ST 551 Homework 4

Luis Garcia-Lamas

Fall 2022

Instructions

This assignment is due by 11:59 PM, November 7th on Canvas via Gradescope. **You should submit your assignment as a typed PDF which you can compile using the provide .Rmd (R Markdown) template.** Include your code in your solutions and indicate where the solutions for individual problems are located when uploading into Gradescope. You should also use complete, grammatically correct sentences for your solutions.

Question 1 (12 points)

What is the power of a sign test which tests that the population median is zero ($H_0 : m = 0$) if the true population distribution is Normal(1, 1) [Note: This isn't the standard normal as $\mu = 1$ here] and the sample has size $n = 30$? Answer this question initially by using theoretical arguments and the normal cdf. Then, perform a simulation to confirm your answer.

```
p <- 1-pnorm(0, mean = 1, sd = 1)

n <- 30
mu = 1
sd = 1
sample <- rnorm(n, mean = mu, sd = sd)

x <- sum(sample > 0)

binom.test(x = x, n = n, p = .5)
```

```
##
##  Exact binomial test
##
## data:  x and n
## number of successes = 24, number of trials = 30, p-value = 0.001431
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.6143335 0.9228645
## sample estimates:
## probability of success
##                    0.8
```

```
N <- 1000
n <- 30
mu = 1
sd = 1
m <- 0
p_null <- .5

p_values <- rep(0,N)


for (i in 1:N) {

  p <- mean(rnorm(n, mean = mu, sd = sd) > m)
  z <- (p - p_null)/sqrt((p_null*(1-p_null))/n)

  p_values[i] <- 1-pnorm(z)

}

mean(p_values <= .05)
```

```
## [1] 0.997
```

2

Question 2 (12 points)

Perform the following using the `ex0222` data from the `Sleuth3` R package. The data set contains test scores for a sample of 2,584 men and women.

a. (3 points) Make a histogram of the scores on the `Math` component test for all individuals in this sample. Then, answer whether it would be reasonable to assume that the underlying population distribution is symmetric.

```r
# Write the code for your histogram below
library(Sleuth3)

data <- ex0222
math <- data$Math

hist(math)
```

Histogram of math
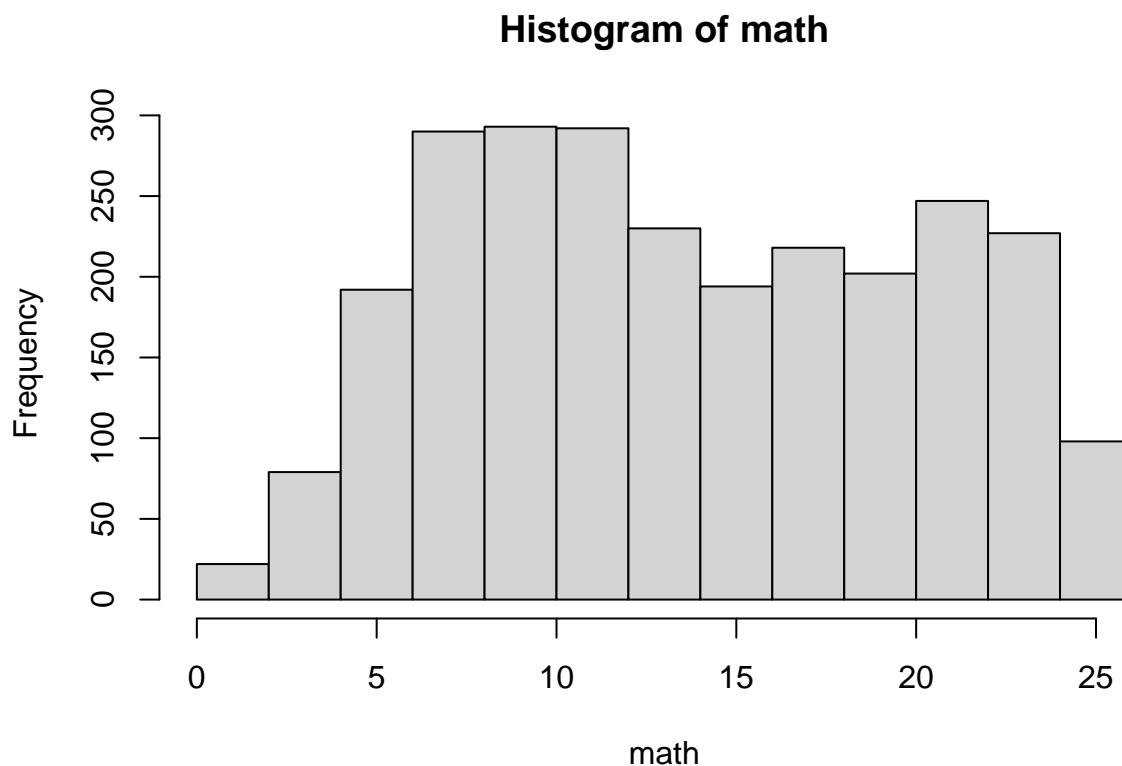

```r
summary(math)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0     9.0    13.0    14.2    20.0    25.0
```

It looks fairly symmetric around the mean of 14.2 even though the 2 lumps are slightly different, one could definetely argue symmetry here.

b. (3 points) Construct a 95% confidence interval for the population mean `Math` score using the t-test (You can do this by hand if you want but personally I'd just use R).

```
t.test(math, conf.level = .95)
```

```
##
##  One Sample t-test
##
## data:  math
## t = 114.83, df = 2583, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  13.95303 14.43784
## sample estimates:
## mean of x
##  14.19543
```

c. (3 points) Construct an approximate 95% confidence interval for the population *center* `Math` score using the Wilcoxon Signed-Rank test. Explain how you determined the range of values that were in the interval or include the code you used (Again, you can do this by hand or use R).

Be aware, rumor has it that the R function `wilcox.test()` has a "bug" in the code which results in incorrect confidence intervals. Not sure if the bug was fixed in recent updates to R, but fear not; so long as you use a reasonable function call you'll get full credit. For those feeling adventurous, take a look at the source code for the `wilcox.test()` function (i.e. run: `getAnywhere("wilcox.test.default")`) and see if you find any errors.

```
wilcox.test(math, conf.int = T, exact = F, correct = F, conf.level = .95)
```

```
##
##  Wilcoxon signed rank test
##
## data:  math
## V = 3329490, p-value < 2.2e-16
## alternative hypothesis: true location is not equal to 0
## 95 percent confidence interval:
##  13.99991 14.49996
## sample estimates:
## (pseudo)median
##       14.00003
```

d. (3 points) Compare the two intervals from parts (b) and (c).

t test: (13.9530, 14.4378) wilcox: (13.9999, 14.5)

They are actually fairly similar to eachother but have different lengths. The wilcox calculated confidence interval length is about .50005, and t test is .48481 in length. The wilcox confidence interval has a center of 14.249935, while the t test has a center (mean) of 14.19543. These confidence intervals are honestly fairly close to eachother overall, with even similar centers!

Question 3 (10 points): Open-ended Question

Explore the performance of the Wilcoxon signed-rank test of data from asymmetric population distributions. This is an open-ended question so what I'd like you to do is: (1) Write a short description of the explorations you performed, (2) Explain how you generated your data and what values you tested (Test several different null hypotheses), and (3) how you evaluated the exactness and power of your tests. Remember to discuss your findings (though don't feel compelled to try and "generalize" your results).

This is likely to be a first attempt at threading writing, code, and outputs together for many students. A few tips:

- Only include relevant code and outputs. Meaning, remove extraneous code from the document and only include outputs if you comment about them in the writing.
- Break up the code into distinct pieces. This way, you can comment in the writing about what it is you've done, or will do, in terms of the code. This will help your written explanations and code "flow" together.
- Use code comments to describe, for your future self, what it is you're doing in the code.

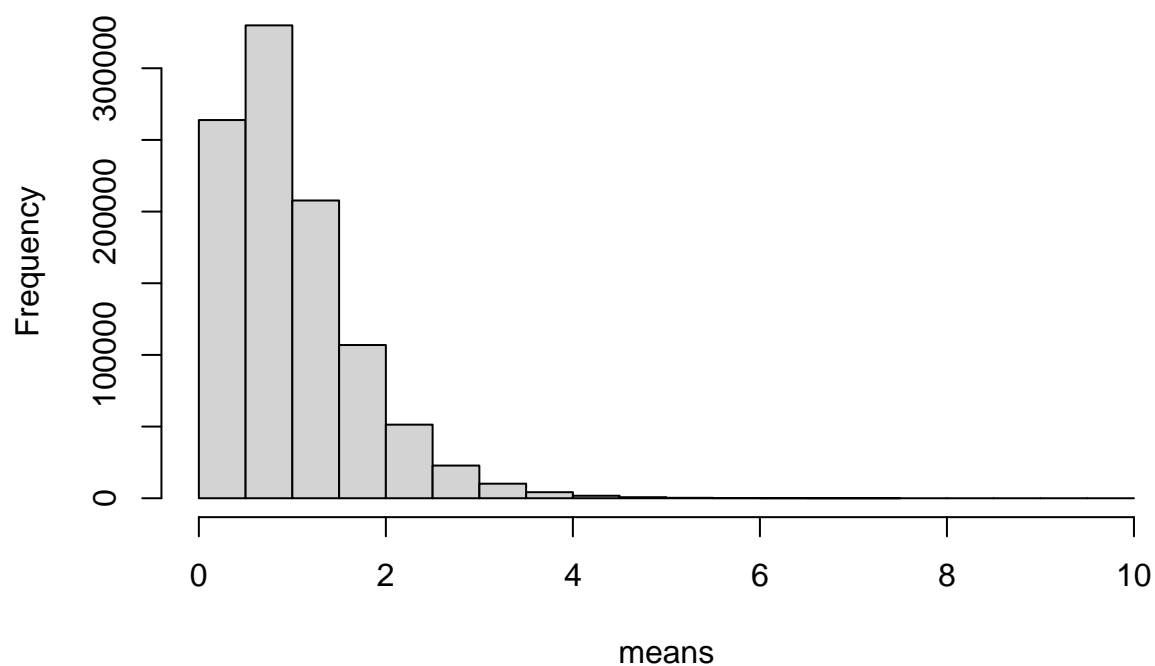Let's try figuring out the psuedomedian for an exponential distribution with with parameter = 1.

```
means <- rep(0,1000000)


for (i in 1:1000000) {
  pair <- rexp(2, rate = 1)
  means[i] <- mean(pair)

}

pseudomedian <- median(means)

hist(means)
```

Histogram of means

Now let's say, for n = 50 we do a wilcox sign test, and simulate for about 10000 times:

```r
N <- 10000
n <- 50
pseudomedian <- median(means)
p_values <- rep(0,N)

for (i in 1:N) {

  sample <- rexp(n, rate = 1)
  test <- wilcox.test(sample, mu = pseudomedian, exact = F, correct = F)
  p_values[i] <- test$p.value

}
sum(p_values < .05)/N
```
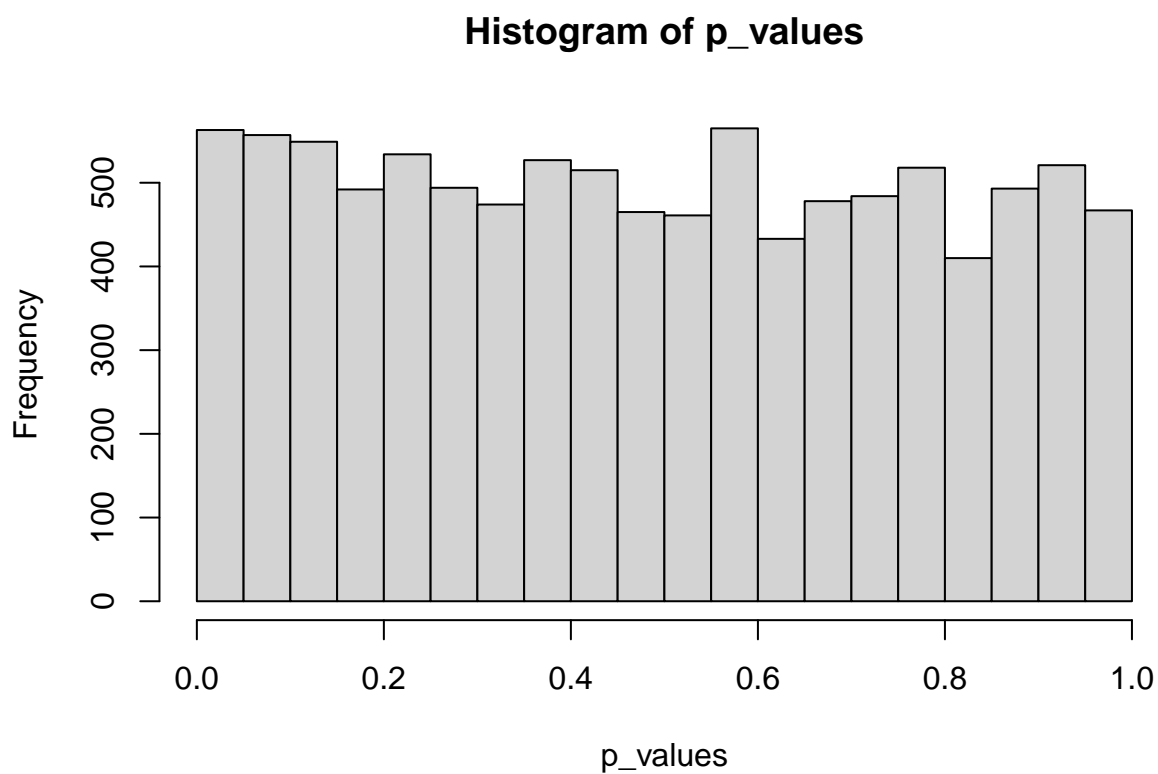
```
## [1] 0.0563
```
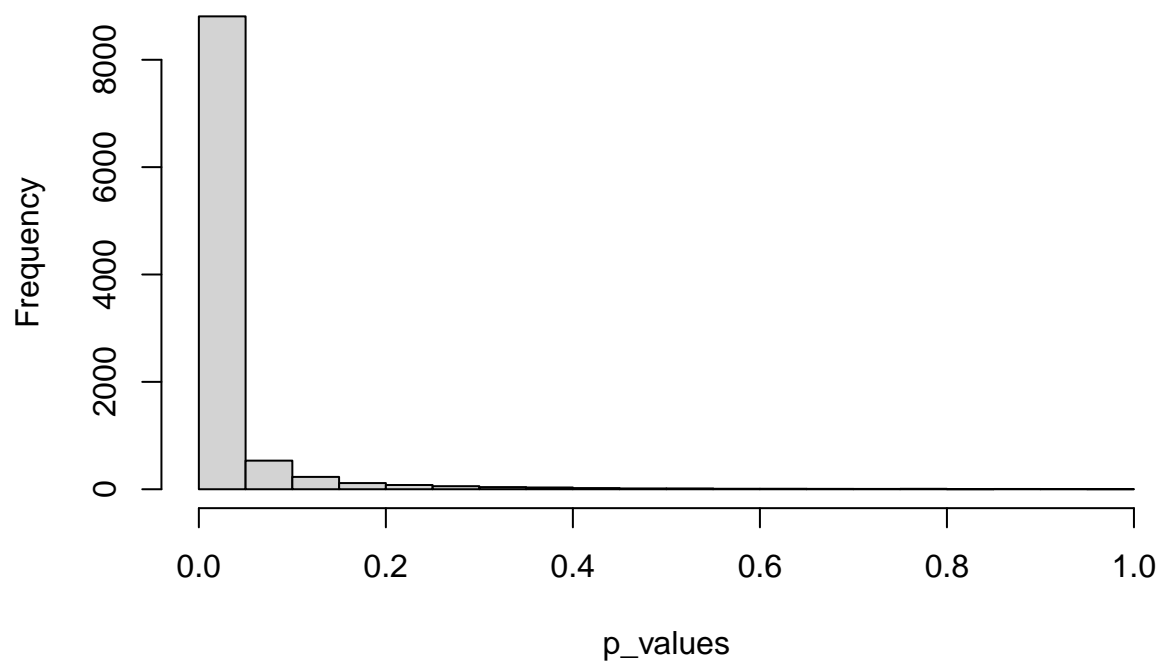
```r
hist(p_values)
```

Histogram of p_values


```r
pseudomedian <- median(means) + .5

for (i in 1:N) {

  sample <- rexp(n, rate = 1)
  test <- wilcox.test(sample, mu = pseudomedian, exact = F, correct = F)
  p_values[i] <- test$p.value

}

sum(p_values < .05)/N
```

```
## [1] 0.8809
```

```r
hist(p_values)
```

Histogram of p_values


```r
pseudomedian <- median(means) - .5


for (i in 1:N) {

  sample <- rexp(n, rate = 1)
  test <- wilcox.test(sample, mu = pseudomedian, exact = F, correct = F)
  p_values[i] <- test$p.value

}

sum(p_values < .05)/N
```

```
## [1] 0.998
```

```r
hist(p_values)
```

Histogram of p_values

My approach to this was to calculate the true pseudomedian in the first part of the code I showed. From there, I wanted to see what type of confidence I could get from estimating at the pseudomedian and slightly away from it. As we can see, the further away we were from the actual psuedomedian, the more confident (ie power approaching 1) we were to reject it. This means the test has been pretty consistent. Also when the true population psuedomedian is used, we can see the reject region is what we would expect it to be which is around .05. If you also look at the distribution of p values, it is fairly uniform. Therefore, also exact.

Question 4 (10 points)

Prove that the chi-squared test is a consistent test of the population variance, no matter the underlying distribution of the data (Hint: We know that s^2 is a consistent estimator for the population variances, so what happens to the test statistics as the sample size n increases to infinity?).

We know that the sample variance s^2 is consistent ie $s^2 \to \sigma^2$ as $n \to \infty$.

Now, considering a one-sided upper test for this case. Our null would be that $H_0 : \sigma^2 = \sigma_0^2$ and assuming that $\sigma^2 > \sigma_0^2$. We would reject the null if:

$$\frac{(n-1)s^2}{\sigma_0^2} > \chi_{(n-1)}^2(1-\alpha)$$

We can approximate the right side because of the fact that a chi squared distribution converges to a normal distribution as our sample size increases. The converged equivalence is:

$$\chi_{(n-1)}^2(1-\alpha) \stackrel{.}{\sim} N((n-1), 2(n-1))$$

Because of this fact, we can approximate the critical value for the chi square distribution using the normal distribution with the specified parameters. By properties of the normal distribution, we know the corresponding critical value for $(1-\alpha)$ is $\mu + \sigma z_{1-\alpha}$. In this case that would look like: $(n-1) + \sqrt{2(n-1)}z_{1-\alpha}$. Hence, using the converged equivalence above:

$$\chi_{(n-1)}^2(1-\alpha) \approx (n-1) + \sqrt{2(n-1)}z_{1-\alpha}$$

Now, going back to the original inequality, let's calculate the probability of it using our new found information.

$$P(\frac{(n-1)s^2}{\sigma_0^2} > \chi_{(n-1)}^2(1-\alpha)) \approx P(\frac{(n-1)s^2}{\sigma_0^2} > (n-1) + \sqrt{2(n-1)}z_{1-\alpha})$$

Multiplying $(n-1)$ on both sides:

$$P(\frac{(n-1)s^2}{\sigma_0^2} > (n-1) + \sqrt{2(n-1)}z_{1-\alpha}) = P(\frac{s^2}{\sigma_0^2} > 1 + \sqrt{\frac{2}{(n-1)}}z_{1-\alpha})$$

Now since we know that the sample variance approaches the true population variance as the sample size increases, we would have $\frac{s^2}{\sigma_0^2} > 1$ and $1 + \sqrt{\frac{2}{(n-1)}}z_{1-\alpha} \to 1$ as $n \to \infty$. This result can be shown similarly, variance being less than null, or not equal to. Clearly this shows that the power is converging to 1. Therefore, consistent.

Question 5 (12 points)

Kolmogorov-Smirnov Test: Explore how the test performs if you estimate the mean and standard deviation from the data, and then test that the distribution is normal. Generate 1,000 datasets, each consisting of $n = 50$ observations from the standard normal distribution.

Part A. (3 Points) Perform the Kolmogorov-Smirnov test of the null hypothesis that the data are normally distributed with the correct mean 0 and standard deviation 1, and evaluate the rejection frequency. How does this test perform? At level $\alpha = 0.05$ does it reject too often? Not often enough?

```
# Run your simulations below:

N <- 1000
n <- 50

p_values <- rep(0,N)

for (i in 1:N) {
    p_values[i] <- ks.test(rnorm(n, 0, 1), pnorm, 0, 1)$p.value

}

sum(p_values < .05)/N
```
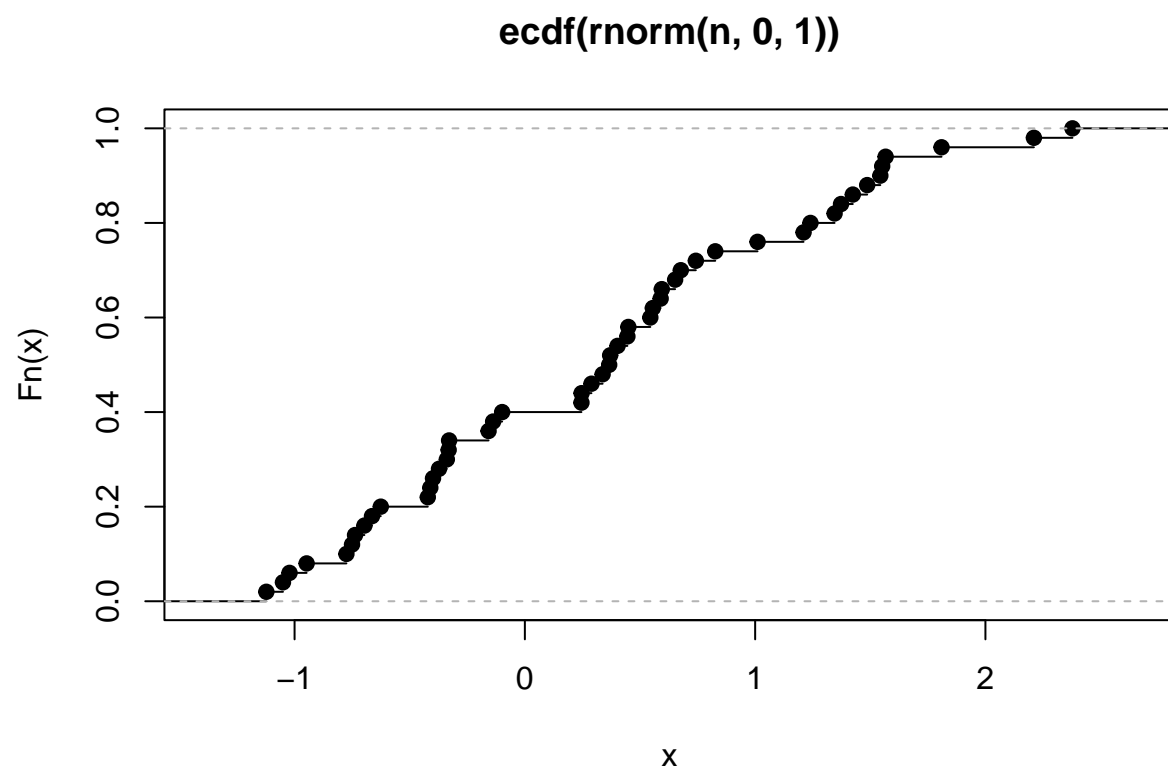
```
## [1] 0.045
```

Useful graphs for me:
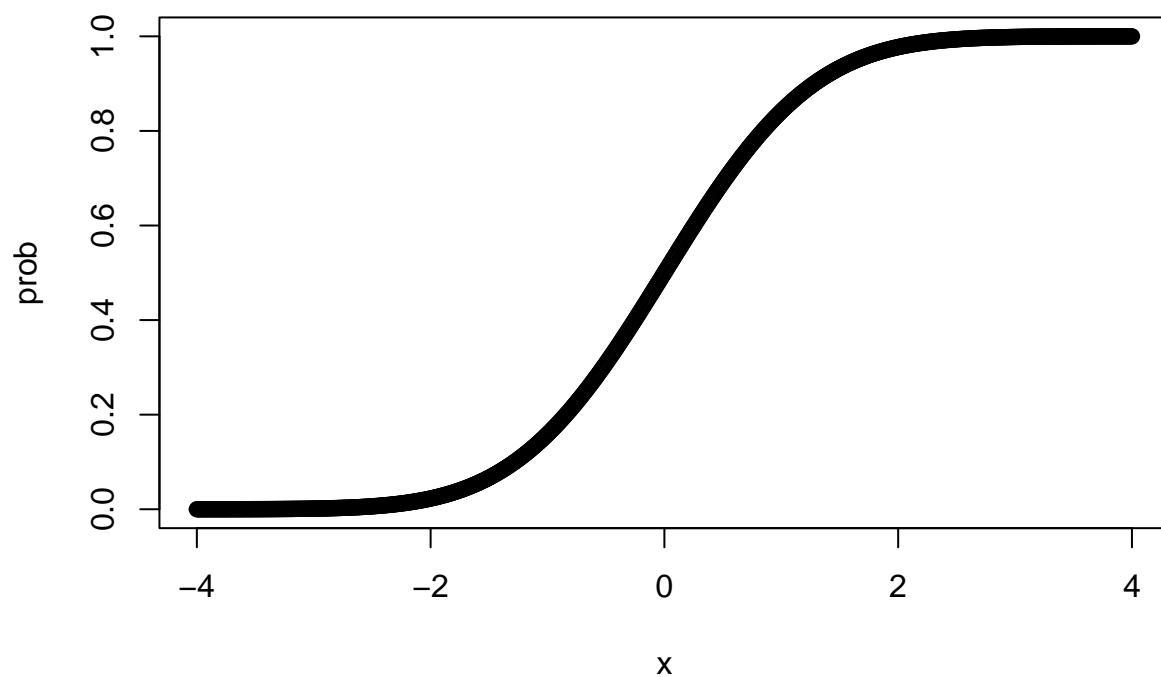
```
plot(ecdf(rnorm(n, 0, 1)))
```

ecdf(rnorm(n, 0, 1))
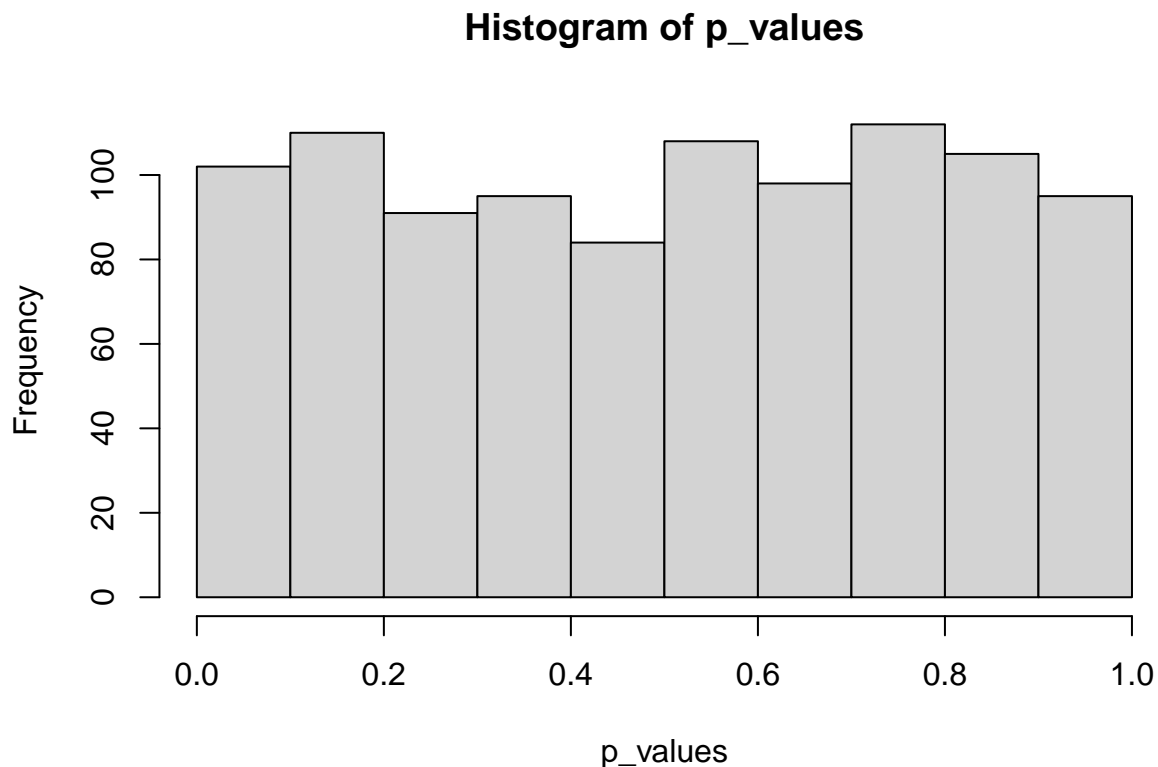


```
x <- seq(-4,4,.01)
prob <- pnorm(x)

plot(x,prob)
```

This test performs well and it's close to the type 1 error we would expect.

Part B. (3 Points) Produce a histogram of the resulting p-values. What should this histogram look like if the test were well-calibrated (exact)? Does it look like that?

```
# Create your histogram below
hist(p_values)
```

Histogram of p_values

It should look perfectly uniform if it was exact.

Part C. (3 Points) Now perform the Kolmogorov-Smirnov test, but this time use the sample mean and the sample standard deviation as the parameters of the null hypothesis distribution (that is, test that the data are normally distributed with mean equal to the sample mean, and standard deviation equal to the sample standard deviation). How does this test perform? At level $\alpha = 0.05$ does it reject too often? Not often enough?

```
# Run your simulations below:
N <- 1000
n <- 50

p_values <- rep(0,N)

for (i in 1:N) {
  sample <- rnorm(n, 0, 1)
  samp_mean <- mean(sample)
  samp_sd <- sd(sample)
  p_values[i] <- ks.test(sample, pnorm, samp_mean, samp_sd)$p.value

}

sum(p_values < .05)/N
```

```
## [1] 0
```

```
sum(p_values > .05)/N
```

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

It seems to almost never catch a significant p value, which makes sense as our predicted means are the actual sample means. A very interseting result.

Part D. (3 Points) Produce a histogram of the resulting p-values. What should this histogram look like if the test were well-calibrated (exact)? Does it look like that?

```
# Create your histogram below
hist(p_values)
```

Histogram of p_values
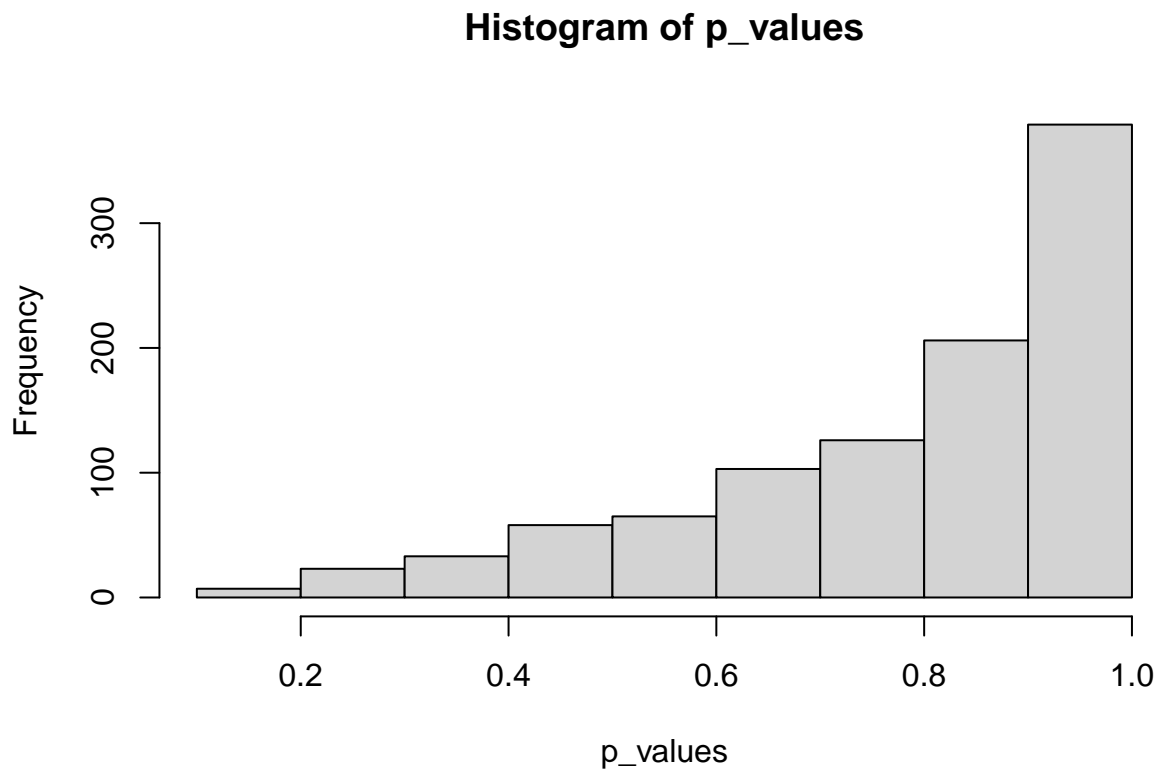
I am honestly not sure, this is an interesting result that I will go into office hours for. It seems that the distribution of p values matters less, but the fact that all of them are > .05 is important. It makes sense by construction as the distribution is always testing its own parameters therefore in theory it should never reject.