STAT 2509B4

Assignment 1

Krystian Wojcicki, 101001444

Winter 2020

1. **Define the following:**
   (i) **variable:** a characteristic that varies from a person or a thing to another one, or over time.
   (ii) **experimental unit:** indiviudals or objects on which a variable is measured
   (iii) **a single measurement:** a single measurement is the value of a variable measured on a an experimental unit.
   (iv) **population:** the set of all measurements of a variable of interest to the investigator
   (v) **sample:** a subset of measurements selected and observed from the population of interest

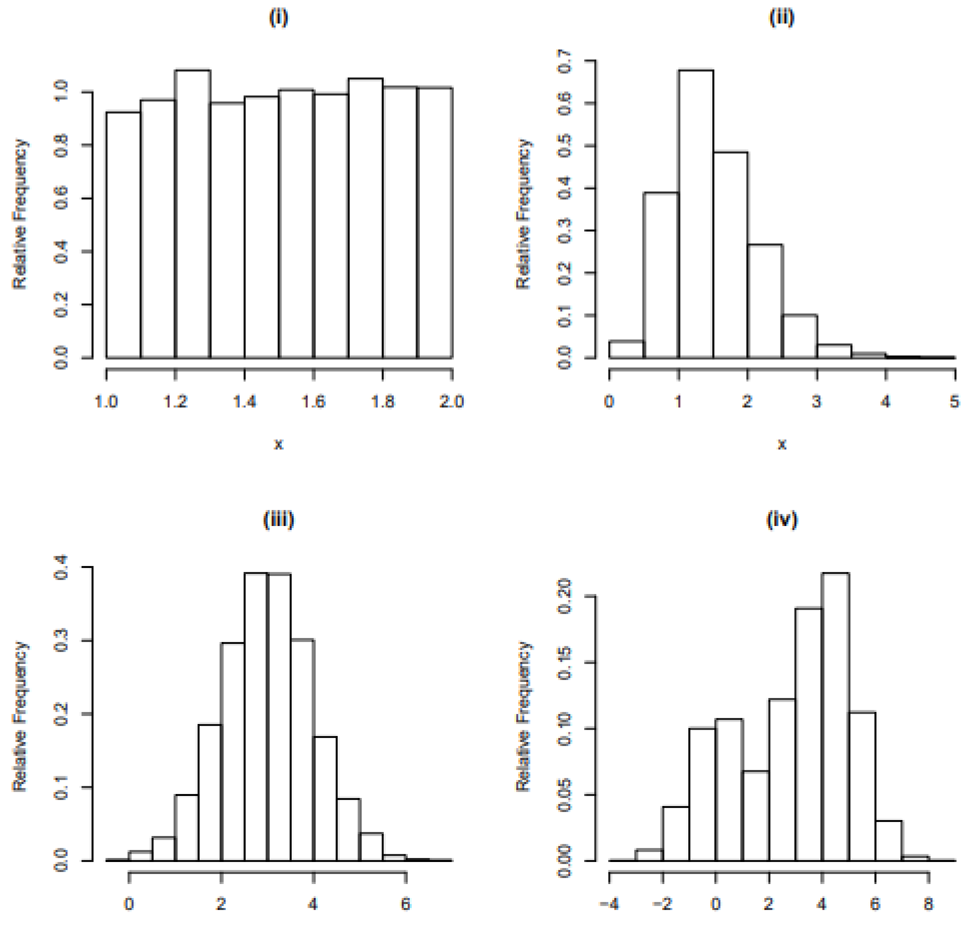
2. (a) **List the possible types of variables:** The possible types of variables are qualitative and quantitative. Qualitative can be split up into pure qualitative and qualitative ranked. Quantitative can be split up into quantitative & discrete and quantitative & continuous.
   (b) **Identify the following variables as either "pure qualitative" (or "pure categorical"),"qualitative & ranked" (or "categorical & ranked)", "quantitative & discrete", or "quantitative & continuous":**
      (i) **Time until a bulb burns out:** quantitative & continuous
      (ii) **Beer tasting ranking (excellent, good, fair, or poor):** qualitative & ranked
      (iii) **Student ID number:** pure qualitative
      (iv) **Number of cars entering Carleton each day:** quantitative & discrete
      (v) **Average daily temperature in Ottawa during January:** quantitative & continuous
      (vi) **Letter grade of a course (A, B, C, D, E, or F):** qualitative & ranked
      (vii) **Number of M&M candies in a bag:** quantitative & discrete
      (viii) **Blood type of a person:** pure qualitative

3. **Consider a normal population distribution with the value of the standard deviation σ known**
   (a) **What are the confidence level for the following confidence intervals about the population mean:**
      (i) $\bar{x} \pm 1.96\sigma/\sqrt{n}$**:** $\rightarrow z_{\alpha/2} = 1.96 \rightarrow \alpha/2 = 0.025 \rightarrow \alpha = 0.05 \rightarrow 1 - \alpha = 0.95$. Therefore 95% CI for population mean.
      (ii) $\bar{x} \pm 2.65\sigma/\sqrt{n}$**:** $\rightarrow z_{\alpha/2} = 2.65 \rightarrow \alpha/2 = 0.004 \rightarrow \alpha = 0.008 \rightarrow 1 - \alpha = 0.992$. Therefore 99.2% CI for population mean.
      (iii) $\bar{x} \pm 3.34\sigma/\sqrt{n}$**:** N/A as value not present in the given z table.
   (b) **What value of z in the confidence interval formula**

$$(\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\sigma/\sqrt{n})$$

   results in a confidence level of
      (i) **97.96%:** $\rightarrow 1 - \alpha = 0.9796 \rightarrow \alpha = 0.0204 \rightarrow \alpha/2 = 0.0102 \rightarrow z_{\alpha/2} = 2.32$
      (ii) **78.88%:** $\rightarrow 1 - \alpha = 0.7888 \rightarrow \alpha = 0.2112 \rightarrow \alpha/2 = 0.1056 \rightarrow z_{\alpha/2} = 1.25$
      (iii) **99.94%:** $\rightarrow 1 - \alpha = 0.9994 \rightarrow \alpha = 0.0006 \rightarrow \alpha/2 = 0.0003 \rightarrow z_{\alpha/2} =$ N/A as value not present in given z table.

4. **Which of the following histograms looks like a histogram for data from a normal distribution?**



<u>Explain</u>

Graph #3 looks like a histogram displaying a normal distribution. This is because a normal distribution is characterized by a symmetric, bell-shaped curve. All the other graphs are either non-symmetric or not bell shaped, while #3 is bell-shaped and symmetric.

5. **Let X be a random variable having a normal distribution with mean μ and standard deviation σ. Let c be a constant. Find the <u>distribution</u>, <u>mean</u> and <u>standard deviation</u> of each of the following random variables:**

   (i) $X + c$**:** as taught in class if X is a random variable with a normal distribution then $X \sim N(\mu, \sigma^2)$. Meaning $X + c$ (where c is a constant) gives us $X + c \sim N(\mu + c, \sigma^2)$. Therefore for $X + c$ is a normal distribution, with a mean of $\mu + c$ and standard deviation of σ.

   (ii) $(X - \mu)/\sigma$**:** as $X \sim N(\mu, \sigma^2)$ then $z := (X - \mu)/\sigma \sim N(0, 1)$. Meaning $(X - \mu)/\sigma$ is a standard normal distribution, with a mean of 0 and standard deviation of 1.

6. **Find the following values from the upper-tail z and t tables:**

   (i) $z_{0.0154} =$ 2.16

   (ii) $z_{0.9846} =$ $z_{0.0154} = 2.16$

   (iii) $z_{0.1215} =$ 1.17

   (iv) $t_{6;0.05} =$ 1.943

   (v) $-t_{10;0.025} =$ -2.228

   (vi) $t_{10;0.975} =$ $-t_{10;0.025} = -2.228$

7. a **Define two-sided and one-sided hypotheses about a parameter θ and give the steps involved in their testing:**
2-sided hypothesis is a 2 tailed test for testing parameter $\theta \neq 0$. For example $H_0 : \theta = 0$ or $H_a : \theta \neq 0$.
1-sided hypothesis is a 1 tailed test for testing parameter $\theta < 0$ or $\theta > 0$. For example $H_0 : \theta \leq 0$ vs $H_a : \theta > 0$ or the other side, $H_0 : \theta \geq 0$ vs $H_a : \theta < 0$.
The steps involved for either hypothesis is as follows:
   1) State H_0 and H_a
   2) Find the test statistic for the test
   3) Find the rejection or critical region (or p-value)
   4) Draw conclusion

   b **For any hypothesis test, what are the two types of errors that may be made? Explain.**
Type I error is an error when we reject H_0 when it is in fact true. The probabiltiy of type I error is α.
Type II error is an error when we do not reject H_0 when it is in fact false. The probability of type II error is β.

8. **Classify each of the following quantities as either a parameter or a statistic:**
   (i) σ^2**:** parameter
   (ii) $\hat{\beta}_1$**:** statistic
   (iii) s^2**:** statistics
   (iv) μ**:** parameter
   (v) β_0**:** parameter
   (vi) $\bar{x}$**:** statistic

9. **Show that:**
$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}$

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$
$$= \sum_{i=1}^{n}(x_i y_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y})$$
$$= \sum_{i=1}^{n}(x_i y_i - x_i\bar{y} - \bar{x}y_i) + n\bar{x}\bar{y}$$
$$= \sum_{i=1}^{n}(x_i y_i) - \bar{y}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} y_i + n\bar{x}\bar{y}$$
$$= \sum_{i=1}^{n}(x_i y_i) - \bar{y}\bar{x}n - \bar{x}\bar{y}n + n\bar{x}\bar{y}$$
$$= \sum_{i=1}^{n}(x_i y_i) - 2\bar{y}\bar{x}n + n\bar{x}\bar{y}$$
$$= \sum_{i=1}^{n}(x_i y_i) - \bar{y}\bar{x}n$$
$$= \sum_{i=1}^{n}(x_i y_i) - \frac{\sum_{i=1}^{n} y_i}{n}\frac{\sum_{i=1}^{n} x_i}{n} n$$
$$= \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}$$

Therefore $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}$