Stat 140: Pizza Slices Revisited

Evan Ray

December 6, 2017

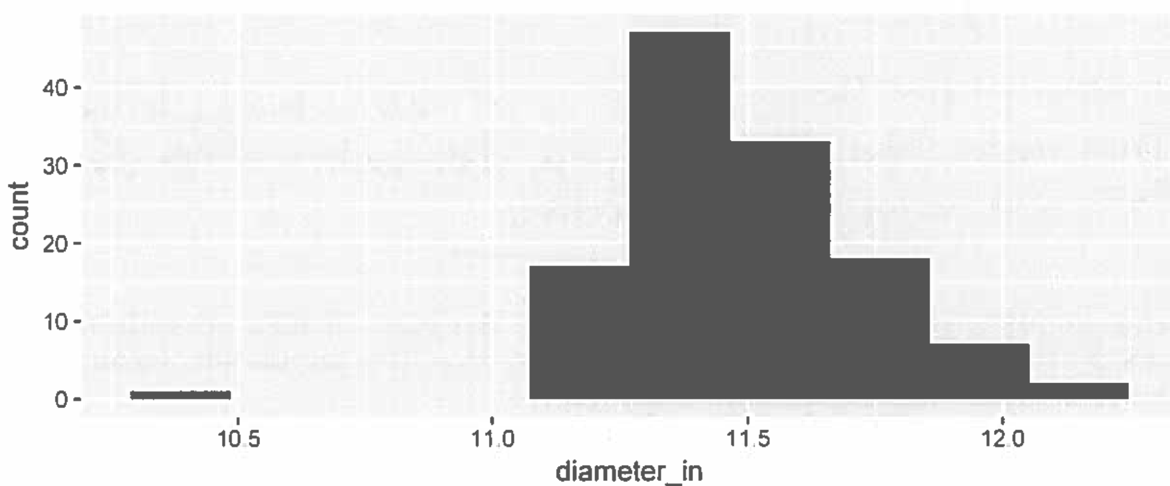
Pizza Slices

Reminder of context from HW7 problem: In 2012 a big Australian pizza chain, Eagle Boys, ran an advertising campaign saying their pizza was better than the competition's pizza because it was bigger. They made several specific claims, and one of them was that the average size of their pizzas was greater than 12 inches. (You can see a screenshot of their website making these claims here: https://mhc-stat140-2017.github.io/data/jse/pizza/pizza_website.png). Interestingly, the company also provided the data to "back up" their claims.

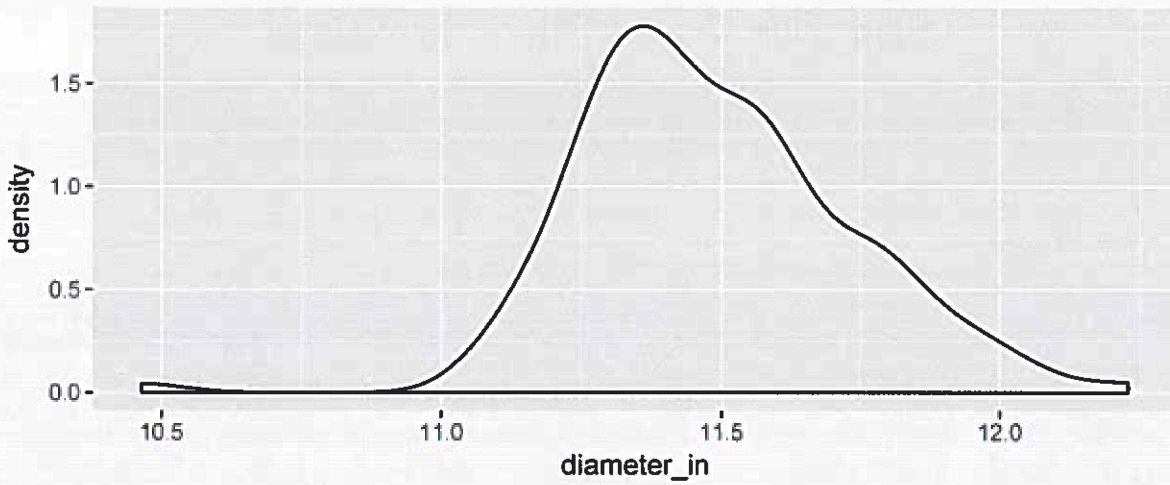
```
pizza <- read_csv("https://mhc-stat140-2017.github.io/data/jse/pizza/pizzasize.csv")
```

```
## Parsed with column specification:
## cols(
##   ID = col_integer(),
##   Store = col_character(),
##   CrustDescription = col_character(),
##   Topping = col_character(),
##   Diameter = col_double()
## )
```

```
pizza_eagle <- filter(pizza, Store == "EagleBoys") %>%
  mutate(diameter_in = Diameter * 0.39370079)
ggplot() +
  geom_histogram(mapping = aes(x = diameter_in), bins = 10, data = pizza_eagle)
```




```
ggplot() +
  geom_density(mapping = aes(x = diameter_in), data = pizza_eagle)
```



Check conditions for inference

1) Independence. We can't really check this using the information given. We would have to assume that the pizzas were selected randomly, from a representative sample of different stores at different times. The sample size of 125 pizzas is less than 10% of all pizzas ever made by this large chain.

2) Data distribution nearly normal.

There is one outlier, which is concerning. I do think the mean will be representative of the center of this distribution though.

3) Sufficient sample size. The sample size of 125 is fairly large. This means that the sampling distribution of the sample mean will be approximately normal.

Being skeptical consumers, we don't believe the pizza chain's advertisements. Let's perform a hypothesis test of the claim that the population mean pizza diameter is less than 12 inches.

(a) State the null and alternative hypotheses

$H_0: \mu = 12$

$H_A: \mu < 12$

Here, μ denotes the population mean pizza diameter for all pizzas made by Eagle Boys pizza.

(b) **What do you conclude?**

SOLUTION:

```
t.test(pizza_eagle$diameter_in, mu = 12, alternative = "less")

##
##  One Sample t-test
##
## data:  pizza_eagle$diameter_in
## t = -23, df = 120, p-value <2e-16
## alternative hypothesis: true mean is less than 12
## 95 percent confidence interval:
##   -Inf 11.52
## sample estimates:
## mean of x
##     11.49
```

The p-value is less than $2 * 10^{-16}$. If we use a significance level of $\alpha = 0.001$, then the p-value is less than α. This means that the data offer enough evidence to conclude that the population mean pizza diameter is less than 12 inches. We can reject the null hypothesis.

I will not ask you to do this on a quiz or final.

(c) **Conduct the hypothesis test again by comparing the test statistic to a critical value**

```
mean(pizza_eagle$diameter_in)
## [1] 11.49
sd(pizza_eagle$diameter_in)
## [1] 0.2466
nrow(pizza_eagle)
## [1] 125
qt(0.05, df = 124)
## [1] -1.657
```

I'm doing this for $\alpha = 0.05$

The test statistic is

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{11.49 - 12}{0.2466/\sqrt{125}} = -23.12$$

value from the null hypothesis

Reject H_0 if the test statistic is in this region: the p-value will be less than $\alpha = 0.05$

Fail to reject H_0 if the test statistic is in this region: the p-value will not be less than $\alpha = 0.05$

Since -23.12 is < -1.657, we can reject H_0. The data offer enough evidence to conclude that.....

Type I Errors

Type I Error: We incorrectly reject the null hypothesis – the null hypothesis is actually true, but we say it's not.

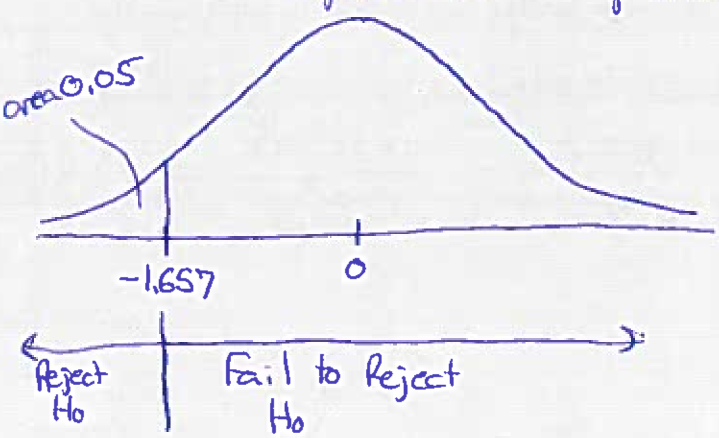
(a) State what a Type I Error would be in this problem.

A type I error would occur if the population ~~mean~~ pizza diameter was actually 12 inches, but we (incorrectly) rejected the null hypothesis and concluded that the population mean pizza diameter was less than 12 inches.

(b) Is it possible that we made a Type I Error in this case? If so, is there a way to tell for sure whether or not we made this error?

Since we rejected the null hypothesis, it's possible that we made a type I error. However, since we don't know the population mean there's no way to be sure.

(c) If the null hypothesis was true, for what proportion of samples would we commit a Type I Error? Here's a reproduction of the picture in the last page:



This is a picture of the test statistic: it describes the distribution of values of the test statistic that we would get from different samples.

If the null hypothesis is true 5% of samples will have a test statistic that is less than −1.657, in the "Reject Ho" region.

This means that if the null hypothesis is true, we will incorrectly reject the null hypothesis 5% of the time.

If the null hypothesis is true, P(we make a type I error) = α

significance level of the test.

Type II Errors

Type II Error: We incorrectly fail to reject the null hypothesis – the null hypothesis is actually wrong, but we don't find conclusive evidence that it is wrong at the specified significance level α.

(a) State what a Type II Error would be in this problem.

A type II error would occur if the population mean pizza diameter was actually less than 12 inches but we ~~[illegible]~~ did not find enough evidence to conclude that the population mean pizza diameter was less than 12 inches.

(b) Is it possible that we made a Type II Error in this case? If so, is there a way to tell for sure whether or not we made this error?

Since we rejected the null hypothesis, there is no way we could have made a type II error.

If we had failed to reject the null hypothesis, it would be possible that we could have made a type II error – but there would be no way to be sure.

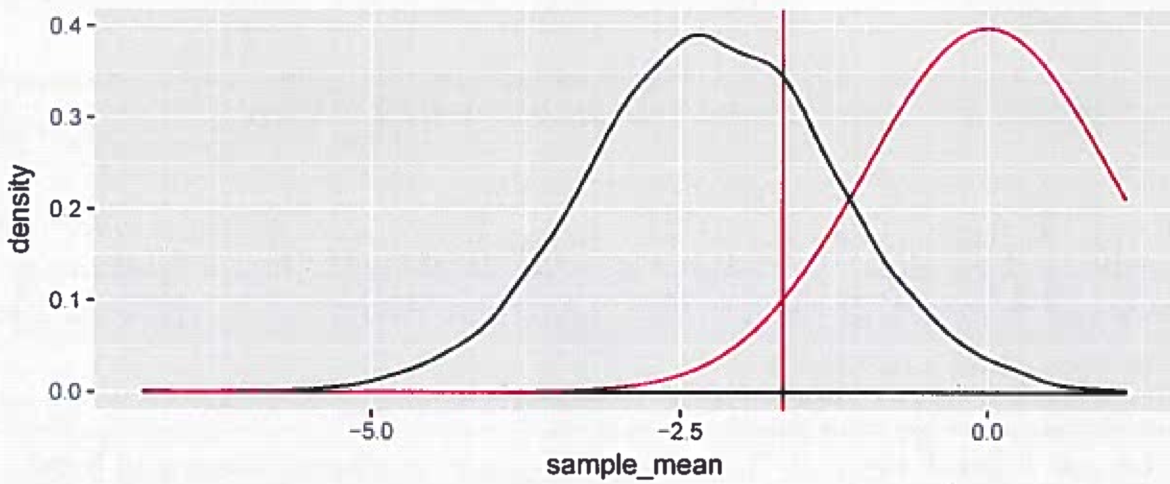
(c) If the null hypothesis ~~was~~ NOT true, for what proportion of samples ~~would~~ we commit a Type I Error?

Suppose pizza diameters follow a normal distribution with mean 11.9 inches and standard deviation 0.5 inches, and we take a sample of size $n = 125$. How often will we (incorrectly) fail to reject the null hypothesis?

```
sim_results <- do(10000) * {
  sample_x <- rnorm(125, mean = 11.9, sd = 0.5)
  sample_mean <- mean(sample_x)
  sample_sd <- sd(sample_x)
  test_stat <- (sample_mean - 12) / (sample_sd / sqrt(125))
}

plot_df <- data.frame(
  sample_mean = sim_results[[1]]
)
ggplot(mapping = aes(x = sample_mean), data = plot_df) +
  geom_density() +
  stat_function(fun = dt, color = "red", args = list(df = 124)) +
  geom_vline(xintercept = qt(0.05, df = 124), color = "red")
```

Main point of all of this: the probability of making a Type II error depends on the (unknown) population mean so we can't really say how likely we are to make this mistake.



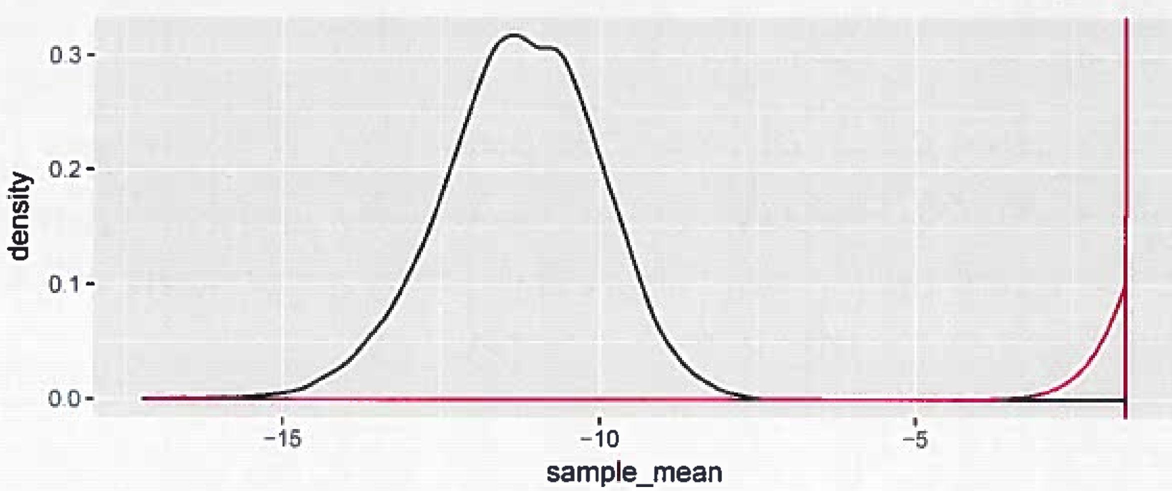

```
mean(sim_results >= qt(0.05, df = 124))
```

```
## [1] 0.2799
```

Suppose pizza diameters follow a normal distribution with mean 11.5 inches and standard deviation 0.5 inches, and we take a sample of size $n = 125$. How often will we (incorrectly) fail to reject the null hypothesis?

```
sim_results <- do(10000) * {
  sample_x <- rnorm(125, mean = 11.5, sd = 0.5)
  sample_mean <- mean(sample_x)
  sample_sd <- sd(sample_x)
  test_stat <- (sample_mean - 12) / (sample_sd / sqrt(125))
}

plot_df <- data.frame(
  sample_mean = sim_results[[1]]
)
ggplot(mapping = aes(x = sample_mean), data = plot_df) +
  geom_density() +
  stat_function(fun = dt, color = "red", args = list(df = 124)) +
  geom_vline(xintercept = qt(0.05, df = 124), color = "red")
```



```
mean(sim_results >= qt(0.05, df = 124))

## [1] 0
```