

# Lab practice 1

## Instructions

- Place your answers in a Word document.
- Make sure to include graphs where needed. Adjust them so they are readable. Clean them up where needed.
- Where tables are called for, don't just use the tables from SPSS. Make nice-looking tables using Word's abilities and only using the values from SPSS that you need to include.
- You are allowed to work with others.
- This will count as a homework.

## Questions

1. Data containing information about the various counties can be downloaded at: <https://skiadas.github.io/AppliedStatsCourse/site/datasets/countyComplete.txt>. You can also find them by navigating to the Datasets folder in the course website. Use the Read-Text-Data mechanism to import the data into SPSS. Each row in the data corresponds to a county.
  - How many counties do we have?
  - Which counties are the top 3 by population (2010)? Does it make sense that those counties have so much population?
  - Which counties are the bottom 3 by population?
2. Which three states have the most number of counties? Which three have the least number of counties? Describe how you found your answer.
3. Draw a suitable histogram of the distribution of population by county. You will need to possibly zoom in a region and leave out some outliers, to offer a better view. Describe the overall distribution.
4. Compute the median population for each state. List the top three and bottom three states.
5. Draw a suitable histogram of the percent of females in each county (the variable is called female). Describe the distribution.
6. Describe (including suitable graph and/or equations) the relationship between the population in 2010 and the population in 2000. Is there a general pattern? Does it make sense?
7. Compute (by creating a new variable) the **relative** population change on each county from 2000 to 2010: This should be the change (difference) in population divided by the 2000 population.
  - Draw a suitable histogram and describe the resulting pattern and reasons for it. Also compare to the overall US growth rate from 2000 to 2010 of 9.7%.
  - Locate some of the outliers (on both ends) and discuss their particular circumstances, where possible, and the reasons for these extreme changes from 2000 to 2010.

- Which two states in Indiana had the highest relative population change? Make sure to identify them geographically.
8. Draw a suitable graph and describe the relationship between a county's poverty rate and the percent of people in the county with high school degrees. Include a regression line if suitable.
  9. Describe the distribution of the area of different counties. You may need to restrict the histogram to see the main pattern.
    - Describe the overall pattern, including a discussion of what the typical values for the area of a county are.
    - Describe deviations from that pattern, and their reasons.
    - Which states have a high median county area?