

# Home Work 4 — STA104

Filip Wilhelm Sjostrand

2022-12-08

## P1: Exercises 1 141/142

```
# Packages -----
library(dplyr)
library(ggplot2)

# Data -----
treat1 <- c(250, 50, 80, 55, 188)
treat2 <- c(240, 48, 72, 47, 230)
diff <- treat1-treat2
absidff <- abs(diff)

# Permutation
permutation <- c(1:32)

pair1 <-
  c(rep(absidff[1],32)) *
  c(1,1,1,1,1, -1, 1,1,1, -1, 1,1, -1, 1, -1,-1, 1,1,1,1, -1,-1,-1,-1,-1,-1, 1, -1,-1,-1,-1,-1)

pair2 <-
  c(rep(absidff[2],32)) *
  c(1,1,1,1, -1, 1,1,1, -1, 1,1, -1, 1, -1, 1, -1, 1, -1,-1,-1, 1,1,1, -1,-1,-1,-1, 1, -1,-1,-1,-1 )

pair3 <-
  c(rep(absidff[3],32)) *
  c(1,1,1, -1, 1,1,1, -1, 1,1, -1, 1,1, -1,-1, 1, -1, 1, -1,-1, 1, -1,-1, 1,1, -1,-1,-1, 1, -1,-1, -1)

pair4 <-
  c(rep(absidff[4],32)) *
  c(1,1, -1, 1,1,1, -1, 1,1,1, -1,-1,-1, 1,1,1, -1,-1, 1, -1,-1, 1, -1, 1, -1, 1,1 -1,-1,-1, 1, -1,-1)

pair5 <-
  c(rep(absidff[5],32)) *
  c(1, -1, 1,1,1,1, -1,-1,-1,-1, 1,1,1,1,1,1, -1,-1,-1, 1, -1,-1, 1, -1, 1,1, -1,-1,-1,-1, 1, -1)

# Table
df <- data.frame(permutation, pair1, pair2, pair3, pair4, pair5)

df <- df %>%
  mutate(mean=(pair1+pair2+pair3+pair4+pair5)/5)
```

permutation	pair1	pair2	pair3	pair4	pair5	mean
1	10	2	8	8	42	14.0
2	10	2	8	8	-42	-2.8
3	10	2	8	-8	42	10.8
4	10	2	-8	8	42	10.8
5	10	-2	8	8	42	13.2
6	-10	2	8	8	42	10.0
7	10	2	8	-8	-42	-6.0
8	10	2	-8	8	-42	-6.0
9	10	-2	8	8	-42	-3.6
10	-10	2	8	8	-42	-6.8
11	10	2	-8	-8	42	7.6
12	10	-2	8	-8	42	10.0
13	-10	2	8	-8	42	6.8
14	10	-2	-8	8	42	10.0
15	-10	2	-8	8	42	6.8
16	-10	-2	8	8	42	9.2
17	10	2	-8	-8	-42	-9.2
18	10	-2	8	-8	-42	-6.8
19	10	-2	-8	8	-42	-6.8
20	10	-2	-8	-8	42	6.8
21	-10	2	8	-8	-42	-10.0
22	-10	2	-8	8	-42	-10.0
23	-10	2	-8	-8	42	3.6
24	-10	-2	8	8	-42	-7.6
25	-10	-2	8	-8	42	6.0
26	-10	-2	-8	8	42	6.0
27	10	-2	-8	0	-42	-8.4
28	-10	2	-8	-8	-42	-13.2
29	-10	-2	8	-8	-42	-10.8
30	-10	-2	-8	8	-42	-10.8
31	-10	-2	-8	-8	42	2.8
32	-10	-2	-8	-8	-42	-14.0

permutation	pair1	pair2	pair3	pair4	pair5	SR+
1	4	1	2.5	2.5	5	3.0
2	4	1	2.5	2.5	-5	1.0
3	4	1	2.5	-2.5	5	2.0
4	4	1	-2.5	2.5	5	2.0
5	4	-1	2.5	2.5	5	2.6
6	-4	1	2.5	2.5	5	1.4
7	4	1	2.5	-2.5	-5	0.0
8	4	1	-2.5	2.5	-5	0.0
9	4	-1	2.5	2.5	-5	0.6
10	-4	1	2.5	2.5	-5	-0.6
11	4	1	-2.5	-2.5	5	1.0
12	4	-1	2.5	-2.5	5	1.6
13	-4	1	2.5	-2.5	5	0.4
14	4	-1	-2.5	2.5	5	1.6
15	-4	1	-2.5	2.5	5	0.4
16	-4	-1	2.5	2.5	5	1.0
17	4	1	-2.5	-2.5	-5	-1.0
18	4	-1	2.5	-2.5	-5	-0.4
19	4	-1	-2.5	2.5	-5	-0.4
20	4	-1	-2.5	-2.5	5	0.6
21	-4	1	2.5	-2.5	-5	-1.6
22	-4	1	-2.5	2.5	-5	-1.6
23	-4	1	-2.5	-2.5	5	-0.6
24	-4	-1	2.5	2.5	-5	-1.0
25	-4	-1	2.5	-2.5	5	0.0
26	-4	-1	-2.5	2.5	5	0.0
27	4	-1	-2.5	0.0	-5	-0.9
28	-4	1	-2.5	-2.5	-5	-2.6
29	-4	-1	2.5	-2.5	-5	-2.0
30	-4	-1	-2.5	2.5	-5	-2.0
31	-4	-1	-2.5	-2.5	5	-1.0
32	-4	-1	-2.5	-2.5	-5	-3.0

## P2: Exercise 4 on page 142

```
# Data -----
df <- data.frame(
  subject=c(1,2,3,4,5,6,7,8),
  before=c(89, 90, 87, 98, 120, 85, 97, 110),
  after=c(76, 101, 84, 86, 105, 84, 93, 115)
)

# Wilcoxon signed-rank test -----
pvalW <- wilcox.test(df$before, df$after, alternative='two.sided', paired=TRUE)$p.value

# Normal approximation -----
n <- nrow(df)
diff <- df$before-df$after
abs_diff_rank <- rank(abs(diff))
signed <- c()
for(i in 1:n){
  if(diff[i]<0){
    signed[i] <- abs_diff_rank[i]*(-1)
  }
  else{
    signed[i] <- abs_diff_rank[i]
  }
}

SR <- sum(signed[signed>0])
E_SR <- (n*(n+1))/4
VAR_SR <- (n*(n+1)*(2*n+1))/24
Z <- (SR - E_SR)/sqrt(VAR_SR)

pvalZ <- 2*pnorm(Z, lower.tail = FALSE)
```

The p-value of the Wilcoxon Signed-rank test is 0.25 and from the normal approximation it is 0.21. In either case, there is an insignificant difference between LDH before and after fasting.

### P3: Exercises 1 and 2 on page 189

```
# Data -----
height <- c(57, 65, 70, 78)
weight <- c(120, 145, 153, 162)
model_obs <- lm(weight~height)
slope_obs <- model_obs$coefficients[2]
spear_obs <- cor(height, weight, method="spearman")
tau_obs <- cor(height, weight, method="kendall")

# Permutation
perm_y <- combinat::permn(c(120, 145, 153, 162))
perm_y_df <- as.data.frame(perm_y)
colnames(perm_y_df) <- c(1:24)
perm_y_df <- as.data.frame(t(perm_y_df))
colnames(perm_y_df) <- c("y1", "y2", "y3", "y4")

slopes <- c()
spearman <- c()
tau <- c()

for(i in 1:length(perm_y)){
  slopes[i] <- lm(unlist(perm_y[i])~height)$coefficients[2]
  spearman[i] <- cor(height, unlist(perm_y[i]), method="spearman")
  tau[i] <- cor(height, unlist(perm_y[i]), method="kendall")
}

perm_y_df <- perm_y_df %>%
  cbind(slopes, spearman, tau) %>%
  arrange(desc(slopes))

row.names(perm_y_df)[row.names(perm_y_df)==1] <- "Observed"
```

Table 3: x1=57, x2=65, x3=70, x4=78

	y1	y2	y3	y4	slopes	spearman	tau
Observed	120	145	153	162	1.9785408	1.0	1.0000000
8	120	153	145	162	1.8068670	0.8	0.6666667
2	120	145	162	153	1.6695279	0.8	0.6666667
3	120	162	145	153	1.3047210	0.4	0.3333333
7	120	153	162	145	1.2231760	0.4	0.3333333
24	145	120	153	162	1.1201717	0.8	0.6666667
6	120	162	153	145	1.0300429	0.2	0.0000000
23	145	120	162	153	0.8111588	0.6	0.3333333
9	153	120	145	162	0.6738197	0.4	0.3333333
17	145	153	120	162	0.4120172	0.4	0.3333333
16	153	145	120	162	0.1373391	0.2	0.0000000
10	153	120	162	145	0.0901288	0.0	0.0000000
22	145	162	120	153	-0.0901288	0.0	0.0000000
4	162	120	145	153	-0.1373391	-0.2	0.0000000
5	162	120	153	145	-0.4120172	-0.4	-0.3333333
21	162	145	120	153	-0.6738197	-0.4	-0.3333333
11	153	162	120	145	-0.8111588	-0.6	-0.3333333
18	145	153	162	120	-1.0300429	-0.2	0.0000000
12	162	153	120	145	-1.1201717	-0.8	-0.6666667
19	145	162	153	120	-1.2231760	-0.4	-0.3333333
15	153	145	162	120	-1.3047210	-0.4	-0.3333333
14	153	162	145	120	-1.6695279	-0.8	-0.6666667
20	162	145	153	120	-1.8068670	-0.8	-0.6666667
13	162	153	145	120	-1.9785408	-1.0	-1.0000000

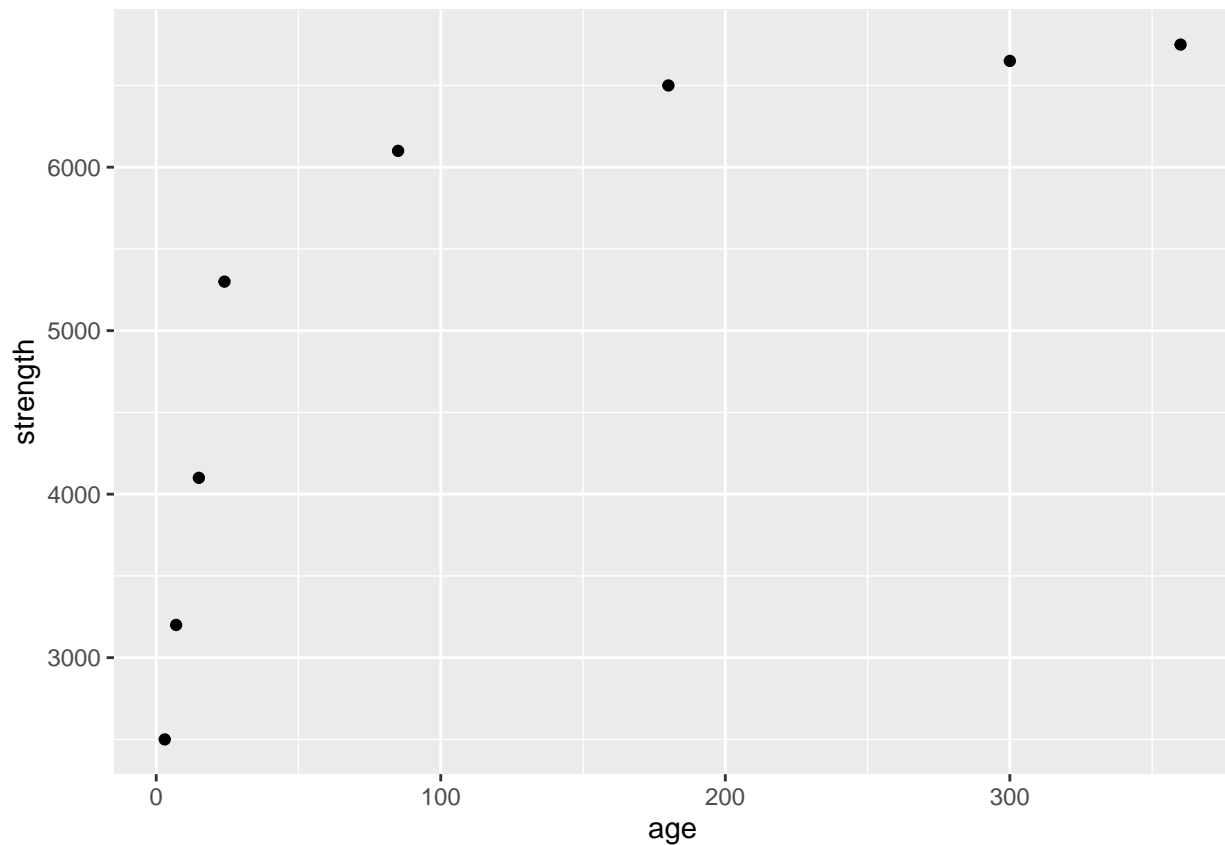
## P4: Exercise 3 on page 189

a)

```
# Data -----
age <- c(3, 7, 15, 24, 85, 180, 300, 360)
strength <- c(2500, 3200, 4100, 5300, 6100, 6500, 6650, 6750)
df <- data.frame(age, strength)

# Associations -----
pearson <- cor(age, strength, method="pearson")
spearman <- cor(age, strength, method="spearman")
tau <- cor(age, strength, method="kendall")
table <- data.frame(pearson, spearman, tau)

# Visualize
df %>%
  ggplot(aes(age, strength)) +
  geom_point()
```



pearson	spearman	tau
0.7999108	1	1

b)

```
# Test for bivariate normal -----  
BIV <- mvnrmtest::mshapiro.test(t(as.matrix(df))) # FTR
```

Since the data is neither Bivariate normal or of a larger size we will rely on permutation.

```
# Permutation  
n <- nrow(df)  
perm_y <- combinat::permn(strength)  
perm_y_df <- as.data.frame(perm_y)  
colnames(perm_y_df) <- c(1:factorial(n))  
perm_y_df <- as.data.frame(t(perm_y_df))  
colnames(perm_y_df) <- c("y1", "y2", "y3", "y4", "y5", "y6", "y7", "y8")  
  
pearsons <- c()  
spearman <- c()  
taus <- c()  
  
for(i in 1:length(perm_y)){  
  pearsons[i] <- cor(age, unlist(perm_y[i]), method="pearson")  
  spearman[i] <- cor(age, unlist(perm_y[i]), method="spearman")  
  taus[i] <- cor(age, unlist(perm_y[i]), method="kendall")  
}  
  
perm_y_df <-  
  perm_y_df %>%  
  cbind(pearsons, spearman, taus)  
  
row.names(perm_y_df)[perm_y_df$pearsons==pearson] <- "Observed"
```

Table 5: Head of Permutation

	pearsons	spearman	taus
Observed	0.7999108	1.0000000	1.0000000
2	0.7962972	0.9761905	0.9285714
3	0.7782294	0.9285714	0.8571429
4	0.7410396	0.8571429	0.7857143
5	0.6877694	0.7619048	0.7142857
6	0.6734055	0.6428571	0.6428571
7	0.6563012	0.5000000	0.5714286
8	0.6460627	0.3333333	0.5000000
9	0.6406423	0.3095238	0.4285714
10	0.6508808	0.4761905	0.5000000

## Pearson & Spearman

$H_0 : \rho = 0$  VS  $H_a : \rho \neq 0$



**Tau**

$$H_0 : \tau = 0 \text{ VS } H_a : \tau \neq 0$$

**Result**

Using either we derive the same permutation p-value of  $\frac{1}{4.032 \times 10^4} = 2.4801587 \times 10^{-5}$

Hence, in every case, thew association is significantly different from 0.

## P5: Exercise 7 on page 190

```
# Data -----
df <- data.frame(c(4,7), c(3,2))
df <- df %>%
  rbind(c(11, 5)) %>%
  cbind(c(7, 9, 16))
colnames(df) <- c("Nearby", "Not Nearby", "Row Totals")
rownames(df) <- c("Low", "High", "Column Totals")

# Fishers Exact test -----
fisher <- function(xmax, ndot1, ndot2, ndotdot, n1dot, xobs){
  P <- c()
  j <- 1
  for(i in 0:xmax){
    P[j] <- (choose(ndot1, i) * choose(ndot2, n1dot-i))/choose(ndotdot, n1dot)
    j=j+1
  }
  lowtail <- sum(P[1:(xobs+1)])
  uptail <- sum(P[(xobs+1):(length(P)]])
  P <- c(lowtail, uptail)
  P <- append(P, 2*min(P))
  return(P)
}

pval_twotail <- fisher(7, df[3,1], df[3,2], df[3,3], df[1,3], 4)[1]
```

The p-value is 0.3653846. Thus for a left-tail test, there is insufficient evidence that there is a significant association.

## P6: Exercise 12 on page 192

a)

$$H_0 : P_{MadeMiss} = P_{MissMade} \text{ VS. } H_a : P_{MadeMiss} \neq P_{MissMade}$$

$$\alpha = 0.05$$

```
# Data -----
df <- data.frame(c(5,14), c(3,8))
colnames(df) <- c("Missed second", "Made second")
rownames(df) <- c("Missed first", "Made first")
madeMiss <- df[2,1]
missMade <- df[1,2]
Tn <- ((madeMiss-missMade)**2)/(madeMiss+missMade)
p <- 2*pchisq(Tn,1, lower.tail = FALSE)
```

With a p-value of 0.0152658, we reject  $H_0$ . Thus, there is sufficient evidence that there is significance difference in the probability of making the first shot and making the second shot.

b)

$$H_0 : p_{ij} = p_{i.}p_{.j} \quad \forall (i,j) \text{ VS. } H_a : p_{ij} \neq p_{i.}p_{.j}$$

```
# Unpair -----
missed <- c(df[1,1]+df[2,1], df[1,1]+df[1,2])
made <- c(df[1,2]+df[2,2], df[2,1]+df[2,2])
unpair <- data.frame(missed,made)

# Permutation -----
obs <- chisq.test(as.matrix(unpair), simulate.p.value = TRUE)$statistic
tot <- 10000
mats <- list()
p <- c()

for (i in 1:tot){
  mat <- matrix(nrow=2, ncol=2)
  mat[1,1] <- sample(0:30, 1)
  mat[1,2] <- 30-mat[1,1]
  if(mat[1,1]==0){ # X11=0
    mat[2,1] <- sample(1:30, 1)
  } else if(mat[1,2]==0){ # X12=0
    mat[2,1] <- sample(0:29, 1)
  } else{
    mat[2,1] <- sample(0:30, 1)
  }
  mat[2,2] <- 30-mat[2,1]
  mat
  if(any(sapply(mats, identical, y=mat))){
  }else{
    res <- as.numeric(chisq.test(mat, simulate.p.value = TRUE)$statistic)
    p <- append(p, ((res >= obs)+0))
    mats <- append(mats, list(mat))
  }
}
```

```
    }  
  }  
  
pval <- sum(p)/tot
```

The p-value from the permutation is 0.0444, and therefore we reject  $H_0$ . Thus, there is sufficient evidence that there is an association between making first shot and making the second shot. Comparing this value to the one from a), we see that they are approximately the same. However, the chi-square test assume independence, but since we are getting the data from paired data this is a violated.