

AP STATISTICS

Investigative Task #1 – Is the data Normal?

Due Date – 5 February 2010 (Friday)

A Normal model can be a useful tool for interpreting what data have to say – sometimes. Your task is to check the usefulness of such a model for data you collect or create. This task has three parts.

1. Collect the data
2. Describe the data
3. Check the Normal model

Collect the data

You need a data set with 30 – 50 values. Find something you are interested in. Use existing data or better yet, create some data yourself. Here are a few suggestions:

- Put 10 pennies in a glass, put your hand over the top, shake well, then dump them out on a table and count the number that came up heads.
- Roll two dice and record the total.
- Deal cards from a well-shuffled deck one at a time. Count the number of cards you have to turn over until you find an ace.
- Use some data from another class – a science experiment, a Biology lab, etc.
- Look something up in an almanac. For example, there are lots of tables of data about states – crime rates, population density, median income, etc.
- Use sports statistics – number of wins for baseball teams, scores in a golf tournament, weights of players on a football team, etc.
- Find something on the internet – see www.census.gov as an example.

You must attach the raw data to your report. **You may not share data.** All data that you do not create should be appropriately referenced – meaning I should be able to go to that website or to that book and find the data that you used. Do not say “I got the data from Google” even if you do use Google to initially find a website.

Describe the data

The first paragraph of your paper should describe the four W's of your data (see p. 12 of your textbook). Give a numerical summary of your data (five number summary, mean, standard deviation). Give a graphical summary of your data (histogram, stemplot, or boxplot). Give a verbal summary of your data (shape, center, spread, unusual features).

Check the Normal model

Draw a picture of the Normal model for the mean and standard deviation of your data. Do this even if your data is not Normal. Compare your data to the Normal model using the 68–95–99.7 rule. Present this comparison in an easy to understand table. Compare your data to the Normal model using a Normal probability plot. Your determination of whether or not your data is Normal should be very similar to problem 2.39 or 2.40 in your textbook. See pages 148–155 for a discussion of how to determine if a set of data can be approximated by a Normal model.

Keep in mind that you must make a decision. The concluding paragraph of your write-up must clearly answer the following question: **Is a Normal model an appropriate approximation to my data?** Your answer must also clearly indicate your reasoning. Your data does not need to be approximately Normal.

Requirements and suggestions

The report must be typed (10-12 point, professional font, 1-inch margins, single-spaced). The summary tables (five-number summary, compare your data to 68–95–99.7 rule, raw data) should also be typed.

The graphs can be hand drawn but must be drawn to scale. The graphs can also be generated using a computer. Either way, the graphs should be professional-looking. Consider using the website statcrunch.com to generate your graphs. If you want to use MS Excel to create histograms and similar graphs, search on the web for the “Analysis add-in” in Excel to see how to access some tools that are built-in to Excel for statistics.

The graphs can either be attached to the report or incorporated into the body of the report. Using either method, each graph and table should include a caption “Figure 1. Stemplot of Hillgrove GPAs.” and “Table 1. Summary statistics of Hillgrove GPAs.” Your text should refer to the graph as “In Figure 1, a histogram of GPAs show that it is clearly skewed to the left.” Use your textbook as a guide. As always, your figures should clearly label the axes and identify the subject of the graph.

Use correct, professional English. Proofread your paper.

The document should should contain about 1 page of single-spaced text excluding all tables and figures. You should have a minimum of 3 graphs (data summary, Normal model, normality plot) and 3 tables (raw data, summary statistics, comparison between 68–95–99.7 rule and your data).