MATH 204: Principles of Statistics 2

Winter 2026

Dr David A. Stephens

Department of Mathematics & Statistics
Room 1225, Burnside Hall

david.stephens@mcgill.ca
www.math.mcgill.ca/dstephens/204/

Textbook: McClave and Sincich, Statistics (13th Edition) Chapters 10-14.

Note: Recent editions of McClave and Sincich contain essentially the same material.

Prerequisites: MATH 203 (or equivalent)

Some statistical computing knowledge useful.

Method of Assessment:

- Assignments
- Quizzes
- Mid-Term
- Final

See syllabus handout for precise details.

Course Objectives

- Extension of MATH 203 topics to other practical experimental contexts
- Basic statistical computation using software (R)
- Practice in the use of statistical methods, in particular, hypothesis testing and linear modelling.

Course Objectives

From the Course Catalogue:

The concept of degrees of freedom and the analysis of variability. Planning of experiments. Experimental designs. Polynomial and multiple regressions. Statistical computer packages (no previous computing experience is needed). General statistical procedures requiring few assumptions about the probability model.

Three main parts

I. The Analysis of Variance and Designed Experiments
   - *The concept of degrees of freedom*
   - *The analysis of variability*
   - *Planning of experiments*
   - *Experimental designs*

II. Linear Regression Modelling
   - *Polynomial and multiple regressions*

III. Non-Parametric Testing
   - *General statistical procedures requiring few assumptions about the probability model*

Statistical computer packages (no previous computing experience is needed)

Recap of hypothesis testing

Typical experimental scenario:

- two different groups of subjects
- single observation/measurement made on each subject
- scientific question of interest

Example: Pre-Natal Care

Comparison of the birthweights of babies in two groups of mothers:

- GROUP A: Received five or fewer pre-natal visits
- GROUP B: Received more than five pre-natal visits

Do the GROUP A babies have

significantly different

birthweights from those from GROUP B ?

Data: Birthweights (grammes)

- GROUP A: 10 subjects

2164	2600	2184	2080	1820
2496	2184	2080	2184	2576

- GROUP B: 7 subjects

3224	2704	2912	2444	3120
		2912	3848	

First step in analysis: statistical summary

- GROUP A:
  - Sample size: $n_A = 10$
  - Sample mean: $\overline{x}_A = 2236.8$
  - Sample variance: $s_A^2 = 61190.4$
- GROUP B:
  - Sample size: $n_B = 7$
  - Sample mean: $\overline{x}_B = 3023.429$
  - Sample variance: $s_B^2 = 198679.6$

First step in analysis: statistical summary

Recall, for data $x_1, \ldots, x_n$

$$\begin{aligned} \overline{x} &= \frac{x_1 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \\ s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \end{aligned}$$

$\overline{x}$ measures the average value of sample
s^2 measures the amount of variability around the average.

First step in analysis: statistical summary

Implementation in R:

```
> load("204DataSets.rda")
> Birthweight

##    Group gp   BW
## 1      A  1 2164
## 2      A  1 2600
## 3      A  1 2184
## 4      A  1 2080
## 5      A  1 1820
## 6      A  1 2496
## 7      A  1 2184
## 8      A  1 2080
## 9      A  1 2184
## 10     A  1 2576
## 11     B  2 3224
## 12     B  2 2704
## 13     B  2 2912
## 14     B  2 2444
## 15     B  2 3120
## 16     B  2 2912
## 17     B  2 3848
```

First step in analysis: statistical summary

```
> xA<-Birthweight$BW[Birthweight$Group=='A']
> c(length(xA),mean(xA),var(xA))

## [1]    10.0  2236.8 61190.4

> xB<-Birthweight$BW[Birthweight$Group=='B']
> c(length(xB),mean(xB),var(xB))

## [1]      7.000   3023.429 198679.619
```

First step in analysis: statistical summary

In the birthweight example

$$\overline{x}_A = 2236.8 \qquad \overline{x}_B = 3023.429$$

so it appears that Group B birthweights are higher....

... but are they significantly higher ?

i.e. is the difference due to chance alone ? Clearly

- sample sizes quite small
- birthweights quite variable

Statistical testing

We adopt the following procedure to assess the significance of the difference between $\overline{x}_A$ and $\overline{x}_B$.

1. Define a test statistic, T, that permits comparison of the two groups.

2. Predict how T will behave assuming that the two groups are not significantly different.

3. Compare the prediction with what was actually observed.

Statistical testing

Formally, we

- assume a Normal distribution for the data

  Assume $x_{A1}, \ldots, x_{An_A}$ are drawn from a population that is well-modelled by a

  $$Normal(\mu_A, \sigma_A^2)$$

  distribution. Similarly

  $$x_{B1}, \ldots, x_{Bn_B} \sim Normal(\mu_B, \sigma_B^2)$$

  We might initially assume that

  $$\sigma_A^2 = \sigma_B^2.$$

Statistical testing

- consider the two hypotheses

$$H_0 : \quad \mu_A = \mu_B$$
$$H_a : \quad \mu_A \neq \mu_B$$

H_0 is the null hypothesis
H_a is the alternative hypothesis

  - here H_a is a two-sided alternative
  - this leads to a two-sided test

Statistical testing

- find the distribution of T assuming that H_0 is true
  - this distribution is known as the null distribution.
  - for probability calculations associated with the null distribution, we use the notation P_0: for example, we might compute

$$P_0(T > 2.04) \qquad P_0(T \leqslant -1.13)$$

and so on.

Statistical testing

- compute the observed test statistic

$$T_{obs} = \frac{\overline{x}_A - \overline{x}_B}{s_P \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

where

$$s_P^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}$$

Statistical testing

s_P^2 is an estimate of the common population variance

$$\sigma^2 = \sigma_A^2 = \sigma_B^2.$$

Here

$$s_P^2 = \frac{(10-1)61190.4 + (7-1)198679.6}{10+7-2} = 116186.1$$

so that

$$s_P = 340.8608.$$

Thus

$$T_{obs} = \frac{2236.8 - 3023.429}{340.8608\sqrt{\frac{1}{10} + \frac{1}{7}}} = -4.683$$

Statistical testing

If the null hypothesis was true, so that $\mu_A = \mu_B$, the computed test statistic T_{obs} should look like an observation from a

Student-t or t

distribution with

$$n_A + n_B - 2 = 15$$

degrees of freedom. We write t_{15} to denote this distribution.

If H_0 is true, T_{obs} should lie somewhere in a

high probability region

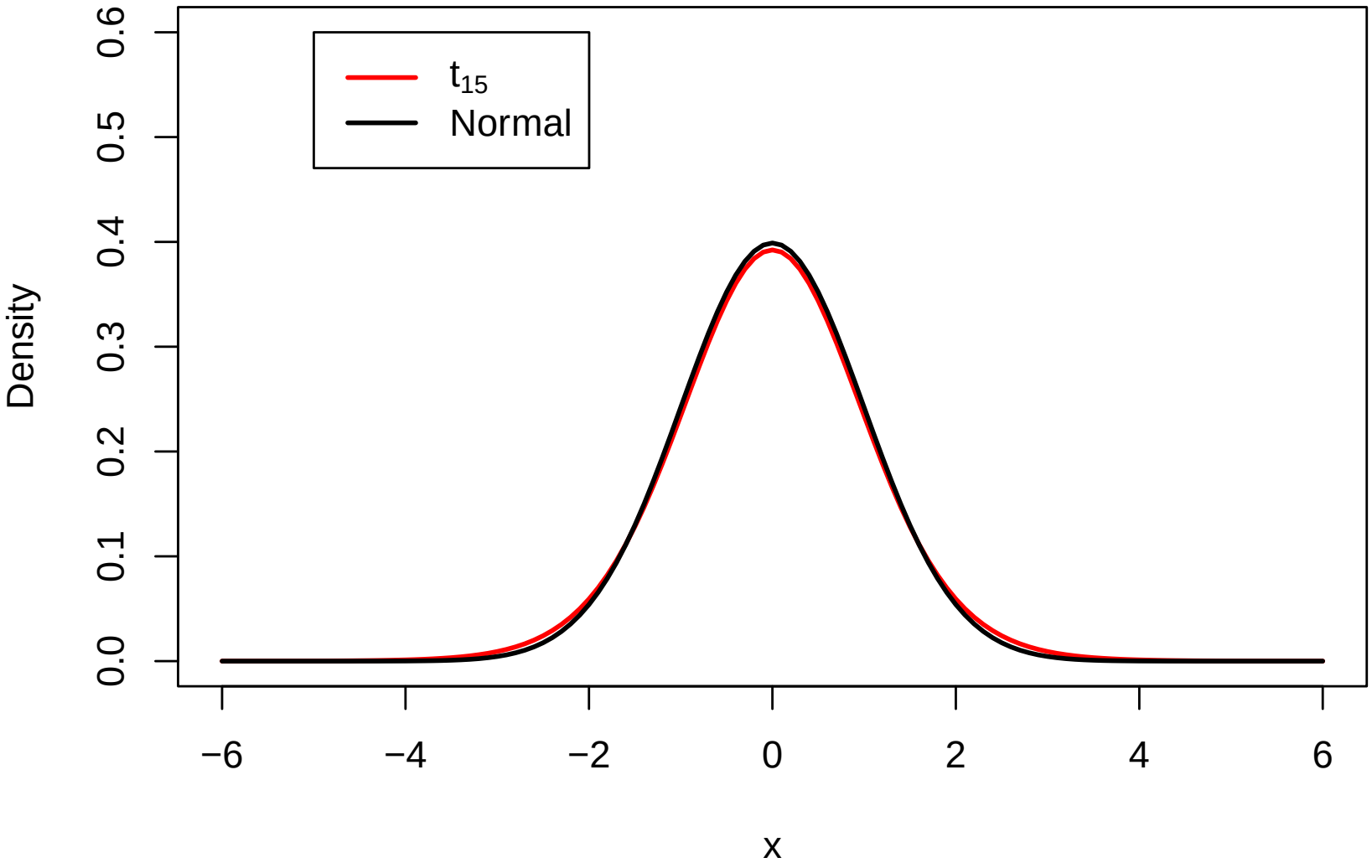
of the t_{15} probability distribution.

Statistical testing

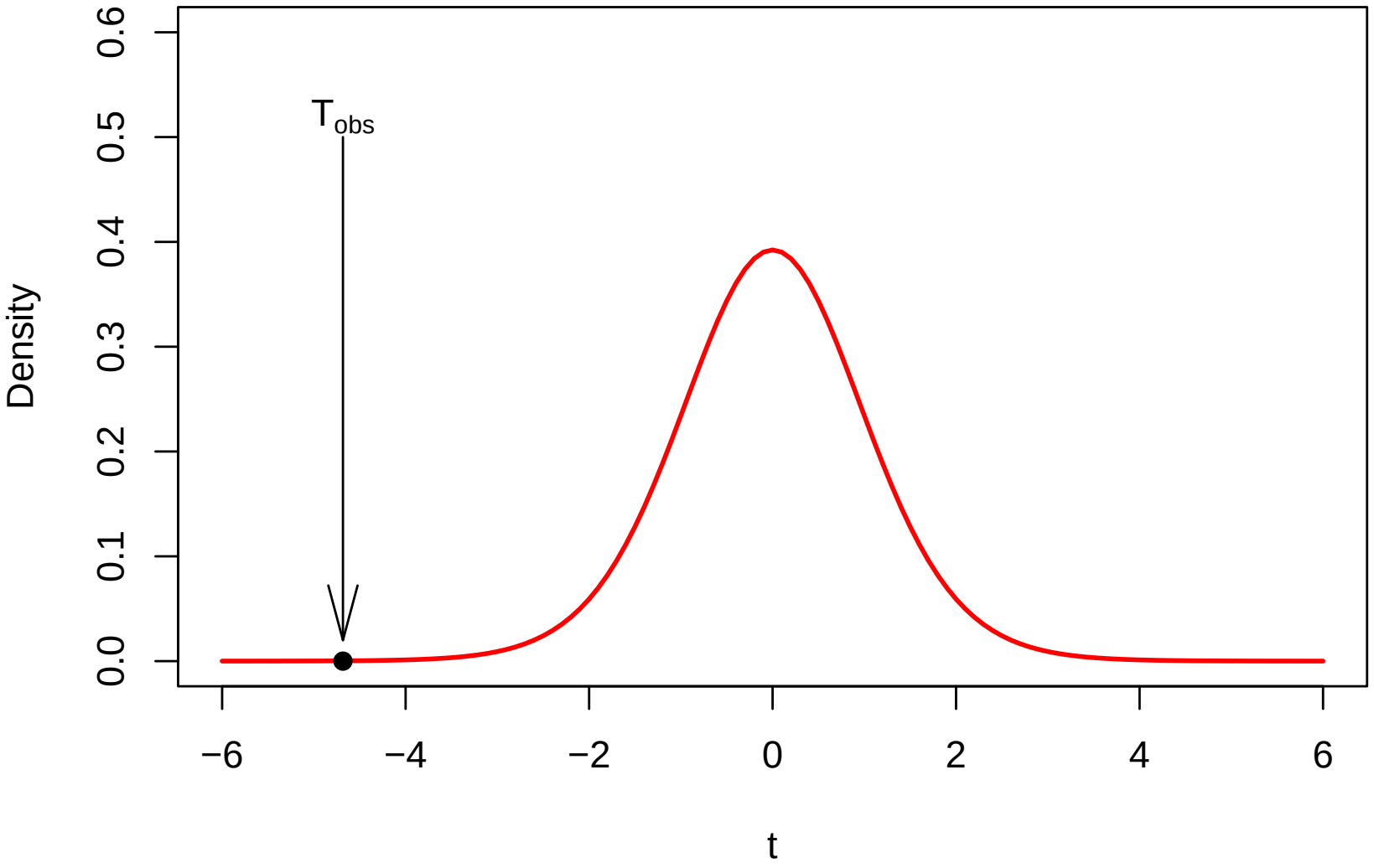
Calculation in R:

```
> nA<-length(xA)
> nB<-length(xB)
> sP.sq<-((nA-1)*var(xA)+(nB-1)*var(xB))/(nA+nB-2)
> sP.sq

+ [1] 116186.1

> sP<-sqrt(sP.sq)
> sP

+ [1] 340.8608

> Tobs<-(mean(xA)-mean(xB))/(sP*sqrt(1/nA+1/nB))
> Tobs

+ [1] -4.682925
```

Statistical testing



Statistical testing



Statistical testing

In this case, T_{obs} does not lie in a high probability region.

i.e. we are surprised to see T_{obs} so far away from zero.

The predicted behaviour of T assuming that H_0 is true

does not match the observed behaviour !

and so the assumption that H_0 is true must be incorrect. We

Reject H_0

in favour of H_a.

Statistical significance

How do we quantify the statistical significance ?

Two approaches:

1. Define the high probability region, and reject H_0 if T_{obs} does not lie in this region.

2. Compute the level of surprise at observing T_{obs} under the assumption that H_0 is true.

Statistical significance

For 1: Set significance level α, with $0 < \alpha < 1$, and find the central $1 - \alpha$ high probability region, between the two values $-C_R$ and C_R.

- α is the probability of a Type I Error, which occurs when we reject H_0, given that it is true

Statistical significance

- $\pm C_R$ are the critical values
  - here, C_R is the upper $\alpha/2$ point of the t_{15} distribution, denoted $t_{\alpha,15}$, where

$$P_0(T > t_{\alpha/2,15}) = \alpha/2.$$

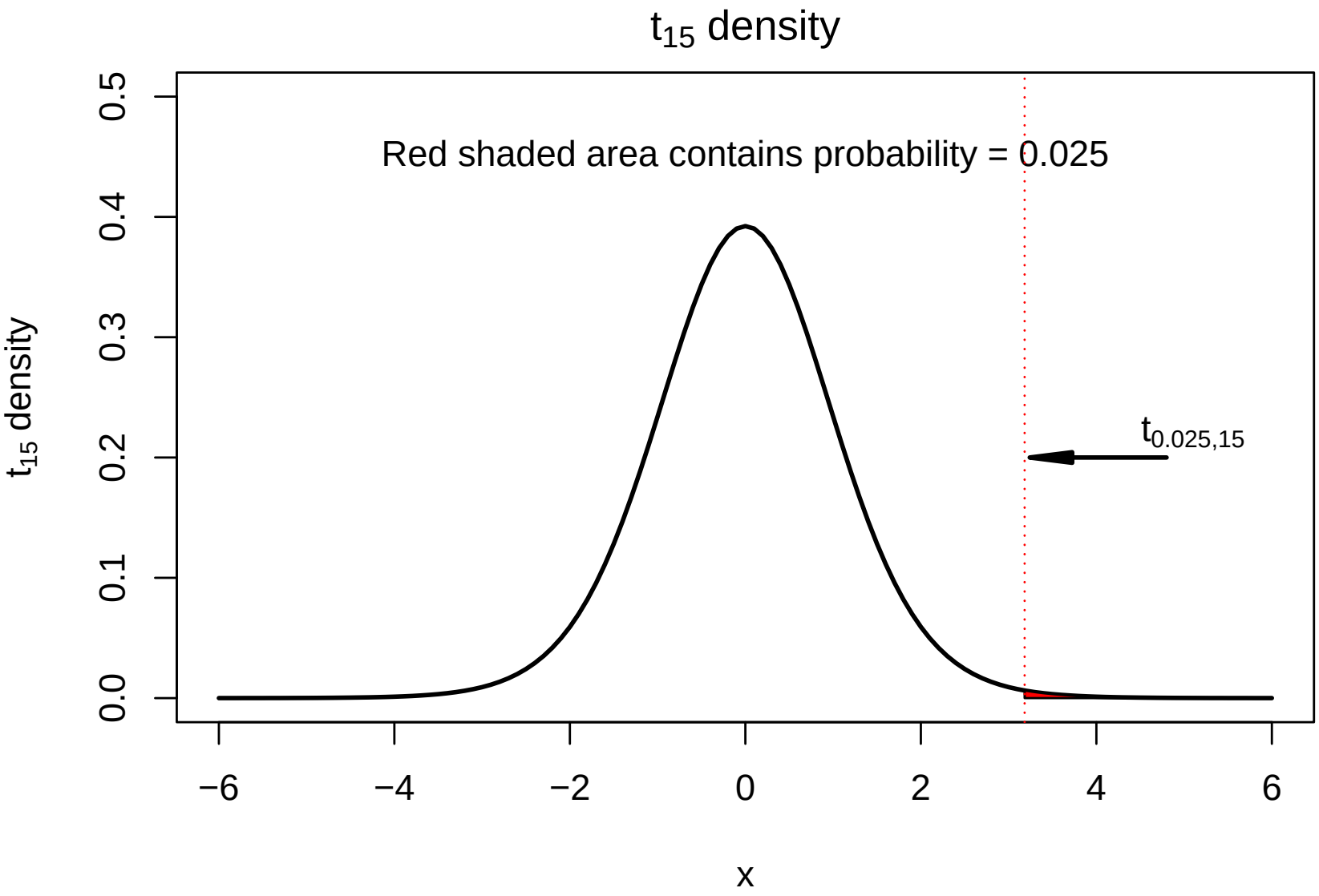
  - we use $\alpha/2$ as we are carrying out a two-tailed test
  - the total probability content of the high probability region is

$$1 - \alpha/2 - \alpha/2 = 1 - \alpha.$$

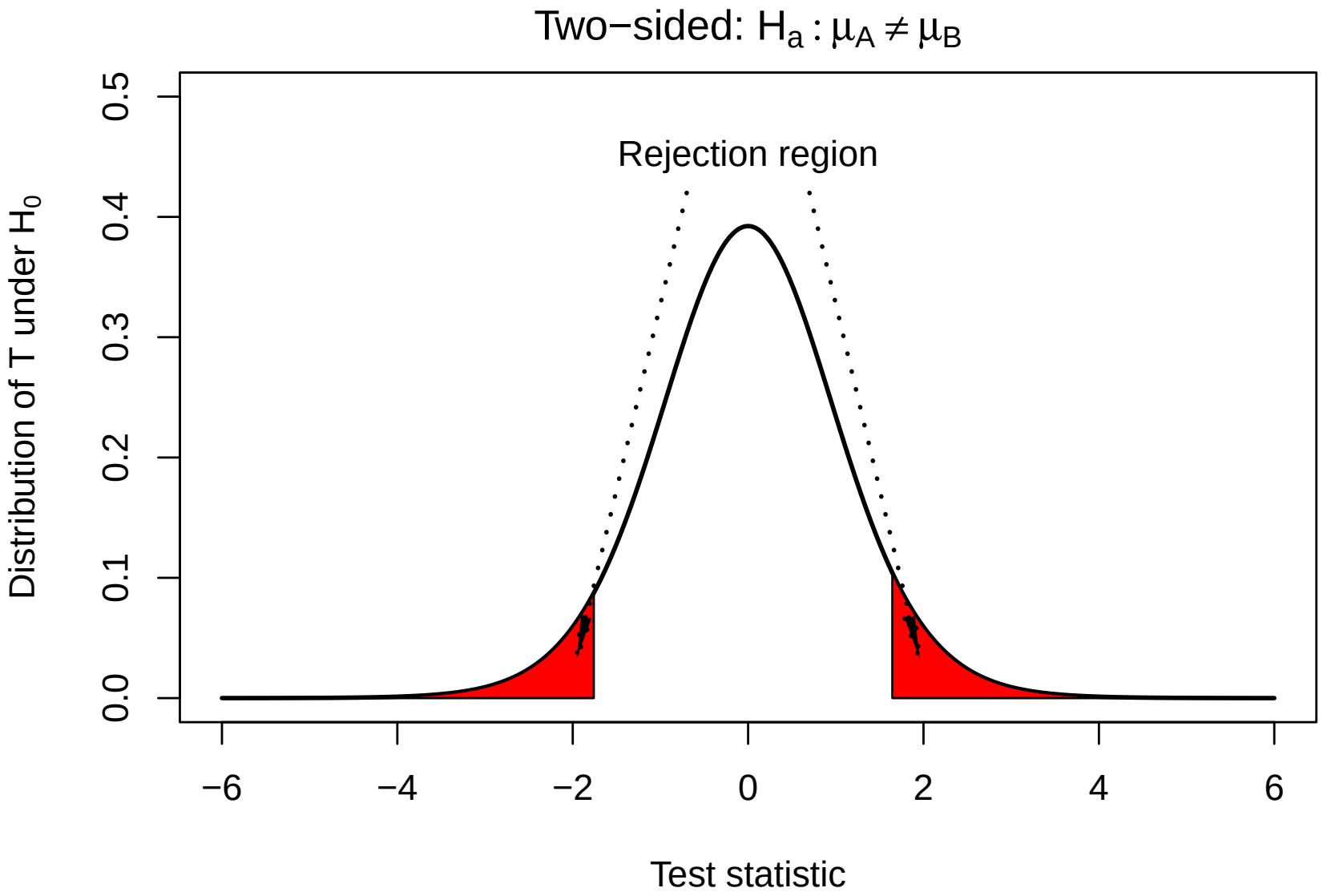
- the region outside the high probability region $(-C_R, C_R)$ is the rejection region $\mathcal{R}$.

If $T_{obs} \leqslant -C_R$ or $T_{obs} \geqslant C_R$, we reject H_0.

Statistical significance



Statistical significance



Statistical significance

Typically, $\alpha = 0.05$ (or 0.01), so for the t_{15} distribution

$$C_R = 2.131 \qquad (\text{or } 2.947)$$

As $T_{obs} = -4.683$ lies in the rejection region, we reject H_0.

Statistical significance

For 2: To compute the level of surprise, we evaluate the probability of observing a more extreme test statistic under the assumption that H_0 is true.

For the two-sided test, we report

$$2\Pr[T \geqslant |T_{obs}| | H_0 \text{ is true}] = 2P_0(T \geqslant |T_{obs}|)$$

that is, in this case

$$P_0(T \leqslant -4.683) + P_0(T \geqslant 4.683)$$

which we can compute using tables or in software.

Statistical significance

Here, this probability is

$$p = 0.00029.$$

```
> Tobs

+ [1] -4.682925

> 2*(1-pt(abs(Tobs),15))    #P-value

+ [1] 0.0002944777
```

Statistical significance

This probability is very small, so we are very surprised by the observed result.

- p is termed the p-value.
- if $p \leqslant \alpha = 0.05$ (or 0.01), we reject H_0.

Confidence Intervals

Can also compute a 95 % confidence interval for $\mu_A - \mu_B$

$$(\overline{x}_A - \overline{x}_B) \pm t_{\alpha/2, n_A + n_B - 2} s_P \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$$

where

$$t_{\alpha/2, n_A + n_B - 2} \equiv t_{0.025, 15} = 2.1314$$

is the upper 0.025 point of the t_{15} distribution.

Hence the 95 % confidence interval is

$$(-1144.59, -428.67)$$

which does not contain zero !

Further questions

Some general questions:

- How do we choose the test statistic ?
- How do we choose α ?
- Why is the distribution of T a t_{15} distribution ?
- How do we compute C_R and p ?

Equal Variances ?

Is the assumption of equal population variances

$$\sigma_A = \sigma_B$$

fair in this case ?

$$\begin{aligned} s_A^2 &= 61190.4 \\ s_B^2 &= 198679.6 \end{aligned}$$

so that

$$\frac{s_A^2}{s_B^2} = 0.3080.$$

Can we test $\sigma_A = \sigma_B$ formally ?

Equal Variances ?

Yes (see section 9.6 of textbook):

$$H_0: \quad \sigma_A = \sigma_B$$
$$H_a: \quad \sigma_A \neq \sigma_B$$

or equivalently

$$H_0: \quad \frac{\sigma_A^2}{\sigma_B^2} = 1$$

$$H_a: \quad \frac{\sigma_A^2}{\sigma_B^2} \neq 1$$

Equal Variances ?

Test statistic is

$$F = \frac{s_A^2}{s_B^2}$$

If H_0 is true, F should look like an observation from an

F-distribution

with

$$(n_A - 1, n_B - 1)$$

degrees of freedom.

Equal Variances ?

The F-distribution is just another probability distribution:

- the two degrees of freedom parameters determine the shape
- here, we have

$$n_A - 1 = 9 \qquad n_B - 1 = 6$$

  so we need the $F_{9,6}$ distribution.

Equal Variances ?

Here the observed value of the test statistic is

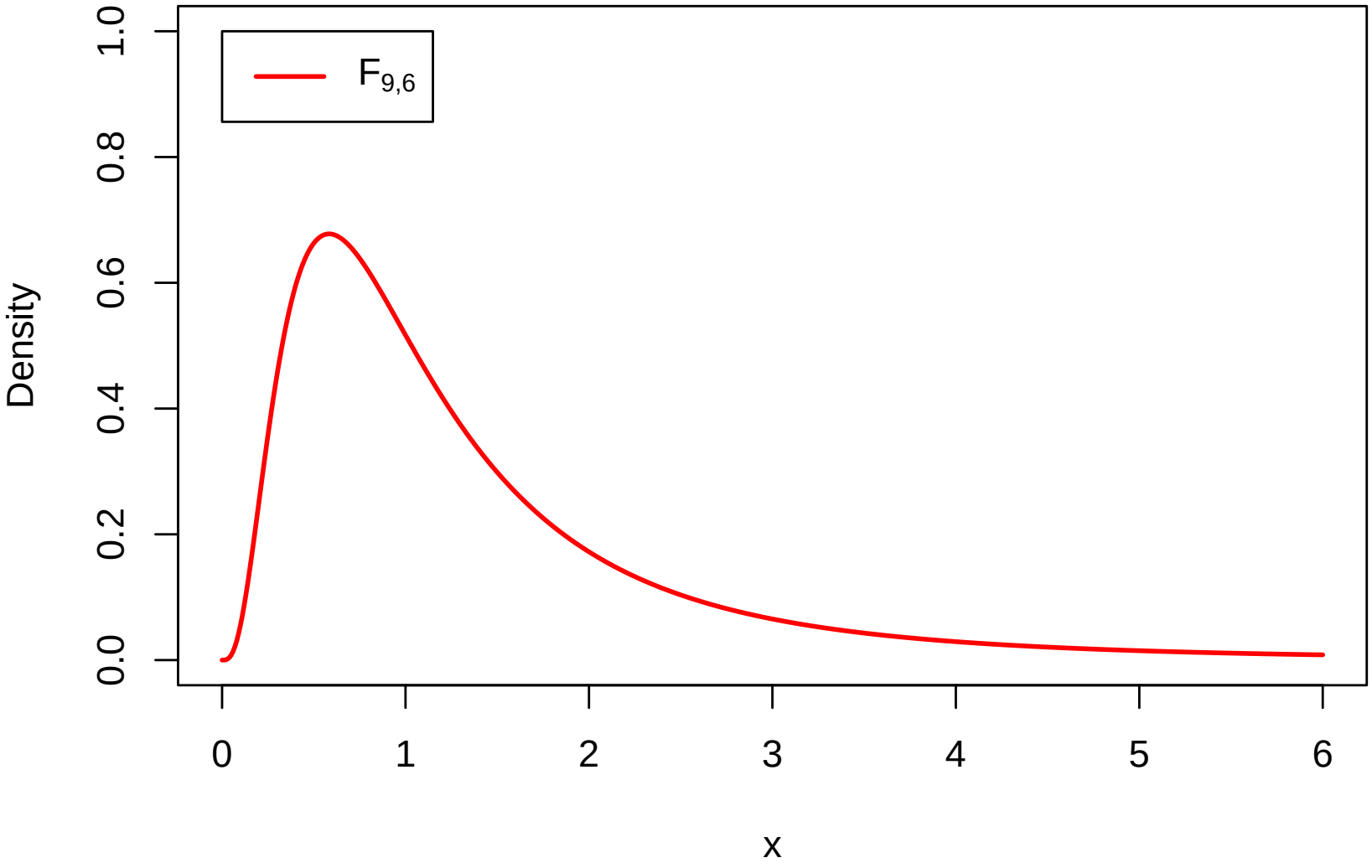
$$F_{obs} = \frac{61190.400}{198679.619} = 0.308$$

```
> Fval<-var(xA)/var(xB)
> Fval

+ [1] 0.3079853
```

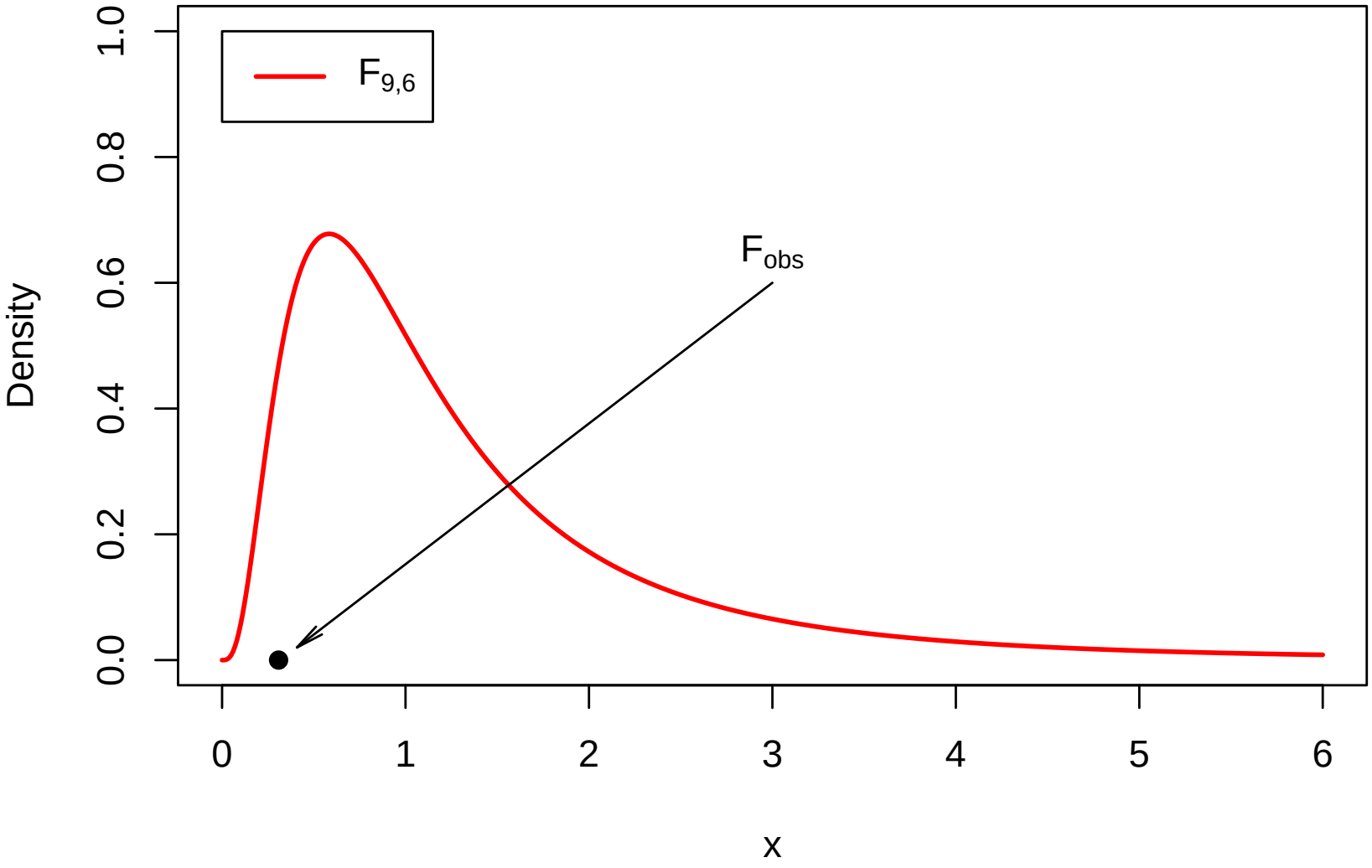
Values near 1 support H_0

- how far away can F_{obs} be before we start to prefer H_a ?
- we need to assess using the null distribution, here the $F_{9,6}$ distribution.

Equal Variances ?



Equal Variances ?



Equal Variances ?

Table: Upper $a = 0.05$ points of F_{ν_1,ν_2} distribution

	ν_1								
ν_2	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65

From textbook, p 479. Reported to 2 dp.

Equal Variances ?

Note that this is the upper tail probability: that is,

$$P_0(F > F_{a,\nu_1,nu_2}) = a$$

and entries in the table are the numerical values denoted

$$F_{a,\nu_1,\nu_2}$$

for $a = 0.05$. For example,

$$F_{0.05,9,6} = 4.10$$

from the column headed 9, row headed 6.

Equal Variances ?

However, here we are doing a two-tailed test, so we need a rejection region that reflects this:

- if F_{obs} is very small or very large, this supports H_a !
- we need a lower tail region as well as an upper tail region
- we need to consider $a = \alpha/2 = 0.025$.

Equal Variances ?

Table: Upper $a = 0.025$ points of F_{ν_1,ν_2} distribution

	ν_1								
ν_2	1	2	3	4	5	6	7	8	9
1	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21

Reported to 2 dp.

Equal Variances ?

We discover that,

$$F_{0.025,9,6} = 5.52$$

from the column headed 9, row headed 6. This handles the upper tail.

What about the lower tail ?

Equal Variances ?

It turns out that for the F-distribution we have the following result: if

$$X \sim F_{\nu_1,\nu_2}$$

and

$$Y = \frac{1}{X},$$

then

$$Y \sim F_{\nu_2,\nu_1}.$$

Equal Variances ?

This means that, for our test, for any x value

$$P_0(F \leqslant x) \equiv P_0(1/F \geqslant 1/x) \equiv P_0(1/F > 1/x)$$

where the symbol '$\equiv$' means 'is identical to'. In this expression, F is the variable representing the test statistic.

Equal Variances ?

But from the above result, we know that

$$\frac{1}{F} \sim F_{\nu_2,\nu_1}$$

which is enough information for us to compute the lower tail critical value.

Equal Variances ?

We

- Look up the $a = 0.025$ upper point of the $F_{6,9}$ distribution, and find out that

$$F_{0.025,6,9} = 4.32$$

  from the column headed 6, row headed 9.

- Take the reciprocal, to obtain that

$$F_{0.975,9,6} = \frac{1}{F_{0.025,6,9}} = \frac{1}{4.32} = 0.23$$

That is, the lower tail critical value is 0.23.

Equal Variances ?

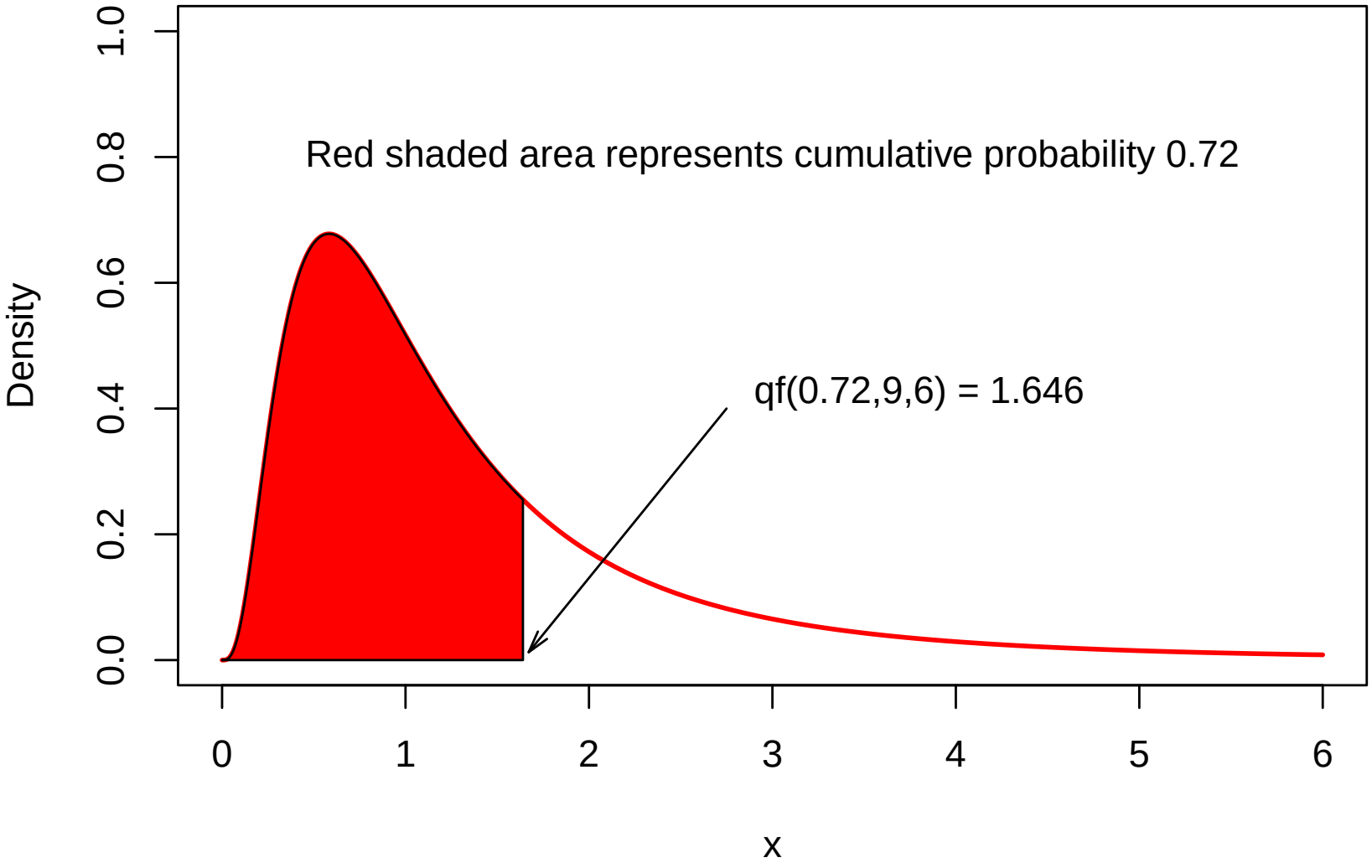
We can compute more precisely using R:

```
> CR.1<-qf(0.025,9,6)
> CR.2<-qf(0.975,9,6)
> c(CR.1,CR.2)

+ [1] 0.2314964 5.5234066

> CR.1<-1/qf(0.975,6,9)
> c(CR.1,CR.2)

+ [1] 0.2314964 5.5234066
```

The function `qf` takes three arguments

- the cumulative probability: for the upper a point, this is $1 - a$.
- the first degree of freedom, ν_1.
- the second degree of freedom, ν_2.

Equal Variances ?



Equal Variances ?

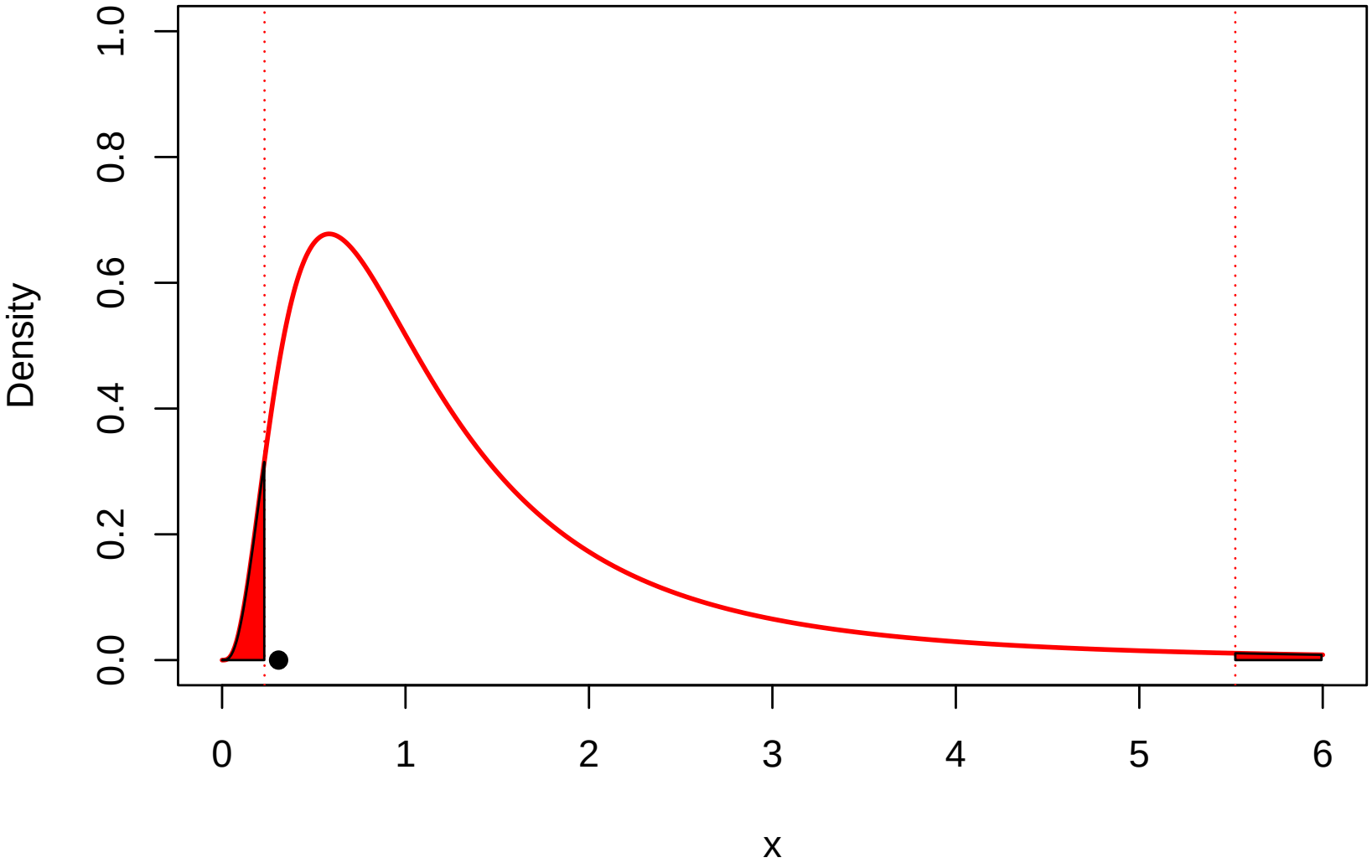
For $\alpha = 0.05$, the critical values in the test are

$$C_{R_1} = 0.231 \qquad C_{R_2} = 5.523$$

The observed value F_{obs} does lie in the high probability region, and there is no reason to reject H_0 at $\alpha = 0.05$.

For these data, there is no evidence that the population variances are different.

Equal Variances ?



Recap

This test of equal variances illustrates the general strategy for hypothesis testing:

- set up the hypotheses H_0 and H_a
- choose a suitable test statistic that allows assessment of the two hypotheses
- find the null distribution
- use the null distribution to define the rejection region
- compare the observed test statistic to the computed rejection region to complete the test

Course data sets

```
> e <- new.env(parent = emptyenv())
> load("204DataSets.rda", envir=e)
> ls(e)

+  [1] "AIDS"         "Batteries"    "Birthweight"  "Butterfat"    "Coleman"
+  [6] "Diabetes"     "DZT"          "Gravel"       "Haemo"        "HairEyes"
+ [11] "Hooker"       "Hypnosis"     "Lyrics"       "Memory"       "Milk"
+ [16] "MilkBac"      "Muco"         "PEFRAsthma"   "Potato"       "PTSD"
+ [21] "Pull"         "RBDExample"   "Sign1"        "Sign2"        "Soil"
+ [26] "TaskDistract" "Treadmill"    "Viscosity"
```

Birthweight data

```
> str(Birthweight)

+ 'data.frame': 17 obs. of  3 variables:
+  $ Group: Factor w/ 2 levels "A","B": 1 1 1 1 1 1 1 1 1 1 ...
+  $ gp   : int  1 1 1 1 1 1 1 1 1 1 ...
+  $ BW   : int  2164 2600 2184 2080 1820 2496 2184 2080 2184 2576 ...

> table(Birthweight$Group)

+
+  A  B
+ 10  7

> tapply(Birthweight$BW,Birthweight$Group,mean)

+        A        B
+ 2236.800 3023.429

> tapply(Birthweight$BW,Birthweight$Group,var)

+        A        B
+  61190.4 198679.6
```

Two-sample t-test

```
> t.test(BW~Group,data=Birthweight,var.equal=T)

+
+  Two Sample t-test
+
+ data:  BW by Group
+ t = -4.6829, df = 15, p-value = 0.0002945
+ alternative hypothesis: true difference in means between group A and group B is not equal to 0
+ 95 percent confidence interval:
+  -1144.6653  -428.5919
+ sample estimates:
+ mean in group A mean in group B
+        2236.800        3023.429
```

Test for equal variances

```
> var.test(BW~Group,data=Birthweight)

+
+  F test to compare two variances
+
+ data:  BW by Group
+ F = 0.30799, num df = 9, denom df = 6, p-value = 0.1106
+ alternative hypothesis: true ratio of variances is not equal to 1
+ 95 percent confidence interval:
+  0.05576002 1.33041078
+ sample estimates:
+ ratio of variances
+          0.3079853
```

Chapter 1
The Analysis of Variance

In this section

- Textbook Chapter 10
- introduction to the terminology of designed experiments
- extension of statistical testing theory to comparison of more than two population means
- The Analysis Of Variance (ANOVA) F-test

Designed Experiments

Studies typically fall into one of two categories:

(i) Observational studies: the experimenter has no control over the variables under study, and can only measure outcomes.

- The IQs of MAC and PC users
- The relationship between environmental exposure to toxins and health status.

i.e. The experimenter does not control the exposure to variables that may cause changes in the outcome of interest.

Designed Experiments

This type of study is common in medicine and epidemiology as it is relatively cheap to carry out.

Common type of observational study:

CASE-CONTROL STUDY

Designed Experiments

Example: Smoking and Lung Cancer

A study (Doll and Hill, 1950) investigated 649 lung cancer cases and 649 matched healthy controls, both drawn from a population of men in the UK. They found out what proportion in each group were smokers.

Neither health status nor smoking status were controlled by the experimenter, but were merely observed.

	Smokers	Non-smokers	Total
Lung cancer	647	2	649
Controls	622	27	649

Designed Experiments

This type of study can be unreliable, and cannot uncover all the relationships of interest.

A preferred approach involves the experimenter controlling the variables that cause variation in the other variables.

Note that this may not be ethical in a smoking/lung cancer study.

Designed Experiments

(ii) Designed experiments: the experimenter control the levels of variables that may affect the variable of interest.

For designed experiments, sometimes called experimental studies, the experimenter intervenes to assign the levels of influential variables such as

- dose levels of therapeutic treatments
- types of treatment

Designed Experiments

Example: Birthweight study

Recall

$$\begin{aligned} \text{GROUP A} &: \text{5 or fewer visits} \\ \text{GROUP B} &: \text{More than 5 visits.} \end{aligned}$$

If at the control of

(a) mothers, then this is an observational study

(b) doctors, this is a designed experiment

- for example, after each mother is recruited to take part in the study, they are randomly assigned to either GROUP A or GROUP B.

Designed Experiments

This is termed a

randomized experimental study

This type of study is preferable, but can be more difficult to implement.

Terminology

- Response variable (dependent variable):
  - the variable of interest in the study
- Factors : the variables that may have an effect on the response variable
  - quantitative if measured on a numerical scale
  - qualitative otherwise

Terminology

- Factor Levels: the values of the factors utilized in the experiment
- Treatments: the factor-level combinations utilized.
- Experimental Units (subjects):
  - the objects on which the factors are measured or observed.

Terminology

Therefore:

- A designed experiment is one for which the analyst or experimenter controls the specification of treatments and the method of assigning units to treatments.

- An observational experiment or study is one for which the analyst simply observes the treatments and response on a sample of experimental units.

Terminology

Example: Birthweight study

- Response: Birthweight (g)
- Factor: Pre-natal treatment group
- Factor levels: GROUP A or GROUP B

that is, we have a single factor with two factor levels.

Terminology

Example: Test scores

Test scores of students in four schools for two types of test.

- Response: Test score
- Factors: TEST and SCHOOL (both qualitative)
- Factor levels:
  - TEST: 1, 2
  - SCHOOL: A, B, C, D

that is, we have a two factors, TEST with two factor levels and SCHOOL with four factor levels.

There are 8 possible treatments:

$$(1, A), (1, B), (1, C), (1, D), (2, A), (2, B), (2, C), (2, D)$$

Terminology

https://cihr-irsc.gc.ca/e/50833.html

Terminology

Example: Pain Relief

Different pain relief remedies are to be compared.

- REMEDY (quantitative/qualitative, 3 levels)
  - Dose levels 0, 1, 2
- AGE GROUP (quantitative/qualitative, 4 levels)
  - 0-16 years
  - 17-40 years
  - 41-65 years
  - 66 years and over
- SEX (qualitative, 2 levels)
  - Female
  - Male

Terminology

In this example, there are a total of $3 \times 4 \times 2 = 24$ possible treatments;

- REMEDY is the only factor that can be assigned by the analyst
- We might consider stratification into the 4×2 strata defined by the other two factors.

The Completely Randomized Design

A completely randomized design (CRD) is a design for which treatments are randomly assigned to experimental units, or in which random samples of experimental units are selected for each treatment.

- if the 'treatments' are DOSE 0/DOSE 1, a CRD assigns experimental units independently to the two treatment groups at random.

Statistical Objectives

The experimental units assigned to different treatments (that is, factor-level combinations) form

independent samples

from

different populations

in a CRD.

We wish to compare treatments: specifically, we wish to compare the treatment-specific population means.

A Multiple Group Comparison of Means !

Comparing k Treatments

Suppose that there are k treatments:

TMT 1	Mean μ_1
TMT 2	Mean μ_2
$\vdots$	$\vdots$
TMT k	Mean μ_k

We wish to test the hypotheses

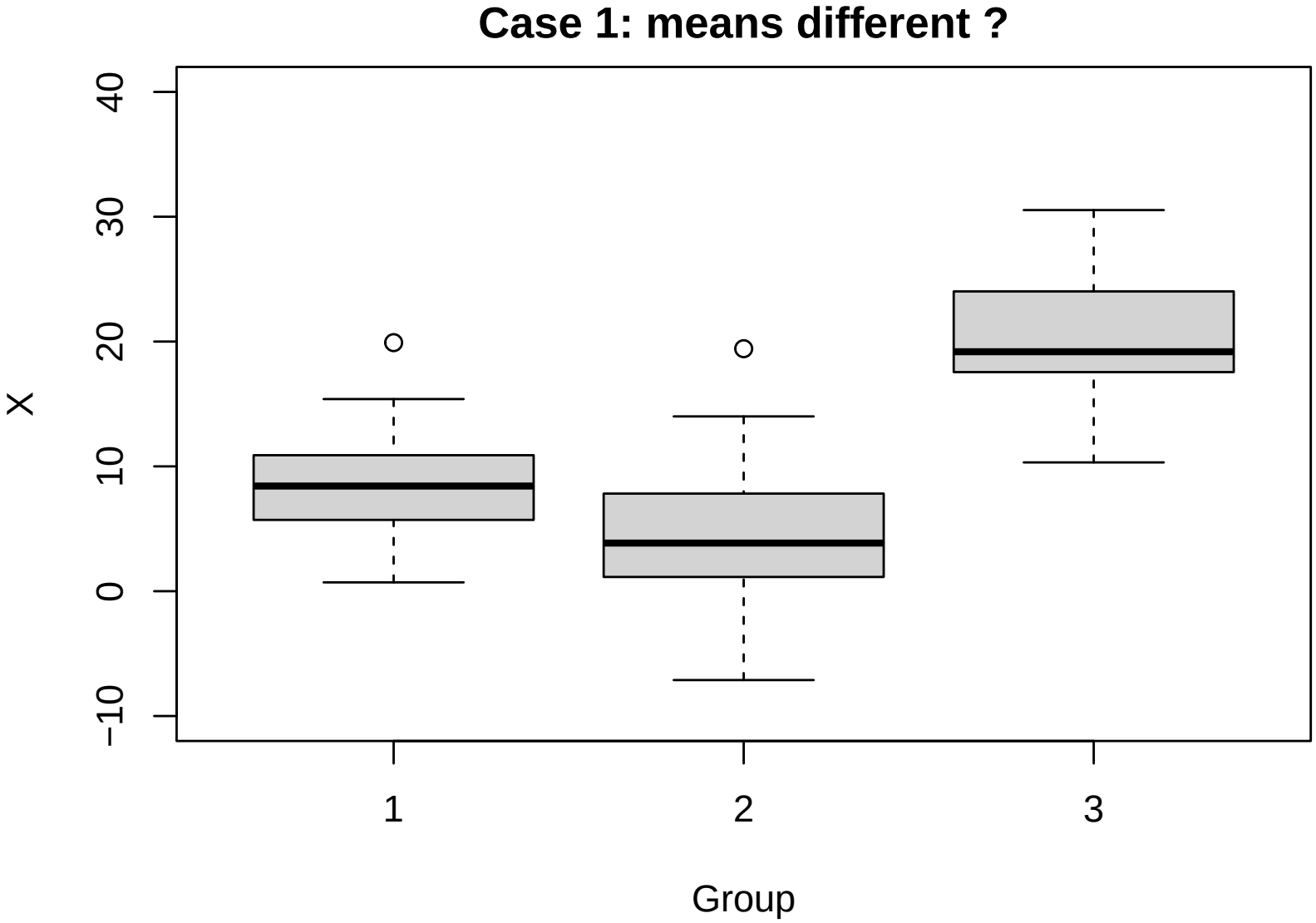
$$H_0 \quad : \quad \mu_1 = \mu_2 = \cdots = \mu_k$$

$$H_a \quad : \quad \text{At least two of the } k \text{ treatment means are different}$$

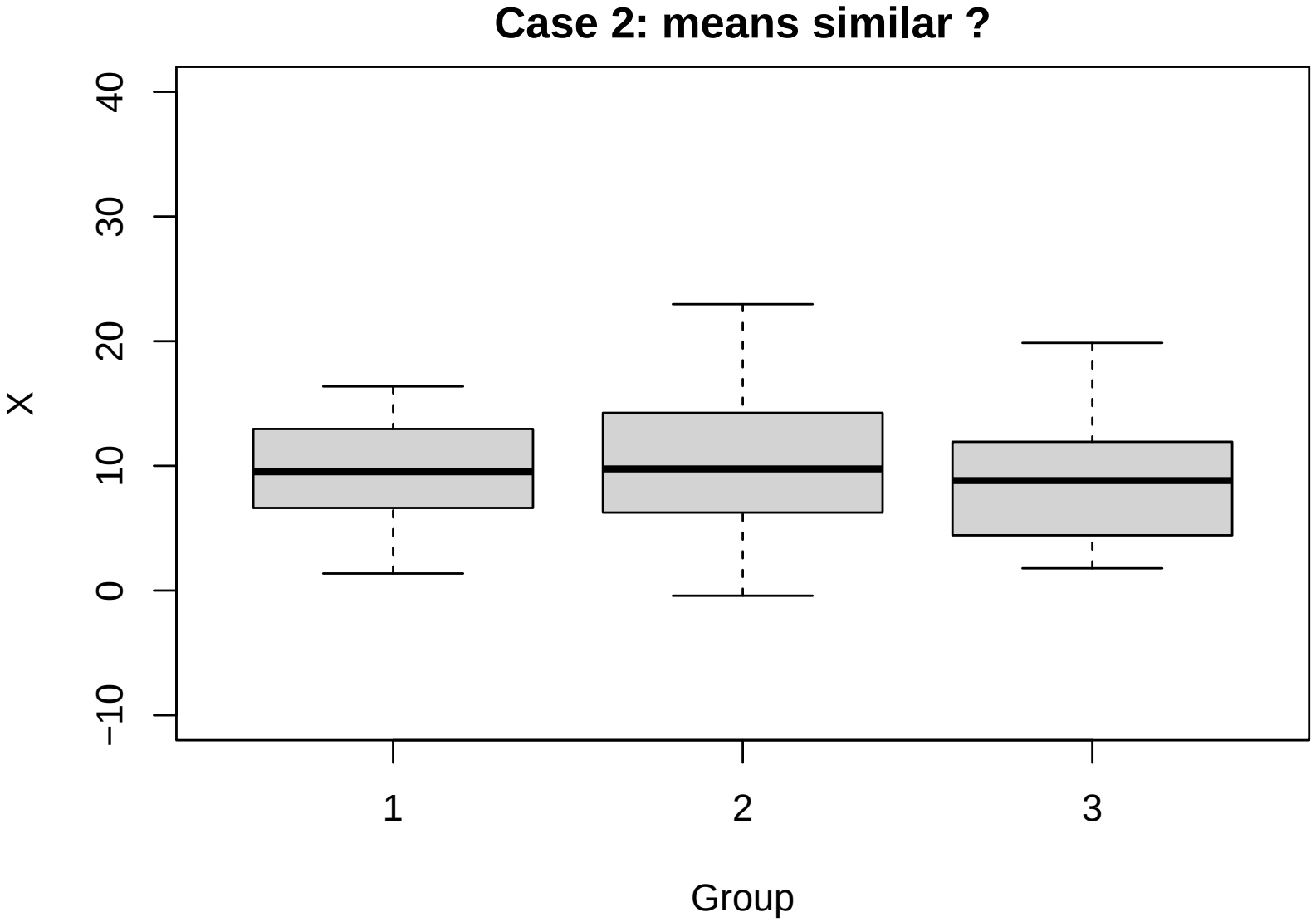
How do we do this ?

What is the relevant test statistic ?

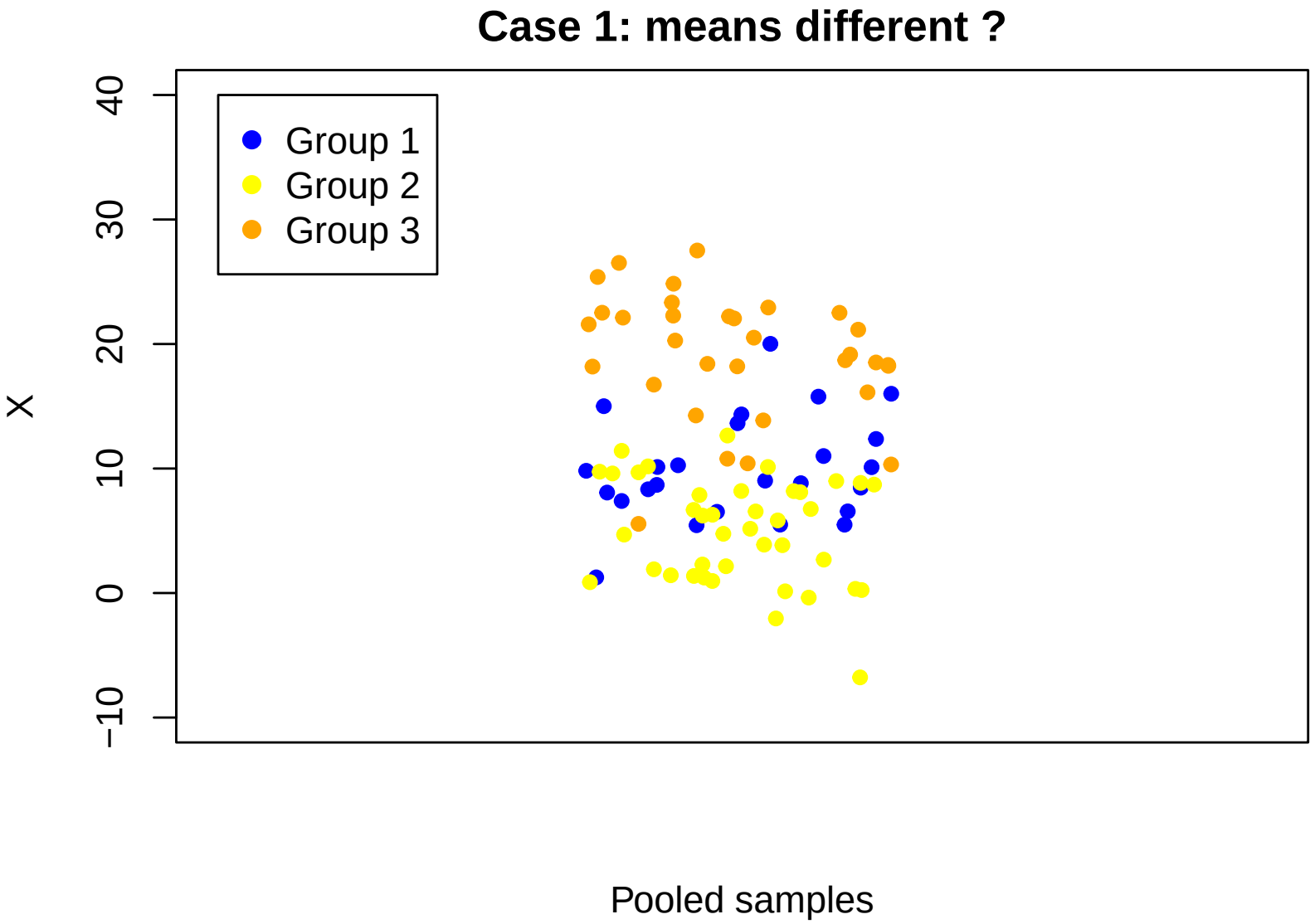
Comparing k Treatments



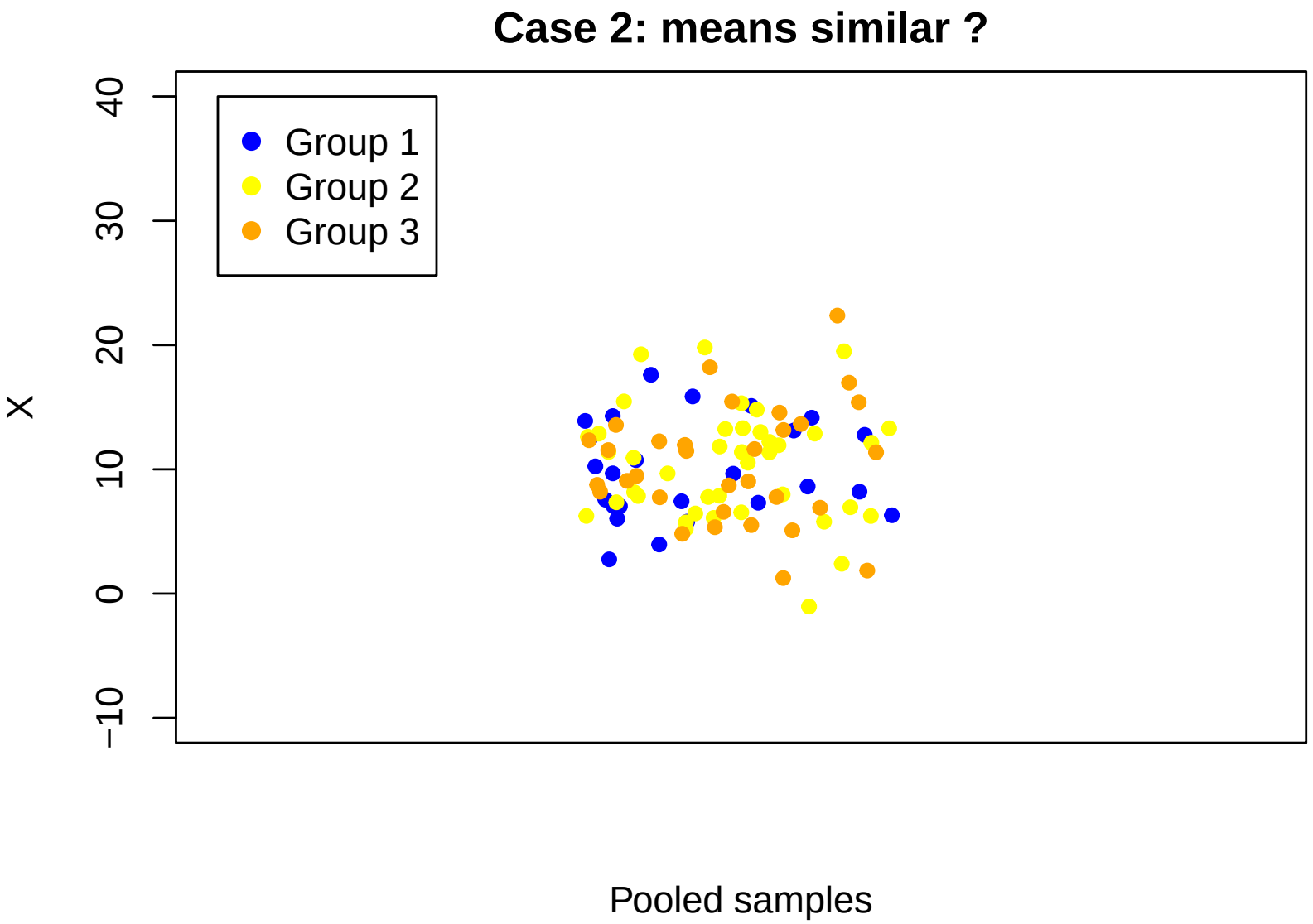
Comparing k Treatments



Comparing k Treatments



Comparing k Treatments



Comparing k Treatments

We want to find a way to compare

the between group variability

with the

the overall variability

we observe in the data.

How much of the overall variability is due to the fact that the means are different ?

The ANOVA F-test for CRDs

We attempt to generalize the earlier two sample comparisons:

- find a suitable test statistic
- predict how it should behave if H_0 is true
- compute the observed value of the test statistic
- compare the observed value to the rejection region.

The ANOVA F-test for CRDs

Suppose

$$\begin{array}{ll} \text{TMT } 1 & \text{has } n_1 \text{ experimental units} \\ \text{TMT } 2 & \text{has } n_2 \text{ experimental units} \\ \vdots & \vdots \\ \text{TMT } k & \text{has } n_k \text{ experimental units} \end{array}$$

Denote by x_{ij} the response for unit j in treatment group i, for $j = 1, \ldots, n_i$ and $i = 1, \ldots, k$.

The ANOVA F-test for CRDs

Our model for how the data are generated is

$$X_{ij} = \mu_i + \epsilon_{ij}$$

where ϵ_{ij} are independent residual errors.

We typically assume that $\epsilon_{ij} \sim Normal(0, \sigma_i^2)$, so that

$$X_{ij} \sim Normal(\mu_i, \sigma_i^2).$$

The ANOVA F-test for CRDs

Let

$$\overline{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

denote the sample mean for treatment i, and

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \overline{x}_i)^2$$

denote the sample variance for treatment i.

Now we consider pooling, that is, combining all units into a single group.

The ANOVA F-test for CRDs

Define

- the total sample size

$$n = n_1 + \cdots + n_k = \sum_{i=1}^{k} n_i$$

- the overall sample mean

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}$$

- the overall sample variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(x_{ij} - \overline{x}\right)^2$$

The ANOVA F-test for CRDs

Finally, consider the pooled sample variance

$$s_P^2 = \frac{1}{n-k} \sum_{i=1}^{k} (n_i - 1) s_i^2$$

This is the extension of the pooled estimate of the population variance in a two-sample t-test.

Using these quantities, we can derive a test statistic for multiple group comparison.

The ANOVA F-test for CRDs

We wish to compare how much variation is due to the

A difference between treatments

and how much is due to

B random variation within treatments

The ANOVA F-test for CRDs

We measure $\boxed{\text{A}}$ using the statistic

$$SST = \sum_{i=1}^{k} n_i \left(\overline{x}_i - \overline{x}\right)^2$$

SST - <u>S</u>um of <u>S</u>quares for <u>T</u>reatments

The ANOVA F-test for CRDs

We measure $\boxed{\text{B}}$ using the statistic

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \overline{x}_i)^2 = (n-k)s_P^2$$

SSE - <u>S</u>um of <u>S</u>quares for <u>E</u>rror

The ANOVA F-test for CRDs

NOTE: This measure of random or error variability implicitly assumes that the variability within the treatment groups is the same for each group. That is, population variances

$$\sigma_1^2, \ldots, \sigma_k^2$$

are equal, so that

$$X_{ij} \sim Normal(\mu_i, \sigma^2)$$

for all i and j.

In practice this assumption must be checked.

The ANOVA F-test for CRDs

We define the test statistic using the average levels of variability:

- MST - Mean Square for Treatments

$$MST = \frac{SST}{k-1} = \frac{1}{k-1}\sum_{i=1}^{k} n_i\left(\overline{x}_i - \overline{x}\right)^2$$

- MSE - Mean Square for Error

$$MSE = \frac{SSE}{n-k} = \frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \overline{x}_i\right)^2 = s_P^2$$

The ANOVA F-test for CRDs

Then the test statistic is

$$F = \frac{MST}{MSE} = \frac{\text{Average Variation due to Treatments}}{\text{Average Variation due to Errors}}$$

F large $\Longrightarrow$ Treatments Different !
F small $\Longrightarrow$ Treatments Similar !

The behaviour of F is given by the following Theorem.

The ANOVA F-test for CRDs

Theorem: ANOVA F-test to compare k treatments in a Completely Randomized Design

To test the hypothesis of equal treatment means,

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \cdots = \mu_k \\ H_a &: \text{At least two of the } k \text{ treatment means are different} \end{aligned}$$

the test statistic is

$$F = \frac{MST}{MSE}$$

The ANOVA F-test for CRDs

Theorem: ANOVA F-test to compare k treatments in a Completely Randomized Design

If H_0 is true, then

$$F \sim F_{k-1,n-k}$$

and the rejection region for a test at significance level α is the region to the right of the upper α point of this F distribution.

The ANOVA F-test for CRDs

Note: The ANOVA F-test is a one-tailed test – we reject H_0 if

$$F_{obs} \geqslant F_{\alpha,k-1,n-k}$$

- large values of F_{obs} support H_a
- small values of F_{obs} support H_0.

The ANOVA F-test for CRDs

Note: If

$$SS = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \overline{x})^2$$

is the overall or total sum of squares, then we have that

$$SS = SST + SSE$$

so we can decompose the overall variation (SS) into the variation due to treatments (SST) and the variation due to the errors (SSE).

Assumptions behind the ANOVA F-test

There are three assumptions:

1. The samples are randomly selected in an independent manner from the k treatment populations.

2. All k treatment groups have distributions that are Bormal (at least approximately).

3. The k treatment group variances are equal:

$$\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2.$$

Example

Example: Milk Quality Data

The impact on milk protein level of three different diets is being studied.

Data: Measurements of protein levels for $n = 1337$ samples.

- Response: Milk Protein Level (%)
- Factor: DIET
- Factor levels: $k = 3$
  - 1: Barley
  - 2: Barley + Lupins
  - 3: Lupins

Example

Data summary:

	TMT 1	TMT 2	TMT 3
n_i	425	459	453
$\overline{x}_i$	3.532	3.430	2.312
s_i^2	0.102	0.091	0.114

$$\begin{aligned} SST &= 10.606 \\ SSE &= 136.432 \\ SS &= 147.038 \end{aligned}$$

$$\begin{aligned} k-1 &= 2 \\ n-k &= 1334 \end{aligned}$$

Example

Therefore

$$MST = \frac{SST}{k-1} = \frac{10.606}{2} = 5.303$$

$$MSE = \frac{SSE}{n-k} = \frac{136.432}{1334} = 0.102$$

and

$$F_{obs} = \frac{MST}{MSE} = 51.851$$

Example

If H_0 is true, that is,

$$\mu_1 = \mu_2 = \mu_3$$

then F should look like an observation from a

$$F_{k-1,n-k}$$

distribution. Here we are dealing with the

$$F_{2,1334}$$

distribution.

Example

From R, we discover that if $\alpha = 0.05$, then the relevant upper point is

$$F_{0.05,2,1334} = 3.002$$

and thus we

$$\text{Reject } H_0$$

and conclude that there is a significant impact on milk protein level due to diet.

Example

Note: Table VI in the textbook (p 822) only gives

$$\begin{aligned} F_{0.05,2,120} &= 3.07 \\ F_{0.05,2,\infty} &= 3.00 \end{aligned}$$

so we cannot look up $F_{0.05,2,1334}$ exactly. However, we do know that

$$3.00 < F_{0.05,2,1334} < 3.07$$

and here the test statistic is $F_{obs} = 51.851$.

Are the assumptions met ?

1. Independent samples ? Not possible to tell with current information. In fact, data comprise repeated measurements on 79 cows - potentially not independent, as observations on the same cow are likely to be more similar.

2. Normal Distributions ? Visual inspection of boxplots indicates that this may be valid.

3. Equal variances ? : group sample variances are

$$s_1^2 = 0.102 \qquad s_2^2 = 0.091 \qquad s_3^2 = 0.114$$

so assumption appears to be valid.

Milk data: Analysis in R

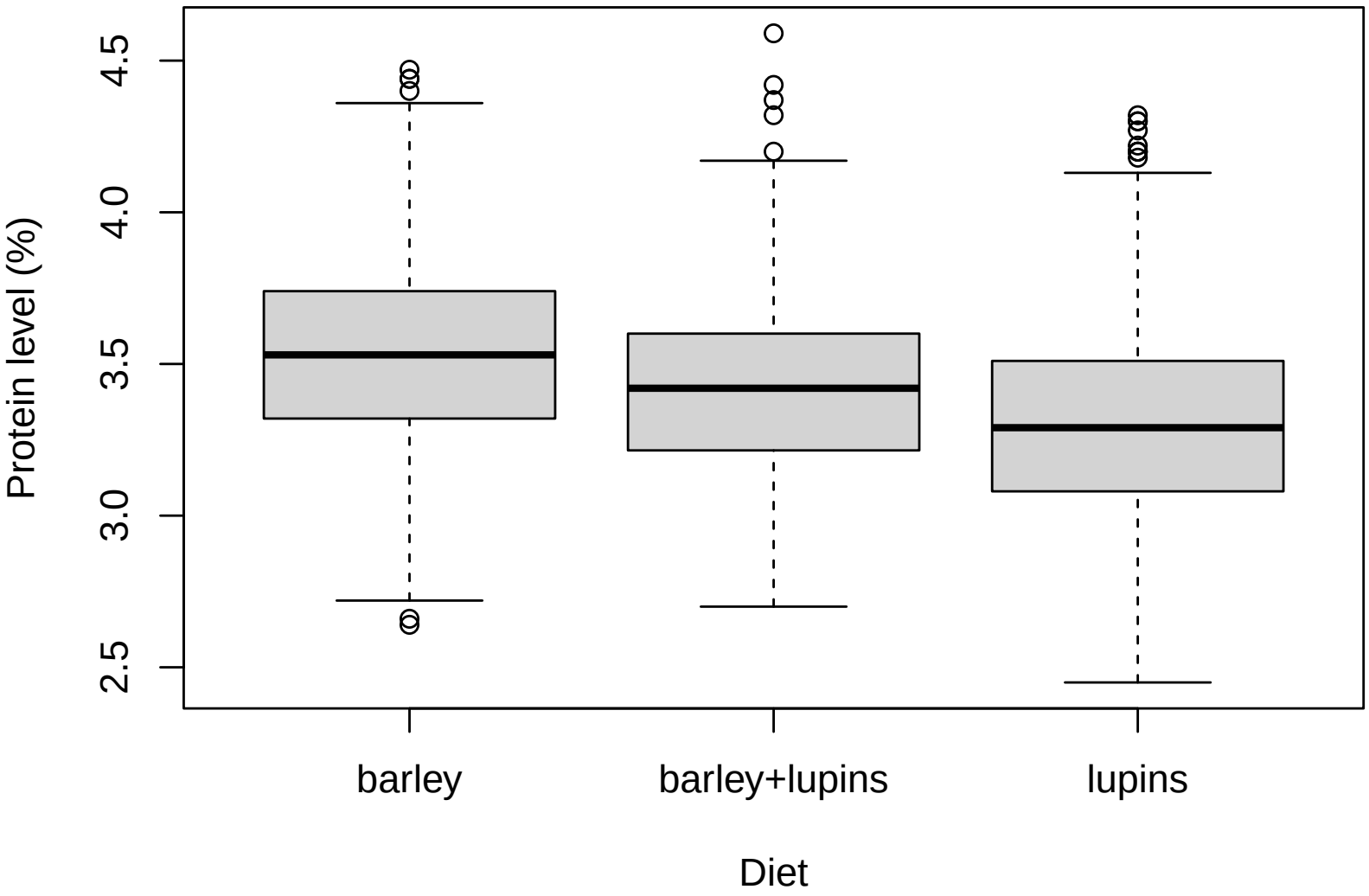
```
> head(Milk)

+   protein Time Cow   Diet
+ 1    3.63    1 B01 barley
+ 2    3.57    2 B01 barley
+ 3    3.47    3 B01 barley
+ 4    3.65    4 B01 barley
+ 5    3.89    5 B01 barley
+ 6    3.73    6 B01 barley
```

Milk data: Analysis in R

Summary statistics by treatment group:

```
> tapply(Milk$protein,Milk$Diet,length)

+        barley barley+lupins        lupins
+           425           459           453

> tapply(Milk$protein,Milk$Diet,mean)

+        barley barley+lupins        lupins
+      3.531929      3.429695      3.312384

> tapply(Milk$protein,Milk$Diet,var)

+        barley barley+lupins        lupins
+    0.10189721    0.09141126    0.11363103

> #Boxplot
> par(mar=c(4,4,2,2))
> boxplot(protein~Diet,data=Milk, ylab='Protein level (%)')
```

Milk data: Analysis in R



Milk data: Analysis in R

Compute the total sum of squares (SS):

```
> xbar<-mean(Milk$protein)
>
> SS<-sum((Milk$protein-xbar)^2)
> SS

+ [1] 147.0379
```

Milk data: Analysis in R

Compute the sum of squares for treatments (SST):

```
> nvec<-tapply(Milk$protein,Milk$Diet,length)
> mvec<-tapply(Milk$protein,Milk$Diet,mean)
> SST<-sum(nvec*(mvec-xbar)^2)
> SST

+ [1] 10.6059
```

Milk data: Analysis in R

Compute the sum of squares for error (SSE):

```
> svec<-tapply(Milk$protein,Milk$Diet,var)
> SSE<-sum((nvec-1)*svec)
> SSE

+ [1] 136.432
```

Note that

$$SS = SST + SSE$$

that is

$$147.0379023 = 10.605902 + 136.4320003$$

Milk data: Analysis in R

Using aov and anova functions

```
> summary(aov(protein~1,data=Milk))

+               Df Sum Sq Mean Sq F value Pr(>F)
+ Residuals   1336    147  0.1101

> anova(lm(protein~1,data=Milk))

+ Analysis of Variance Table
+
+ Response: protein
+             Df Sum Sq Mean Sq F value Pr(>F)
+ Residuals 1336 147.04 0.11006
```

Milk data: Analysis in R

The result of the F-test is also computed

```
> summary(aov(protein~Diet,data=Milk))

+               Df Sum Sq Mean Sq F value Pr(>F)
+ Diet           2  10.61   5.303   51.85 <2e-16 ***
+ Residuals   1334 136.43   0.102
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Milk data: Analysis in R

Look up the critical value using the `qf` function, and compute the p-value using the `pf` function.

```
> qf(0.95,2,1334)

+ [1] 3.00247

> 1-pf(51.85,2,1334)

+ [1] 0
```

Note that the `pf` function computes the cumulative probability (that is, in the left-tail of the distribution); for the p-value, we need the right-tail probability.

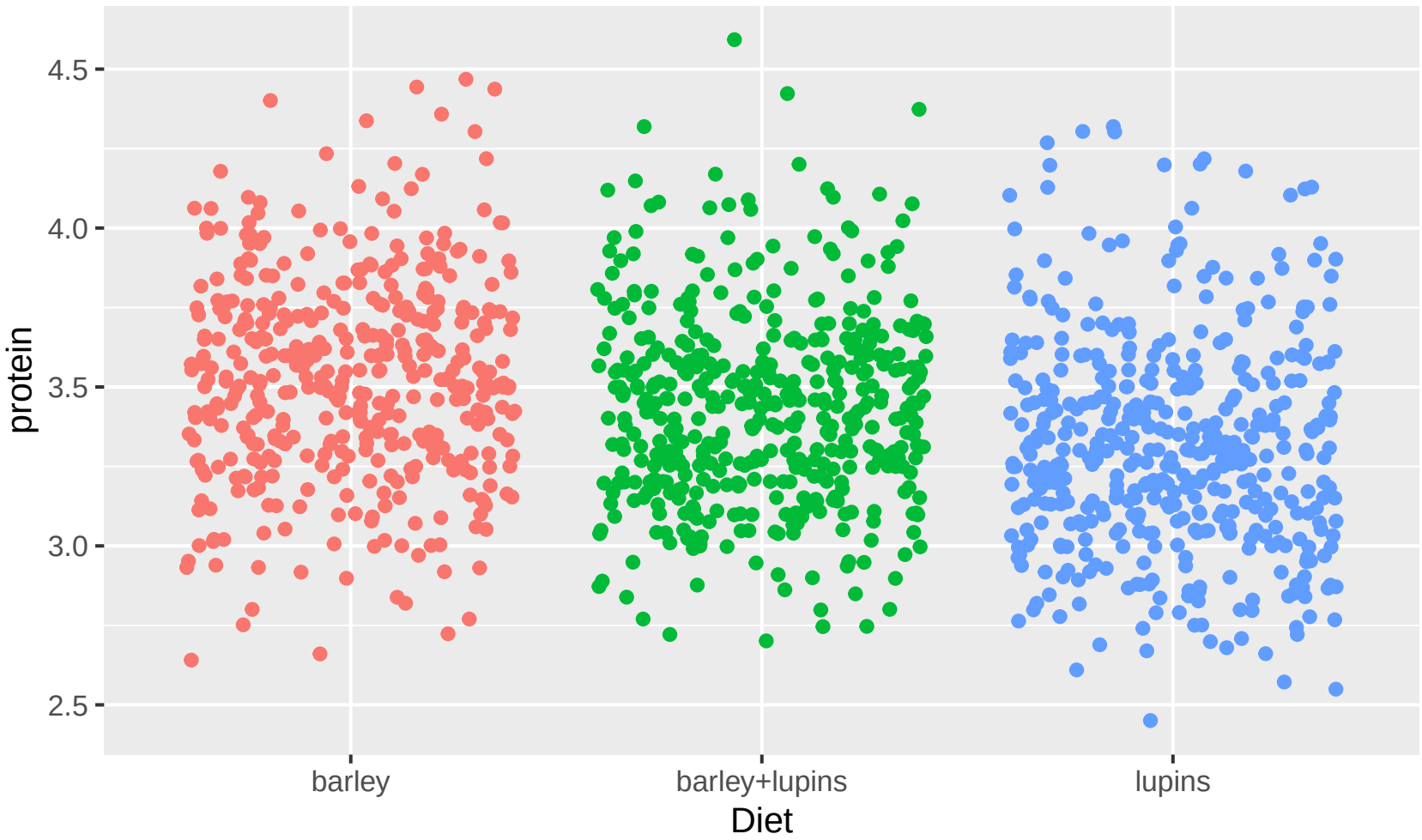
Milk data: Analysis in R

```
> library(supernova)
> supernova(lm(protein~Diet,data=Milk))

+  Analysis of Variance Table (Type III SS)
+  Model: protein ~ Diet
+
+                              SS   df    MS      F   PRE     p
+  ----- --------------- | ------- ---- ----- ------ ----- -----
+  Model (error reduced) |  10.606    2 5.303 51.851 .0721 .0000
+  Error (from model)    | 136.432 1334 0.102
+  ----- --------------- | ------- ---- ----- ------ ----- -----
+  Total (empty model)   | 147.038 1336 0.110
```

Milk data: Analysis in R

```
> library(ggplot2)
> ggplot(Milk) +
+   aes(x = Diet, y = protein, color = Diet) +
+   geom_jitter() +
+   theme(legend.position = "none")
```

Milk data: Analysis in R



Example: Anxiety Response

Example: Anxiety Response Treatment

In a study of Alzheimer's disease and care of its sufferers, a medication designed to improve anxiety relief has been developed.

In a lab experiment, $n = 20$ rats were assigned to one of four ($k = 4$) treatment groups corresponding to dose-level of the medication.

A measure of response to a "flee stimulus" was recorded.

Example: Anxiety Response

Example: Anxiety Response Treatment

- Response: Pull response to stimulus (units of force)
- Factor: DOSE-LEVEL
- Factor levels: $k = 4$
  - Dose 0 (zero units)
  - Dose 1 (one unit)
  - Dose 2 (two units)
  - Dose 3 (three units)

Example: Anxiety Response

Data:

0	1	2	3
27.0	22.8	21.9	23.5
26.2	23.1	23.4	19.6
28.8	27.7	20.1	23.7
33.5	27.6	27.8	20.8
28.8	24.0	19.3	23.9

We have $k = 4$ and $n = 20$, so

$$k - 1 = 3 \qquad n - k = 16.$$

Example: Anxiety Response

We find that

$$SST = 140.094 \qquad SSE = 116.324 \qquad SS = 256.418$$

$$MST = 46.698 \qquad MSE = 7.270$$

and

$$F_{obs} = 6.423$$

which we need to compare with the $F_{3,16}$ distribution.

Example: Anxiety Response

For $\alpha = 0.05$, from textbook tables

$$F_{0.05,3,16} = 3.24$$

and so we

$$\text{Reject } H_0$$

at $\alpha = 0.05$ and conclude that there is a significant difference between treatment groups.

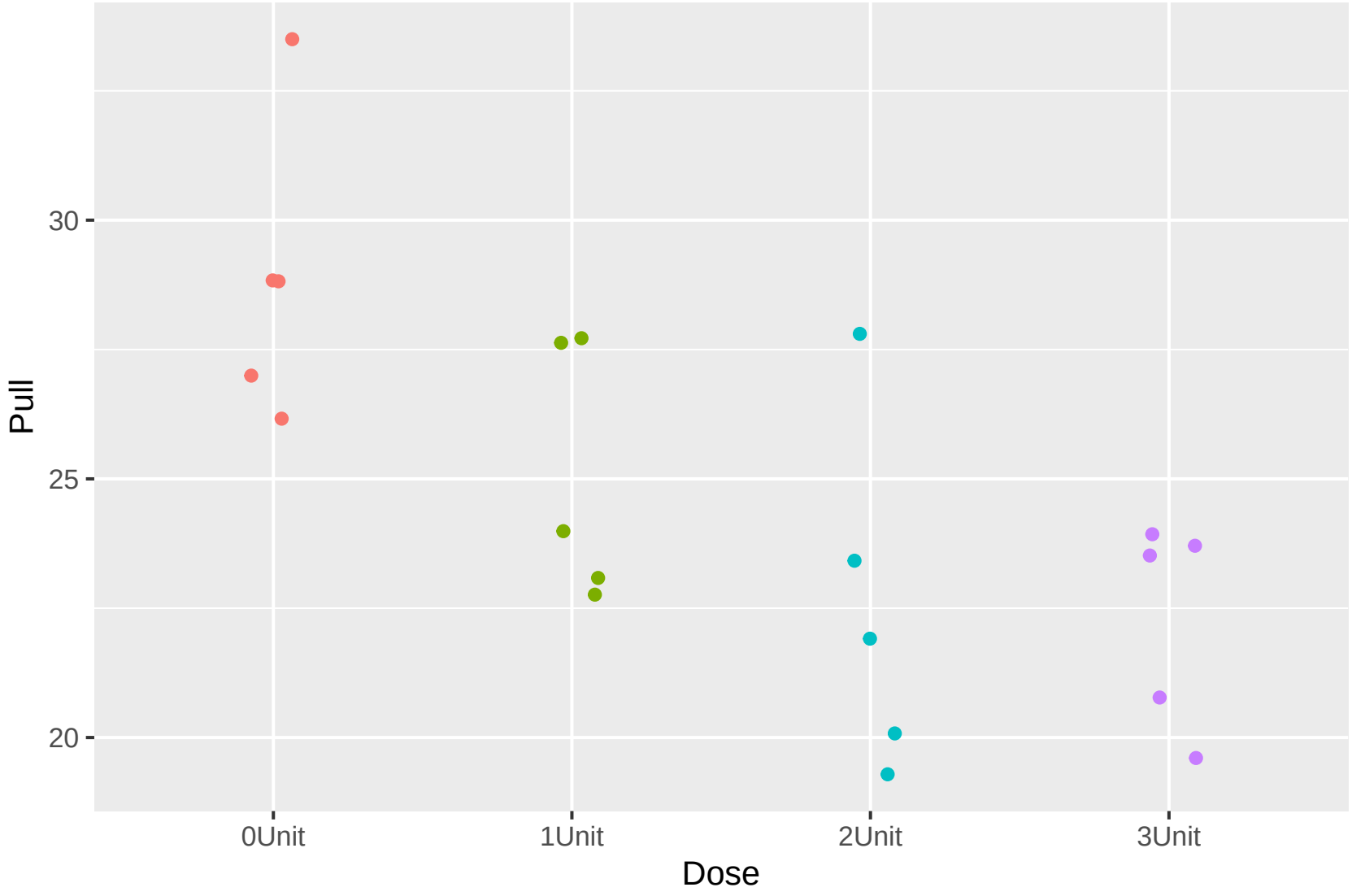
p-value is 0.0046.

Example: Anxiety Response

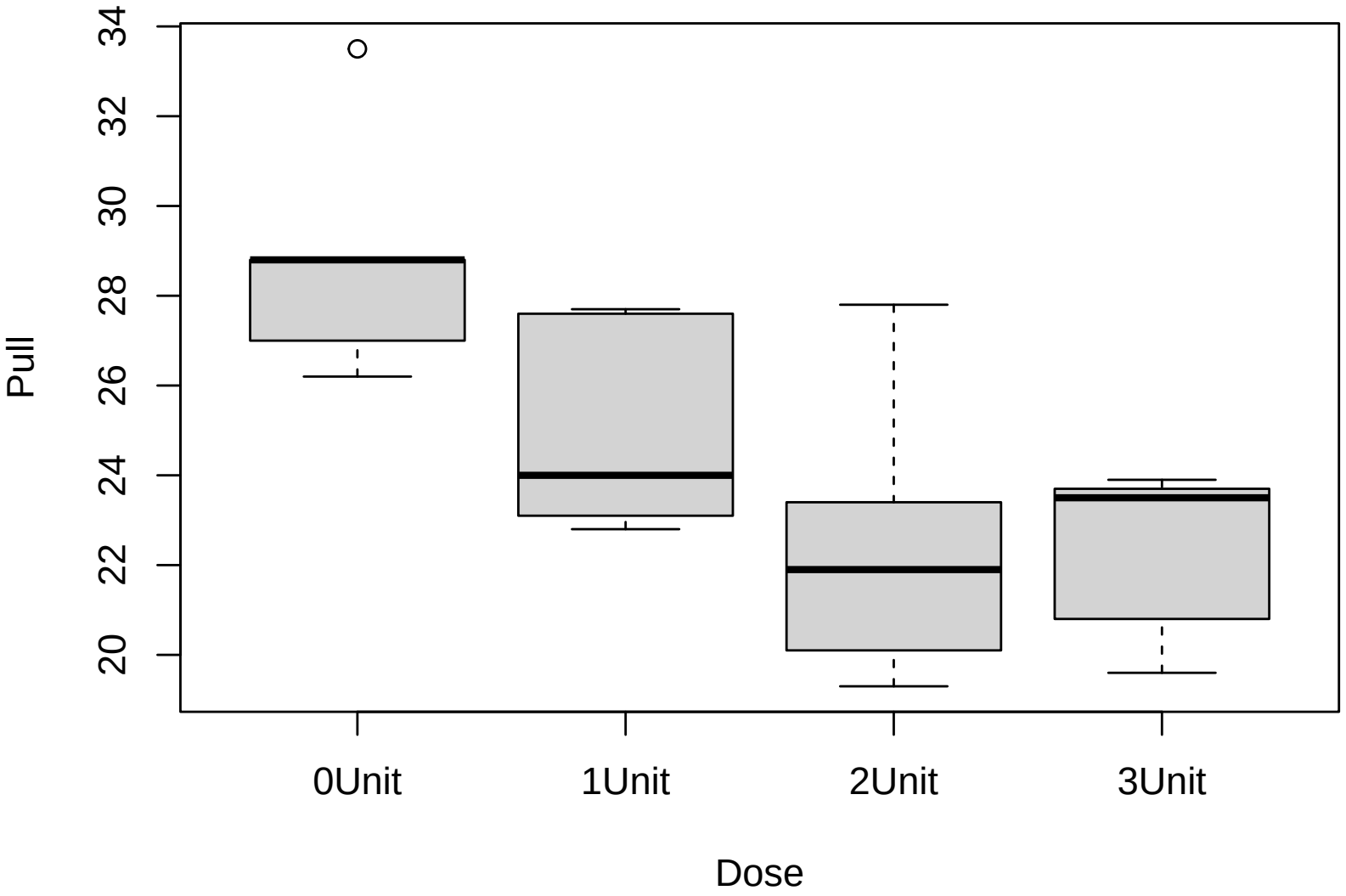
Summaries:

```
> tapply(Pull$Pull,Pull$Dose,mean)  #Group Means

+ 0Unit 1Unit 2Unit 3Unit
+ 28.86 25.04 22.50 22.30

> tapply(Pull$Pull,Pull$Dose,var)   #Group Variances

+  0Unit  1Unit  2Unit  3Unit
+  8.018  5.873 11.315  3.875
```

Example: Anxiety Response



Example: Anxiety Response



Example: Anxiety Response

```
> summary(aov(Pull~Dose,data=Pull))

+             Df Sum Sq Mean Sq F value  Pr(>F)
+ Dose         3  140.1   46.70   6.423 0.00462 **
+ Residuals   16  116.3    7.27
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> qf(0.95,3,16)

+ [1] 3.238872

> 1-pf(6.423,3,16)

+ [1] 0.004623413
```

Assessing unequal group variances

Here

	DOSE 0	DOSE 1	DOSE 2	DOSE 3
s_i^2	8.018	5.873	11.315	3.875

so we might suspect that the treatment variances $\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$ are not equal. We may test this formally using

Levene's test.

Assessing unequal group variances

Note: Visual inspection can give an idea of whether the equal variance assumption is valid, or whether the populations are normal.

However, the sample sizes may be small, so that visual inspection or testing may not detect deviations from these assumptions.

Ideally we would like to be able to relax these assumptions.

Assessing unequal group variances

Levene's Test: To test

$$\begin{aligned} H_0 &: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 \\ H_1 &: \text{At least one pair of } \sigma^2 \text{ different.} \end{aligned}$$

Test statistic

$$W = \frac{(n-k)}{(k-1)} \frac{SST_Z}{SSE_Z} = \frac{MST_Z}{MSE_Z}$$

where SST_Z and SSE_Z are the usual sums of squares evaluated for the new data z_{ij} where now

$$z_{ij} = \left| x_{ij} - \overline{x}_i \right|.$$

If H_0 is true

$$W \sim F_{k-1,n-k}.$$

Assessing unequal group variances

```
> library(car)

+ Loading required package:  carData

> leveneTest(Pull~Dose,center='mean',data=Pull)

+ Levene's Test for Homogeneity of Variance (center = "mean")
+       Df F value Pr(>F)
+ group  3  0.2948 0.8286
+       16

> leveneTest(Pull~Dose,center='median',data=Pull)

+ Levene's Test for Homogeneity of Variance (center = "median")
+       Df F value Pr(>F)
+ group  3  0.1955 0.8979
+       16
```

The ANOVA Table

For a completely randomized design, we may report the results of the ANOVA F-test in a stylized form.

SOURCE	DF	SS	MS	F
TREATMENTS	$k-1$	SST	$MST = \dfrac{SST}{(k-1)}$	$F = \dfrac{MST}{MSE}$
ERROR	$n-k$	SSE	$MSE = \dfrac{SSE}{(n-k)}$	
TOTAL	$n-1$	SS		

The ANOVA Table

Note

(i) $(k-1)+(n-k)=(n-1)$

(ii) $SST + SSE = SS$

i.e. we can fill in missing values if they are not given.

The ANOVA Table

Sometimes an extra column is added at the right of the table to give the p-value of the ANOVA F-test.

SOURCE	DF	SS	MS	F	p
TMT	$k-1$	SST	MST	$F = \dfrac{MST}{MSE}$	p
ERROR	$n-k$	SSE	MSE		
TOTAL	$n-1$	SS			

where p solves

$$\frac{MST}{MSE} = F_{p,k-1,n-k}$$

and F_{a,ν_1,ν_2} is the upper a point of the $F_{k-1,n-k}$ distribution.

The ANOVA Table

```
> supernova(lm(Pull~Dose,data=Pull))

+  Analysis of Variance Table (Type III SS)
+  Model: Pull ~ Dose
+
+                              SS df     MS     F   PRE     p
+  ----- --------------- | ------- -- ------ ----- ----- -----
+  Model (error reduced) | 140.093  3 46.698 6.423 .5463 .0046
+  Error (from model)    | 116.324 16  7.270
+  ----- --------------- | ------- -- ------ ----- ----- -----
+  Total (empty model)   | 256.417 19 13.496
```

PRE is the proportional reduction in error

$$PRE = \frac{SST}{SS}$$

Further Examples and R code

- Milk: milk-protein level example
- Pull: pull-strength in Alzheimer's example
- AIDS:
  - RESPONSE: gut permeability of drug mannitol
  - FACTOR: AIDS HIV Status
  - FACTOR LEVELS: $k = 4$

    AIDS - Full AIDS

    ARC - AIDS-related conditions

    HIV+ - HIV positive

    HIV- - HIV negative

Further Examples and R code

- MilkBac: bacteria level in different batches of milk
  - RESPONSE: Bacteria level count per million
  - FACTOR: Batch number
  - FACTOR LEVELS: 1,2,3,4,5 ($k = 5$)

Further Examples and R code

- PTSD: PTSD score in different treatment groups.
  - RESPONSE: PTSD score
  - FACTOR: Therapeutic treatment method
  - FACTOR LEVELS: $k = 4$

    SIT - "Stress Innoculation Therapy"

    RE - "Relive Experience"

    SC - "Standard Counselling"

    WL - "Waiting List" (Control)

Further Examples and R code

Example: PTSD Example

$n = 45, k = 4.$

F-statistic	$F_{obs} = 3.046$
Critical Value	$F_{0.05,3,41} \simeq 2.84$
	$F_{0.025,3,41} \simeq 3.46$
	$F_{0.01,3,41} \simeq 4.31$

Textbook tables give $F_{\alpha,3,40}$.

$$\Longrightarrow \text{Reject } H_0 \text{ at } \alpha = 0.05 \quad (p = 0.039).$$

but Levene's Test suggests that the assumption of equal variances is not valid.

Relaxing the assumptions: permutation tests

Why do we need the three assumptions ?

- independence
- Normality
- equal variances

- so that we can predict (under H_0) that

$$F \sim F_{k-1,n-k}$$

and complete the test (compute p-values and the rejection region).

Relaxing the assumptions: permutation tests

But our null hypothesis of interest is (essentially)

$$H_0 \ : \ \text{No difference between treatments}$$

Under this hypothesis, the treatment labels

should not matter

that is, we should be able to exchange the group labels, and not notice any major difference in the test statistic.

Relaxing the assumptions: permutation tests

This leads us to consider permutation or randomization tests.

i.e. we compute the test statistic for all possible relabellings consistent with H_0, retaining the group sample sizes, and use these values to compute the rejection region.

Suppose that there are N possible relabellings that give rise to test statistics

$$F_1, F_2, \ldots, F_N$$

Relaxing the assumptions: permutation tests

Then the rejection region for significance level α is the interval to the right of

$$N(1-\alpha)\text{th largest of the values } F_1, F_2, \ldots, F_N$$

and the p-value is

$$\frac{\text{Number of } F_1, F_2, \ldots, F_N \geqslant F_{obs}}{N}$$

where the observed test statistic is

$$F_{obs} = \frac{MST}{MSE}.$$

Relaxing the assumptions: permutation tests

Note: If the group sample sizes are $n_1, n_2, \ldots, n_k$ then

$$N = \frac{n!}{n_1! n_2! \ldots n_k!}$$

where

$$n! = n(n-1)(n-2)\ldots 3.2.1$$

("n factorial") - potentially very large.

Relaxing the assumptions: permutation tests

Example: PTSD Example

$$k = 4, n = 45 \qquad (n_1 = 14, n_2 = 10, n_3 = 11, n_4 = 10)$$

There are

$$\frac{45!}{14!10!11!10!} = 2.610 \times 10^{24}$$

possible relabellings: a very big number. We must compute

$$F = \frac{MST}{MSE}$$

for each relabelling. Here, $F_{obs} = 3.046$.

Relaxing the assumptions: permutation tests

Example: PTSD Example

Instead of computing all of the F values for this large number of permutations, we may approximate the distribution using Monte Carlo procedures

- we sample $N = 10000$ (say) data sets where the labels are permuted
- we approximate the null distribution using this sample

Relaxing the assumptions: permutation tests

Example: PTSD Example

Using the exact permutation approach, we compute for $\alpha = 0.05$

$$\begin{aligned} \text{CRITICAL VALUE} &: \quad C_R = 2.844 \\ \text{p-VALUE} &: \quad p = 0.040 \end{aligned}$$

Compare this with the ANOVA F-test values

$$\begin{aligned} \text{CRITICAL VALUE} &: \quad C_R = 2.833 \\ \text{p-VALUE} &: \quad p = 0.039 \end{aligned}$$

using the $F_{3,41}$ distribution.

Relaxing the assumptions: permutation tests

Example: PTSD Example

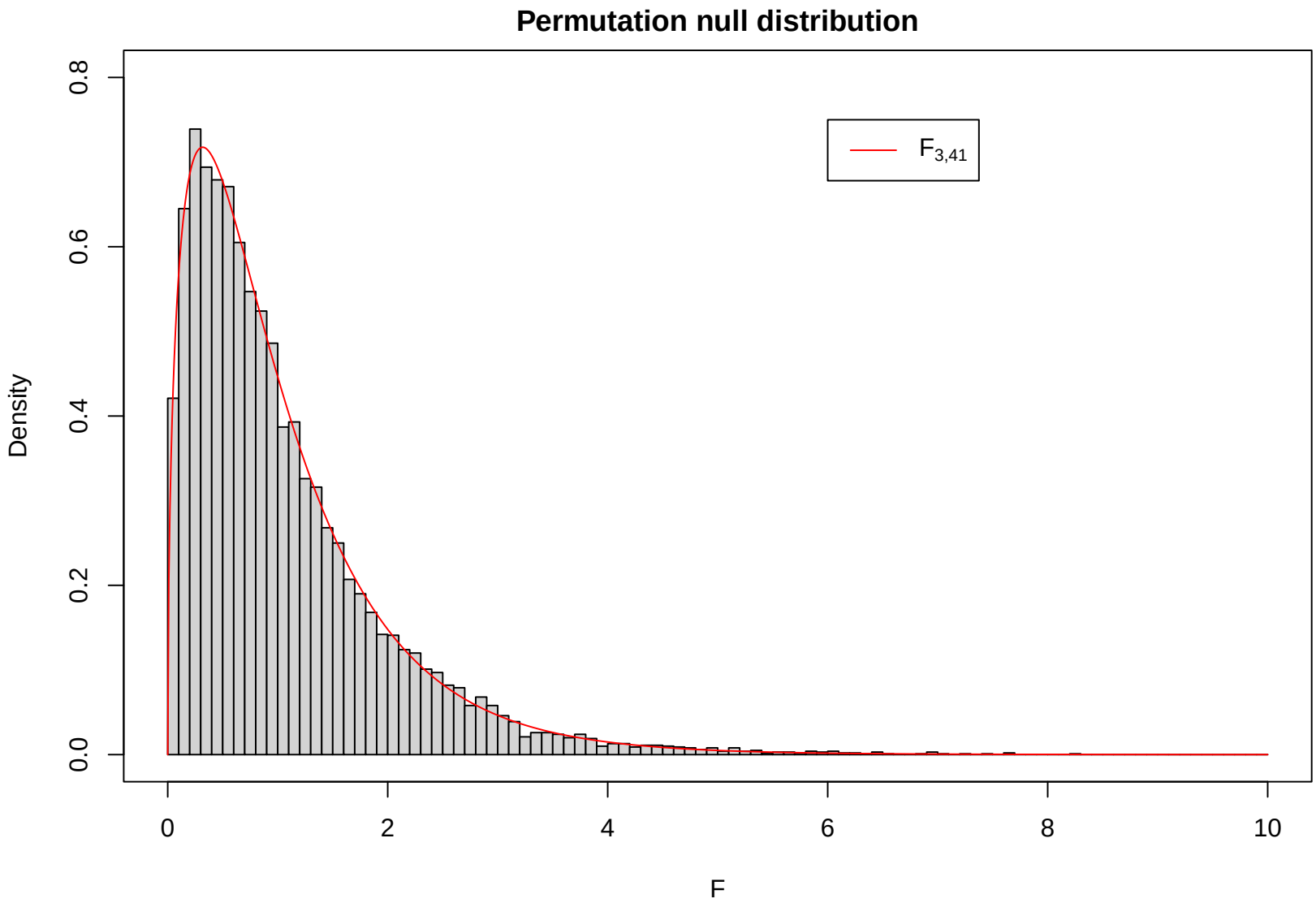
Thus we obtain virtually identical results; however

- the randomization test does not need the assumptions of Normality or equal variances.

Relaxing the assumptions: permutation tests

```
> N<-10000
> Fstat<-rep(0,N)
> n<-nrow(PTSD)
> x<-PTSD$PTSD
> z<-PTSD$GROUP
> for(iperm in 1:N){
+     gp<-z[sample(1:nrow(PTSD))]
+     Fstat[iperm]<-summary(aov(x~gp))[[1]][1,4]
+ }
> quantile(Fstat,prob=0.95)

+      95%
+ 2.846929
```

Relaxing the assumptions: permutation tests



Relaxing the assumptions: permutation tests

Example: PTSD Example

Thus the null hypothesis (of equal means) is

$$\text{rejected}$$

under both procedures at the $\alpha = 0.05$ significance level.

In this case, the computations give similar conclusions. Here the truth or otherwise of the normality/equal variance assumptions does not matter.

Final Note on ANOVA F-test for a CRD

If $k = 2$, consider $F = MST/MSE$;

$$
\begin{aligned}
MST &= \frac{1}{k-1}\sum_{i=1}^{k} n_i(\overline{x}_i - \overline{x})^2 = n_1(\overline{x}_1 - \overline{x})^2 + n_2(\overline{x}_2 - \overline{x})^2 \\
&= \frac{n_1 n_2}{n_1 + n_2}(\overline{x}_1 - \overline{x}_2)^2
\end{aligned}
$$

$$
\begin{aligned}
MSE &= \frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_i)^2 = s_P^2 \\
&= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}
\end{aligned}
$$

Final Note on ANOVA F-test for a CRD

Therefore

$$F_{obs} = \frac{\left(\frac{n_1 n_2}{n_1 + n_2}\right)(\overline{x}_1 - \overline{x}_2)^2}{s_P^2} = \left(\frac{(\overline{x}_1 - \overline{x}_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right)^2$$

Thus $F = t^2$, where t is the two-sample t-test statistic.

Final Note on ANOVA F-test for a CRD

Thus if $k = 2$, the ANOVA F-test and the two sample t-test are equivalent

$$\begin{aligned} t &\sim \text{Student-t}(n-2) \\ F &\sim F(1, n-2) \end{aligned}$$

and we must get the same conclusion (to reject H_0 or otherwise) using either statistic.

Summary

If the assumptions

- independence (holds by design in a CRD)
- Normal populations
- equal variances

hold, use

ANOVA F-test

If the assumptions do not hold

- use Randomization/Permutation test
- use Non-parametric test (see Section 3)

Multiple Comparison of Means

If the ANOVA F-test null hypothesis

$$H_0 : \mu_1 = \cdots = \mu_k$$

is rejected, then it is of interest to discover which of the means are different. For k groups, there are $c = k(k-1)/2$ pairs of group means that can be compared.

Multiple Comparison of Means

Consider a “family" of hypothesis tests - a collection of tests of different hypotheses carried out independently on different data sets. For each test in the family, we consider testing the hypothesis at significance level α.

Label the tests $i = 1, \ldots, c$, and for each i, label

- the null hypotheses H_{0i}
- the test statistics T_i
- the rejection regions $\mathcal{R}_i$

that are potentially different for each i.

Multiple Comparison of Means

We specify for each i,

$$\alpha = \Pr(T_i \in \mathcal{R}_i | H_{0i} \text{ is true})$$

which implicitly defines $\mathcal{R}_i$. Note that α is the

"Test Type I Error Rate" or "Comparisonwise Error Rate"

Now consider the results of all tests in the family; what is the "Familywise" Type I error rate ?

Multiple Comparison of Means

Using the laws of probability

$$\Pr\big(T_i \in \mathcal{R}_i \big| H_{0i} \text{ is true}\big) = \alpha$$

means that

$$\Pr\big(T_i \notin \mathcal{R}_i \big| H_{0i} \text{ is true}\big) = 1 - \alpha$$

giving the probability that the test does not reject H_{0i}, if H_{0i} is in fact true, is $1 - \alpha$.

Multiple Comparison of Means

Now we consider all tests together;

$$P\big[\text{Each } T_i \notin \mathcal{R}_i \big| \text{Each } H_{0i} \text{ is true}\big] = (1-\alpha)^c$$

This is the probability that each test results in the null hypothesis not being rejected

- that is, the probability that we never commit a Type I error across the c tests.

Multiple Comparison of Means

Therefore the probability of at least one Type I error is

$$\alpha_F = 1 - (1 - \alpha)^c$$

α_F is the Familywise Error Rate.

	$\alpha = 0.05$	$\alpha = 0.01$
c	α_F	α_F
5	0.226	0.049
10	0.401	0.096
50	0.923	0.395
100	0.994	0.634

Therefore, when we carry out a “family" of tests, we should not use the traditional choices of $\alpha = 0.05$ or 0.01 on each test.

The Bonferroni Method

To fix $\alpha_F = 0.05$, say, we need to use α on each test where

$$\alpha_F = 1 - (1 - \alpha)^c \quad \Longleftrightarrow \quad \alpha = 1 - (1 - \alpha_F)^{1/c}$$

For example, if $\alpha_F = 0.05$ and $c = 10$, use

$$\alpha = 1 - (1 - 0.05)^{1/10} = 0.0051$$

The Bonferroni Method

It can be shown that

$$1 - (1 - \alpha)^c \approx c\alpha.$$

If α_F is the required familywise error rate, we must set the comparisonwise error rate to be $\alpha = \alpha_F/c$.

α_F/c is known as the Bonferroni Correction.

Confidence Intervals

For the $k = 2$ comparison of means, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is

$$(\overline{x}_1 - \overline{x}_2) \pm t_{\alpha/2, n_1+n_2-2} s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where $t_{\alpha,\nu}$ is the upper α point of the t-distribution with ν degrees of freedom (under the assumptions of independence, Normality and equal group variances).

Confidence Intervals

Note: If we move to a family of c tests, to get simultaneous confidence intervals for the differences in means $\mu_i - \mu_j$ for all pairs of i and j, we should adjust α to α_F when computing the $100(1-\alpha)\%$ confidence interval.

Confidence Intervals

There are many different methods for correcting the confidence interval for use in different experimental situations.

For example, we might consider

- planned comparisons $\mu_1 = \mu_3$, $\mu_7 = \mu_{10}$ etc.
- all comparisons

of group means.

Confidence Intervals

Three methods are recommended:

- Bonferroni's Method
- Tukey's HSD Method for all pairwise comparisons
- Scheffé's Method for all possible contrast comparisons.

Having selected a multiple comparison correction method, we compute simultaneous confidence intervals for each comparison of means, and identify

- which means are significantly different
- the ranking of differences $\mu_i - \mu_j$ in terms of magnitude.

Randomized Block Designs

A randomized block design used matched experimental units organized into sets known as blocks and assigns one member from the set to each treatment.

For k treatments

1. Compile b blocks of k experimental units, with each block comprising units that are similar.
2. Assign one unit from each block to each treatment at random.

Then there are a total of $n = bk$ measured responses.

Randomized Block Designs

We wish to compare treatments whilst acknowledging that there may be differences between the blocks.

That is, the observed variation is due to

treatments and blocks and error

rather than merely

treatments and error

as was the case for the CRD.

Randomized Block Designs

Example: Test scores

- Response : Measured test Score
- Factor : Test type
- Factor-levels : $k = 2$ (Type 1, Type 2)
- Blocks : $b = 5$ (Previous GPA)

i.e. $k = 2, b = 5$, so $n = 10$.

Randomized Block Designs

Example: Test scores

Block	GPA	Type 1	Type 2
1	2.75	540	530
2	3.00	570	550
3	3.25	590	580
4	3.50	640	620
5	3.75	690	690

Randomized Block Designs

Example: Test scores

This design recognizes that GPA score is likely to explain some variation in test score, but is not directly related to the factor of interest (test type).

i.e. the blocking variable removes systematic variation in response that is not of primary interest.

Randomized Block Designs

Example: Test Scores

To implement, we

- pick two students from each GPA band (block)
- randomly assign one to take test type 1, the other to test type 2.

Randomized Block Designs

Example: Treatment for Hypertension

- Response : Blood Pressure (mgHg)
- Factor : Drug Type
- Factor-levels : $k = 3$ (Drug 1, Drug 2, Drug 3)
- Blocks : $b = 4$ Age/Sex combinations
  - Female/Under 50
  - Male/Under 50
  - Female/Over 50
  - Male/Over 50

i.e. $k = 3, b = 4 \therefore n = 12$.

Testing for Equal Treatment Means

An ANOVA F-test can be constructed for a RBD. Let

- $i = 1, \ldots, k$ index treatments
- $j = 1, \ldots, b$ index blocks

i.e. x_{ij} is the response for the ith treatment in the jth block. Let

- $\overline{x}_i$ be the ith treatment mean
- $\overline{x}_j^{(B)}$ be the jth block mean
- $\overline{x}$ be the overall mean

Testing for Equal Treatment Means

Let

$$
\begin{aligned}
SST &= \sum_{i=1}^{k} b\left(\overline{x}_i - \overline{x}\right)^2 \\
SSB &= \sum_{j=1}^{b} k\left(\overline{x}_j^{(B)} - \overline{x}\right)^2 \\
SS &= \sum_{i=1}^{k}\sum_{j=1}^{b}\left(x_{ij} - \overline{x}\right)^2
\end{aligned}
$$

SST: Sum of Squares for Treatments

SSB: Sum of Squares for Blocks

SS: Total Sum of Squares

Testing for Equal Treatment Means

Finally

$$SS = SST + SSB + SSE \quad \therefore \quad SSE = SS - SST - SSB$$

SSE: Sum of Squares for Errors

Test statistic is

$$F = \frac{MST}{MSE}$$

where

$$MST = \frac{SST}{k-1} \qquad MSE = \frac{SSE}{n-b-k+1}$$

Testing for Equal Treatment Means

The presumed model is

$$X_{ij} = \mu_i + \mu_j^{(\mathrm{B})} + \epsilon_{ij}$$

where

- μ_i is the mean contribution from treatment level i
- $\mu_j^{(\mathrm{B})}$ is the mean contribution from block j

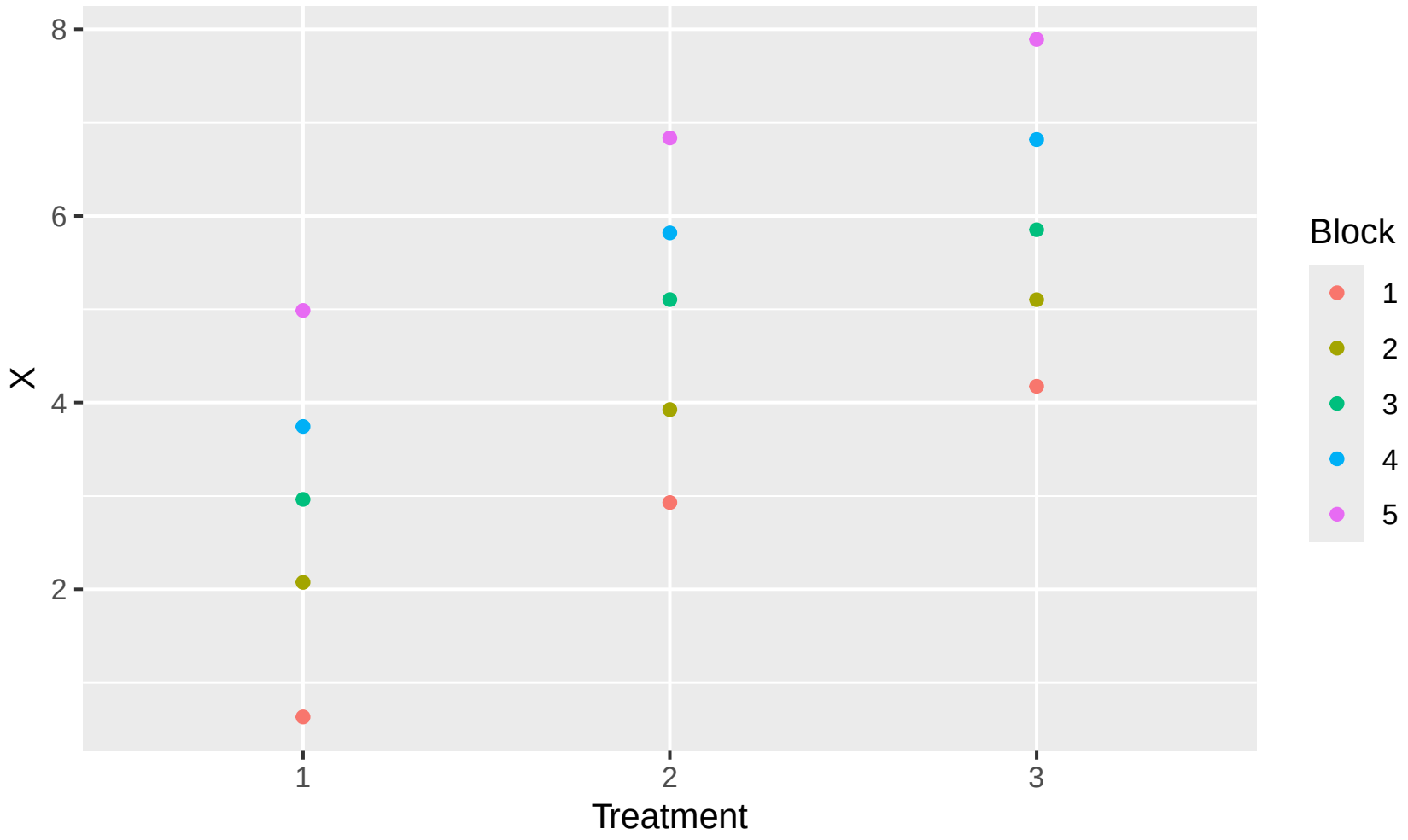
and the effects combine additively.

Testing for Equal Treatment Means

Here μ_i is not simply the population mean of the ith treatment group:

- the expected response for any experimental unit depends on the block they are in
- μ_i is the contribution to this expected response that is due to being assigned to treatment level i.

Testing for Equal Treatment Means



ANOVA F-test for a randomized block design

Theorem: ANOVA F-test for a RBD

To test

$$\begin{aligned} H_0 &: \mu_1 = \cdots = \mu_k \\ H_a &: \text{At least one pair of } \mu_i\text{s different.} \end{aligned}$$

use the test statistic

$$F = \frac{MST}{MSE}$$

If H_0 is true

$$F \sim F_{k-1,n-b-k+1}$$

This defines the rejection region for significance level α, and the p-value, in the usual way.

ANOVA F-test for a randomized block design

Assumptions:

1. Experimental units are independent, and treatments are allocated at random (within blocks).
2. Normality
3. bk block/treatment combinations correspond to populations with equal variances.

ANOVA F-test for a randomized block design

ANOVA Table

SOURCE	DF	SS	MS	F
TMTS	$k-1$	SST	MST	F = MST/MSE
BLOCKS	$b-1$	SSB	MSB	
ERROR	$n-k-b+1$	SSE	MSE	
TOTAL	$n-1$	SS		

ANOVA F-test for a randomized block design

After the ANOVA test is complete, and the hypothesis

$$H_0 : \mu_1 = \cdots = \mu_k$$

is rejected, we can proceed with the “post-hoc" tests of hypotheses $\mu_i = \mu_j$ for $i \neq j$.

ANOVA F-test for a randomized block design

Notes:

1. In a RBD, it is not (in general) possible to estimate individual treatment means: $\overline{x}_i$ does not estimate μ_i as it is an average across blocks, which are believed to be different in terms of response.

ANOVA F-test for a randomized block design

2. Testing the Block Means However, we can test whether the block means $\mu_1^{(B)}, \ldots, \mu_b^{(B)}$ are significantly different. For

$$H_0 : \mu_1^{(B)} = \cdots = \mu_b^{(B)}$$

we use the F statistic

$$F = \frac{MSB}{MSE}$$

where

$$MSB = \frac{SSB}{b-1}$$

ANOVA F-test for a randomized block design

If H_0 is true

$$F \sim F_{b-1,n-k-b+1}$$

That is, we treat the blocks as levels of another factor, and test to see whether this factor affects response.

ANOVA F-test for a randomized block design

Example: Soil Analysis (see handout)

Results of two ANOVA F-tests:

Test of	F	p	Conclusion
SOLVENT	0.673	0.585	No Difference
SOIL	10.568	0.001	Difference

Here SOLVENT is the treatment variable, SOIL is the blocking variable.

ANOVA F-test for a randomized block design

We must remember to check the assumptions (independence, normality, equal variances in each treatment/block combination)

Equal variances may be hard to check as we only have one observation per treatment/block comparison.

ANOVA F-test for a randomized block design

In the sums-of-squares decompositions

$$CRD: \quad SS = [SST] + SSE$$

$$RBD: \quad SS = [SST + SSB] + SSE$$

are both of the form

$$\begin{array}{ccccc} \text{TOTAL} & = & \text{SYSTEMATIC} & + & \text{RANDOM} \\ \text{VARIATION} & & \text{VARIATION} & & \text{VARIATION} \end{array}$$

ANOVA F-test for a randomized block design

$$\text{“SYSTEMATIC"} \quad \begin{cases} \text{For the CRD:} & SST \\ \text{For the RBD:} & SST + SSB \end{cases}$$

$$\text{“RANDOM"} \quad \text{For both:} \quad SSE$$

ANOVA F-test for a randomized block design

We have studied the

Randomized Complete Block Design

where each block/treatment combination has one experimental unit.

An incomplete design could also be considered, where some block/treatment combinations are omitted. However, this design does not lead to straightforward ANOVA analysis.

Factorial Experiments

Designs studied so far:

- CRD - one factor
- RBD - one factor, plus one blocking variable, so two factors in total, where one (the blocking variable) is a known source of systematic variation.

However, in the RBD, we must assume that the treatments behave in a similar way across blocks.

Factorial Experiments

Let i index treatments ($1 \leqslant i \leqslant k$) and consider block j, and two treatment (factor levels) i_1 and i_2.

In an RBD, we assume that

$$\mathbb{E}[X_{i_1 j} - X_{i_2 j}] = \mu_{i_1} - \mu_{i_2}$$

which does not depend on j.

That is, the expected difference in response due to the two treatments does not depend on the block.

Factorial Experiments

But perhaps the difference does depend on block; perhaps we have interaction:

- an interaction is present if the effect of one variable on the outcome is modified in the presence of a second variable
- for example, the effect of treatment may be different in different blocks.

In the current RBD, we do not have enough data to look for this. We now seek to extend the RBD to allow for tests for interaction; we do this by using replication.

RBD with balanced replication

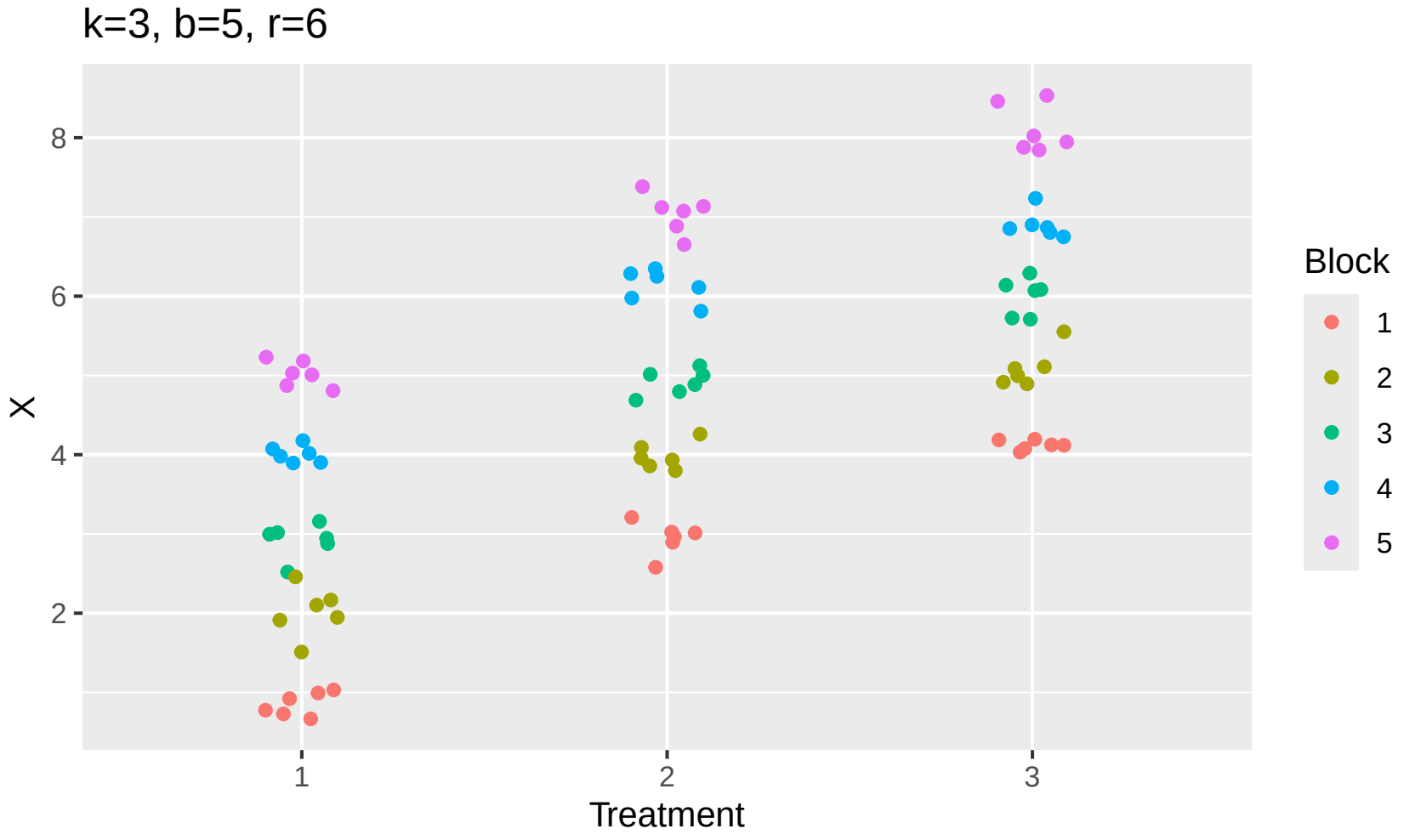
Suppose we have r observations per block/treatment combination (termed replicates), so that we have

$$n = bkr$$

experimental units in total.

Balanced designs have equal numbers of replicates in each block/treatment combination.

RBD with balanced replication



RBD with balanced replication

In this design, all the quantities

$$SST, SSB, SSE, SS$$
$$MST, MSB, MSE$$

can be defined, and an ANOVA F-test can be carried out - the only difference is that $n = bkr$.

RBD with balanced replication

- Sum of Squares for Treatments (SST)

$$SST = \sum_{i=1}^{k} br(\overline{x}_i - \overline{x})^2$$

- Sum of Squares for Blocks (SSB)

$$SSB = \sum_{j=1}^{b} kr(\overline{x}_j^{(\mathrm{B})} - \overline{x})^2$$

- Overall Sum of Squares (SS)

$$SS = \sum_{i=1}^{k} \sum_{j=1}^{b} \sum_{t=1}^{r} (x_{ijt} - \overline{x})^2$$

and $SSE = SS - SST - SSB$. The third index t indexes the replicates.

RBD with balanced replication

The RBD with replication does allow the investigation of interaction. The new test is based on the decomposition

$$SS = SST + SSB + SSI + SSE$$

where SSI is the Sum of Squares for Interaction.

RBD with balanced replication

We have SST, SSB and SS as before, and

$$SSI = \sum_{i=1}^{k} \sum_{j=1}^{b} r\left(\overline{x}_{ij} - \overline{x}_i - \overline{x}_j^{(B)} + \overline{x}\right)^2$$

where

$$\overline{x}_{ij} = \frac{1}{r} \sum_{t=1}^{r} x_{ijt} \qquad i = 1, \ldots, k,\ j = 1, \ldots, b$$

is the sample mean for replicates in (i, j)th treatment/block combination.

RBD with balanced replication

The presumed model here is

$$X_{ijt} = \mu_i + \mu_j^{(\mathrm{B})} + \mu_{ij}^{(\mathrm{I})} + \epsilon_{ijt}$$

where

- μ_i is the mean contribution from treatment level i
- $\mu_j^{(\mathrm{B})}$ is the mean contribution from block j
- $\mu_{ij}^{(\mathrm{I})}$ is the additional mean contribution for treatment i in block j

Note: we need to take care in interpreting these parameters.

Testing in the RBD with Replication

The three F statistics

$$F = \frac{MST}{MSE} \qquad F = \frac{MSB}{MSE} \qquad F = \frac{MSI}{MSE}$$

can be used to test for significant Treatment, Block and Interaction effects respectively.

Testing in the RBD with Replication

Now

$$MSE = \frac{SSE}{\text{Error d.f.}}$$

But what is “Error d.f." ? It is a constant that dictates how large SSE should be on average.

- the error d.f. depends on the complexity of the model fitted.

Testing in the RBD with Replication

The general rule for computing the error d.f. for any model is

$$\text{Error d.f.} = n - p$$

where n is the total sample size and p is the total number of parameters fitted.

Testing in the RBD with Replication

How many parameters do we fit ?

- No Interaction

$$p = 1 + (b - 1) + (k - 1)$$

  that is, the overall mean μ, plus the $b-1$ differences from μ due to the blocks, plus the $k-1$ differences from μ due to the treatments.

- Interaction

$$p = bk$$

  that is, one parameter in each cell of the two-way table of blocks by treatments.

Testing in the RBD with Replication

Thus

- No Interaction

$$p = 1 + (b - 1) + (k - 1) = b + k - 1$$

  parameters, so

$$\text{Error d.f.} = n - p = n - b - k + 1$$

- Interaction: we fit $p = bk$ parameters, so

$$\text{Error d.f.} = n - p = n - bk$$

Testing in the RBD with Replication

It transpires that if

$$MSI = \frac{SSI}{(b-1)(k-1)}$$

is the Mean Square for Interaction, then

$$F = \frac{MSI}{MSE}$$

yields a test statistic suitable for testing interaction.

Testing in the RBD with Replication

Under the null hypothesis of no interaction, then

$$F \sim F_{(b-1)(k-1),n-bk}$$

where $n = bkr$.

Why $(b-1)(k-1)$? This is the number of extra parameters we fit to include the interaction.

Testing in the RBD with Replication

All of the ANOVA F-tests here are tests of model simplification

$$\begin{aligned} H_0 &: \text{NULL MODEL} \\ H_a &: \text{FULL MODEL} \end{aligned}$$

where we obtain the null model from the full model by setting some parameters to zero.

Testing in the RBD with Replication

For the CRD:

$$\begin{array}{rcl} H_a & & H_0 \\ \text{FULL MODEL} & \longrightarrow & \text{NULL MODEL} \\ k \text{ parameters} & \longrightarrow & 1 \text{ parameter} \end{array}$$

so there are $(k-1)$ extra parameters, and so SST varies on $(k-1)$ degrees of freedom

$$\mu_1, \ldots, \mu_k \longrightarrow \mu$$

Testing in the RBD with Replication

The (i,j)th treatment/block combination has mean

$$\mu_i + \mu_j^{(\mathrm{B})}$$

so for testing for a TREATMENT effect

$$\begin{array}{rcl} H_a & & H_0 \\ \text{FULL MODEL} & \longrightarrow & \text{NULL MODEL} \\ 1 + (k-1) + (b-1) \text{ param.} & \longrightarrow & 1 + (b-1) \text{ param.} \end{array}$$

so there are $(k-1)$ extra parameters, and so SST varies on $(k-1)$ degrees of freedom.

$$\mu_1, \ldots, \mu_k \longrightarrow \mu$$

Testing in the RBD with Replication

For testing for a BLOCK effect

$$\begin{array}{rcl} H_a & & H_0 \\ \text{FULL MODEL} & \longrightarrow & \text{NULL MODEL} \\ 1+(k-1)+(b-1)\text{ param.} & \longrightarrow & 1+(k-1)\text{ param.} \end{array}$$

so there are $(b-1)$ extra parameters, and so SSB varies on $(b-1)$ degrees of freedom

$$\mu_1^{(\mathrm{B})}, \ldots, \mu_k^{(\mathrm{B})} \longrightarrow \mu^{(\mathrm{B})}$$

These models and tests can be fitted and carried out even if we do not have replication.

Testing in the RBD with Replication

With replication, we can investigate the interaction, that is the model where the (i,j)th treatment/block combination has mean

$$\mu_i + \mu_j^{(\mathrm{B})} + \mu_{ij}^{(\mathrm{I})}$$

rather than the model where

$$\mu_i + \mu_j^{(\mathrm{B})}$$

that is, we wish to test

$$\begin{aligned} H_0 &: \mu_{ij}^{(\mathrm{I})} = 0 \qquad \text{for all } i \text{ and } j \\ H_a &: \mu_{ij}^{(\mathrm{I})} \neq 0 \end{aligned}$$

Testing in the RBD with Replication

In the full interaction model: we fit bk parameters

In the restricted, no interaction model: we fit

$$1 + (b - 1) + (k - 1) = b + k - 1$$

parameters. Therefore the differences is

$$bk - (b + k - 1) = bk - b - k + 1 = (b - 1)(k - 1)$$

and SSI varies on $(b - 1)(k - 1)$ degrees of freedom.

Testing in the RBD with Replication

If we do not have replication, we CANNOT fit the interaction. Recall that

$$\text{Error d.f.} = n - bk$$

but if $r = 1$, $n = rbk = bk$, so the error d.f. is zero.

In fact, $SSE = 0$ also, so the MSE is not defined.

ANOVA Table

SOURCE	DF	SS	MS	F
TMTS	$k-1$	SST	MST	F_T
BLOCKS	$b-1$	SSB	MSB	F_B
INTERACTION	$(b-1)(k-1)$	SSI	MSI	F_I
ERROR	$(n-bk)$	SSE	MSE	
TOTAL	$n-1$	SS		

where

$$MST = \frac{SST}{k-1} \qquad MSB = \frac{SSB}{b-1}$$

$$MSI = \frac{SSI}{(b-1)(k-1)} \qquad MSE = \frac{SSE}{n-bk}$$

and

$$F_T = \frac{MST}{MSE} \qquad F_B = \frac{MSB}{MSE} \qquad F_I = \frac{MSI}{MSE}$$

Example

Simulated data with $k = 5, b = 3, r = 6$:

```
> k<-3
> b<-5
> r<-6
> tmt<-rep(rep(1:k,each=b),r)
> blk<-rep(rep(1:b,k),r)
> rept<-rep(1:r,b*k)
> mui<-c(0,2,3)
> mub<-c(1:5)
> X<-mui[tmt]+mub[blk]+rnorm(r*b*k)*0.2
> RBDdat<-data.frame(Treatment=as.factor(tmt),Block=as.factor(blk),X)
> supernova(lm(X~Treatment+Block+Treatment:Block,data=RBDdat), type=2)

+  Analysis of Variance Table (Type II SS)
+  Model: X ~ Treatment + Block + Treatment:Block
+
+                                          SS df     MS        F   PRE     p
+  --------------- --------------- | ------- -- ------ -------- ----- -----
+            Model (error reduced) | 321.762 14 22.983  533.074 .9901 .0000
+        Treatment                 | 136.912  2 68.456 1587.787 .9769 .0000
+            Block                 | 184.673  4 46.168 1070.838 .9828 .0000
+  Treatment:Block                 |   0.177  8  0.022    0.515 .0520 .8419
+            Error (from model)    |   3.234 75  0.043
+  --------------- --------------- | ------- -- ------ -------- ----- -----
+            Total (empty model)   | 324.995 89  3.652
```

Example

Interaction model:

```
> summary(aov(X~Treatment+Block+Treatment:Block,data=RBDdat))

+                 Df Sum Sq Mean Sq  F value Pr(>F)
+ Treatment        2 136.91   68.46 1587.787 <2e-16 ***
+ Block            4 184.67   46.17 1070.838 <2e-16 ***
+ Treatment:Block  8   0.18    0.02    0.515  0.842
+ Residuals       75   3.23    0.04
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example

No interaction model:

```
> summary(aov(X~Treatment+Block,data=RBDdat))

+             Df Sum Sq Mean Sq F value Pr(>F)
+ Treatment    2 136.91   68.46    1666 <2e-16 ***
+ Block        4 184.67   46.17    1123 <2e-16 ***
+ Residuals   83   3.41    0.04
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example: Batteries Data (see handout)

```
> summary(aov(life~temp+material+temp:material, data=Batteries))

+               Df Sum Sq Mean Sq F value      Pr(>F)
+ temp           2  39083   19542  29.344 0.000000169 ***
+ material       2  10633    5317   7.983     0.00189 **
+ temp:material  4   9438    2359   3.543     0.01897 *
+ Residuals     27  17981     666
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For $\alpha = 0.05$, there is a significant temp effect ($p < 0.001$), and a significant material effect ($p < 0.002$), and a significant interaction ($p < 0.019$)

Multifactor designs

We now study multifactor designs, to assess the effects and interactions of several factors simultaneously.

We consider all possible combinations of

$$\begin{array}{lll} \text{FACTOR A} & \text{with} & a \text{ levels} \\ \text{FACTOR B} & \text{with} & b \text{ levels} \\ \text{FACTOR C} & \text{with} & c \text{ levels} \\ & \vdots & \end{array}$$

to define the treatments in a factorial design.

Factorial Experiments

A complete factorial experiment is one in which every combination of the factors is utilized.

- the number of treatments is equal to the total number of factor-level combinations.

Consider two factor experiments:

FACTOR A	with	a levels
FACTOR B	with	b levels

so there are ab treatments in total.

Factorial Experiments

The data can be arranged in a table where the cells define the factor-level combinations.

Factor B

Factor A		1	2	3	4	5
	1					
	2					
	3					

This table displays a two-way layout with $a = 3$ and $b = 5$. Here

- Levels of A will be in the rows
- Levels of B will be in the columns

Factorial Experiments

This design is very similar to the RBD, but now the second factor is not a blocking factor:

- the ab treatment populations are constructed independently from the same base population, and are believed to be not systematically different.
- individuals from the same base population are assigned at random to one of the ab treatments.

Factorial Experiments

In this design we can study the effect of

- Factor A and Factor B (the main effects)
- the interaction

provided we have balanced replication.

Factorial Experiments

We construct ANOVA F-tests based on the decomposition

$$SS = SST_A + SST_B + SSI_{AB} + SSE$$

- Sum of Squares for Treatments due to factor A (SST_A)

$$SST_A = \sum_{i=1}^{a} br(\overline{x}_{i.} - \overline{x}_{..})^2$$

- Sum of Squares for Treatments due to factor B (SST_B)

$$SST_B = \sum_{j=1}^{b} ar(\overline{x}_{.j} - \overline{x}_{..})^2$$

- Sum of Squares for Interaction (SSI_{AB})

$$SSI_{AB} = \sum_{i=1}^{a} \sum_{j=1}^{b} r(\overline{x}_{ij} - \overline{x}_{i.} - \overline{x}_{.j} + \overline{x}_{..})^2$$

Factorial Experiments

New notation:

- sample mean for Factor A level i

$$\overline{x}_{i.} = \frac{1}{br} \sum_{j=1}^{b} \sum_{t=1}^{r} x_{ijt} \qquad i = 1, \ldots, a$$

  which is a row mean

- sample mean for Factor B level j

$$\overline{x}_{.j} = \frac{1}{ar} \sum_{i=1}^{a} \sum_{t=1}^{r} x_{ijt} \qquad j = 1, \ldots, b$$

  which is a column mean

Factorial Experiments

- sample mean for replicates in (i, j)th factor combination

$$\overline{x}_{ij} = \frac{1}{r} \sum_{t=1}^{r} x_{ijt}$$

  for each $i = 1, \ldots, a$, $j = 1, \ldots, b$, which is a cell mean

- overall sample mean

$$\overline{x}_{..} = \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{t=1}^{r} x_{ijt}$$

  which is termed the overall mean

Factorial Experiments

These allow computation of

$$SST_A, SST_B, SSI_{AB}, SS, SSE$$

$$MST_A, MST_B, MSI_{AB}, MSE$$

using the degrees of freedom identical to those in the RBD with replication.

Tests for

- significant effect for Factor A
- significant effect for Factor B
- significant interaction

will be carried out as before using an ANOVA table.

ANOVA Table

SOURCE	DF	SS	MS	F
FACTOR A	$a-1$	SST_A	MST_A	F_A
FACTOR B	$b-1$	SST_B	MST_B	F_B
INTERACTION	$(a-1)(b-1)$	SSI_{AB}	MSI_{AB}	F_{AB}
ERROR	$(n-ab)$	SSE	MSE	
TOTAL	$n-1$	SS		

ANOVA Table

If Factor A is not influential (H_0 specifying no difference between responses at different levels of factor A), then

$$F_A \sim F_{a-1,n-ab}$$

Similarly,

$$\begin{aligned} \text{No effect of Factor B} &: \quad F_B \sim F_{b-1,n-ab} \\ \text{No Interaction} &: \quad F_{AB} \sim F_{(a-1)(b-1),n-ab} \end{aligned}$$

Notation

The main effects plus interaction model can be written

$$A + B + A{:}B$$

whereas the main effects only can be written

$$A + B$$

The models

$$A + A{:}B \qquad\qquad B + A{:}B$$

do not make sense.

Notation

For a two factor design, the only models that should be considered and reported are

MODEL	FACTOR	INTERACTION
NULL	NONE	NO
A	A	NO
B	B	NO
A+B	A,B	NO
A+B+A:B	A,B	YES

that is, if the interaction is significant, the only model you should report is

$$A + B + A{:}B$$

Two variable analysis

ANOVA analysis for the RBD and FD (both with replication) are identical. The only difference lies in the interpretation of the factors

- RBD: one blocking, one treatment factor
- FD: two treatment factors

“Blocking” factors are known or strongly believed to have a significant effect on the response.

Estimating Effect Size

In multifactor designs, parameter estimation can be carried out in different parameterizations

For the CRD (one-way layout):

- Natural parameters: $\mu_1, \ldots, \mu_k$
- Contrast parameters: $\beta_1, \beta_2, \ldots, \beta_k$ where

$$\begin{aligned} \beta_1 &= \mu_1 \\ \beta_i &= \mu_i - \mu_1, \quad i = 2, \ldots, k \end{aligned}$$

Estimating Effect Size

The $(k-1)$ parameters

$$\beta_2, \ldots, \beta_k$$

measure the differences in mean from baseline group (level 1 of the factor).

The null hypothesis assumes that

$$\beta_2 = \cdots = \beta_k = 0.$$

Estimating Effect Size

For the Milk protein data set: using the natural parameters

```
> summary(lm(protein~-1+Diet,data=Milk))$coef

+                   Estimate Std. Error  t value Pr(>|t|)
+ Dietbarley        3.531929 0.01551264 227.6807        0
+ Dietbarley+lupins 3.429695 0.01492705 229.7638        0
+ Dietlupins        3.312384 0.01502558 220.4497        0

> tapply(Milk$protein,Milk$Diet,mean)

+        barley barley+lupins        lupins
+      3.531929      3.429695      3.312384
```

Here the `Estimate` column contains the means for the three groups which are the estimates of

$$\mu_1, \mu_2, \mu_3$$

Estimating Effect Size

Using the contrast parameters:

```
> summary(lm(protein~Diet,data=Milk))$coef

+                      Estimate Std. Error    t value     Pr(>|t|)
+ (Intercept)         3.5319294 0.01551264 227.680728 0.000000e+00
+ Dietbarley+lupins -0.1022344 0.02152809  -4.748885 2.266185e-06
+ Dietlupins        -0.2195453 0.02159653 -10.165769 1.963302e-23
```

which reveals the estimates of β_2 and β_3 as

$$-0.1022344 \quad \text{and} \quad -0.2195453$$

Estimating Effect Size

For the two-factor designs (RBD/FD): The cell means can be denoted m_{ij} where

$$m_{ij} = \mu_{i.} + \mu_{.j} + \mu_{ij}$$

for cells (i, j), $i = 1, \ldots, a$, $j = 1, \ldots, b$, where

- $\mu_{i.}$ gives the Factor A contribution
- $\mu_{.j}$ gives the Factor B contribution
- μ_{ij} gives the interaction

and for the replicates in cell (i, j), we have

$$X_{ijt} \sim Normal(m_{ij}.\sigma^2).$$

Estimating Effect Size

By default, R uses the contrast parameterization:

$$
\begin{aligned}
m_{ij} &= \beta_1 && i = 1, j = 1 \\
&= \beta_1 + \beta_i^{(A)} && i = 2, \ldots, a, j = 1 \\
&= \beta_1 + \beta_j^{(B)} && i = 1, j = 2, \ldots, b \\
&= \beta_1 + \beta_i^{(A)} + \beta_j^{(B)} + \gamma_{ij}^{(AB)} && i = 2, \ldots, a \\
& && j = 2, \ldots, b
\end{aligned}
$$

where

$$
\begin{aligned}
\beta_i^{(A)} &: \text{contrasts for factor A} \\
\beta_j^{(B)} &: \text{contrasts for factor B} \\
\gamma_{ij}^{(AB)} &: \text{interaction}
\end{aligned}
$$

Estimating Effect Size

By default, R takes the first level of each factor as the baseline level, and looks at differences compared to this baseline.

The ab parameters are

β_1	1
$\beta_2^{(A)}, \ldots, \beta_a^{(A)}$	$(a-1)$
$\beta_2^{(B)}, \ldots, \beta_b^{(B)}$	$(b-1)$
$\gamma_{ij}^{(AB)}, i = 2, \ldots, a, j = 2, \ldots, b$	$(a-1)(b-1)$
Total	ab

Estimating Effect Size

For example: $a = 3, b = 4$.

Factor B

Factor A		1	2	3	4
	1	β_1	$\beta_1 + \beta_1^{(B)}$	$\beta_1 + \beta_2^{(B)}$	$\beta_1 + \beta_3^{(B)}$
	2	$\beta_1 + \beta_1^{(A)}$	①	②	③
	3	$\beta_1 + \beta_2^{(A)}$	④	⑤	⑥

where

$$① = \beta_1 + \beta_2^{(A)} + \beta_2^{(B)} + \gamma_{22}^{(AB)}$$

$$⑥ = \beta_1 + \beta_3^{(A)} + \beta_4^{(B)} + \gamma_{34}^{(AB)}$$

and so on.

Estimating Effect Size

Estimation is still straightforward:

PARAMETER	ESTIMATE
β_1	$\overline{x}_{11}$
$\beta_i^{(A)}$	$\overline{x}_{i.} - \overline{x}_{11}$
$\beta_j^{(B)}$	$\overline{x}_{.j} - \overline{x}_{11}$
$\gamma_{ij}^{(AB)}$	$\overline{x}_{ij} - \overline{x}_{i.} - \overline{x}_{.j} + \overline{x}_{11}$

for $i = 2, \ldots, a, j = 2, \ldots, b$.

Note: other contrast parameterizations can be used.

Example

For example: $a = 4, b = 3$.

- Factor A: levels $1, 2, \ldots, a$
- Factor B: levels $1, 2, \ldots, b$

Most complicated model: Main Effects plus Interaction

A + B + A:B

Example

That is, we have

- a baseline mean: β_1
- an effect for each level of Factor A: $\beta_i^{(A)}$
- an effect for each level of Factor B: $\beta_j^{(B)}$
- an interaction that modifies the effect of changing levels of Factor A at each level of Factor B: $\gamma_{ij}^{(AB)}$

Example

In R, the baseline group is the one where Factor A has level 1 and Factor B has level 1, but this choice is arbitrary

- changing this assumption should have no effect on the results we obtain.

Example

We adopt the following modelling strategy:

- Establish a baseline
- Look for changes from baseline introduced by Factor A
- Look for changes from baseline introduced by Factor B

Example

- Look for changes from baseline introduced by Factor A and Factor B acting additively, so that the effect of changing the level of Factor A is identical in each level of Factor B, and vice versa).

Example

- Look for changes from baseline introduced by Factor A and Factor B additively with interaction, so that the effect of changing the level of Factor A is different in each level of Factor B, and vice versa).

Two-way table: 4×3

Factor B

Factor A	1	2	3
1			
2			
3			
4			

Null Model : Baseline Mean Only

Null Model: no effect of either factor

Factor B

Factor A	1	2	3
1	β_1	β_1	β_1
2	β_1	β_1	β_1
3	β_1	β_1	β_1
4	β_1	β_1	β_1

Effect of Factor A only

Main Effect Only: A

Factor B

Factor A	1	2	3
1	β_1	β_1	β_1
2	$\beta_1 + \beta_2^{(A)}$	$\beta_1 + \beta_2^{(A)}$	$\beta_1 + \beta_2^{(A)}$
3	$\beta_1 + \beta_3^{(A)}$	$\beta_1 + \beta_3^{(A)}$	$\beta_1 + \beta_3^{(A)}$
4	$\beta_1 + \beta_4^{(A)}$	$\beta_1 + \beta_4^{(A)}$	$\beta_1 + \beta_4^{(A)}$

Effect of Factor B only

Main Effect Only: B

Factor B

Factor A	1	2	3
1	β_1	$\beta_1 + \beta_2^{(B)}$	$\beta_1 + \beta_3^{(B)}$
2	β_1	$\beta_1 + \beta_2^{(B)}$	$\beta_1 + \beta_3^{(B)}$
3	β_1	$\beta_1 + \beta_2^{(B)}$	$\beta_1 + \beta_3^{(B)}$
4	β_1	$\beta_1 + \beta_2^{(B)}$	$\beta_1 + \beta_3^{(B)}$

Effect of Factor A plus Effect of Factor B

Main Effects Only: A + B

Factor B

Factor A	1	2	3
1	β_1	$\beta_1 + \beta_2^{(B)}$	$\beta_1 + \beta_3^{(B)}$
2	$\beta_1 + \beta_2^{(A)}$	$\beta_1 + \beta_2^{(A)} + \beta_2^{(B)}$	$\beta_1 + \beta_2^{(A)} + \beta_3^{(B)}$
3	$\beta_1 + \beta_3^{(A)}$	$\beta_1 + \beta_3^{(A)} + \beta_2^{(B)}$	$\beta_1 + \beta_3^{(A)} + \beta_3^{(B)}$
4	$\beta_1 + \beta_4^{(A)}$	$\beta_1 + \beta_4^{(A)} + \beta_2^{(B)}$	$\beta_1 + \beta_4^{(A)} + \beta_3^{(B)}$

Main effects plus Interaction between A and B

Main Effects Plus Interaction: A + B + A:B

Factor B

Factor A	1	2	3
1	β_1	$\beta_1 + \beta_2^{(B)}$	$\beta_1 + \beta_3^{(B)}$
2	$\beta_1 + \beta_2^{(A)}$	$\beta_1 + \beta_2^{(A)} + \beta_2^{(B)} + \gamma_{22}^{(AB)}$	$\beta_1 + \beta_2^{(A)} + \beta_3^{(B)} + \gamma_{23}^{(AB)}$
3	$\beta_1 + \beta_3^{(A)}$	$\beta_1 + \beta_3^{(A)} + \beta_2^{(B)} + \gamma_{32}^{(AB)}$	$\beta_1 + \beta_3^{(A)} + \beta_3^{(B)} + \gamma_{33}^{(AB)}$
4	$\beta_1 + \beta_4^{(A)}$	$\beta_1 + \beta_4^{(A)} + \beta_2^{(B)} + \gamma_{42}^{(AB)}$	$\beta_1 + \beta_4^{(A)} + \beta_3^{(B)} + \gamma_{43}^{(AB)}$

Null Model : Baseline Mean Only

Null Model: no effect of either factor

Factor B

Factor A	1	2	3
1	10	10	10
2	10	10	10
3	10	10	10
4	10	10	10

Effect of Factor A only

Main Effect Only: A

Factor B

Factor A	1	2	3
1	10	10	10
2	10 + 2	10 + 2	10 + 2
3	10 + 5	10 + 5	10 + 5
4	10 + 8	10 + 8	10 + 8

Effect of Factor B only

Main Effect Only: B

Factor B

Factor A	1	2	3
1	10	10 + (-1)	10 + (-4)
2	10	10 + (-1)	10 + (-4)
3	10	10 + (-1)	10 + (-4)
4	10	10 + (-1)	10 + (-4)

Main effects of factors A and B only

Main Effects Only: A + B

Factor B

Factor A	1	2	3
1	10	10 + (-1)	10 + (-4)
2	10 + 2	10 + 2 + (-1)	10 + 2 + (-4)
3	10 + 5	10 + 5 + (-1)	10 + 5 + (-4)
4	10 + 8	10 + 8 + (-1)	10 + 8 + (-4)

Main effects plus Interaction between A and B

Main Effects Plus Interaction: A + B + A:B

Factor B

Factor A	1	2	3
1	10	10 + (-1)	10 + (-4)
2	10 + 2	10 + 2 + (-1) + (-2)	10 + 2 + (-4) + 1
3	10 + 5	10 + 5 + (-1) + (-5)	10 + 5 + (-4) + 8
4	10 + 8	10 + 8 + (-1) + 2	10 + 8 + (-4) + 7

Illegal models

Q. Why are the following models

- A:B
- A + A:B
- B + A:B

not considered ?

A. Because they make specific, perhaps unrealistic assumptions about the data, and they imply that the levels of the factors are not arbitrarily labelled.

R will not fit such models, although it appears that it does !

Illegal models

Recall the definition of interaction:

- Variation in the effect of changing levels of one factor at the different levels of the other factor.
- For example, the effect on the response mean of moving from level 1 to level 2 for Factor B is different at different levels of Factor A.

Consider the model

$$\text{A:B}$$

this model implies that all parameters apart from the baseline and the interaction parameters are zero.

Illegal models

Interaction: A:B

Factor B

Factor A	1	2	3
1	β_1	$\beta_1 + 0$	$\beta_1 + 0$
2	$\beta_1 + 0$	$\beta_1 + 0 + 0 + \gamma_{22}^{(AB)}$	$\beta_1 + 0 + 0 + \gamma_{23}^{(AB)}$
3	$\beta_1 + 0$	$\beta_1 + 0 + 0 + \gamma_{32}^{(AB)}$	$\beta_1 + 0 + 0 + \gamma_{33}^{(AB)}$
4	$\beta_1 + 0$	$\beta_1 + 0 + 0 + \gamma_{42}^{(AB)}$	$\beta_1 + 0 + 0 + \gamma_{43}^{(AB)}$

Illegal models

In this set-up,

- for Factor A, Level 1: the effect of moving from Level 1 to Level 2 of factor B is zero
- for Factor A, Level 3: the effect of moving from Level 1 to Level 2 of factor B is $\gamma_{32}^{(AB)}$.

Therefore, there is a fundamental difference between the way that we regard the levels of Factor A.

Illegal models

Main Effects of Factor A Plus Interaction: A + A:B

Factor B

Factor A	1	2	3
1	β_1	$\beta_1 + 0$	$\beta_1 + 0$
2	$\beta_1 + \beta_2^{(A)}$	$\beta_1 + \beta_2^{(A)} + 0 + \gamma_{22}^{(AB)}$	$\beta_1 + \beta_2^{(A)} + 0 + \gamma_{23}^{(AB)}$
3	$\beta_1 + \beta_3^{(A)}$	$\beta_1 + \beta_3^{(A)} + 0 + \gamma_{32}^{(AB)}$	$\beta_1 + \beta_3^{(A)} + 0 + \gamma_{33}^{(AB)}$
4	$\beta_1 + \beta_4^{(A)}$	$\beta_1 + \beta_4^{(A)} + 0 + \gamma_{42}^{(AB)}$	$\beta_1 + \beta_4^{(A)} + 0 + \gamma_{43}^{(AB)}$

Illegal models

In this set-up,

- for Factor A, Level 1: the effect of moving from Level 1 to Level 2 of Factor B is zero
- for Factor A, Level 3: the effect of moving from Level 1 to Level 2 of Factor B is $\gamma_{32}^{(AB)}$.

Illegal models

Therefore, there is a fundamental difference between the way that we regard the levels of Factor A. If we rearrange the labels of the levels of Factor A

we may get a different result.

Therefore, although it is possible in general to fit such models, it is no longer possible to talk of the effect of “Factor A".

How does R Handle Such Models ?

It is possible to use the models

A + A:B B + A:B A:B

in an ANOVA in R.

How does R Handle Such Models ?

A + B + A:B

```
> a<-3
> b<-5
> r<-6
> A<-rep(rep(1:a,each=b),r)
> B<-rep(rep(1:b,a),r)
> rept<-rep(1:r,b*a)
> mua<-c(0,2,3)
> mub<-c(1:5)
> X<-mua[A]+mub[B]+rnorm(r*b*a)*0.2
> FDdat<-data.frame(A=as.factor(A),B=as.factor(B),X)
> summary(aov(X~A+B+A:B, data=FDdat))

+             Df Sum Sq Mean Sq  F value Pr(>F)
+ A            2 141.50   70.75 1895.568 <2e-16 ***
+ B            4 181.50   45.37 1215.734 <2e-16 ***
+ A:B          8   0.37    0.05    1.229  0.294
+ Residuals   75   2.80    0.04
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

How does R Handle Such Models ?

A + A:B

```
> summary(aov(X~A+A:B, data=FDdat))

+             Df Sum Sq Mean Sq F value Pr(>F)
+ A            2  141.5   70.75  1895.6 <2e-16 ***
+ A:B         12  181.9   15.16   406.1 <2e-16 ***
+ Residuals   75    2.8    0.04
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

How does R Handle Such Models ?

This produces the usual ANOVA table, with terms including

A and A:B

However, in fact the model

A + B + A:B

has been fitted

- The results are just reported differently
- The terms B and A:B are reported together

How does R Handle Such Models ?

This is the default behaviour in the standard functions. However, it is not universal

- `supernova` exhibits different default behaviour
- `Anova` from the `car` library also has different behaviour

How does R Handle Such Models ?

```
> supernova(lm(X~A+B+A:B, data=FDdat))

+  Analysis of Variance Table (Type III SS)
+  Model: X ~ A + B + A:B
+
+                              SS df     MS       F   PRE     p
+  ----- --------------- | ------- -- ------ ------- ----- -----
+  Model (error reduced) | 323.364 14 23.097 618.850 .9914 .0000
+      A                 |  27.853  2 13.927 373.135 .9087 .0000
+      B                 |  57.798  4 14.449 387.143 .9538 .0000
+    A:B                 |   0.367  8  0.046   1.229 .1159 .2943
+  Error (from model)    |   2.799 75  0.037
+  ----- --------------- | ------- -- ------ ------- ----- -----
+  Total (empty model)   | 326.163 89  3.665

> summary(aov(X~A+B+A:B, data=FDdat))    #Previous analysis

+             Df Sum Sq Mean Sq  F value Pr(>F)
+ A            2 141.50   70.75 1895.568 <2e-16 ***
+ B            4 181.50   45.37 1215.734 <2e-16 ***
+ A:B          8   0.37    0.05    1.229  0.294
+ Residuals   75   2.80    0.04
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

How does R Handle Such Models ?

These results are different. To recover the original results, we need to specify that we want the

Type II sum of squares

which accounts for the presence of the interaction terms in a different way.

How does R Handle Such Models ?

```
> supernova(lm(X~A+B+A:B, data=FDdat), type=2)

+  Analysis of Variance Table (Type II SS)
+  Model: X ~ A + B + A:B
+
+                              SS df     MS        F   PRE     p
+  ----- --------------- | ------- -- ------ -------- ----- -----
+  Model (error reduced) | 323.364 14 23.097  618.850 .9914 .0000
+      A                 | 141.497  2 70.749 1895.568 .9806 .0000
+      B                 | 181.500  4 45.375 1215.734 .9848 .0000
+    A:B                 |   0.367  8  0.046    1.229 .1159 .2943
+  Error (from model)    |   2.799 75  0.037
+  ----- --------------- | ------- -- ------ -------- ----- -----
+  Total (empty model)   | 326.163 89  3.665
```

How does R Handle Such Models ?

```
> supernova(lm(X~A+A:B, data=FDdat))

+  Analysis of Variance Table (Type III SS)
+  Model: X ~ A + A:B
+
+                              SS df     MS        F   PRE     p
+  ----- --------------- | ------- -- ------ -------- ----- -----
+  Model (error reduced) | 323.364 14 23.097  618.850 .9914 .0000
+      A                 |  27.853  2 13.927  373.135 .9087 .0000
+    A:B                 | 181.867  4 45.467 1218.192 .9848 .0000
+  Error (from model)    |   2.799 75  0.037
+  ----- --------------- | ------- -- ------ -------- ----- -----
+  Total (empty model)   | 326.163 89  3.665

> supernova(lm(X~A+A:B, data=FDdat), type=2)

+  Analysis of Variance Table (Type II SS)
+  Model: X ~ A + A:B
+
+                              SS df     MS        F   PRE     p
+  ----- --------------- | ------- -- ------ -------- ----- -----
+  Model (error reduced) | 323.364 14 23.097  618.850 .9914 .0000
+      A                 | 141.497  2 70.749 1895.568 .9806 .0000
+    A:B                 | 181.867  4 45.467 1218.192 .9848 .0000
+  Error (from model)    |   2.799 75  0.037
+  ----- --------------- | ------- -- ------ -------- ----- -----
+  Total (empty model)   | 326.163 89  3.665
```

How does R Handle Such Models ?

Anova from the car package

```
> library(car)
> Anova(lm(X~A+B+A:B, data=FDdat))

+ Anova Table (Type II tests)
+
+ Response: X
+            Sum Sq Df   F value Pr(>F)
+ A         141.497  2 1895.5681 <2e-16 ***
+ B         181.500  4 1215.7341 <2e-16 ***
+ A:B         0.367  8    1.2288 0.2943
+ Residuals   2.799 75
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

How does R Handle Such Models ?

```
> Anova(lm(X~A+B+A:B, data=FDdat), type="III")

+ Anova Table (Type III tests)
+
+ Response: X
+             Sum Sq Df  F value Pr(>F)
+ (Intercept)  7.402  1 198.3253 <2e-16 ***
+ A           27.853  2 373.1349 <2e-16 ***
+ B           57.798  4 387.1432 <2e-16 ***
+ A:B          0.367  8   1.2288 0.2943
+ Residuals    2.799 75
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> Anova(lm(X~-1+A+B+A:B, data=FDdat), type="III")

+ Anova Table (Type III tests)
+
+ Response: X
+            Sum Sq Df   F value Pr(>F)
+ A         161.454  3 1441.9444 <2e-16 ***
+ B          57.798  4  387.1432 <2e-16 ***
+ A:B         0.367  8    1.2288 0.2943
+ Residuals   2.799 75
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

How does R Handle Such Models ?

This analysis tells us that

- `aov` and `Anova` report the Type II sums of squares by default in these multifactor settings
- `supernova` reports the Type III sums of squares

How does R Handle Such Models ?

Note: for models without interactions, Type II and Type III sums of squares are `identical` for the balanced design.

```
> supernova(lm(X~A+B, data=FDdat), type=2)

+  Analysis of Variance Table (Type II SS)
+  Model: X ~ A + B
+
+                              SS df     MS        F   PRE     p
+  ----- --------------- | ------- -- ------ -------- ----- -----
+  Model (error reduced) | 322.997  6 53.833 1411.221 .9903 .0000
+      A                 | 141.497  2 70.749 1854.664 .9781 .0000
+      B                 | 181.500  4 45.375 1189.500 .9829 .0000
+  Error (from model)    |   3.166 83  0.038
+  ----- --------------- | ------- -- ------ -------- ----- -----
+  Total (empty model)   | 326.163 89  3.665
```

How does R Handle Such Models ?

```
> supernova(lm(X~A+B, data=FDdat), type=3)

+  Analysis of Variance Table (Type III SS)
+  Model: X ~ A + B
+
+                              SS df     MS        F   PRE     p
+  ----- --------------- | ------- -- ------ -------- ----- -----
+  Model (error reduced) | 322.997  6 53.833 1411.221 .9903 .0000
+      A                 | 141.497  2 70.749 1854.664 .9781 .0000
+      B                 | 181.500  4 45.375 1189.500 .9829 .0000
+  Error (from model)    |   3.166 83  0.038
+  ----- --------------- | ------- -- ------ -------- ----- -----
+  Total (empty model)   | 326.163 89  3.665
```

Example: Batteries Data

A - Material

B - Temperature

Model A + B + A:B

```
> summary(aov(life~material*temp,data=Batteries))

+               Df Sum Sq Mean Sq F value      Pr(>F)
+ material       2  10633    5317   7.983     0.00189 **
+ temp           2  39083   19542  29.344 0.000000169 ***
+ material:temp  4   9438    2359   3.543     0.01897 *
+ Residuals     27  17981     666
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$SS = SST_A + SST_B + SSI_{AB} + SSE$$

Example: Batteries Data

A - Material

B - Temperature

Model A + A:B

```
> summary(aov(life~material+material:temp,data=Batteries))

+               Df Sum Sq Mean Sq F value     Pr(>F)
+ material       2  10633    5317   7.983    0.00189 **
+ material:temp  6  48521    8087  12.143 0.00000135 ***
+ Residuals     27  17981     666
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$SS = SST_A + SSI_{B:AB} + SSE$$

where

$$SSI_{B:AB} = SST_B + SSI_{AB}$$

Example: Batteries Data

A - Material

B - Temperature

Model B + A:B

```
> summary(aov(life~temp+material:temp,data=Batteries))

+               Df Sum Sq Mean Sq F value      Pr(>F)
+ temp           2  39083   19542  29.344 0.000000169 ***
+ temp:material  6  20071    3345   5.023     0.00142 **
+ Residuals     27  17981     666
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$SS = SST_B + SSI_{A:AB} + SSE$$

where

$$SSI_{A:AB} = SST_A + SSI_{AB}$$

Example: Batteries Data

A - Material

B - Temperature

Model A:B

```
> summary(aov(life~material:temp,data=Batteries))

+               Df Sum Sq Mean Sq F value      Pr(>F)
+ material:temp  8  59154    7394    11.1 0.000000861 ***
+ Residuals     27  17981     666
+ ---
+ Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$SS = SSI_{A:B:AB} + SSE$$

where

$$SSI_{A:B:AB} = SST_A + SST_B + SSI_{AB}$$

Final Note on ANOVA

We have studied the simplest design scenarios: extension to

- incomplete
- unbalanced
- nested
- random effect

designs are possible, but more complicated statistically.

Furthermore R has greater functionality: for example, it has the capability to carry out ANOVA-like analyses even for the case of non-equal variances (when Levene's test rejects the hypothesis of equal variances).