

## Homework 2

Each part of the problems 5 points

Due on Blackboard before 11:45am on Thursday January 28.

Note: Use Sweave or Markdown to prepare the report. For problem 2 use the functions `pnorm` and `qnorm`.

1. We will be again working with the `diamonds` dataset in the library `ggplot2`.
  - (a) Use function `qqnorm` to produce quantile-quantile plots of price of the diamonds, and also of the natural logarithm of price. Interpret the plots, and state your conclusions
  - (b) Reproduce the results of Figure 3 in the vignette ‘Importance of being uncertain’ by Krzywinski and Altman for the price of the diamonds, as follows:
    - i. Assume that the diamonds in the dataset represent the entire population of the diamonds in nature. Plot the histogram of price.
    - ii. Sample 3 without replacement 3 diamonds from this population (use `sample`), and record their mean price. Repeat this step a large number (say, 100) times. Plot the histogram of the means.
    - iii. Repeat (ii) by randomly sampling 5, 10 and 20 diamonds. Interpret the results.
    - iv. Repeat steps (i)-(iii) while considering the natural logarithm of price rather than the original value of price. Interpret the results.
2. The heights of adult American females are Normally distributed with  $\mu=65\text{in}$  and  $\sigma=2.5\text{in}$ .
  - (a) Characterize the sampling distribution for samples of size 25. (That is, state the shape of the distribution, its expected value and standard error.)
  - (b) If we chose a random sample of 25 American women, what is the probability that their average height would be
    - i. more than 66in?
    - ii. less than 63in?
    - iii. between 63.5in and 66.5in?
  - (c) In samples of size 25, 95% of all possible samples would have averages between which values? (I.e., find a symmetric interval that contains the middle 95% of possible sample means.)
  - (d) If we walked out in the hallway and saw a first woman who came by, what is the probability that she would be more than 66 in tall?
  - (e) Why does your answer to (d) differ from your answer to (b) (i)?

3. Repeat question 2 for samples of size 100.
4. What is the difference between a standard deviation and a standard error? For a given population, which will be larger?
5. What is the relationship between the standard error and the size of the sample?
6. State the Central Limit Theorem and explain its importance.