Exam 2 Practice Problems

```
##              Candy Calories          Type      Maker MiniCalories
## 1       Almond Joy      234     Chocolate    Hershey          200
## 2             Twix      286     Chocolate Mars, Inc.          173
## 3          Reece's      231     Chocolate    Hershey          197
## 4        Milky Way      264     Chocolate Mars, Inc.          202
## 5         Snickers      215     Chocolate Mars, Inc.          118
## 6       Starbursts      200 Not Chocolate Mars, Inc.           75
## 7         Skittles      162 Not Chocolate Mars, Inc.           84
## 8         Twizzler      100 Not Chocolate    Hershey           64
## 9   Jolly Ranchers       70 Not Chocolate    Hershey           52
## 10 Good and Plenty      110 Not Chocolate    Hershey           62
```

Group 1 You want to know if there is a difference in the number of calories in candy based on if it contains chocolate. With a degrees of freedom of 6.2709, the critical value is +/- 2.4.

Group 2 You know that the mini versions of candy have fewer calories than the original versions, but you want to know if they have markedly fewer calories than the original or if the mini versions have roughly a similar amount of calories to the original versions.

Group 3 You are concerned that your results are biased because some makers only make chocolate candy, then those might have more calories than the candy from a brand that doesn't make chocolate. You need to know if chocolate/not chocolate is somehow related to the producer. However, you know that the data shown here are limited. You pool from other sources and get the following information: Hershey's makes 18 types of non-chocolate candy and 46 types of chocolate candy. Mars, Inc. makes 22 types of non-chocolate candy and 37 types of chocolate candy. Assume a critical value of 3.8

Group 1 - Independent Samples *t*

Need standard error:

```
variances = candy %>%
  group_by(Type) %>%
  summarize(vars = var(Calories))

variances
```

```
## # A tibble: 2 × 2
##   Type           vars
##   <fct>         <dbl>
## 1 Chocolate      814.
## 2 Not Chocolate 2703.
```

```
sdChoc = variances$vars[1]
sdNonChoc = variances$vars[2]
N1 = 5
N2 = 5

welchSE = sqrt((sdChoc/N1) + (sdNonChoc/N2))
welchSE
```

```
## [1] 26.51905
```

Group 1 - Independent Samples *t*

Get test statistic

```
means = candy %>%
  group_by(Type) %>%
  summarize(means = mean(Calories))

means
```

```
## # A tibble: 2 × 2
##   Type          means
##   <fct>         <dbl>
## 1 Chocolate      246
## 2 Not Chocolate  128.
```

```
meanChoc = means$means[1]
meanNotChoc = means$means[2]

tStatistic = (meanChoc - meanNotChoc)/welchSE
tStatistic
```

```
## [1] 4.434548
```

Group 1 - Independent Samples *t*

```
t.test(Calories ~ Type, data = candy)
```

```
## 
##     Welch Two Sample t-test
## 
## data:  Calories by Type
## t = 4.4345, df = 6.2079, p-value = 0.004054
## alternative hypothesis: true difference in means between group Chocolate and group Not
## 95 percent confidence interval:
##   53.23328 181.96672
## sample estimates:
##     mean in group Chocolate mean in group Not Chocolate
##                       246.0                       128.4
```

Group 2 - Paired Samples *t*

```
candy$diff = candy$Calories - candy$MiniCalories
meanDiff = mean(candy$diff)
sdDiff = sd(candy$diff)
NPairs = 10

seDiff = sdDiff/(sqrt(10))

tStatisticPaired = meanDiff/seDiff
tStatisticPaired
```

```
## [1] 5.502914
```

```
pt(tStatisticPaired, df = NPairs-1, lower.tail = F)
```

```
## [1] 0.0001893778
```

Group 2 - Paired Samples *t*

```
t.test(candy$Calories, candy$MiniCalories, paired = T, alternative = "greater"
```

```
##
##     Paired t-test
##
## data:  candy$Calories and candy$MiniCalories
## t = 5.5029, df = 9, p-value = 0.0001894
## alternative hypothesis: true mean difference is greater than 0
## 95 percent confidence interval:
##  43.01397      Inf
## sample estimates:
## mean difference
##            64.5
```

Group 3 - Chi-square association

Null = There is no relationship (not associated, are independent)

Alternative = There is a relationship (are associated, are dependent)

```
newCandy = data.frame(Hershey = c(18, 46),
           Mars = c(22, 37))
rownames(newCandy) <- c("Non-Chocolate", "Chocolate")
newCandy
```

```
##               Hershey Mars
## Non-Chocolate      18   22
## Chocolate          46   37
```

Group 3 - Chi-square association

Expected frequencies rowsum * columsum/total n

```
##                  Hershey      Mars
## Non-Chocolate 40*64/123 40*59/123
## Chocolate     83*64/123 83*59/123
```

```
##               Hershey  Mars
## Non-Chocolate   20.81 19.19
## Chocolate       43.19 39.81
```

Group 3 - Chi-square association

Now plug in those expected frequencies

```
##                             Hershey                  Mars
## Non-Chocolate (18-20.81)^2/20.81 (22-19.19)^2/19.19
## Chocolate     (46-43.19)^2/43.19 (37-39.81)^2/39.81
```

```
##               Hershey Mars
## Non-Chocolate    0.38 0.41
## Chocolate        0.18 0.20
```

.38 + .41 + .18 + .2 = 1.17

Group 3 - Chi-square association

df = (r-1)(c-1) df = 1

chi(1) = 1.17

```
pchisq(q = 1.17, df = 1, lower.tail = F)
```

```
## [1] 0.2794012
```

```
chisq.test(newCandy, correct = FALSE)
```

```
##
##     Pearson's Chi-squared test
##
## data:  newCandy
## X-squared = 1.1746, df = 1, p-value = 0.2785
```