Lab 2

1.1 Binomial Distribution

Questions

1. Read `Concrete_Data.csv`
   ```
   > concrete_data <- read.csv("Concrete_Data.csv")
   ```
2. Use `dim()` to determine the dimensions of the concrete data (the number of rows and columns)
   ```
   > dim(concrete_data)
   [1] 1030    9
   ```
3. Use `head()` and `tail()` to view the first few and last few rows, respectively, of the concrete data set
   ```
   > head(concrete_data)
   ...
   > tail(concrete_data)
   ...
   ```
4. Produce a Five-Number Summary of the comprehensive strength of concrete
   ```
   > summary(concrete_data$Concrete_comprehensive_strength)
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      2.33   23.71   34.45   35.82   46.13   82.60
   ```
5. Plot a histogram of the comprehensive strengths of concretes. Add an appropriate title and x- and y-axis labels. Note, comprehensive strength is measured in MPa
   ```
   > hist(concrete_data$Concrete_compressive_strength, main="Comprehensive Strengths of Concretes", xlab="Comprehensive Strength (MPa)", ylab="Number of Concretes")
   ```
6. Produce a boxplot of comprehensive strengths of concretes. Add an appropriate title and y-axis label
   ```
   > boxplot(concrete_data$Concrete_compressive_strength, horizontal = TRUE, main="Comprehensive Strengths of Concretes", xlab="Comprehensive Strength (MPa)")
   ```

Lab 3

1.1 Binomial Distribution

Questions

1. Simulate tossing a coin 1000 times. Are the results what you would expect?
   ```
   > tosses <- rbinom(n=1000, size=1, prob=0.5)
   ```

   Refer to whether `mean(tosses)`, `std(tosses)`, `var(tosses)` *coincide with the theoretical parameters of Binomial Distributions*

2. Suppose that n_1 items are to be inspected from one production line and n_2 items are to be inspected from another production line. Let $p_1 =$ The probability of a defective from line 1 and $p_2 =$ The probability of a defective from line 2. Let X be a Binomial Random Variable with parameters n_1 and p_1. Let Y be a Binomial Random Variable with parameters n_2 and p_2. A variable of interest is W, which is the total number of defective items observed in both production lines. Let $W = X + Y$. Use simulation to see how the distribution of W will behave. Useful information could be obtained by looking at the histogram of W_is generated and also considering the sample mean and the sample variance. In your simulation use the following random variables X and Y: X is Binomial with $n_1 = 7$ and $p_1 = 0.2$; and Y is Binomial with $n_2 = 8$ and $p_2 = 0.6$.
   ```
   > x = rbinom(n=1000, size=7, prob=0.2)
   > y = rbinom(n=1000, size=8, prob=0.6)
   > w = x + y
   ```

   Refer to `mean(x)`, `sd(x)`, `var(x)` *and* `mean(y)`, `sd(y)`, `var(y)` *and their relationship with* `mean(w)`, `sd(w)`, `var(w)`. *Additionally, consider referring to* `hist(x)`, `hist(y)` *and* `hist(w)` *to support your answer.*

1.3 Normal Distribution

Questions

If X is a Normally distributed random variable with μ = 20 and σ = 5, calculate the following:

1. P(X < 15)
   ```
   > pnorm(q=15, mean=20, sd=5)
   [1] 0.1586553
   ```
2. P(14 < X < 23)
   ```
   pnorm(q=23, mean=20, sd=5) - pnorm(q=14, mean=20, sd=5)
   [1] 0.6106772
   ```
3. Find the value of k such that P(X < k) = 0.9345
   ```
   qnorm(p=0.9345, 20, sd=5)
   [1] 27.55085
   ```

2017 Semester 2 Lab Quiz

Questions

Question 1

Load the dataset Loblolly into R by executing the command `data(“Loblolly”)`. Answer the following questions in regards to the variable `height`.

a. Obtain the Five Number Summary.
```
> fivenum(Loblolly$height)
[1]  3.460 10.455 34.000 51.395 61.100
```
The numbers correspond to minimum, Q1, median, Q3 and maximum

b. Determine the inter-quartile range.
```
> 51.395 – 10.455
[1] 40.94
```

c. Obtain a box plot for the variable and check if there are any outliers
```
> boxplot(Loblolly$height, horizontal=TRUE, main="Loblolly Height Attribute",
xlab="Height")
...
```
There are no outliers

d. Obtain a histogram for the variable.
```
> hist(Loblolly$height, main="Loblolly Height Attribute", xlab="Height")
...
```
There are no outliers

e. Find the 90% confidence interval for the variable assuming that σ is known and is equal to s.
```
> x_bar <- mean(Loblolly$height)
> s <- sd(Loblolly$height)
> z <- qnorm(p=0.95)
> n <- length(Loblolly$height)
> lower_bound <- x_bar - z * (s / sqrt(n))
> lower_bound
[1] 28.65415
> upper_bound <- x_bar + z * (s / sqrt(n))
> upper_bound
[1] 36.07466
```
Therefore, the 90% CI is (`28.65415`, `36.07466`)

f. True or false: “The probability that the mean lies in the 90% CI is 0.9.”
False

Question 2

If X is a normally distributed random variable with $\mu = 25$ and $\sigma = 6$, calculate the following using R:

a. $P(18 < X < 27)$
```
> pnorm(q=27, mean=25, sd=6) - pnorm(q=18, mean=25, sd=6)
[1] 0.5088862
```
Therefore, $P(18 < X < 27)$ = `0.5088862`

b. Find the value of k such that $P(X < k) = 0.7352$
```
> qnorm(p=0.7352, mean=25, sd=6)
[1] 28.7717
```
Therefore, k = `28.7717`

Question 3

Generate 100 means for samples of size 10 from the digits 1 to 6. Plot your results using a histogram. Note your observations
```
> hist(replicate(100, mean(sample(1:6, 10, replace=TRUE))))
...
```
The histogram is roughly symmetric and approximately normal.