

# STAT 423 Review Guide

Shaandro Sarkar



# Contents

<b>About</b>	<b>5</b>
<b>1 Paired Two-Sample Tests</b>	<b>7</b>
1.1 Paired $t$ -test . . . . .	7
1.2 Sign test . . . . .	9
1.3 Wilcoxon Signed-Rank Test . . . . .	10
1.4 Permutation Test . . . . .	11
<b>2 Independent Two-Sample Tests</b>	<b>13</b>
2.1 Two-Sample $t$ -Test . . . . .	14
2.2 Permutation Test . . . . .	15
2.3 Mann-Whitney Test . . . . .	17
2.4 Kolmogorov-Smirnov Test . . . . .	17
2.5 Tests on Deviances . . . . .	18
<b>3 Independent <math>K</math>-Sample Tests</b>	<b>21</b>
3.1 Classic $F$ -Test . . . . .	21
3.2 Permutation $F$ -Test . . . . .	22
3.3 Kruskal-Wallis Test . . . . .	23
3.4 Jonkheere-Terpstra Test . . . . .	23
3.5 Bartlett Test . . . . .	24
3.6 Multiple Comparisons and Multiple Testing . . . . .	25
<b>4 In-Class Review Questions</b>	<b>29</b>
4.1 Estimator . . . . .	29
4.2 One-Proportion $z$ -Test . . . . .	29
4.3 Paired Two-Sample Tests . . . . .	30
4.4 Paired Permutation Test . . . . .	31
4.5 ANOVA and Classic $F$ -Test . . . . .	31
4.6 Kruskal-Wallis Test . . . . .	32
4.7 SSE (from ANOVA) . . . . .	32
4.8 SStr (from ANOVA) . . . . .	32
4.9 Permutation $F$ -test . . . . .	32

4.10 Multiple Comparison (Pairwise $t$ -Tests) . . . . .	33
--	----

# About

This is my review guide for STAT 423 (Nonparametric Statistics). The course was taught by Dr. Ya Su at Virginia Commonwealth University in the fall 2021 semester.

The lecture slides and accompanying R code is available on Google Drive.



# Chapter 1

## Paired Two-Sample Tests

Data for paired two-sample tests have the following properties:

- each subject  $i$  has two observations  $X_i$  and  $Y_i$  to give a paired observation  $(X_i, Y_i)$
- $(X_1, Y_1), \dots, (X_n, Y_n)$  are independent

Our general strategy is to perform one-sample tests on the paired difference  $Z_i = Y_i - X_i$

For this section, we will use the R dataset `sleep`. In this data, each subject was given both drug 1 and drug 2, and the change in their sleep was measured. The column `group` indicates which drug the subject took, and the column `extra` indicates the change in sleep.

```
data("sleep")

# X: effect on sleep from drug 1
x <- sleep$extra[sleep$group == 1]

# Y: effect on sleep from drug 2
y <- sleep$extra[sleep$group == 2]

# Z: paired difference between X and Y
z <- y - x
```

For each of these tests, we will be testing the hypotheses  $H_0: \mu = 0$  vs  $H_a: \mu \neq 0$ , where  $\mu$  is the mean of  $Z$ .

### 1.1 Paired $t$ -test

- Parametric test

- One-sample  $t$ -test on the mean  $\mu$  of  $Z$
- Assumption:  $Z_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$
- Hypotheses:  $H_0: \mu = \mu_0$  vs  $H_a: \mu \neq \mu_0$  (or  $<$ ,  $>$ )
- Test statistic:

$$T_{\text{obs}} = \frac{\bar{Z} - \mu_0}{s_Z / \sqrt{n}},$$

where  $\bar{Z}$  and  $s_Z$  are the sample mean and sample standard deviation of  $Z$

- Null distribution: Under  $H_0$ ,  $T_{\text{obs}} \sim t_{n-1}$

There are three ways to do this test in R using the function `t.test`.

If the observations for group  $X$  and group  $Y$  are stored in separate vectors `x` and `y`, then we use `t.test` with the argument `paired = TRUE`.

```
t.test(x, y, paired = TRUE, alternative = "two.sided")

##
## Paired t-test
##
## data: x and y
## t = -4.0621, df = 9, p-value = 0.002833
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -2.4598858 -0.7001142
## sample estimates:
## mean of the differences
## -1.58
```

If the paired differences  $Z = Y - X$  are stored in one vector `z`, then we use a one-sample `t.test`.

```
t.test(z, alternative = "two.sided")

##
## One Sample t-test
##
## data: z
## t = 4.0621, df = 9, p-value = 0.002833
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 0.7001142 2.4598858
## sample estimates:
## mean of x
## 1.58
```

If the data is stored in one matrix or dataframe where one column contains the observations and another column contains the groups, we can use the following syntax. For this data, `extra` contains the observations, and `group` indicates which group each observation comes from.



```
t.test(extra ~ group, paired = TRUE, data = sleep, alternative = "two.sided")

##
## Paired t-test
##
## data: extra by group
## t = -4.0621, df = 9, p-value = 0.002833
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -2.4598858 -0.7001142
## sample estimates:
## mean of the differences
## -1.58
```

The result for all three values is the same:  $p$ -value = 0.002833.

## 1.2 Sign test

- Also called binomial test
- Nonparametric test on the median  $\theta$  of  $Z$
- Assumption:  $Z_i$  are iid
- Hypotheses:  $H_0: \theta = \theta_0$  vs  $H_a: \theta \neq \theta_0$  (or  $<$ ,  $>$ )
- Test statistic:

$$X = I\{Z_i > \theta_0\},$$

where  $I\{Z_i > \theta_0\}$  is the number of observations  $Z_i$  that are greater than  $\theta_0$ .

- Null distribution: Under  $H_0$ ,  $X \sim \text{Binom}(n, p = 0.5)$

The alternative hypothesis  $H_a: \theta < 0$  is equivalent to  $H_a: p < 0.5$ , where  $p = P(Z_i > 0)$ . This is because, under  $H_0: \theta = 0$ , the probability of  $Z_i$  being less than the median  $\theta$  is 0.5.

The test is conducted as follows.

```
# calculate paired differences
z <- y - x

# calculate the number of observations greater than 0
X <- sum(z != 0)

# calculate p-value
2 * min(pbinom(X, size = length(z), prob = 0.5),
        1 - pbinom(X, size = length(z), prob = 0.5))

## [1] 0.001953125
```

### 1.3 Wilcoxon Signed-Rank Test

- Nonparametric test on the median  $\theta$  of  $Z$
- Assumptions:  $Z_i$  are iid and symmetric
- Hypotheses:  $H_0: \theta = \theta_0$  vs  $H_a: \theta \neq \theta_0$  (or  $<$ ,  $>$ )
- Test statistic:

$$V = \sum_{i=1}^n I\{Z_i > \theta_0\} \text{rank}(|Z_i - \theta_0|)$$

Like with the paired  $t$ -test, there are three ways to conduct the Wilcoxon test using `wilcox.test`.

```
wilcox.test(x, y, paired = TRUE, alternative = "two.sided")
```

```
## Warning in wilcox.test.default(x, y, paired = TRUE, alternative = "two.sided"):
## cannot compute exact p-value with ties
```

```
## Warning in wilcox.test.default(x, y, paired = TRUE, alternative = "two.sided"):
## cannot compute exact p-value with zeroes
```

```
##
```

```
## Wilcoxon signed rank test with continuity correction
```

```
##
```

```
## data: x and y
```

```
## V = 0, p-value = 0.009091
```

```
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(z, alternative = "two.sided")
```

```
## Warning in wilcox.test.default(z, alternative = "two.sided"): cannot compute
## exact p-value with ties
```

```
## Warning in wilcox.test.default(z, alternative = "two.sided"): cannot compute
## exact p-value with zeroes
```

```
##
```

```
## Wilcoxon signed rank test with continuity correction
```

```
##
```

```
## data: z
```

```
## V = 45, p-value = 0.009091
```

```
## alternative hypothesis: true location is not equal to 0
```

```
wilcox.test(extra ~ group, paired = TRUE, data = sleep, alternative = "two.sided")
```

```
## Warning in wilcox.test.default(x = c(0.7, -1.6, -0.2, -1.2, -0.1, 3.4, 3.7, :
## cannot compute exact p-value with ties
```

```
## Warning in wilcox.test.default(x = c(0.7, -1.6, -0.2, -1.2, -0.1, 3.4, 3.7, :
## cannot compute exact p-value with zeroes
```

```
##
```

```
## Wilcoxon signed rank test with continuity correction
##
## data:  extra by group
## V = 0, p-value = 0.009091
## alternative hypothesis: true location shift is not equal to 0
```

Using all three methods,  $p$ -value = 0.009091.

## 1.4 Permutation Test

- Paired data  $(X_i, Y_i)$ , where  $X_i$  and  $Y_i$  are correlated
- Assumption:  $(X_1, Y_1), \dots, (X_n, Y_n)$  are iid
- For  $n$  subjects, there are  $R = 2^n$  equally likely permutation outcomes
- For each possible outcome  $\ell \in \{1, \dots, R\}$ , we calculate the sample mean of differences

$$z_\ell = \frac{1}{n} \sum_{i=1}^n (Y_{\ell i} - X_{\ell i}),$$

where  $(X_{\ell i}, Y_{\ell i})$  is the  $i$ th observation of the  $\ell$ th permutation outcome.

The  $p$ -value depends on the direction of  $H_a$ .

$H_a$	$p$ -value
$\mu > 0$	$I\{z_\ell \geq \bar{Z}\}/R$
$\mu < 0$	$I\{z_\ell \leq \bar{Z}\}/R$
$\mu \neq 0$	$I\{ z_\ell  \geq  \bar{Z} \}/R$

### 1.4.1 Exact permutation test

The exact permutation test for  $H_a: \mu \neq 0$  is performed as follows.

```
# paired differences
z <- y - x
# sample size
n <- length(z)

# create permutation matrix
lists <- split(matrix(c(-1, 1), n, 2, byrow=TRUE), 1:n)
all.outcomes <- as.matrix(expand.grid(lists))

# observed test statistic
zbar <- mean(z)

# test statistic for permutation samples
z1 <- apply(all.outcomes, 1, function(u) mean(u*abs(z)))
```

```
# p-value
mean(abs(z1) >= abs(zbar))
```

```
## [1] 0.00390625
```

In this example, we used the sample mean as our test statistic, but the choice of test statistic for the permutation test is flexible.

### 1.4.2 Large-sample approximation for permutation test

Instead of using all  $R = 2^n$  possible permutation outcomes, we can randomly select  $R$  permutation outcomes and calculate the approximate  $p$ -value.

```
R <- 1000
z <- y - x
n <- length(z)

# generate permutation matrix
outcomes <- replicate(R, sample(c(-1, 1), n, replace = TRUE))

# observed test statistic
zbar <- mean(z)

# test statistic for permutation samples
z1 <- apply(outcomes, 2, function(u) mean(u*abs(z)))

# p-value
mean(abs(z1) >= abs(zbar))

## [1] 0.005
```

## Chapter 2

# Independent Two-Sample Tests

Data for independent two-sample tests have the following properties:

- two independent samples  $X_1, \dots, X_m$  and  $Y_1, \dots, Y_n$
- $X_1, \dots, X_m$  are mutually independent
- $Y_1, \dots, Y_n$  are mutually independent

We are interested in the mean differences between the two samples.

For this example, we will use the dataset `chickwts`. The column `feed` indicates the feed chicks were fed, and the column `weight` indicates the weight of the chick. We will only consider the `weights` of the `feeds` `horsebean` and `linseed`.

```
data("chickwts")

# subset only horsebean and linseed
chickwts <- subset(chickwts, feed == "horsebean" | feed == "linseed")

# X: weights of chicks with horsebean feed
x <- chickwts$weight[chickwts$feed == "horsebean"]

# Y: weights of chicks with linseed feed
y <- chickwts$weight[chickwts$feed == "linseed"]
```

For each test, we will be testing  $H_a: \Delta \neq \Delta_0$ , where  $\Delta$  is the difference in center (mean or median) between the two groups.

## 2.1 Two-Sample $t$ -Test

- Parametric test
- Assumptions:  $X_i \stackrel{\text{iid}}{\sim} N(\mu_X, \sigma^2)$ ,  $Y_i \stackrel{\text{iid}}{\sim} N(\mu_Y, \sigma^2)$
- Parameter of interest:  $\mu = \mu_X - \mu_Y$  Hypotheses:  $H_0: \mu = \mu_0$  vs  $H_a: \mu \neq \mu_0$  (or  $<$ ,  $>$ )
- Test statistic:

$$T_{\text{obs}} = \frac{\bar{X} - \bar{Y} - \mu_0}{s_p \sqrt{\frac{1}{m} + \frac{1}{n}}}, \quad \text{where} \quad s_p = \frac{(m-1)s_X^2 + (n-1)s_Y^2}{m+n-2}$$

- Null distribution: Under  $H_0$ ,  $T_{\text{obs}} \sim t_{m+n-2}$

There are two ways to conduct this test using `t.test`. For both ways, we use the argument `var.equal = TRUE` and we let the argument `mu` be equal to  $\mu_0$ .

If your two samples are in separate vectors `x` and `y`, you perform the test the way you're probably used to.

```
t.test(x, y, mu = 0, var.equal = TRUE, alternative = "two.sided")
```

```
##
## Two Sample t-test
##
## data: x and y
## t = -2.934, df = 20, p-value = 0.008205
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -100.17618 -16.92382
## sample estimates:
## mean of x mean of y
## 160.20 218.75
```

If your response variable is in one column of a dataset and the groups are indicated in a different column, we can use `~` and the argument `data` like we did for the paired  $t$ -test.

```
t.test(weight ~ feed, data = chickwts,
       mu = 0, var.equal = TRUE, alternative = "two.sided")
```

```
##
## Two Sample t-test
##
## data: weight by feed
## t = -2.934, df = 20, p-value = 0.008205
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -100.17618 -16.92382
## sample estimates:
```

```
## mean in group horsebean    mean in group linseed
##                160.20                218.75
```

Using both methods, we get  $p\text{-value} = 0.008205$ .

## 2.2 Permutation Test

- Nonparametric tests
- Assumptions
  - $X$  and  $Y$  are independent
  - $X_1, \dots, X_m$  are iid
  - $Y_1, \dots, Y_n$  are iid
- Parameters of interest
  - $\Delta_1 = E(X_i) - E(Y_i)$
  - $\Delta_2 = \text{median}(X_i) - \text{median}(Y_i)$
- Hypotheses:  $H_0: \Delta = \Delta_0$  vs  $H_a: \Delta \neq \Delta_0$  (or  $<$ ,  $>$ )
- Test statistic:
  - $\hat{\Delta}_1 = \bar{X} - \bar{Y}$
  - $\hat{\Delta}_2 = \text{median}(X) - \text{median}(Y)$

For the remainder of this section, we will use  $\Delta$  to represent the parameter of interest (difference in population means or medians) and  $\hat{\Delta}$  to represent the corresponding test statistic, since the permutation test works the same way regardless of the choice of test statistic.

- The total number of permutation outcomes is  $R = \binom{m+n}{n}$ .
- For each possible outcome  $\ell \in \{1, \dots, R\}$ , we calculate the sample test statistic (mean or median)  $\delta_\ell$ .

The  $p$ -value depends on the direction of  $H_a$ .

$H_a$	$p\text{-value}$
$\Delta > 0$	$I\{\delta_\ell \geq \hat{\Delta}\}/R$
$\Delta < 0$	$I\{\delta_\ell \leq \hat{\Delta}\}/R$
$\Delta \neq 0$	$I\{ \delta_\ell  \geq  \hat{\Delta} \}/R$

### 2.2.1 Exact permutation test

The exact permutation test for  $H_a: \Delta \neq 0$ , where  $\Delta$  is the difference in population means, is carried out as follows. (You can also do this for medians by replacing `mean` with `median` in the code.)

```
# append x and y
xy <- c(x, y)

m <- length(x) # sample size of x
```

```

n <- length(y) # sample size of y

# construct permutation matrix
permut <- combn(m + n, n)

# observed difference in sample means
Delta <- mean(x) - mean(y)

# sample mean for each permutation outcome
delta <- apply(permut, 2,
               function(u) mean(xy[-u]) - mean(xy[u]))

# p-value
mean(abs(delta) >= abs(Delta))

## [1] 0.009229161

```

This code takes a while to run because there are so many permutation outcomes. Because of that, we can perform a large-sample approximation using fewer permutation outcomes.

### 2.2.2 Large-sample approximation for permutation test

Instead of using all  $R = \binom{m+n}{n}$  possible permutation outcomes, we can randomly select  $R$  permutation outcomes and calculate the approximate  $p$ -value.

```

R <- 1000
xy <- c(x, y)
m <- length(x) # sample size of x
n <- length(y) # sample size of y

# construct permutation matrix
permut <- replicate(R, sample(m + n, n))

# observed difference in sample means
Delta <- mean(x) - mean(y)

# sample mean for each permutation outcome
delta <- apply(permut, 2,
               function(u) mean(xy[-u]) - mean(xy[u]))

# p-value
mean(abs(delta) >= abs(Delta))

## [1] 0.007

```



## 2.3 Mann-Whitney Test

- Equivalent to Wilcoxon rank-sum test
- Nonparametric test
- Assumptions
  - $X$  and  $Y$  are independent
  - $X_1, \dots, X_m$  are iid
  - $Y_1, \dots, Y_n$  are iid
- Parameter of interest:  $\Delta = \text{median}(X_i) - \text{median}(Y_i)$
- Hypotheses:  $H_0: \Delta = 0$  vs  $H_a: \Delta \neq 0$  (or  $<$ ,  $>$ )
- Test statistic: For  $i \in \{1, \dots, m\}$  and  $j = 1, \dots, n$ ,

$$U = I\{X_i > X_j\}$$

Because the Mann-Whitney test and the Wilcoxon rank-sum test are equivalent, we can use `wilcox.test` the same as with paired samples.

```
wilcox.test(x, y, paired = FALSE, alternative = "two.sided")

##
## Wilcoxon rank sum exact test
##
## data: x and y
## W = 20, p-value = 0.007145
## alternative hypothesis: true location shift is not equal to 0

wilcox.test(weight ~ feed, data = chickwts,
             paired = FALSE, alternative = "two.sided")

##
## Wilcoxon rank sum exact test
##
## data: weight by feed
## W = 20, p-value = 0.007145
## alternative hypothesis: true location shift is not equal to 0
```

## 2.4 Kolmogorov-Smirnov Test

- Designed to detect the differences of two distributions in either location, variability, or shape.
- Hypotheses: For two continuous CDFs  $F(u)$  and  $G(u)$ ,  $H_0: F(u) = G(u)$  vs  $H_a: F(u) \neq G(u)$
- Test statistic:

$$KS = \max |\hat{F}(u) - \hat{G}(u)|$$

where  $\hat{F}(u)$  and  $\hat{G}(u)$  are the empirical CDFs of two samples

```
ks.test(x, y)

##
## Two-sample Kolmogorov-Smirnov test
##
## data: x and y
## D = 0.55, p-value = 0.04889
## alternative hypothesis: two-sided
```

## 2.5 Tests on Deviances

Suppose we have two groups of data: -  $X_i = \mu_X + \sigma_X \epsilon_i$ , where  $i \in \{1, \dots, m\}$  -  $Y_i = \mu_Y + \sigma_Y \epsilon_{i+m}$ , where  $i \in \{1, \dots, n\}$

We will test  $H_0: \sigma_X = \sigma_Y$  for the `chickwts` data from the previous section.

### 2.5.1 $F$ -Test

- Assumption:  $\epsilon_i \stackrel{\text{iid}}{\sim} N(0, 1)$
- Test statistic:

$$F_{\text{obs}} = \frac{s_X^2}{s_Y^2}$$

- Null distribution: Under  $H_0$ ,  $F_{\text{obs}} \sim F_{m-1, n-1}$

```
R <- 1000
xy <- c(x, y)
m <- length(x) # sample size of x
n <- length(y) # sample size of y

# construct permutation matrix
permut <- replicate(R, sample(m + n, n))

# sample variances
varx <- var(x)
vary <- var(y)

# test statistic
Fobs <- varx / vary

# p-value
2 * min(pf(Fobs, m - 1, n - 1), 1 - pf(Fobs, m - 1, n - 1))

## [1] 0.3738707
```

### 2.5.2 Nonparametric Test

- Calculate the deviances:
  - $\text{dev}_{iX} = X_i - \bar{X}$
  - $\text{dev}_{iY} = Y_i - \bar{Y}$
- Calculate the absolute mean deviances:
  - $\text{dev}_X = \frac{1}{m} \sum_{i=1}^n |\text{dev}_{iX}|$
  - $\text{dev}_Y = \frac{1}{n} \sum_{i=1}^n |\text{dev}_{iY}|$
- Test statistic for  $H_a: \sigma_X > \sigma_Y$  and  $H_a: \sigma_X < \sigma_Y$

$$r_{\text{obs}} = \frac{\text{dev}_X}{\text{dev}_Y}$$

- Test statistic for  $H_a: \sigma_X \neq \sigma_Y$

$$r_{\text{obs}} = \frac{\min\{\text{dev}_X, \text{dev}_Y\}}{\max\{\text{dev}_X, \text{dev}_Y\}}$$

Let  $R$  be the number of permutation outcomes. For all  $\ell \in \{1, \dots, R\}$ ,  $r_\ell$  is the test statistic for the permutation outcome  $\ell$ . The  $p$ -value depends on the direction of  $H_a$ .

$H_a$	$p$ -value
$\sigma_X > \sigma_Y$	$I\{r_\ell \geq r\}/R$
$\sigma_X < \sigma_Y$	$I\{r_\ell \leq r\}/R$
$\sigma_X \neq \sigma_Y$	$I\{r_\ell \geq r\}/R$

```
R <- 1000
xy <- c(x, y)

# construct permutation matrix
permut <- replicate(R, sample(m + n, n))

# absolute mean deviances
devx <- mean(abs(x - mean(x)))
devy <- mean(abs(y - mean(y)))

# test statistic
robs <- max(devx, devy) / min(devx, devy)

# test statistics for permutation outcomes
devx1 <- apply(permut, 2,
               function(u) mean(abs(xy[u] - mean(xy[u]))))
devy1 <- apply(permut, 2,
               function(u) mean(abs(xy[-u] - mean(xy[-u]))))
r <- apply(cbind(devx1, devy1), 1,
```

```
function(u) {max(u) / min(u)}  
  
# p-value  
mean(r >= robs)  
  
## [1] 0.202
```

## Chapter 3

# Independent $K$ -Sample Tests

- $K$  independent samples
- $n_i$ : number of observations in group  $i$
- $X_{ij}$ :  $j$ th observation from  $i$ th sample
  - $i \in \{1, \dots, K\}$
  - $j \in \{1, \dots, n_i\}$
- $\mu_i$ : mean of group  $i$
- One-way ANOVA model:  $X_{ij} = \mu_i + e_{ij}$
- Hypotheses:  $H_0: \mu_1 = \mu_2 = \dots = \mu_K$  vs  $H_a: H_0$  is not true

For this section, we will use the dataset `iris`. The column `Sepal.Length` indicates the length of the flower's sepal, and the column `Species` indicates its species.

```
data("iris")
```

### 3.1 Classic $F$ -Test

- Assumption:  $e_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$
- Mean of group  $i$ :  $\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$
- Grand mean:  $\bar{X} = \frac{1}{N} \sum_{i=1}^K \sum_{j=1}^{n_i} X_{ij}$
- Treatment sum of squares:

$$\text{SSTr} = \sum_{i=1}^K n_i (\bar{X}_i - \bar{X})^2$$

- Error sum of squares:

$$\text{SSE} = \sum_{i=1}^K \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

- Test statistic:

$$F_{\text{obs}} = \frac{\text{SSTr}/(K-1)}{\text{SSE}/(N-K)}$$

- Null distribution: Under  $H_0$ ,  $F_{\text{obs}} \sim F_{K-1, N-K}$

We can conduct the classic  $F$ -test using `anova` and `lm`.

```
anova(lm(Sepal.Length ~ Species, data = iris))
```

```
## Analysis of Variance Table
##
## Response: Sepal.Length
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Species      2  63.212   31.606   119.26 < 2.2e-16 ***
## Residuals 147  38.956    0.265
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The SSTr is the value in the first row of column Sum Sq. The SSE is the value in the row Residuals of column Sum Sq. The  $p$ -value is  $\text{Pr}(>F)$ .

## 3.2 Permutation $F$ -Test

- Nonparametric test
- Total number of permutation outcomes:  $R = N!/(n_1!n_2!\dots n_K!)$
- Calculate  $F_{\text{obs}}$
- For  $\ell \in \{1, \dots, R\}$ ,  $F_\ell$  is the  $F$ -statistic for the  $\ell$ th permutation outcome
- We can use the simplified calculation

$$F \propto \sum_{i=1}^K n_i \bar{X}_i^2$$

- $p\text{-value} = I\{F_\ell \geq F_{\text{obs}}\}/R$

```
# number of permutation samples
R <- 1000
# number of samples in each group
n <- tapply(iris$Sepal.Length, iris$Species, length)
# total number of observations
N <- sum(n)
# sepal length of each observation
x <- iris$Sepal.Length
```

```

# species of each observation
t <- iris$Species

# F-statistic
Fobs <- sum(n * tapply(x, t, mean)^2)

# permutation sample F-statistics
F <- c()
for (i in 1:R) {
  F <- c(F, sum(n * tapply(x, sample(t), mean)^2))
}

# p-value
mean(F >= Fobs)

## [1] 0

```

### 3.3 Kruskal-Walis Test

- Nonparametric test
- Assumption: the data are iid in each group
- Replace data with ranks
- Test statistic:

$$KW = \frac{12}{N(N+1)} SST^*,$$

where  $SST^*$  is the SST calculated with ranks

- Null distribution: Under  $H_0$ ,  $KW \sim \chi_{K-1}^2$  approximately

```
kruskal.test(Sepal.Length ~ Species, data = iris)
```

```

##
## Kruskal-Wallis rank sum test
##
## data: Sepal.Length by Species
## Kruskal-Wallis chi-squared = 96.937, df = 2, p-value < 2.2e-16

```

### 3.4 Jonkheere-Terpstra Test

- Nonparametric test
- Assumption: the data are iid in each group
- Hypotheses:  $H_0: \mu_1 = \mu_2 = \dots = \mu_K$  vs  $H_a: \mu_1 \leq \mu_2 \leq \dots \leq \mu_K$
- Test-Statistic:

$$JT = \sum_{i < j} U_{ij},$$

where  $U_{ij}$  is the Mann-Whitney test statistic for  $H_a: \mu_i < \mu_j$

- Approximate null distribution: Under  $H_0$ ,

$$\frac{JT - E(JT)}{\sqrt{\text{Var}(JT)}} \sim N(0, 1)$$

$$E(JT) = \frac{N^2 - \sum_{i=1}^K n_i^2}{4}$$

$$\text{Var}(JT) = \frac{N^2(2N + 3) - \sum_{i=1}^K n_i^2(2n_i + 3)}{72}$$

```
# assign numbers to each group
species <- factor(iris$Species, levels=c("setosa","versicolor","virginica"))
groups <- as.numeric(species)

sepal <- iris$Sepal.Length

# number of groups
K <- max(groups)

# number of observations
N <- length(sepal) # total
n <- table(groups) # per group

# compute Uij
JT <- 0
for(i in 1:(K-1)) {
  for(j in (i+1):K) {
    JT <- JT + sum(outer(sepal[groups == i], sepal[groups == j], "<"))
  }
}
JT

## [1] 6686
```

### 3.5 Bartlett Test

- Test for equal variances
- Hypotheses:  $H_0: \sigma_1 = \sigma_2 = \dots = \sigma_K$

```
bartlett.test(iris$Sepal.Length ~ iris$Species)
```

```
##
## Bartlett test of homogeneity of variances
##
## data: iris$Sepal.Length by iris$Species
## Bartlett's K-squared = 16.006, df = 2, p-value = 0.0003345
```



## 3.6 Multiple Comparisons and Multiple Testing

We have three methods of deciding the significance levels of multiple comparisons:

Adjustment	Description	p.adjust.method
none	compare $p_i$ with $\alpha$	none
Bonferroni	compare $p_i$ with $\alpha/h$	"bonferroni"
Benjamini-Hochberg	compare BH-adjusted $p_i$ with $\alpha$	"BH"

The Benjamini-Hochberg procedure leads to improved Type I error rate and power.

### 3.6.1 Multiple Comparisons

- If we reject  $H_0: \mu_1 = \mu_2 = \dots = \mu_K$  for a  $K$ -sample test, we can conduct pairwise two-sample tests to determine which pairs of means are not equal.
- Total number of pairwise tests for  $K$  samples:  $h = K(K-1)/2$
- $P(\text{at least one Type I error}) = 1 - (1 - \alpha)^h$

```
pairwise.t.test(iris$Sepal.Length, iris$Species)
```

```
##
## Pairwise comparisons using t tests with pooled SD
##
## data: iris$Sepal.Length and iris$Species
##
##          setosa versicolor
## versicolor 1.8e-15 -
## virginica  < 2e-16 2.8e-09
##
## P value adjustment method: holm
```

#### 3.6.1.1 Bonferroni Adjustment

- compare  $p$ -value for each pairwise comparison  $i$  with significance level  $\alpha_i = \alpha/h$ 
  - works because  $1 - (1 - \alpha_i)^h \approx h\alpha_i$

```
pairwise.t.test(iris$Sepal.Length, iris$Species, p.adjust.method = "bonferroni")
```

```
##
## Pairwise comparisons using t tests with pooled SD
##
## data: iris$Sepal.Length and iris$Species
##
```

```
##          setosa  versicolor
## versicolor 2.6e-15 -
## virginica  < 2e-16 8.3e-09
##
## P value adjustment method: bonferroni
```

### 3.6.2 Multiple Testing

- Consider  $m$  hypothesis tests together
- The  $p$ -values are  $p_1, \dots, p_h$
- Instead of controlling the Type I error rate  $\alpha$ , we control the False Discovery Rate

$$\text{FDR} = E \left( \frac{\text{number of false discoveries}}{\text{number of discoveries}} \right) \leq \alpha$$

– discovery means rejection of  $H_0$

- Threshold: Reject  $H_0$  if  $p\text{-value} \leq t$
- False Discovery Proportion for  $i \in \{1, \dots, m\}$

$$\text{FDP}(t) = \frac{\sum_{i=1}^m I\{p_i \leq t\} H_i}{\sum_{i=1}^m I\{p_i \leq t\}}$$

- $p_i$  is the  $p$ -value for test  $i$
- $H_i = 1$  if  $H_0$  is true and  $H_i = 0$  otherwise

#### 3.6.2.1 Benjamini-Hochberg Procedure

- Choose the threshold for rejection to be

$$t_{\text{BH}} = \max\{P_{(i)} \mid P_{(i)} \leq \frac{\alpha i}{m}, 0 \leq i \leq m\}$$

–  $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(m)}$

- It can be proved that

$$\text{FDR} = E[\text{FDP}(t_{\text{BH}})] \leq \frac{\alpha \sum_{i=1}^m H_i}{m}$$

```
pairwise.t.test(iris$Sepal.Length, iris$Species, p.adjust.method = "BH")
```

```
##
## Pairwise comparisons using t tests with pooled SD
##
## data:  iris$Sepal.Length and iris$Species
```

```
##  
##          setosa  versicolor  
## versicolor 1.3e-15 -  
## virginica  < 2e-16 2.8e-09  
##  
## P value adjustment method: BH
```



## Chapter 4

# In-Class Review Questions

### 4.1 Estimator

Amazon recommends customers products according to their purchasing history. Among 10,000 independent recommendations, 118 were accepted. The estimator of acceptance rate is  $\hat{p} = 118/10000$ .

### 4.2 One-Proportion $z$ -Test

Let  $p$  be the true acceptance rate. Test the hypotheses  $H_0: p = 0.02$  vs  $H_a: p < 0.02$ .

To find the  $p$ -value for this test, we will use the one-proportion  $z$ -test statistic<sup>1</sup>

$$Z = \frac{n\hat{p} - np_0}{\sqrt{np_0(1 - p_0)}}.$$

```
S <- 118 # number of successes
n <- 10000 # sample size
p0 <- 0.02 # expected probability
ES <- n*p0 # expected number of successes
VarS <- n*p0*(1 - p0) # variance
Z <- (S - ES) / sqrt(VarS) # test statistic
pnorm(Z) # p-value
```

```
## [1] 2.35449e-09
```

Under large-sample approximation, the  $p$ -value of this test is  $2.35449 \times 10^{-9}$ .

---

<sup>1</sup>We can use the  $z$ -test because  $np_0 = 10000(0.02) = 200 > 10$  and  $n(1 - p_0) = 10000(0.98) = 9800 > 10$ . Dr. Su refers to this as a large-sample approximation because, with a large enough sample, we can assume that the data is distributed normally.

### 4.3 Paired Two-Sample Tests

Load the R dataset `sleep`<sup>2</sup>.

```
data("sleep")
head(sleep)
```

```
##      extra group ID
## 1    0.7      1  1
## 2   -1.6      1  2
## 3   -0.2      1  3
## 4   -1.2      1  4
## 5   -0.1      1  5
## 6    3.4      1  6
```

Test  $H_0: \mu_1 = \mu_2$  vs  $H_a: \mu_1 < \mu_2$ , where  $\mu_1$  and  $\mu_2$  are the centers of the two groups. Which method gives the smaller  $p$ -value:  $t$ -test or Wilcoxon signed-rank test?

```
# paired t-test p-value
t.test(extra ~ group, paired = TRUE, data = sleep, alternative = "less")$p.value

## [1] 0.001416445

# paired Wilcoxon test p-value
wilcox.test(extra ~ group, paired = TRUE, data = sleep, alternative = "less")$p.value

## Warning in wilcox.test.default(x = c(0.7, -1.6, -0.2, -1.2, -0.1, 3.4, 3.7, :
## cannot compute exact p-value with ties

## Warning in wilcox.test.default(x = c(0.7, -1.6, -0.2, -1.2, -0.1, 3.4, 3.7, :
## cannot compute exact p-value with zeroes

## [1] 0.004545349
```

Since  $0.0014 < 0.0045$ , the  $t$ -test has a smaller  $p$ -value than the Wilcoxon test.

We now consider the permutation test. The expression that generates the appropriate permutation outcome is

```
z <- abs(sleep$extra[1:10] - sleep$extra[11:20])
sample(c(-1,1), 10, replace = TRUE) * z
```

```
## [1] -1.2  2.4  1.3  1.3  0.0  1.0 -1.8  0.8  4.6 -1.4
```

---

<sup>2</sup>The `sleep` data shows the effect of two drugs on 10 patients. The dataset has three columns: `extra` (increase in hours of sleep), `group` (drug given), and `ID` (patient ID). The data is paired, since each patient was given both drug 1 and drug 2.

## 4.4 Paired Permutation Test

Load the R dataset `sleep`. Test  $H_0: \mu_1 = \mu_2$  vs  $H_a: \mu_1 \neq \mu_2$ , where  $\mu_1$  and  $\mu_2$  are the centers of the two groups.

Consider the permutation test. What is the total number of permutation outcomes?

The `sleep` dataset has 10 paired observations. If  $n$  is the sample size, then the number of permutation outcomes is  $2^n$ . For the `sleep` dataset, the number of permutation outcomes is thus  $2^{10} = 1024$ .

## 4.5 ANOVA and Classic $F$ -Test

Load the R dataset `iris`.

```
data("iris")
head(iris)
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1           5.1         3.5          1.4          0.2   setosa
## 2           4.9         3.0          1.4          0.2   setosa
## 3           4.7         3.2          1.3          0.2   setosa
## 4           4.6         3.1          1.5          0.2   setosa
## 5           5.0         3.6          1.4          0.2   setosa
## 6           5.4         3.9          1.7          0.4   setosa
```

For the first column, consider the hypothesis  $H_0: \mu_1 = \mu_2 = \mu_3$  vs  $H_a: H_0$  is not true, where  $\mu_1$ ,  $\mu_2$ , and  $\mu_3$  are the centers for the sepal lengths of *Iris setosa*, *versicolor*, and *virginica*, respectively.

We will find the  $p$ -value of the  $F$ -test for these hypotheses by generating an anova table from the `lm` model.

```
anova(lm(Sepal.Length ~ Species, data = iris))
```

```
## Analysis of Variance Table
##
## Response: Sepal.Length
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Species       2  63.212   31.606   119.26 < 2.2e-16 ***
## Residuals   147  38.956    0.265
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the ANOVA table, we have  $p\text{-value} < 2.2 \times 10^{-16}$ .

## 4.6 Kruskal-Wallis Test

Load the R dataset `iris`. For the first column, consider the hypothesis  $H_0: \mu_1 = \mu_2 = \mu_3$  vs  $H_a: H_0$  is not true, where  $\mu_1, \mu_2$ , and  $\mu_3$  are the centers for the sepal lengths of *Iris setosa*, *versicolor*, and *virginica*, respectively.

We will find the  $p$ -value of the Kruskal-Wallis test using `kruskal.test`.

```
kruskal.test(iris$Sepal.Length, iris$Species)
```

```
##
## Kruskal-Wallis rank sum test
##
## data: iris$Sepal.Length and iris$Species
## Kruskal-Wallis chi-squared = 96.937, df = 2, p-value < 2.2e-16
```

Thus  $p\text{-value} < 2.2 \times 10^{-16}$ .

## 4.7 SSE (from ANOVA)

Load the R dataset `iris`. Consider the first numerical column and find the SSE.

The SSE can be found on the ANOVA table, which we already generated in 5. The SSE is the value in the row `Residuals` and the column `Sum Sq`, i.e.  $\text{SSE} = 38.956$ .

## 4.8 SStr (from ANOVA)

Load the R dataset `iris`. Consider the first numerical column and find the SStr.

The SStr can also be found on the ANOVA table. The SStr is the value in the row `Species` and the column `Sum Sq`, i.e.  $\text{SStr} = 63.212$ .

## 4.9 Permutation $F$ -test

Load the R dataset `iris`. For the first column, compute the  $p$ -value using the permutation  $F$ -test for  $H_0: \mu_1 = \mu_2 = \mu_3$  vs  $H_a: H_0$  is not true.

We will perform the permutation test using  $R = 5000$  permutation samples. We use the simplified calculation of the  $F$ -statistic<sup>3</sup>

$$F \propto \sum_{i=1}^K n_i \bar{X}_i^2,$$

---

<sup>3</sup>`k-sample.pdf`, slide 9



where  $K$  is the number of groups,  $n_i$  is the number of observations in group  $i$ , and  $\bar{X}_i$  is the mean of group  $i$ .

```
# number of samples in each group
n <- tapply(iris$Sepal.Length, iris$Species, length)
# total number of observations
N <- sum(n)
# sepal length of each observation
y <- iris$Sepal.Length
# species of each observation
t <- iris$Species

# F-statistic
Fobs <- sum(n * tapply(y, t, mean)^2)

# number of permutation samples
R <- 5000
# empty vector for the F-statistic of each permutation sample
F <- c()
# calculate permutation sample F-statistics
for (i in 1:R) {
  F <- c(F, sum(n * tapply(y, sample(t), mean)^2))
}

# p-value = proportion of permutation F-statistics greater than Fobs
mean(F >= Fobs)

## [1] 0
```

Thus the  $p$ -value for the permutation  $F$ -test is 0.

## 4.10 Multiple Comparison (Pairwise $t$ -Tests)

Load the R dataset `iris`. For the first column, compute the  $p$ -values from the multiple comparison based on pairwise  $t$ -tests.

```
pairwise.t.test(y, t)

##
## Pairwise comparisons using t tests with pooled SD
##
## data: y and t
##
##          setosa versicolor
## versicolor 1.8e-15 -
## virginica   < 2e-16 2.8e-09
##
```

```
## P value adjustment method: holm
```

The pair with the smallest  $p$ -value is *I. setosa* vs. *I. virginica*.