in your field of interest. (i.e., 1 to 3 sentences total). Describe a population/process that you are interested in learning about, what type of data might be gathered from that system, and what statistic you might calculate to summarize your data.

Answer

Question 2 (2 pts)

In Lecture 1, we discussed various tradeoffs in statistical analysis: bias vs variance, model complexity vs interpretation, false positives vs false negatives. Briefly discuss (i.e., 1 to 3 sentences total) a situation in your field of interest where one of those trade-offs might occur.

Answer

Question 3 (1 pt)

Go to the website https://guessthecorrelation.com/, and guess the correlation for at least 10 plots. Record your guess and the true correlation for the last plot below.

Answer

Guess:

Truth:

Question 4 (3 pt)

The following data is on the spring snow depth on Mt. Rainier from 1920 - 2013.

- year: year of snow depth measurement (1920-2013)
- snow: maximum snow depth in cm measured in April at 1500m elevation at Paradise Ranger Station on Mt. Rainier
- temp: mean temperature the previous winter (Nov. - Apr. mean, in C)
- prec: mean precipitation the previous winter (liquid water equivalent, in cm)
- nino: Index of El Nino Southern Oscillation (ENSO), (sea surface temperature anomaly (C) from the long term seasonal mean observed in the Nino 3.4 region of the Central Pacific (basically the middle of the Pacific on the equator)

Fit a linear model with **maximum snow depth** as the dependent variable and the El Nino index as the independent variable. What are the estimated intercept and slope?

```
snow_data <-
  read.csv("https://raw.githubusercontent.com/ysamwang/btry6020_sp22/main/lectureData/snow_data.csv")
```

```
names(snow_data)
```

```
## [1] "year" "prec" "temp" "snow" "nino"
```

Answer

Question 5 (2 pt)

Give an interpretation for the estimated slope parameter which you could explain to a colleague.

Answer

Question 6 (2 pt)

The latest El Nino measurement (Dec 2023) was 1.99. What is the predicted maximum snow depth in cm?

Answer

Question 7 (2 pt)

Does the linearity assumption seem to hold for the data? Explain why or why not and include a plot in your explanation
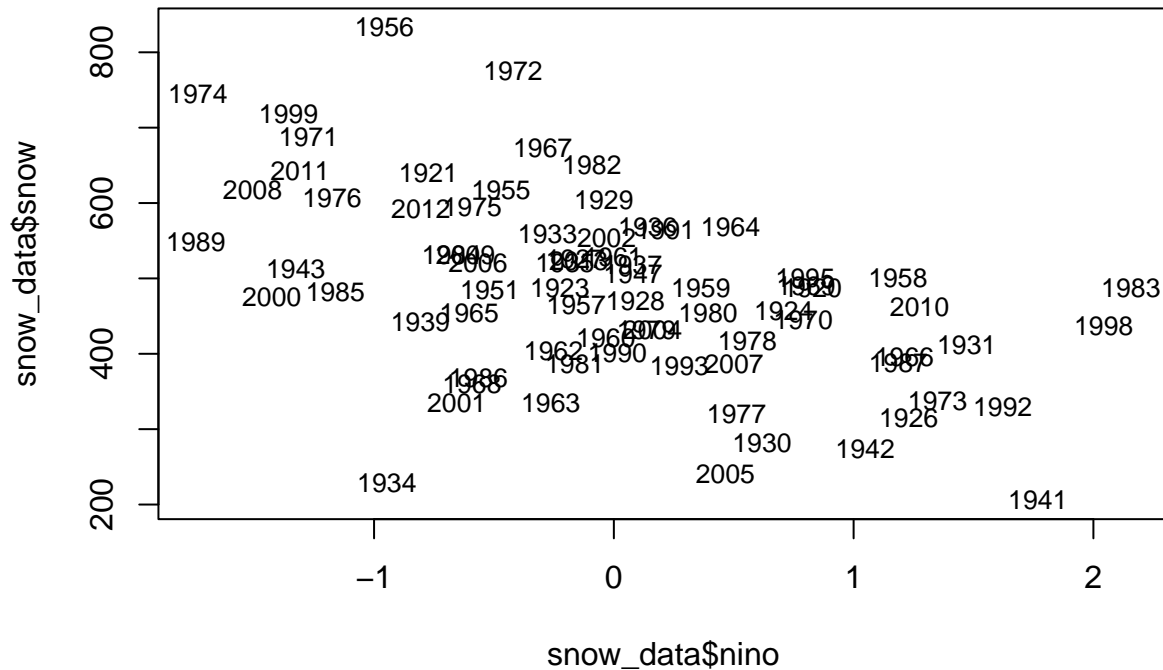
Answer

Question 8 (2 pt)

Does it seem like the variance of errors are constant across the range of the covariate? Explain why or why not and include a plot in your explanation

Answer

Question 9 (3 pt)

The following plot shows the data for snow and el nino index. Identify a point which has large leverage and may also have large influence. Fit a regression for the data with the identified year removed and compare the estimates to the regression you previously fit which included all data.

Answer

Question 10 (3 pt)

In the lecture, we considered data which showed the price of a bottle of wine as well as the rating of the bottle of wine. A plot is shown in the last slide of Lecture 1. There are two points in the upper right hand corner of the plot that might be considered an outlier. Describe a question you might investigate with this data. For that question of interest, explain whether you would choose to include or exclude those points from your analysis.

Answer