Tutorial Sheet - Z-test

```
knitr::opts_chunk$set(echo = TRUE)
library(ggplot2)
```

Question 1

1. To investigate climate change meteorologists in Ireland, wish to examine the rain- fall in November. They have established that the normal mean daily rain fall in November is 2.35 mm. They observed rainfall for the 30 days of November in 2015 and recorded a mean rainfall of 4.05mm and a standard deviation of 2.25mm. Conduct a hypothesis test to determine if the observed data from November 2015 gives evidence (α= 0.01) that there is a change in rainfall.

ANSWER:

```
mu<-2.35
mean<-4.05
SD<-2.25
N<-30
z<-(mean-mu)/sqrt(SD*SD/N)
z
```

```
## [1] 4.138348
```

```
SEM<-1.96*sqrt(SD*SD/N)
Z_criteria_upper<-qnorm(0.995)
Z_criteria_lower<-qnorm(0.005)
```

1. State an the null Hypothesis H_0 :

$$H_0 : \bar{x} = \mu$$

$$H_0 : \bar{x} = 2.35$$

2. State an Alternative Hypothesis H_α :

$$H_\alpha : \bar{x} \neq \mu$$

$$H_\alpha : \bar{x} \neq 2.35$$

3. Calculate a Test Statistic

$$Z = \frac{(\bar{x} - \mu)}{(SD/\sqrt{n})},$$

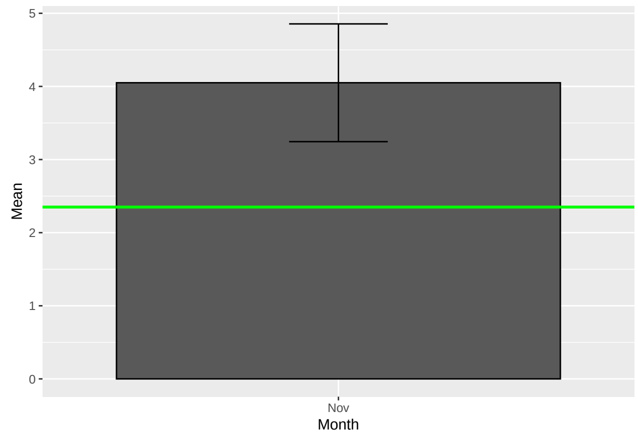
$$Z = \frac{(4.05 - 2.35)}{(2.25/\sqrt{30})} = 4.1383482.$$

4. Calculate a p-value and/or set a rejection region:

Reject for Z<-2.5758293 and Z>2.5758293.

5. State your conclusions:

There was statistically more rainfall this November



Question 2

2. Researchers wished to investigate differences in TV viewing practices between children in Ireland versus America. The researchers established that the mean TV viewing per week is 30.5 hours for American children. They observed 115 Irish children for a week and recorded a mean TV viewing of 28 hours and a standard deviation of 2.5 hours. Conduct a hypothesis test to determine if the observed data gives evidence that Irish children watch less TV than American children ($\alpha = 0.05$).

ANSWER

```
mu<-30.5
mean<-28
SD<-2.5
N<-115
z<-(mean-mu)/sqrt(SD*SD/N)
z
```

```
## [1] -10.72381
```

```
SEM<-1.96*sqrt(SD*SD/N)

#Z_criteria_upper<-qnorm(0.95)
Z_criteria_lower<-qnorm(0.05)
```

1. State an the null Hypothesis H_0 :

$$H_0 : \bar{x} = \mu$$

$$H_0 : \bar{x} = 30.5$$

2. State an Alternative Hypothesis H_α :

$$H_\alpha : \bar{x} < \mu$$

$$H_\alpha : \bar{x} < 30.5$$

3. Calculate a Test Statistic

$$Z = \frac{(\bar{x} - \mu)}{(SD/\sqrt{n})},$$

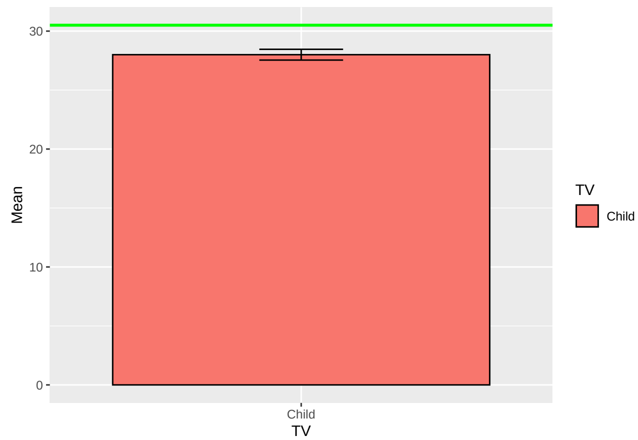
$$Z = \frac{(28 - 30.5)}{(2.5/\sqrt{115})} = -10.7238053.$$

4. Calculate a p-value and/or set a rejection region:

Reject for Z<-1.6448536.

5. State your conclusions:

```
df<-data.frame(Mean=c(mean),
               sd=c(SEM),
               TV=as.factor(c("Child")))
# Default bar plot
ggplot(df, aes(x=TV, y=Mean, fill=TV)) +
  geom_bar(position=position_dodge(), stat="identity",
           colour='black') +
  geom_errorbar(aes(ymin=Mean-sd, ymax=Mean+sd), width=.2) +geom_hline(yintercept = mu, color = "green"
```



Question 3

3. The CEO of a large electric utility claims that 80 percent of his customers are very satisfied with the service they receive. To test this claim, the local newspaper surveyed 140 customers, using simple random sampling. Among the sampled customers, 73 percent say they are very satisfied. Based on these findings, can we reject the CEO's hypothesis that 80% of the customers are very satisfied? Use a 0.05 level of significance.

ANSWER

```
p0<-0.8
p_hat<-0.73

N<-140
z<-(p_hat-p0)/sqrt(p0*(1-p0)/N)
z
```

```
## [1] -2.070628
```

```
SEM<-1.96*sqrt(p0*(1-p0)/N)

Z_criteria_upper<-qnorm(0.975)
Z_criteria_lower<-qnorm(0.025)
```

1. State an the null Hypothesis H_0 :

$$H_0 : \hat{p} = p_0$$

$$H_0 : \hat{p} = 0.8$$

2. State an Alternative Hypothesis H_α :

$$H_\alpha : \hat{p} \neq p_0$$

$$H_\alpha : \hat{p} \neq 0.8$$

3. Calculate a Test Statistic

$$Z = \frac{(\hat{p} - p_0)}{(\sqrt{q_0 p_0 / n})},$$

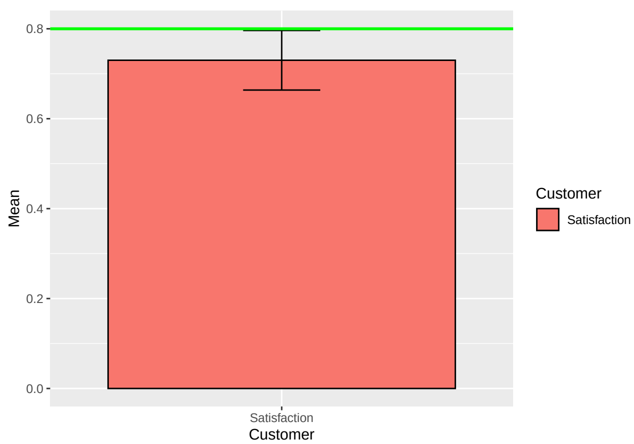
$$Z = \frac{(0.73 - 0.8)}{(\sqrt{0.2 \times 0.8/140})} = -2.0706279.$$

4. Calculate a p-value and/or set a rejection region:

Reject for Z<-1.959964 or Z>1.959964.

5. State your conclusions:

```
df<-data.frame(Mean=c(p_hat),
               sd=c(SEM),
               Customer=as.factor(c("Satisfaction")))
# Default bar plot
ggplot(df, aes(x=Customer, y=Mean, fill=Customer)) +
  geom_bar(position=position_dodge(), stat="identity",
           colour='black') +
  geom_errorbar(aes(ymin=Mean-sd, ymax=Mean+sd), width=.2) +geom_hline(yintercept = p0, color = "green"
```



Question 4

4. A four-sided (tetrahedral) die is tossed 1000 times, and 290 fours are observed. Is there evidence to conclude that the die is biased, that is, say, that more fours than expected are observed? Conduct a hypothesis test to determine if the observed data gives evidence that the die is fair (α= 0.05).

   ANSWER

```
p0<-0.25
p_hat<-0.29

N<-1000
z<-(p_hat-p0)/sqrt(p0*(1-p0)/N)
z
```

```
## [1] 2.921187
```

```
SEM<-1.96*sqrt(p0*(1-p0)/N)
```

1. State an the null Hypothesis H_0 :

$$H_0 : \hat{p} = p_0$$

$$H_0 : \hat{p} = 0.25$$

2. State an Alternative Hypothesis H_α :

$$H_\alpha : \hat{p} \neq p_0$$

$$H_\alpha : \hat{p} \neq 0.25$$

3. Calculate a Test Statistic

$$Z = \frac{(\hat{p} - p_0)}{(\sqrt{q_0 p_0 / n})},$$

$$Z = \frac{(0.29 - 0.25)}{(\sqrt{0.75 \times 0.25/1000})} = 2.921187.$$

4. Calculate a p-value and/or set a rejection region:

Reject for Z<-1.959964 or Z>1.959964.

5. State your conclusions:

```
df<-data.frame(Mean=c(p_hat),
               sd=c(SEM),
               Die=as.factor(c("Roll")))
# Default bar plot
ggplot(df, aes(x=Die, y=Mean)) +
  geom_bar(position=position_dodge(), stat="identity",
           colour='black') +
  geom_errorbar(aes(ymin=Mean-sd, ymax=Mean+sd), width=.2) +geom_hline(yintercept = p0, color = "green"
```