Problem-set_ Descriptive Statistics

Pablo E. Gutiérrez-Fonseca

2023-08-13

Using the Problem Set Snow Data down-loadable from this blackboard folder, answer the following questions. Be sure to round and report ALL answers to 2 decimal places and DO NOT ENTER TEXT into the BB submission for numerical questions. You will also need to include the sign for negative values. Good luck!

For the Yearly Mean Snow Depth data, **using excel calculate**:

1. the **MEAN Yearly Mean Snow Depth**.
2. the **MEDIAN Yearly Mean Snow Depth**.
3. If you round **Yearly Mean Snow Depth** to the nearest integer (whole number with no decimals), **what is the MODE**? (note that this will require first making a new column with the rounded value, use the round function in EXCEL and then venture into the world of pivot tables)
4. Still in Excel, now calculate the **STANDARD DEVIATION** for the Yearly Mean Snow Depth.

You can do the following in excel, **but it might be easier to import your data into R**. Blackboard should be agreeable. But I'd suggest moving to R for the rest of these questions to save yourself time and get used to the software.

5. What is the **INTER-QUARTILE RANGE** for **Yearly Mean snow Depth**?
6. Using the Interquartile range technique, how many **OUTLIER** years are there in your **Yearly Mean Snow Depth Data**?
7. List any **outlier years** here. (just enter the number for the years, e.g. 2003)
8. Now calculate **Pearson's skew for the Yearly Mean Snow Depth Data**. Enter your answer rounded to 2 decimal places. (do this in R; you've already proved your excel skills above)
9. Calculate the **standard error of skew (ses)** for this data so we can determine the significance of our skew. (you have to do this by hand)
10. Based on the ses technique, is this **skew significant**?
11. Now determine the **kurtosis for the Yearly Mean Snow Depth data**.
12. Calculate the **standard error of kurtosis (sek)** for this data so we can determine the significance of our skew. (you have to do this by hand)
13. Based on the sek technique, is this **kurtosis significant**?

14. Based on your statistical summary values, do you think that your data is normally distributed? (just practice the manual approach to assessing normality). What do you think without running an actual goodness of fit test?)

15. Is your data normally distributed? How can you tell? Be sure to report your certainty of this conclusion.

16. Based on these results, which measure of central tendency would be best describe central tendency for this dataset?

17. Now use the **Boardstrength data** down-loadable from this blackboard folder. This datafile contains measurements of board strength for several different types of wood. Your task is to describe this data (**Strength variable**) as though you were presenting it in a paper, poster or presentation.

In other words, We don't want a bunch of words listing off every possible descriptive metric.

You could convey all of the KEY information about board strength across all species in a simple table, with a succinct caption to capture things like normality, outliers, etc. The key is give us enough information to figure out what type of analysis you would have to run to compare them.

A screenshot of the **FULL R output is NOT OK**.

UPLOAD a screenshot of your data description table and caption for the Strength variable.

18. If you wanted to highlight the differences between species, and describe strength for each of them, how would your table look different? Consider that the actual statistical analysis to test for differences among species is run on the full sample of 20 observations, so metrics like kurtosis, skew and goodness of fit tests would be calculated on the full data set.

But we can still describe how the observations for each species differ.
Feel free to use a table to describe how board strength differs across species, or a figure.

Just be sure we have information on both central tendency and variance for each species.

UPLOAD a screenshot of your data description of the strength variable by species.