QSCI 482 Story 5

Simon Hans Edasi

2023-11-06

```r
# quick function to make plotting faster
t_plot <- function(alpha=0.05, dof = 5,t_obs = 0){
# define rejection regions
rej_left <- qt(alpha/2,df = dof)
rej_right <- qt(alpha/2,df = dof,lower.tail = F)
#plot the curve with rejection regions
x <- seq(-4,4,length.out = 100)
plot(
x,
y = dt(x,df = dof),
type = 'l')
## x and y for the whole area
xReject <- c(seq(-4,4,by=0.01))
yReject <- dt(xReject,df = dof)
yReject[xReject > rej_left & xReject < rej_right] <- 0
polygon(c(xReject,xReject[length(xReject)],xReject[1]),
c(yReject,0, 0), col='red')
abline(v = t_obs)
}

library(glue)
```

Load in the data

```r
df <- read.csv('bestroute.csv')
hmin <- df$HighwayMin
hsec <- df$HighwaySec / 60
h <- hmin + hsec

bmin <- df$BackRouteMin
bsec <- df$BackRouteSec / 60
b <- bmin + bsec
```

Question 1

a. Conduct a paired-sample parametric test the long way around to determine whether the highway time is significantly different from the back route time
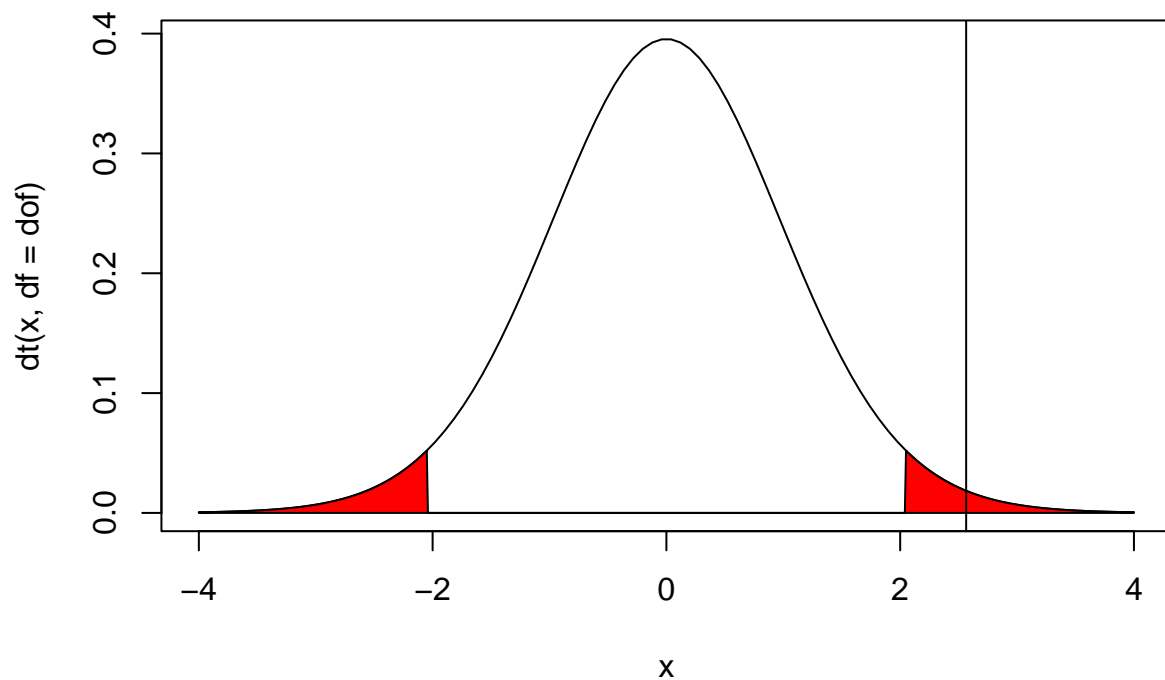
$H_0 : \mu_h = \mu_b$

$H_0 : \mu_h \neq \mu_b$

```r
alpha <- 0.05
n <- length(h)
```

```
d <- h-b
dbar <- mean(d)
s_d <- var(d)
s_dbar <- sqrt(s_d / n)
dof <- n - 1
t <- dbar / s_dbar
t_crit <- abs(qt(alpha/2, df = dof, lower.tail = F))
p_value <- 2*pt(t,df = dof, lower.tail = F)
t_plot(alpha, dof = dof, t_obs = t)
```



```
glue('degrees of freedom: {dof}\nt_obs: {t}\nt_crit: {t_crit}\np-value: {p_value}')
```

```
## degrees of freedom: 29
## t_obs: 2.56449856760162
## t_crit: 2.0452296421327
## p-value: 0.0157743290340233
```

We see significantly higher times on the highway than the backroads, leading us to conclude the backroute is faster.

b. Conduct the test the short way around. Your answers should match.

```
t.test(h,b,paired = T)
```

```
##
##  Paired t-test
##
## data:  h and b
## t = 2.5645, df = 29, p-value = 0.01577
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  0.173236 1.537875
## sample estimates:
## mean difference
##       0.8555556
```

Yay, same values!

Question 2

a. Conduct a Wilcoxon paired-sample test (this is non-parametric), the long way around, to determine whether the highway time is significantly difference from the back route time. Report the test statistic, p-value etc.

b. Conduct the same test but obtain the p-value using the normal approximation (use twice the area for a two-tailed test).

c. Conduct a Wilcoxon paired-sample test the short way around in R. If this differs from 2a) or 2b), explain why.

Question 3

a. Test whether the differences in time are normally distributed using qqplot, histogram, and the Shapiro-Wilk test. What do you conclude about normality?

b. Given the answer in 3a), which of the tests used in Q1 or Q2 is appropriate, and what do you conclude about travel time on the two routes? Was your roommate correct or were you correct?

Question 4

```
k <- c('banana', 'vinegar','honey','sugarwater')
Nobserved <- c(23, 18, 5, 10)
n <- sum(Nobserved)
K <- length(k)
musn <- 3000
preds <- vector()
for (i in k)
  preds[i] <- (1/K) * musn
Npredicted <- (preds)

fobs <- Nobserved / n
fpred <- Npredicted / musn

cbind(fobs, fpred)
```

```
##                  fobs fpred
## banana      0.41071429  0.25
## vinegar     0.32142857  0.25
## honey       0.08928571  0.25
## sugarwater  0.17857143  0.25
```

```
chisq <- sum(  (fobs - fpred)^2 / fpred  )
p <- pchisq(abs(chisq), df = K-1, lower.tail = F)

glue('\nChiSQ goodness of fit\nchisq: {chisq}\np_value: {p}')
```

```
## ChiSQ goodness of fit
## chisq: 0.247448979591837
## p_value: 0.969588605146169
```

```
# chisq.test(fobs,fpred)
```