Statistics 324 Homework #7

Student Name Here

*Submit your homework to Canvas by the due date and time. Email your instructor if you have extenuating circumstances and need to request an extension.

*If an exercise asks you to use R, include a copy of the code and output. Please edit your code and output to be only the relevant portions.

*If a problem does not specify how to compute the answer, you many use any appropriate method. I may ask you to use R or use manual calculations on your exams, so practice accordingly.

*You must include an explanation and/or intermediate calculations for an exercise to be complete.

*Be sure to submit the HWK7 Auto grade Quiz which will give you ~20 of your 40 accuracy points.

*50 points total: 40 points accuracy, and 10 points completion

Exercise 1 From extensive testing, the mean drying time of a company's spray paint product is thought to be normally distributed with mean $\mu = 100$ seconds and standard deviation of $\sigma = 4$ seconds. The manager of the research division of the company that produces the paint considers using a different colorant and wants to know whether the mean drying time will change.

    a. Write out the manager's question in terms of a null H_o and alternative hypotheses H_a about the population mean μ drying time using the new colorant

$$H_0 : \mu = 100$$

The null hypothesis states that the mean drying time μ using the new colorant is equal to the mean drying time of 100 seconds with the original spray paint product.

$$H_a : \mu \neq 100$$

The alternative hypothesis states that the mean drying time μ using the new colorant is not equal to 100 seconds with the original spray paint product.

    b. If the managers choose to use a significance level of $\alpha = 0.10$ and assume $\sigma = 4$, identify the power of a Z test to detect a mean increase of 2 seconds ($\mu_a = 102$). They plan to look at a sample of 25 drying times and do a two-sided hypothesis test. Also identify the probability of making a type 2 error if the true mean drying time with the new colorant is 102 seconds, $\mu_a = 102$. (Compute these quantities "by hand" and then use the power simulation to check your work)

```
calculate_z_test_power <- function(effect_size, sample_size, standard_deviation, alpha) {
  z_critical <- qnorm(1 - alpha)  # Z critical value for one-tailed test

  # Calculate the standard error
  standard_error <- standard_deviation / sqrt(sample_size)
```

```r
  # Calculate the Z-score under the alternative hypothesis
  z_alternative <- (effect_size - 0) / standard_error

  # Calculate the power (probability of rejecting the null hypothesis)
  power <- 1 - pnorm(z_alternative - z_critical)

  return(power)
}

effect_size <- 2  # mean difference
sample_size <- 25  # sample size
standard_deviation <- 4  # standard deviation
alpha <- 0.10  # significance level
calculate_z_test_power(effect_size, sample_size, standard_deviation, alpha)
```

```
## [1] 0.1115268
```

```r
print(power)
```

```
## function (lambda = 1)
## {
##     if (!is.numeric(lambda) || is.na(lambda))
##         stop("invalid argument 'lambda'")
##     if (lambda <= 0)
##         return(make.link("log"))
##     if (lambda == 1)
##         return(make.link("identity"))
##     linkfun <- function(mu) mu^lambda
##     linkinv <- function(eta) pmax(eta^(1/lambda), .Machine$double.eps)
##     mu.eta <- function(eta) pmax((1/lambda) * eta^(1/lambda -
##         1), .Machine$double.eps)
##     valideta <- function(eta) all(is.finite(eta)) && all(eta >
##         0)
##     link <- paste0("mu^", round(lambda, 3))
##     structure(list(linkfun = linkfun, linkinv = linkinv, mu.eta = mu.eta,
##         valideta = valideta, name = link), class = "link-glm")
## }
## <bytecode: 0x7fbef2033a38>
## <environment: namespace:stats>
```

c. Draw pictures of the null and alternative distributions of the sampling distribution of the sample mean ($\bar{X}$) and shade and label the areas that correspond to (i) Type 1 error, (ii) Type 2 error, and (iii) Power from part (b). (You can also include a screenshot of the online power applet as long as you describe where the relevant sections are.)

```r
# Generate x values for the x-axis
x <- seq(98, 104, by = 0.01)

# Calculate the normal densities for null and alternative hypotheses
```
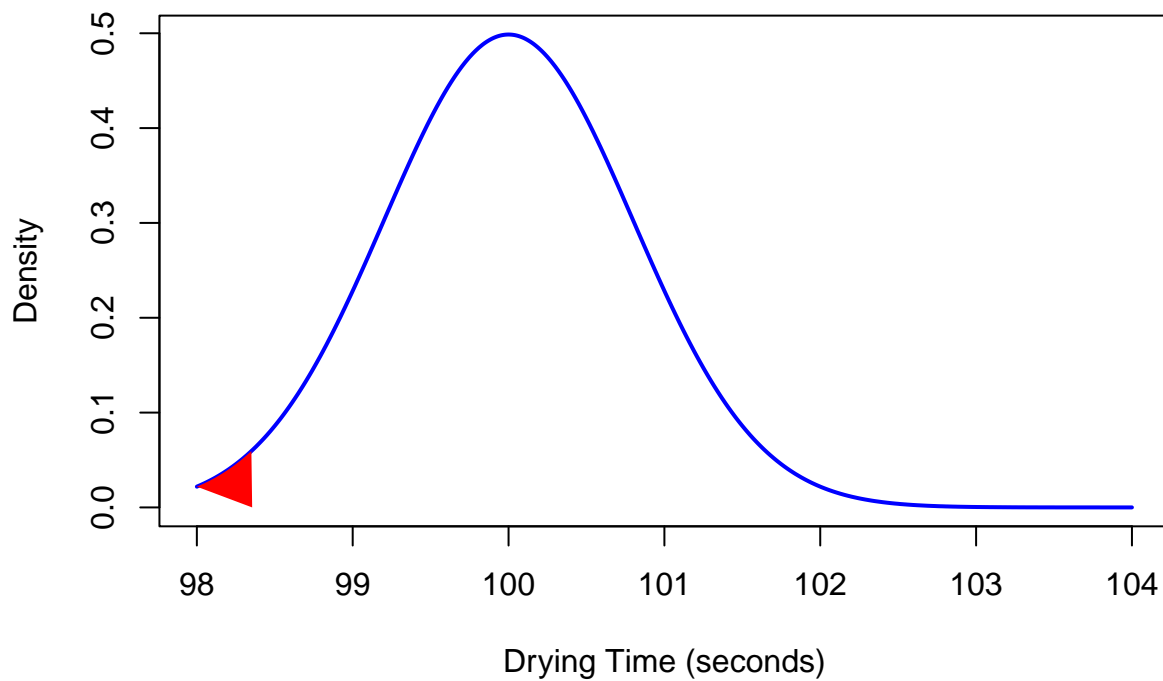
```
null_density <- dnorm(x, mean = 100, sd = 0.8)
alternative_density <- dnorm(x, mean = 102, sd = 0.8)

# Plot the null hypothesis distribution
plot(x, null_density, type = "l", lwd = 2, col = "blue", xlab = "Drying Time (seconds)", ylab = "Density
polygon(c(x[x < 98.355], 98.355), c(null_density[x < 98.355], 0), col = "red", border = NA)  # Type 1 e
```
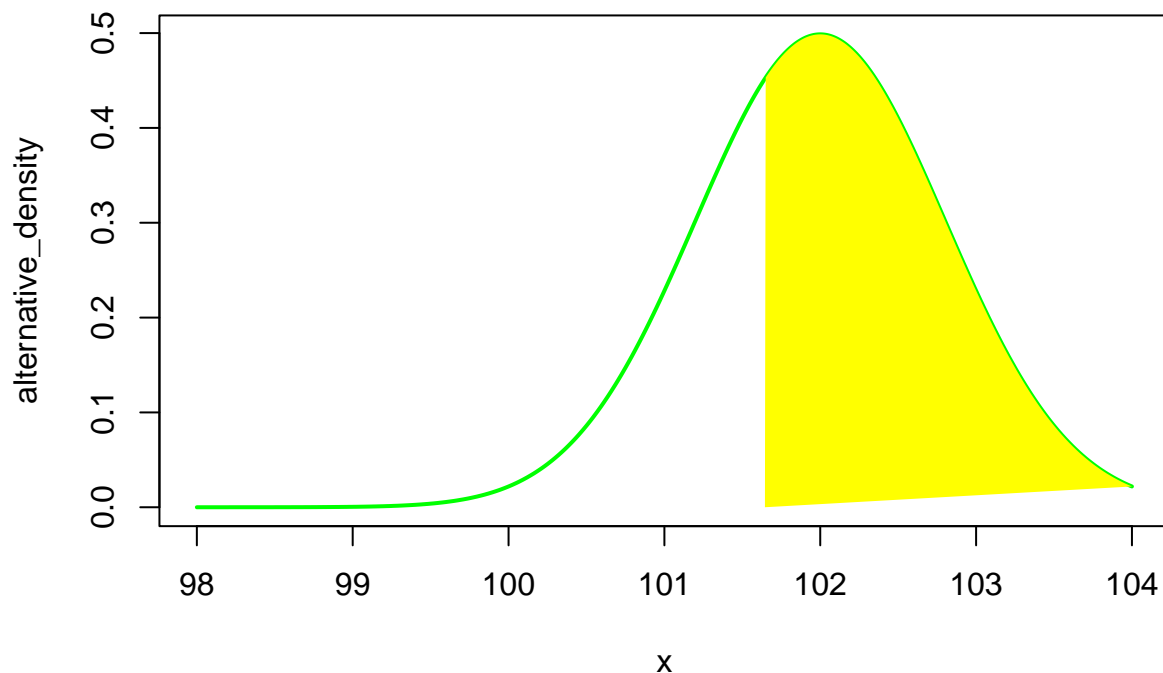


```
# Plot the alternative hypothesis distribution
plot(x, alternative_density, type = "l", lwd = 2, col = "green")
polygon(c(x[x > 101.645], 101.645), c(alternative_density[x > 101.645], 0), col = "yellow", border = NA
polygon(c(x[x > 104], 104), c(alternative_density[x > 104], 0), col = "orange", border = NA)  # Power r
```

d. Describe Type 1 and Type 2 errors of the test in context.

In the spray paint scenario, a Type 1 error would happen if the research division concludes that the new colorant has a significant effect on drying time (rejects H_0) when, in reality, the new colorant has no effect, and the drying time is still 100 seconds (H_0 is true).

In the spray paint scenario, a Type 2 error would happen if the research division fails to reject H_0 (fails to recognize the effect of the new colorant) when, in reality, the new colorant does increase the drying time to 102 seconds (H_a is true).

e. What sample size should the managers use to ensure their 10% level 2-sided Z test has power of at least 0.9 to detect a true mean drying time of 102 ounces (assuming $\sigma = 4$)?

```
((4 * (1.645 + 1.282))/(2))^2
```

```
## [1] 34.26932
```

f. Use the R function power.t.test() to recompute the power and type 2 error rate in (b) and the sample size in (e) assuming a **t test** will be run. Discuss how these values compare to the values computed assuming a Z test statistic and why that makes sense based on what we claim to "know" with each of them.

4

```
# Parameters for the t-test
delta <- 2   # True mean difference (102 - 100)
sigma <- 4   # Population standard deviation
alpha <- 0.10   # Significance level for a two-tailed test
power <- 0.9   # Desired power

# Compute the sample size for the t-test
sample_size <- power.t.test(delta = delta, sd = sigma, sig.level = alpha, power = power, type = "two.sa

# Compute the power and type 2 error rate for the calculated sample size
result <- power.t.test(n = sample_size, delta = delta, sd = sigma, sig.level = alpha, type = "two.sampl

# Print the results
print(paste("Sample Size for t-test:", sample_size))
```

```
## [1] "Sample Size for t-test: 69.197819339444"
```

```
print(paste("Actual Power for t-test:", result$power))
```

```
## [1] "Actual Power for t-test: 0.899999990607993"
```

```
print(paste("Type 2 Error Rate for t-test:", 1 - result$power))
```

```
## [1] "Type 2 Error Rate for t-test: 0.100000009392007"
```

The sample size for the t-test might be slightly larger than for the Z-test. This is because the t-distribution has fatter tails than the normal distribution, making it less efficient for the same sample size. Hence, to achieve the same power, a slightly larger sample size might be needed when using a t-test.

The actual power and type 2 error rate for the t-test could be very close to the specified values (in this case, 0.9). However, they might not be exactly the same due to the discrete nature of sample sizes. The t-test accounts for the variability in the sample mean due to the smaller sample size, which results in a more accurate estimation of power and type 2 error rate compared to the Z-test.

Suppose the managers collect a random sample of 41 drying times from the paint with the new colorant to test the hypotheses: $H_o : \mu = 100$ versus $H_a : \mu \neq 100$ at a $\alpha = 10\%$ significance level. They still believe it is reasonable to assume the distribution of drying times should be approximately normal, but they are not going to assume they know σ. The collected drying times can be found in the "Drytimes.csv" file on Canvas.

     g. Evaluate whether the assumptions of the t test are reasonably met.
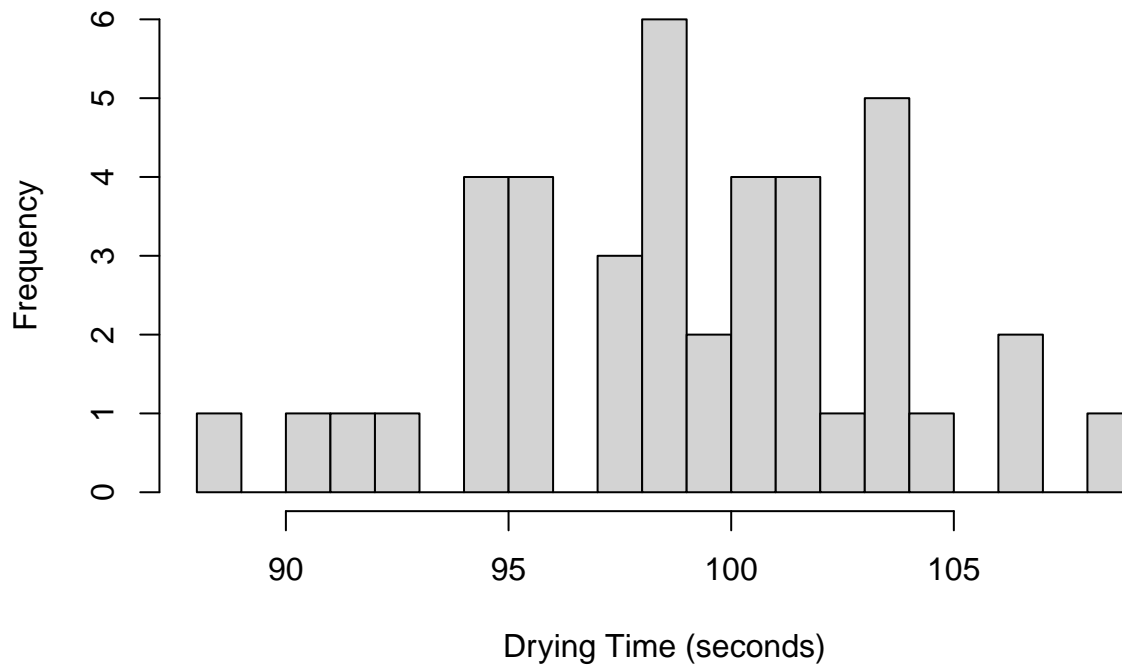
```
# Load the data from the CSV file
drying_times <- read.csv("Drytimes.csv")$seconds

# Plot a histogram to visualize the distribution
hist(drying_times, breaks = 15, main = "Histogram of Drying Times", xlab = "Drying Time (seconds)", yla
```

Histogram of Drying Times


```r
# Perform the Shapiro-Wilk test for normality
shapiro.test(drying_times)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  drying_times
## W = 0.98855, p-value = 0.9486
```

The assumptions of the t-test are reasonably met.

h. Compute a t test statistic and p value and draw a conclusion for the hypothesis test H_o : $\mu = 100$ versus $H_a : \mu \neq 100$ at a 10% level. (I would recommend you do this "by hand" and also with an R function to check your work)

```r
sample_mean = mean(drying_times)
sample_sd = sd(drying_times)
n = length(drying_times)
mu = 100
t = (sample_mean - mu)/(sample_sd / sqrt(n))
df = n - 1
p_value <- 2 * pt(abs(t), df, lower.tail = FALSE)
print(paste("t-statistic:", t))
```

```
## [1] "t-statistic: -1.43374181003101"
```

```
print(paste("p-value:", p_value))
```

```
## [1] "p-value: 0.15941725811784"
```

I do not have enough evidence to reject the null hypothesis that $\mu = 100$.

      i. Perform a bootstrap hypothesis test of $H_o : \mu = 100$ versus $H_a : \mu \neq 100$ at the 10% level.

```
observed_mean <- mean(drying_times)

# Number of bootstrap samples
num_bootstraps <- 10000

# Initialize vector to store bootstrap means
bootstrap_means <- numeric(num_bootstraps)
set.seed(42)  # Set seed for reproducibility
for (i in 1:num_bootstraps) {
  bootstrap_sample <- sample(drying_times - observed_mean, replace = TRUE) + observed_mean
  bootstrap_means[i] <- mean(bootstrap_sample)
}
p_value_bootstrap <- mean(abs(bootstrap_means - 100) > abs(observed_mean - 100))

# Print the p-value
print(paste("Bootstrap p-value:", p_value_bootstrap))
```

```
## [1] "Bootstrap p-value: 0.5008"
```

I do not reject the null hypothesis at this significance level.

      j. Explain why using the t, z, or bootstrap tools are all reasonable in this testing scenario.

You can use all these tests because it is assumed that the distribution of the samples is approximately normal. Additionally, there is a reasonable amount of samples, and the data is sufficiently complex.

Exercise 2 The p-value for a two-sided t test of the hypotheses $H_0 : \mu = 12$, $H_A : \mu \neq 12$ is 0.03.

      a. Does the 99% [two-sided] t confidence interval for μ using this same sample include 12? Why or why not?

If the p-value for the two-sided t-test is 0.03, it means that the observed sample mean is not significantly different from 12 at the 0.01 significance level (since $0.03 > 0.01$). In other words, the null hypothesis is not rejected at the 0.01 significance level. Since it's not rejected, the confidence interval will include 12.

      b. Does the 95% [two-sided] t confidence interval for μ using this same sample include 12? Why or why not?

The null hypothesis is rejected in this instance because $0.03 < 0.05$. Since it's rejected at this level, the confidence interval does not include 12.

Exercise 3 According to the IRS (Internal Revenue Service), the amount of time to fill out an 1040 tax form is about 150 minutes. A researcher believes it actually takes people longer to complete the form and collects the time (in minutes) in takes 25 people to complete the form. The times are listed below in increasing order.

$$16, 24, 25, 48, 68, 91, 112, 121, 145, 150, 172, 181, 194, 201, 205, 221, 242, 250, 252, 315, 365, 382, 422, 511, 614$$

    a. Construct a histogram of the sample data. Explain why the median time to complete the form may be prefered over the mean for a measure of center for all individuals' times to complete the form.

```
times = c(16, 24, 25, 48, 68, 91, 112, 121, 145 ,150, 172, 181, 194, 201, 205, 221, 242, 250, 252, 315,
hist(times)
```

Histogram of times
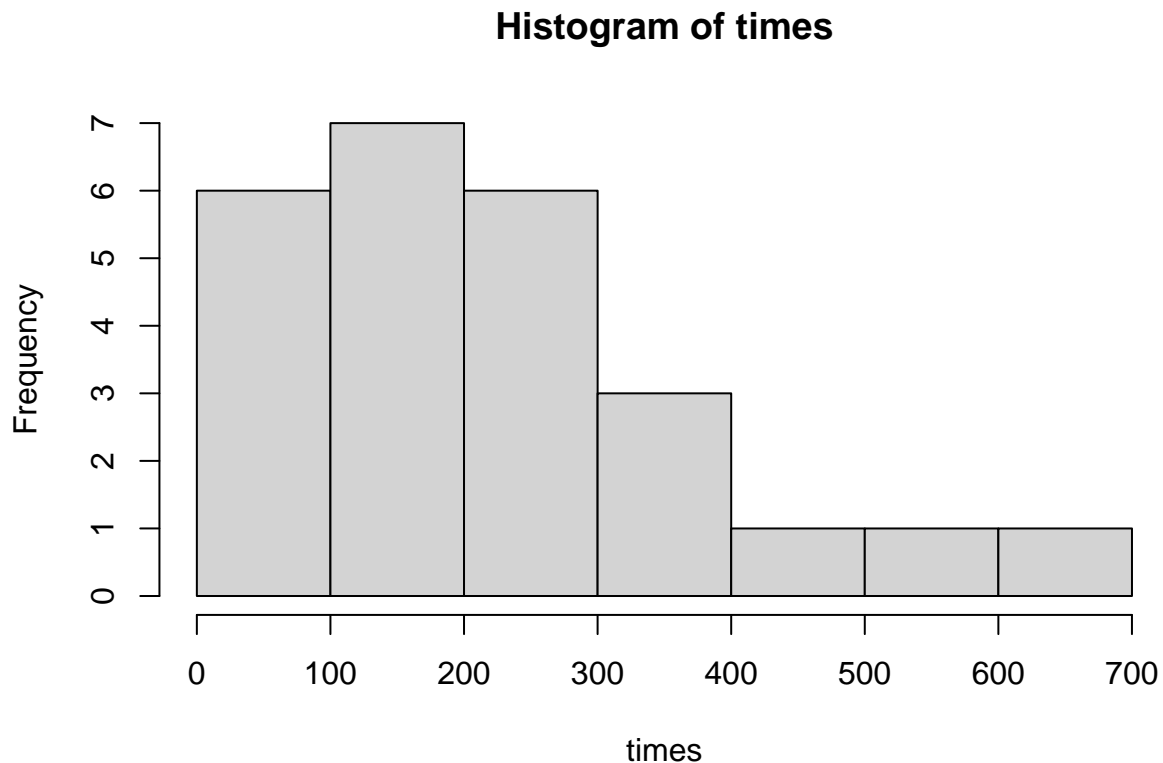
Since there are outliers in this dataset, using the mean might not be the most accurate measure of center because the mean will be bigger than the true center. However, the median will not be affected as much.

b. Let M be the population median time to complete the form. Use the sign test to test:
$H_o : M = 150$ vs $H_A : M > 150$ at $\alpha = 0.05$. Compute the **observed test statistic** and
p-value. Draw a conclusion in the context of the problem.

```r
# Number of observations greater than 150 minutes
obs_greater <- sum(times > 150)
obs_greater
```

```
## [1] 15
```

```r
# Number of observations less than 150 minutes
obs_less <- sum(times < 150)
# Total number of observations
total_obs <- obs_greater + obs_less

# Perform the sign test
d <- binom.test(obs_greater, total_obs, conf.level = 0.95, alternative = "greater")

# Print the p-value
d$statistic
```

```
## number of successes
##                  15
```

```r
d$p.value
```

```
## [1] 0.1537281
```

I fail to reject the null hypothesis.

c. Explain how, if at all, your observed test statistic and p value for the sign test of your
hypotheses would change if the 16 minute form completion time had actually been 5 minutes.

If the 16-minute form completion time had actually been 5 minutes, the observed test statistic would increase
from 9 to 10, and the p-value would likely decrease, providing stronger evidence against the null hypothesis
that the median completion time is 150 minutes.

Exercise 4 An experiment looked at the effectiveness of mushroom compost to counteract petroleum contaminants in soil. The same contaminated soil was divided up into 600 small containers, 200 of each with differing levels of mushroom compost by weight. Six hundred (600) of the same type of seeds were planted into the small containers. The number of seeds that germinated are reported by mushroom compost percentage in the table below. The six hundred containers were housed in the same greenhouse.

Mushroom Compost percentage	3%	4%	5%
Number of seeds that do germinate:	128	136	148
Number of seeds that do not germinate:	72	64	52
Total	200	200	200

a. A seller of Mushroom Compost advertises that "Adding mushroom compost to contaminated soil raises germination rates above 60%!". Use a one-sample large sample Z test for π_{3G}: the proportion of the seeds that germinate in 3% mushroom compost to see if there is significant evidence at the $\alpha = 0.05$ level for the seller's claim at that compost level.

Hypotheses:

$$H_0 : \pi_{3G} \leq 0.60$$

$$H_a : \pi_{3G} > 0.60$$

Assumptions Check:

Do we have SRS of iid observations?

If 600 seeds were randomly assigned to containers with different compost levels, and the data provided includes the number of germinated seeds for each compost level, then it would be an SRS of iid observations.

Do we have sample size large enough for $\hat{P_W} \approx N$?

$$np = 200 \times 0.64 = 128 > 5$$

$$n(1 - p) = 200 \times 0.36 = 72 > 5$$

The sample size is large enough for $\hat{P_W} \approx N$. » *Calculations*:

```
# Number of germinated seeds in 3% mushroom compost
germinated_3percent <- 128

# Total number of seeds in 3% mushroom compost
total_seeds_3percent <- 200

# Null hypothesis proportion
null_prop <- 0.60

# Perform one-sample z-test
result <- prop.test(germinated_3percent, total_seeds_3percent, p = 0.60, conf.level = 0.95)

# Extract observed test statistic and p-value
```

```
observed_statistic <- result$statistic
p_value <- result$p.value

# Print observed test statistic and p-value
print(paste("Observed Test Statistic:", observed_statistic))
```

[1] "Observed Test Statistic: 1.171875"

```
print(paste("P-Value:", p_value))
```

[1] "P-Value: 0.279016313211364"

Conclusion in Context

Since the p-value is greater than 0.05, I do not have enough evidence to reject the null hypothesis.

    b. OPTIONAL REVIEW (Not graded): Compute a [two-sided] 95% CI for π_{3G}: the proportion of the seeds that germinate in 3% mushroom compost. Compare the processes and conclusions to what you did in 4a.