

Homework 2

Math 530/630 - Fall 2018

1. A coin is tossed three times and the sequence of heads and tails is recorded.
 - a. List the sample space.
 - b. List the elements that make up the following events:
 - A = at least two heads;
 - B = the first two tosses are heads;
 - C = the last toss is a tail.
 - c. List the elements of the following events:
 - complement of A ,
 - $A \cap B$,
 - $A \cup C$.
2. In a city, 65% of people drink coffee, 50% drink tea, and 25% drink both.
 - a. What is the probability that a person chosen at random will drink at least one of coffee or tea?
 - b. Will drink neither?
3. In this problem, we'll explore quantile-quantile (Q-Q) plots. A quantile is the proportion of cases we find below a certain value, calculated from the inverse of the cumulative distribution function (CDF) of a random variable, X . The p th quantile of X is the value q_p such that $P(X \leq q_p) = p$. In other words, p is the amount of area under the density curve of X that is to the left of q_p . So the smallest observation in X corresponds to a probability of 0 and the largest to a probability of 1. A Q-Q plot displays quantiles of one distribution against quantiles of another. What this means is that the data are ranked and sorted. A normal Q-Q plot displays quantiles of the normal distribution on the x -axis against quantiles of the empirical (i.e., the observed) distribution on the y -axis. A straight line is typically plotted through the points corresponding to the 1st and 3rd quantiles of each variable. If the empirical data is normally distributed, all the points on the normal Q-Q plot will form a perfectly straight line.
 - a. Draw a random sample of size $n = 15$ from $N(0, 1)$ and plot both the normal quantile plot and the histogram. Do the points on the quantile plot appear to fall on a straight line? Is the histogram symmetric, unimodal, and mound-shaped? Do this several times.
 - b. Repeat part(a) for of size $n = 30$, $n = 60$, and $n = 100$.
 - c. What lesson do you draw about using graphs to assess whether or not a data set follows a normal distribution?

```
library(tibble)
x <- as_data_frame(rnorm(15)) # draw random sample of size 15 from N(0,1)
library(ggplot2)
ggplot(x, aes(sample = value)) +
  stat_qq() +
  stat_qq_line()
ggplot(x, aes(x = value)) +
  geom_histogram()
```