Problem set 1

## Chapter 1 problems

### Problem 1: Hypothermia and stroke

In animal stroke models, neuronal damage during stroke is profoundly affected by body temperature. In humans, hypothermia reportedly improves outcomes in traumatic brain injury, and has been shown to be neuroprotective during surgery. Some work has been done on anesthesia in conjunction with hypothermia for stroke patients, and lower body temperature at admission is associated with improved outcomes among stroke patients. Researchers are interested in whether achieving modest hypothermia without anesthesia is feasible, and if so whether it improves patient outcomes. One of the outcomes measured was mortality at 28 days.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mortality within 28 days |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yes | No | Total |
| Treatment | 1 | 15 | 16 |
| Control | 6 | 44 | 50 |
|  | 7 | 59 | 66 |

1. What percent of patients in the treatment group died within 28 days?
2. What percent died within 28 days in the control group?
3. In which group did a higher percentage of patients die within 28 days?
4. Your findings so far might suggest a real difference in the effect of modest hypothermia on mortality outcomes in stroke patients. However, this is not the only possible conclusion that can be drawn based on your findings so far. What is one other possible explanation for the observed difference between the percentage of patients in the hypothermia and regular care groups that died within 28 days?

### Problem 2: Association

1. Would you expect ice cream sales and shark attack rate to be associated or independent? Why or why not?
2. If associated, do you think that a causal relationship exists?

### Problem 3: Socio-economic class and unethical behavior

In one of seven studies on the relationship between socio-economic class and unethical behavior, researchers studied the relationship between socio-economic class and unethical driving behavior among 274 California drivers. Drivers’ vehicle status (make, age, and appearance) was used as a proxy for drivers’ social class. Observers noted each vehicle status, and also recorded whether or not each driver cut off other vehicles at a busy 4-way stop by crossing the intersection before it was their turn.

1. Identify the main research question of the study.
2. Who are the subjects in this study, and how many are included?
3. The study found that students who were identified as upper-class cut off more vehicles than others. What variables were recorded for each subject in the study in order to conclude these findings? State the variables and their types.
4. What potential problems exist with the assumption that vehicle can be used as a proxy for drivers’ social class?

### Problem 4: Socio-economic class and unethical behavior (part two)

The previous problem introduced a study on the relationship between socio-economic class and unethical behavior.

1. Identify the population of interest and the sample in this study.
2. Comment on whether or not the results of the study can be generalized to the population, and if the findings of the study can be used to establish causal relationships.

### Problem 5: School is a pain in the neck

A study surveying a random sample of college students found that they are more likely to have neck pain, drink more coffee, and stay awake later when they are stressed.

1. What type of study is this?
2. Can this study be used to conclude a causal relationship between increased stress and neck pain?
3. State possible confounding variables that might explain the observed relationship between increased stress and neck pain.

### Problem 6: Vitamin supplements.

[Note: This is problem 1.30 from the OIS book.] To assess the effectiveness of taking large doses of vitamin C in reducing the duration of the common cold, researchers recruited 400 healthy volunteers from staff and students at a university. A quarter of the patients were assigned a placebo, and the rest were evenly divided between 1g Vitamin C, 3g Vitamin C, or 3g Vitamin C plus additives to be taken at onset of a cold for the following two days. All tablets had identical appearance and packaging. The nurses who handed the prescribed pills to the patients knew which patient received which treatment, but the researchers assessing the patients when they were sick did not. No significant differences were observed in any measure of cold duration or severity between the four groups, and the placebo group had the shortest duration of symptoms.

1. Was this an experiment or an observational study? Why?
2. What are the explanatory and response variables in this study?
3. Were the patients blinded to their treatment?
4. Was this study double-blind?
5. Participants are ultimately able to choose whether or not to use the pills prescribed to them. We might expect that not all of them will adhere and take their pills. Does this introduce a confounding variable to the study? Explain your reasoning.

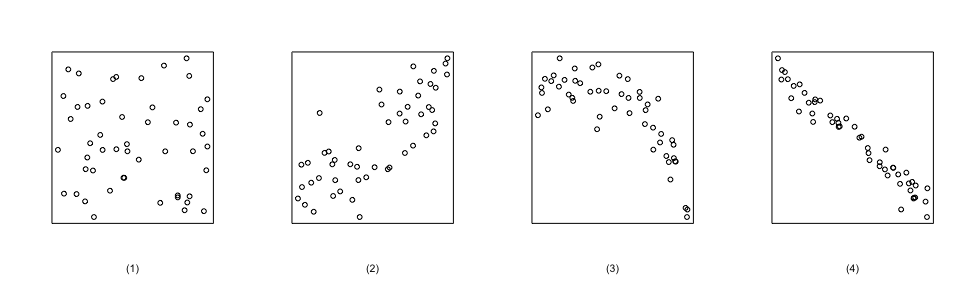
### Problem 7: Aromatherapy and learning

You have been tasked with conducting an experiment among your classmates to see if students learn better if they study without any scents, with essential oils applied just before studying, or with an oil diffuser turned on during studying. Briefly outline a design for this study.

## Chapter 2 problems

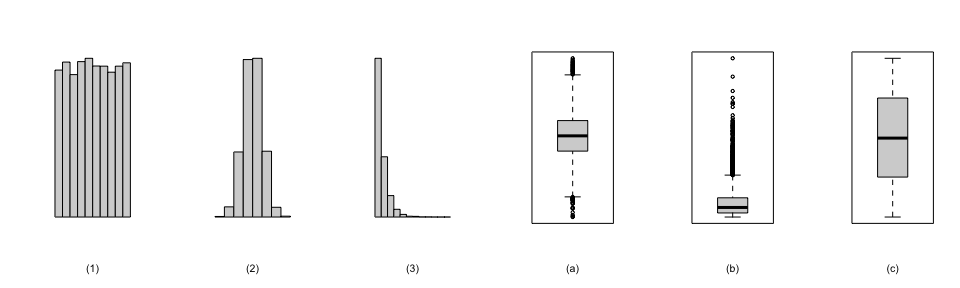
### Problem 8: Associations

Indicate which of the plots show (a) a positive association, (b) a negative association, or (c) no association. Also determine if the positive and negative associations are linear or nonlinear. Each part may refer to more than one plot.



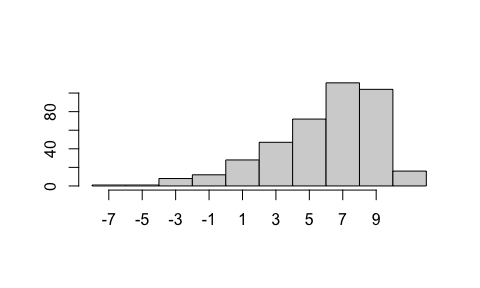
### Problem 9: Mix-and-match

Describe the distribution in the histograms below and match them to the box plots.



### Problem 10: Median vs. mean.

Estimate the median for the 500 observations shown in the histogram, and note whether (and why) you expect the mean to be higher or lower than the median.



### Problem 11:

An informal study of cats showed that 50% of cats nap 10 or more times per day, and that the average daily nap count of cats is 17. What does this study imply about the shape of the distribution of number of cat naps per day?

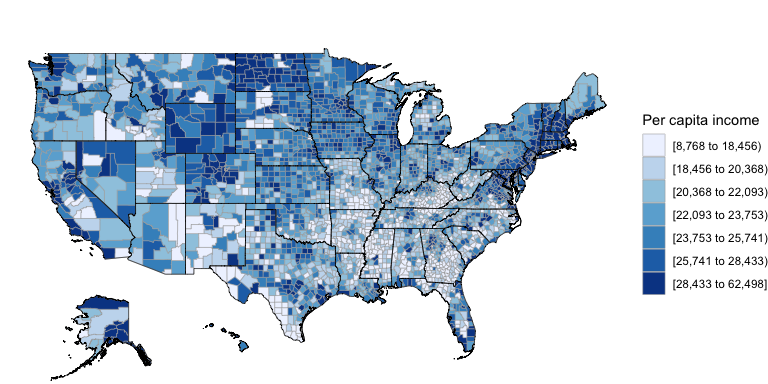
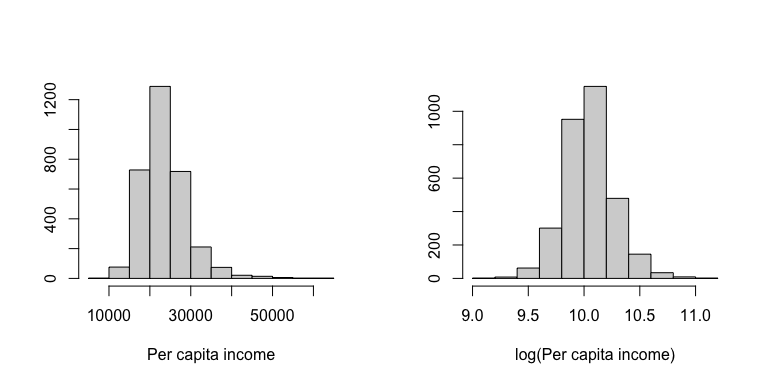
### Problem 12: Distributions and appropriate statistics

For each of the following, state whether you expect the distribution to be symmetric, right skewed, or left skewed. Also specify whether the mean or median would best represent a typical observation in the data, and whether the variability of observations would be best represented using the standard deviation or IQR. Explain your reasoning.

1. Prom gown prices in a store where 25% of the gowns cost below $100, 50% of the gowns cost below $200, 75% of the gowns cost below $300, and very few gowns cost more than $400.
2. Bridal gown prices in a (very fancy) store where 25% of the gowns cost below $2,000, 50% of the gowns cost below $4,000, 75% of the gowns cost below $12,000, and there are a meaningful number of gowns that cost more than $30,000.
3. Ages of medical students, where the majority of students are of typical mid- to late-20s college age, some are in their 30s, and only a very few are returning to medicine as a second career and are in their mid 50s.
4. Number of alcoholic drinks consumed by college students in a given week. Assume that most of these students don’t drink since they are under 21 years old, and only a few drink excessively.

### Problem 13:

The American Community Survey (ACS) records data on the per capita income of Americans, among many other variables. The histogram below shows the distribution of the per capita income of the population in 3,142 counties in the US in 2013. Also shown is a histogram of logs of these values.



1. Describe the numerical distribution and comment on why we might want to use log-transformed values in analyzing or modeling these data.
2. What features of the distribution of income in US counties are apparent in the map but not in the histogram? What features are apparent in the histogram but not the map?
3. Is one visualization more appropriate or helpful than the other? Explain your reasoning.

### Problem 14:

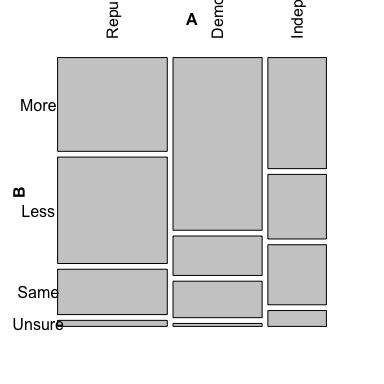
In [SurveyUSA, News Poll #18927](http://www.surveyusa.com/client/PollReport.aspx?g=60d6fa81-2698-4c51-a5f8-714f40976df2) 977 randomly sampled registered voters from Tampa, FL were asked if negative campaign ads (i) made them not want to vote, (ii) made them want to vote even more, or (iii) didn’t affect them. The results of the survey by political ideology are shown below. Note data were collected on Jan 27-29, 2012.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Republican | Democrat | Independent | All |
| (i) Not want to vote | 119 | 132 | 81 | 332 |
| (ii) Want to vote more | 93 | 68 | 40 | 201 |
| (iii) No affect | 204 | 133 | 101 | 438 |
| (iv) Not sure | 0 | 5 | 1 | 6 |
| Total | 416 | 338 | 223 | 977 |

1. What percent of these Tampa, FL voters identify themselves as Republicans?
2. Negative campaign ads make what percent of these Tampa, FL voters not want to vote?
3. What percent of these Tampa, FL voters identify themselves as Republicans and also feel discouraged from voting by negative campaign ads?
4. What percent of these Tampa, FL voters who identify themselves as Republicans are discouraged from voting by negavite campaign ads? What percent of Democrats feel this way? What percent of Independents feel this way?
5. Do political party and voting discouragement by negative campaign ads appear to be independent? Explain your reasoning.

### Problem 15:

In the same SurveyUSA News Poll the voters from Tampa, FL were asked if the government should be doing more, less, or about the same to help the housing market recover. Based on the mosaic plot shown below, do views on housing market recovery and political affiliation appear to be independent? Explain your reasoning.



### Problem 16: OIS Exercise

Complete problem 2.26 on Page 76 of the OIS book.

### Problem 17:

Alspaugh residents at Duke were asked how many hours of video games they play per week. This sample yielded an average of 3.92 hours, with a standard deviation of 3.29 hours. Is the distribution of the number of hours Alspaugh residents game weekly symmetric? If not, what shape would you expect this distribution to have? Explain your reasoning.

### Problem 18:

Suppose you have observations of some quantity. Let be the mean of these observations, and let denote the median. Suppose we have some distribution where all observations are negative, i.e. . What is the expected shape of the distribution under the following conditions?

Kammersgaard, L. P., et al. “Feasibility and safety of inducing modest hypothermia in awake patients with acute stroke through surface cooling: a case-control study: the Copenhagen Stroke Study.” Stroke 31.9 (2000): 2251-2256.

Piff, Paul K., et al. “Higher social class predicts increased unethical behavior.” Proceedings of the National Academy of Sciences 109.11 (2012): 4086-4091.

Audera, Carmen, et al. “Mega‐dose vitamin C in treatment of the common cold: a randomised controlled trial.” Medical journal of Australia 175.7 (2001): 359-362.