# Homework #1.1-Example B

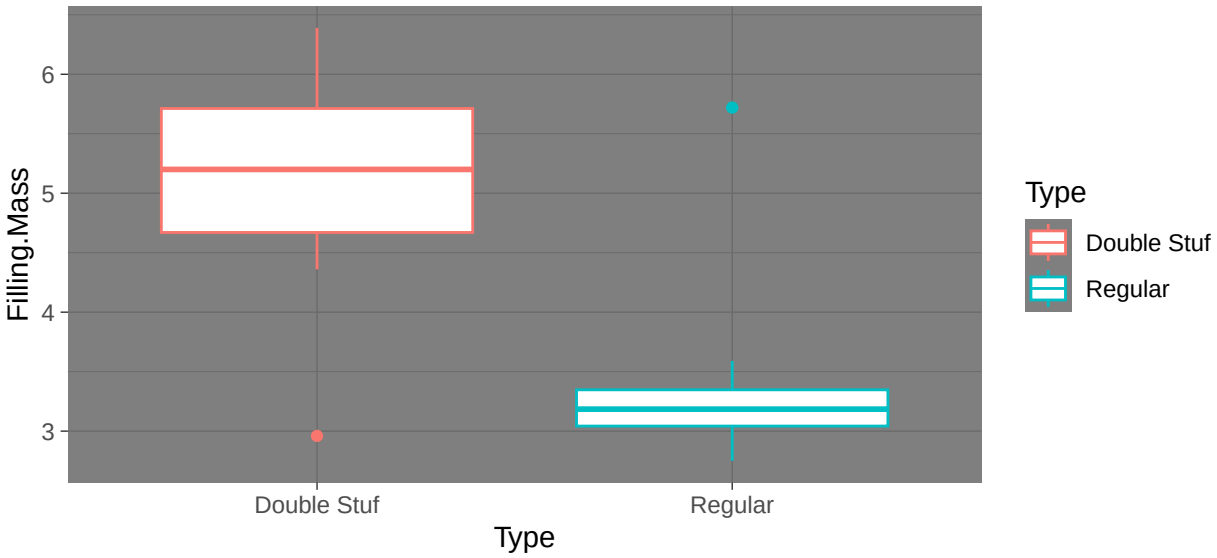
B. Student

1/24/2025

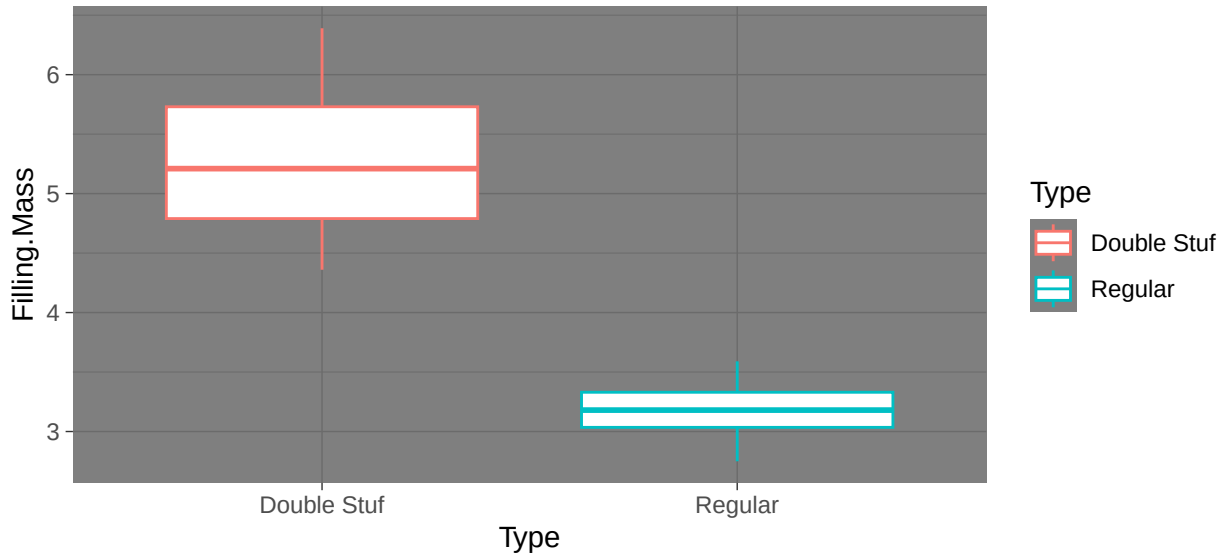
If Nabisco is going to market things as being "double", then consumers should be confident that they are getting twice as much. This brings us to the question of whether Double Stuf Oreos have double the creme filling of regular oreos. We've been given some actual data to explore this question with.

## Exploratory Data Analysis

We will begin by looking at box plots of our data.

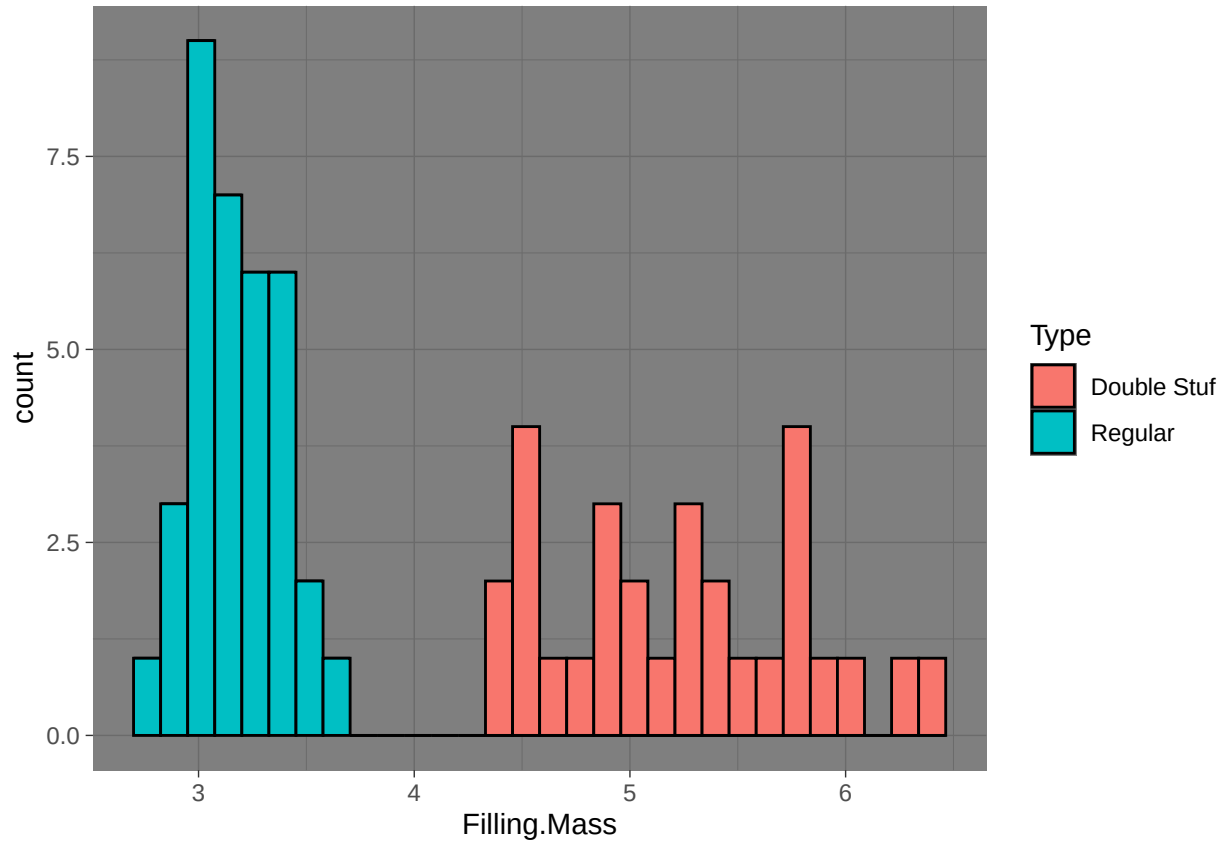



There appear to be two outliers; one in the Regular and one in the Double Stuf. These turn out to be rows 2 and 35; I'm going to drop them since you get rid of the most extreme outliers.




The new box plots look much better (no outliers). We can also look at histograms of creme filling mass.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




There seems to be more variation in filling masses for double stuf oreos than regular. Double stuf oreos have larger amounts of creme filling than Regular.

Let's look at some values of descriptive statistics

```
##     item      group1 vars  n   mean     sd median trimmed    mad  min  max
## X11    1 Double Stuf    1 29 5.2262 0.5860   5.21  5.2052 0.7710 4.36 6.39
## X12    2     Regular    1 35 3.1766 0.2063   3.18  3.1772 0.2372 2.75 3.59
##     range    skew kurtosis     se   IQR Q0.25 Q0.75
## X11  2.03  0.2142  -1.0903 0.1088 0.940 4.790  5.73
## X12  0.84 -0.0074  -0.8879 0.0349 0.295 3.035  3.33
```

The SAV values are 0.343396 and 0.0425597 for double stuf and regular, respectively.

The min tells us the smallest amount of creme filling observed. Thus, the oreo with the least amount of filling had 2.75 grams. The max on the other hand tells us that largest amount of creme filling; no cookie had more than 6.39 grams of filling. Half of the double stuff cookies had less than 5.21 grams (the median) of filling. The first quartile (Q0.25) tells us that 25% of the cookies had no more than 3.035 grams of filling for the Regular oreos. The third quartile tells us that 75% of the cookies had no more than 3.33 grams of filling. The sample arithmetic mean tells us how well the cookies did at amassing creme filling: for double stuf oreos this was 5.2262 grams for each cookie. The sample variance tells us a rough sense of the amount of variation as deviation there is in the set; this value is 0.343 for Double stuf, which says that there is some variation. This is larger than the Regular oreo's value of 0.043, which links up to what we saw in the histograms.

Sample skewness tells us whether we are positively, negatively skewed or symmetric. Double stuf oreo have a positive value, and thus are positively skewed, which matches the histogram. Regular oreos have a slightly negative value, but that value is really close to zero, so that histogram should look symmetric (which it very nearly does).

## Hypothesis Testing

"Are double stuf oreos double the stuf of Regular Oreos?" is our focus. This would be an example of a two sample location problem. Our hypotheses would be:

> Null: Double Stuf oreos have twice the filling of Regular oreos
> $H_0 : D = 2R$
>
> Alternative: Double stuf oreos don't have twice the filling of Regular oreos
> $H_1 : D \neq 2R$

We can transform these hypotheses into the form $D - 2R = 0$, which would be like that for a two sample $t$-test. This would match with our claim that we have a two sample location problem. The assumptions for a two sample t test include independence of observations, and normality. If we use Welch's test, we don't have to worry about the variances being equal, so let's do that. We have independent samples and since we're essentially 30 cookies or more in each group, we can proceed.

Since we have 2R, we need to double the creme filling masses for the regular oreos before we run the test. We will use a threshold of 0.03 to declare an unusual event given the null hypothesis.

```
##
##  Welch Two Sample t-test
##
## data:  newFilling by Type
## t = -8.7181, df = 48.932, p-value = 1.575e-11
## alternative hypothesis: true difference in means between group Double Stuf and group Regular is not
```

```
## 97 percent confidence interval:
##  -1.4158678 -0.8380041
## sample estimates:
## mean in group Double Stuf     mean in group Regular
##                  5.226207                  6.353143
```

From our null hypothesis test we find that $t(48.98) = -8.72$ with a p-value of $1.575 \times 10^{-11}$. This indicates that under the null hypothesis, we would only expect the $t$ statistic to take a value at least as extreme as -8.72 essentially 0 percent of the time. Since is value is less than our threshold of 0.03, we have an unusual event and will take this as evidence against the null hypothesis. The 97% confidence interval for the difference is (-1.42, -0.84). Since 0 isn't in the interval, we can again take this as evidence against the null hypothesis. Thus, we will reject the null hypothesis and decide that double stuf oreos do not have twice the creme filling of regular oreos. Given that the point estimate is -1.127 (double minus 2*regular) and the interval consists of all negative values for the same order of subtraction, this would suggest that double stuff oreos might have less than twice the amount of creme filling of regular oreos.

From our sample, we've decided that double stuf oreos don't have twice the creme filling of regular oreos. There is some evidence that they have less creme filling. However, we should collect a larger sample to test this new hypothesis.

## Code Appendix

```
# Setting Document Options
knitr::opts_chunk$set(
  echo = FALSE,
  warning = FALSE,
  message = FALSE,
  fig.align = "center"
)

# Add additional packages by name to the following list
packages <- c("tidyverse", "knitr", "kableExtra", "psych")
invisible(
  lapply(
    X = packages,
    FUN = library,
    character.only = TRUE,
    warn.conflicts = FALSE,
    quietly = TRUE
  )
)

oreoData <- read.table(
  file = "https://raw.github.com/neilhatfield/STAT461/master/dataFiles/oreo2.dat",
  header = TRUE,
  sep = ","
)

ggplot(
  data = oreoData,
  mapping = aes(x = Type, y = Filling.Mass, color = Type)
) +
  geom_boxplot() +
  theme_dark()
oreoData <- oreoData[-c(2, 35),]
ggplot(
  data = oreoData,
  mapping = aes(x = Type, y = Filling.Mass, color = Type)
) +
  geom_boxplot() +
  theme_dark()
ggplot(
  data = oreoData,
  mapping = aes(x = Filling.Mass, fill = Type)
) +
  geom_histogram(col = "black") +
  theme_dark()

psych::describeBy(
  x = oreoData$Filling.Mass,
  group = oreoData$Type,
  na.rm = TRUE,
  skew = TRUE,
```

```
  ranges = TRUE,
  quant = c(0.25, 0.75),
  IQR = TRUE,
  mat = TRUE,
  digits = 4
)
oreoData <- oreoData %>%
  dplyr::mutate(
    newFilling = ifelse(
      test = Type == "Regular",
      yes = 2 * Filling.Mass,
      no = Filling.Mass
    )
  )

t.test(
  formula = newFilling ~ Type,
  data = oreoData,
  conf.level = 0.97
)
```