Chapter 1 Homework -

1.8 Smoking habits of UK residents –

a. What does each row of the data matrix represent?

Each row represent a case.

b. How many participants were included in the survey?

There were 1691 participants.

c. Indicate whether each variable in the study is numerical or categorical. If numerical, identify as continuous or discrete. If categorical, indicate if the variable is ordinal.

1. Sex – categorical – nominal

2. Age – numerical – discrete

3. Marital Status – categorical –nominal

4. Gross Income – categorical – ordinal

5. Smoke – categorical – nominal

6. amtWeekends – numerical – discrete

7. amtWeekdays – numerical – discrete

1.10 Cheaters, scope and inference –

a. Identify the population of interest and the sample in this study.

The population of interest in this study were children between the ages of 5 and 15.

b. 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.

Since the sample size is relatively small (160 - 5 to 15 year old children), a bigger sample size as well as more repetitions (by other researchers) are needed. In addition, other sound experimental design (controlling, blocking and randomization) has to be followed. Only then can the results of the study be generalized, with confidence, to the population (world population of 5 to 15 year old children).

Causal relationships between variables can be established if we are able to reject the null hypothesis. In this case, the null hypothesis is that giving instructions to 5 to 15 year old children not to cheat outweigh their desire to receive a reward. If it can be proven that the differences observed in the cheating rates between the experimental and control groups are not just by chance, then a causal relationship between the variables can be established. Simulation experiments can be performed to aid in making this conclusion.

1.28 Reading the Paper – Based on this study, can we conclude that smoking causes dementia later in life? Explain your reasoning.

a. Smokers Found More Prone to Dementia.

Since these was an observational study and not experimental, we cannot conclude that smoking causes dementia. At the most, we can only say that there is a correlation between smoking and dementia.

b. The School Bully is Sleepy.

Again, this is an observational study and not an experimental one. As such, we cannot conclude that sleep disorders causes bullying. At the most, we can only say that there is a correlation between sleep disorder and bullying among school children. It may be possible that sleep disorder and bullying for the child may be cause by a confounding variable such as abusive parents or an unstable home environment – but again needs to be proven through experimental methods.

1.36 Exercise and mental health.

a. What type of study is this?

Experimental study.

b. What are treatment and control groups in this study?

Treatment – those who exercise twice a week

Control – those who don’t exercise

c. Does the study make use of blocking? If so, what is the blocking variable?

Yes. The participants were grouped by age.

d. Dose the study make use of blinding?

No. the participants know to what group they belong.

e. Can the results be used to establish a causal relationship between exercise and mental health? Can the conclusions be generalized to the population at large?

Yes, since this is an experimental study with a control and treatment groups. The conclusions of these study can be generalized to the general population only if the sample truly is a representation of the general population without any selection bias. For example, factors such as gender, ethnicity, economic status, etc. may need to be taken into consideration.

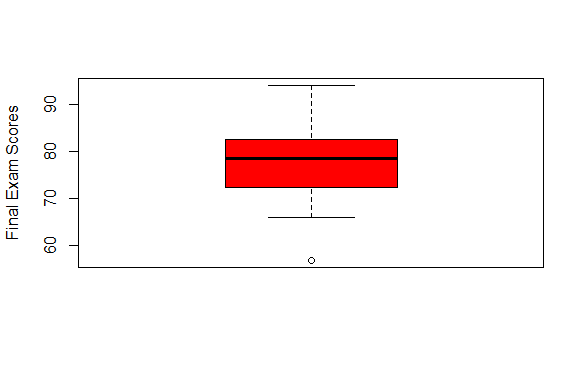
f. Would you have any reservations about funding the study proposal?

This is a useful experiment that would potentially be of great help in understanding of mental health and exercise. But again, to fund the study, I have to make sure that sound principles (randomization, blocking, repetition, etc.) of statistical experimentation are strictly followed. A major consideration too is the length of study.

1.48 Stat Scores

fescores <- c(57,66,69,71,72,73,74,77,78,78,79,79,81,81,82,83,83,88,89,94)

> boxplot(fescores,ylab="Final Exam Scores", col = "red")



1.50 Mix and Match

1. Unimodal, normal distribution – match to 3
2. Multimodal, evenly distributed, high variability, less sensitive to outlier data – match to 2
3. Bimodal, skewed to the right, high variability – match 1

1.56 Distribution and appropriate statistics

a) Skewed to the right. The median for typical observation and IQR for variability. Since there are a meaningful number of outliers (very expensive owned by the upper 1%), the more robust variables are better representations.

b) Symmetric. Since these have very few outliers, the mean and standard deviation would be better representation of a typical observation and variability respectively.

c) Skewed to the right. Median and IQR would be a better representation since there are few outliers who drink excessively can affect the mean and SD.

d) Skewed to the right. Median and IQR for the same reason as c).

1.70 Heart Transplants.

a) No. Survival seems to depend on the patient getting a transplant. From the mosaic plot, a considerably greater proportion of patients in treatment group survived compared to the control group.

b). It clearly shows that the heart transplant increases survival time as shown by the much wider IQR and higher 25, 50 and 75 quartile for the treatment group compared to the control group.

c.) About 90% of control group patients are dead compared to only about 67% in the treatment group. A difference of about 23%.

d.) 1. The experimental study claims that a heart transplant will increase the survival time of a patient.

2. 33 alive = (100-90 + 100-67); 157 dead;

100 size of treatment group; 100 size of control group;

Centered at 0.0

Simulated difference in proportion is 23%

3) There is about a 3% chance that the outcome was by change; this shows with a high degree of confidence that the treatment was effective (we reject the null hypothesis).