

QSCI 482 Story 2

Simon Hans Edasi

2023-10-13

Question 1

a. Read in the data, extract the appropriate column containing the diameters, and make sure that the data you have read in seem reasonable before proceeding.

```
firs <- read.csv('DouglasFirs 2022.csv')
diameters <- firs$diameters
# print(firs)
# plot(firs$treecode, firs$diameters)
# hist(diameters)
# plot(density(diameters))
```

b. Find the sample mean, sample standard deviation, and sample size.

```
dia_mean <- signif(mean(diameters),3)
dia_std <- signif(sd(diameters), 3)
dia_n <- signif(length(diameters), 3)
library(glue)
glue('Mean: {dia_mean}; STD: {dia_std}; Sample Size: {dia_n}')
```

```
## Mean: 95.9; STD: 26.1; Sample Size: 65
```

c. Assuming that the known population SD applies to the sample of trees, calculate the standard error, the 90% confidence interval, 95% confidence interval, and 99% confidence interval for the sample.

```
dia_se <- dia_std / sqrt(dia_n)

Z <- qnorm(0.95)
cil90 <- signif(dia_mean - Z*dia_se, 3)
ciu90 <- signif(dia_mean + Z*dia_se, 3)

Z <- qnorm(0.975)
print(Z)
```

```
## [1] 1.959964
```

```
cil95 <- signif(dia_mean - Z*dia_se, 3)
ciu95 <- signif(dia_mean + Z*dia_se, 3)

Z <- qnorm(0.995)
cil99 <- signif(dia_mean - Z*dia_se, 3)
```

```

ciu99 <- signif(dia_mean + Z*dia_se, 3)
glue('90% CI [{cil90},{ciu90}]; 95% CI [{cil95},{ciu95}]; 99% CI [{cil99},{ciu99}]')

## 90% CI [90.6,101]; 95% CI [89.6,102]; 99% CI [87.6,104]

mean <- 27
samp_std <- 11
n <- 22

se <- samp_std / sqrt(n)
t <- qt(0.95,df = 21)

cil95 <- (mean - t) / se
ciu95 <- (mean + t) / se
print(cil95)

## [1] 10.77911
print(ciu95)

## [1] 12.24657
print(ciu95 - cil95)

## [1] 1.467454

```

d. Conduct a hypothesis test that calculates the probability that the sample comes from a population of 100-year-old trees. Specify your null and alternative hypotheses, outline the test, note whether the test is one-tailed or two-tailed, calculate the test statistic, and report the resulting p-value.

We want to test if our trees are 100 years old which is determined by a diameter of 88.5 cm.

Our null hypothesis is $H_0 : \bar{X} = 88.5$

Our alternative hypothesis is $H_A : \bar{X} \neq 88.5$, which will involve a two-tailed test where because we will reject the null if the mean is higher or lower than our population. We don't know the population standard deviation so we will need to use a t-test instead of a Z-test.

Sticking with a standard confidence of 95%, choose $\alpha = 0.05$. Plot the distribution with rejection regions

```

alpha = 0.05
# define rejection regions
rej_left <- qnorm(alpha/2, 0,1)
rej_right <- qnorm(alpha/2, 0,1, lower.tail = FALSE)
print(rej_left)

## [1] -1.959964

#plot the curve with rejection regions
x <- seq(-4,4,length.out = length(diameters))
plot(
  x,
  y = dnorm(x),
  type = 'l')

## x and y for the whole area
xReject <- c(seq(-4,4,by=0.01))
yReject <- dnorm(xReject, 0, 1)

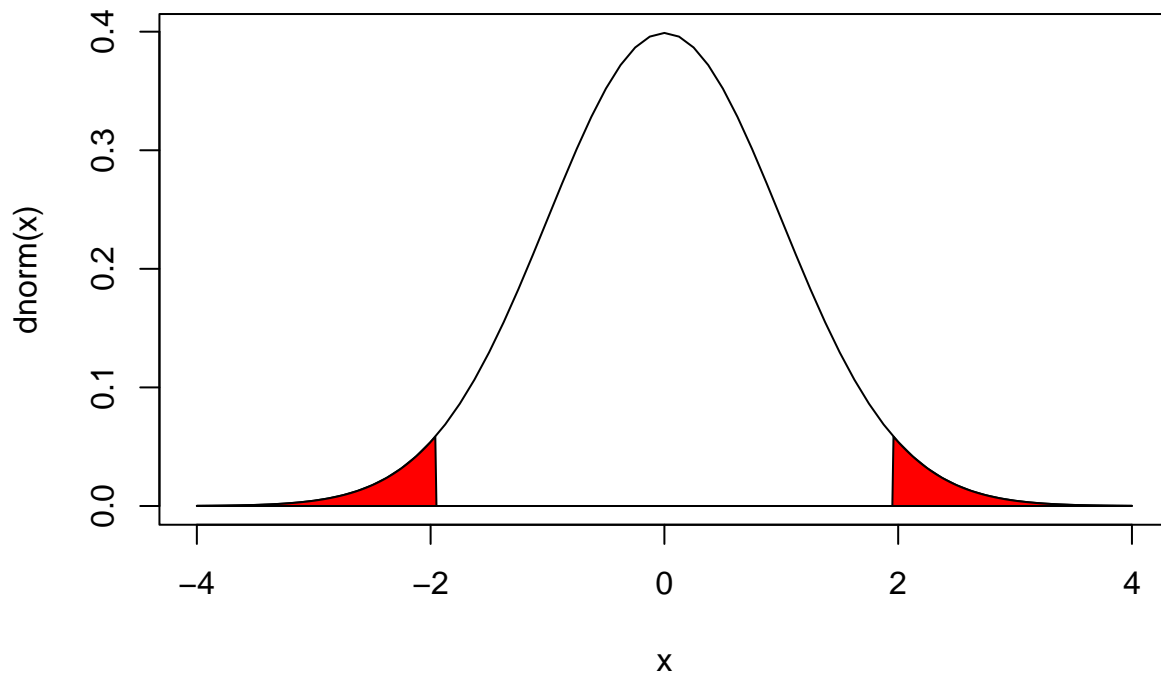
```

```

yReject[xReject > rej_left & xReject < rej_right] <- 0

polygon(c(xReject,xReject[length(xReject)]),xReject[1]),
       c(yReject,0, 0), col='red')

```



Now we want to test if our test statistic falls within our rejection region.

```

mu <- 88.5                                     # null hyp pop mean
dia_se <- dia_std / sqrt(dia_n)                 # standard error
dia_Z <- (dia_mean - mu) / dia_se               # Z statistic

#plot the curve with rejection regions
x <- seq(-4,4,length.out = length(diameters))
plot(
  x,
  y = dnorm(x,0,1),
  type = 'l')
## x and y for the whole area
xReject <- c(seq(-4,4,by=0.01))
yReject <- dnorm(xReject,0,1)

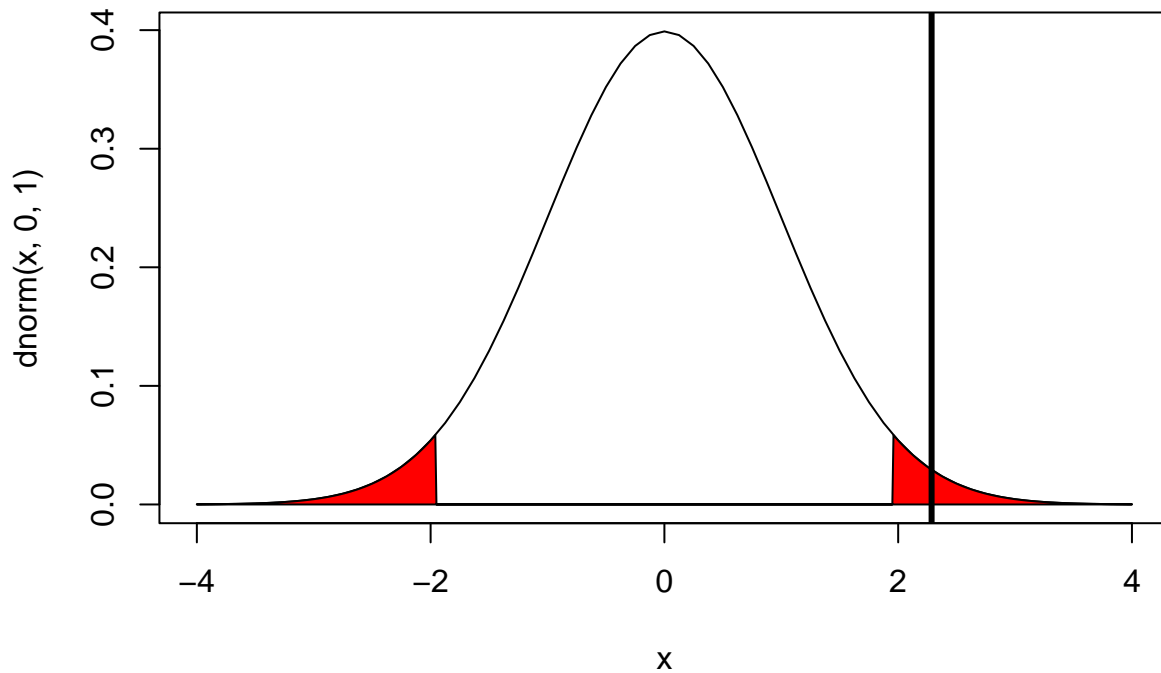
yReject[xReject > rej_left & xReject < rej_right] <- 0

polygon(c(xReject,xReject[length(xReject)]),xReject[1]),
       c(yReject,0, 0), col='red')

# plot test statistic

```

```
abline(v = dia_Z, col = 'black', lwd = 3)
```



What is our p-value? Area of the rejection regions.

```
p_left <- pnorm(dia_Z, 0, 1, lower.tail = FALSE)      # left rejection region
p <- 2 * p_left                                     # p value for a two tail test
print(p)
```

```
## [1] 0.02226299
```

e. If the significance level, α , is 0.05, would you reject the null hypothesis? What if the chosen significance level is 0.005? Given the test results, are the trees on the property owner's land about 100 years old, or are they significantly older or younger?

With $\alpha = 0.05$ we should reject the null hypothesis. For a significance level of 0.005 we can easily repeat the analysis and see that the test statistic falls outside of the rejection region.

```
alpha = 0.005
# define rejection regions
rej_left <- qnorm(alpha/2, 0, 1)
print(rej_left)
```

```
## [1] -2.807034
```

```
rej_right <- qnorm(alpha/2, 0, 1, lower.tail = FALSE)
mu <- 88.5                                     # null hyp pop mean
dia_se <- dia_std / sqrt(dia_n)                # standard error
```

```

dia_Z <- (dia_mean - mu) / dia_se      # Z statistic

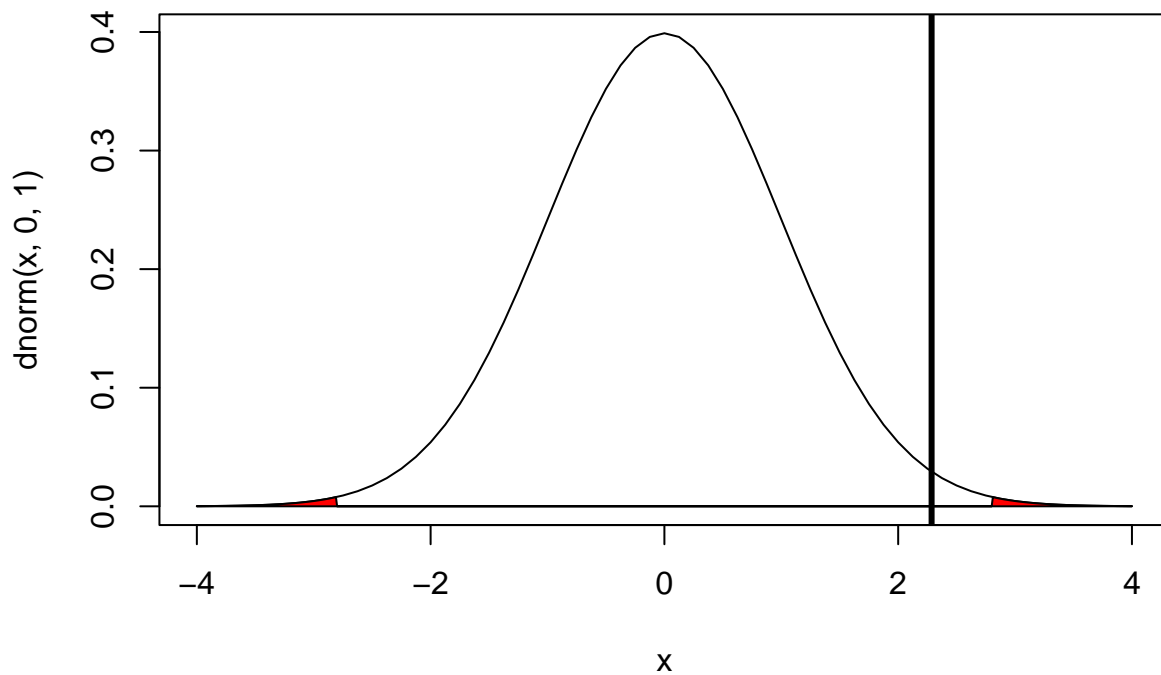
#plot the curve with rejection regions
x <- seq(-4,4,length.out = length(diameters))
plot(
  x,
  y = dnorm(x,0,1),
  type = 'l')
## x and y for the whole area
xReject <- c(seq(-4,4,by=0.01))
yReject <- dnorm(xReject,0,1)

yReject[xReject > rej_left & xReject < rej_right] <- 0

polygon(c(xReject,xReject[length(xReject)],xReject[1]),
       c(yReject,0, 0), col='red')

# plot test statistic
abline(v = dia_Z, col = 'black', lwd = 3)

```



```

# p_left <- pt(-dia_t, df = dia_dof)      # left rejection region
p_left <- pnorm(dia_Z, 0, 1, lower.tail = FALSE )
p <- 2 * p_left                          # p value for a two tail test
print(signif(p,2))

```

```
## [1] 0.022
```

Given these results, I would conclude the trees are significantly older than 100 years. We have a positive Z value so we know our mean is higher than the population mean.

With a lower p-value we end up not rejecting the null hypothesis, however, I think this is incorrect to do. By increasing alpha we crank up the power of our test, which I interpret to mean we are looking for bigger differences between sample and population. Looking for bigger differences means our difference of mean will become insignificant and we will commit a type II error by not rejecting the null when we should.