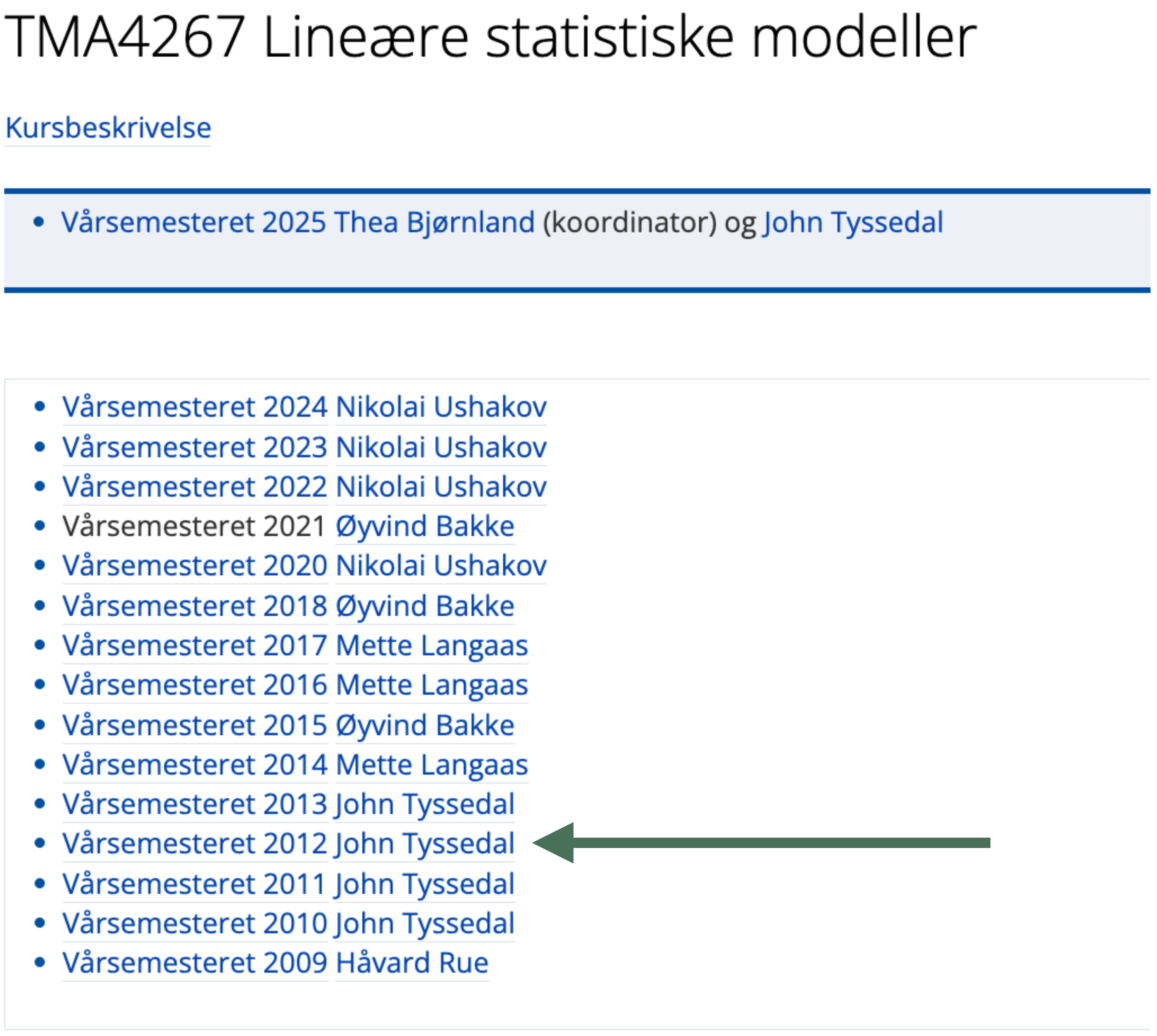
TMA4267 Linear statistical models

6. march 2025

Thea Bjørnland

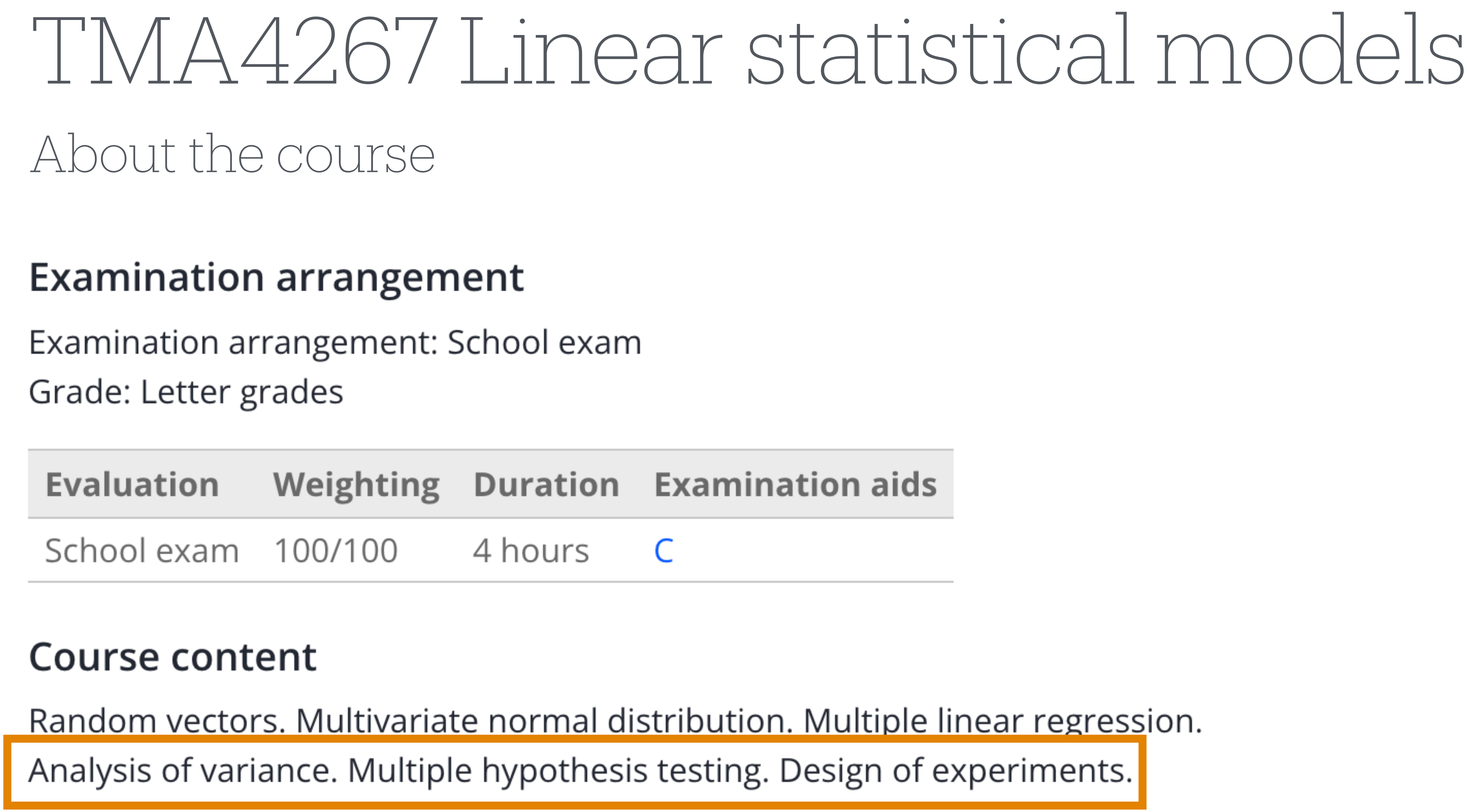


TMA4267 Lineære statistiske modeller

Kursbeskrivelse

- Vårsemesteret 2025 Thea Bjørnland (koordinator) og John Tyssedal

- Vårsemesteret 2024 Nikolai Ushakov
- Vårsemesteret 2023 Nikolai Ushakov
- Vårsemesteret 2022 Nikolai Ushakov
- Vårsemesteret 2021 Øyvind Bakke
- Vårsemesteret 2020 Nikolai Ushakov
- Vårsemesteret 2018 Øyvind Bakke
- Vårsemesteret 2017 Mette Langaas
- Vårsemesteret 2016 Mette Langaas
- Vårsemesteret 2015 Øyvind Bakke
- Vårsemesteret 2014 Mette Langaas
- Vårsemesteret 2013 John Tyssedal
- Vårsemesteret 2012 John Tyssedal
- Vårsemesteret 2011 John Tyssedal
- Vårsemesteret 2010 John Tyssedal
- Vårsemesteret 2009 Håvard Rue



TMA4267 Linear statistical models

About the course

Examination arrangement

Examination arrangement: School exam
Grade: Letter grades

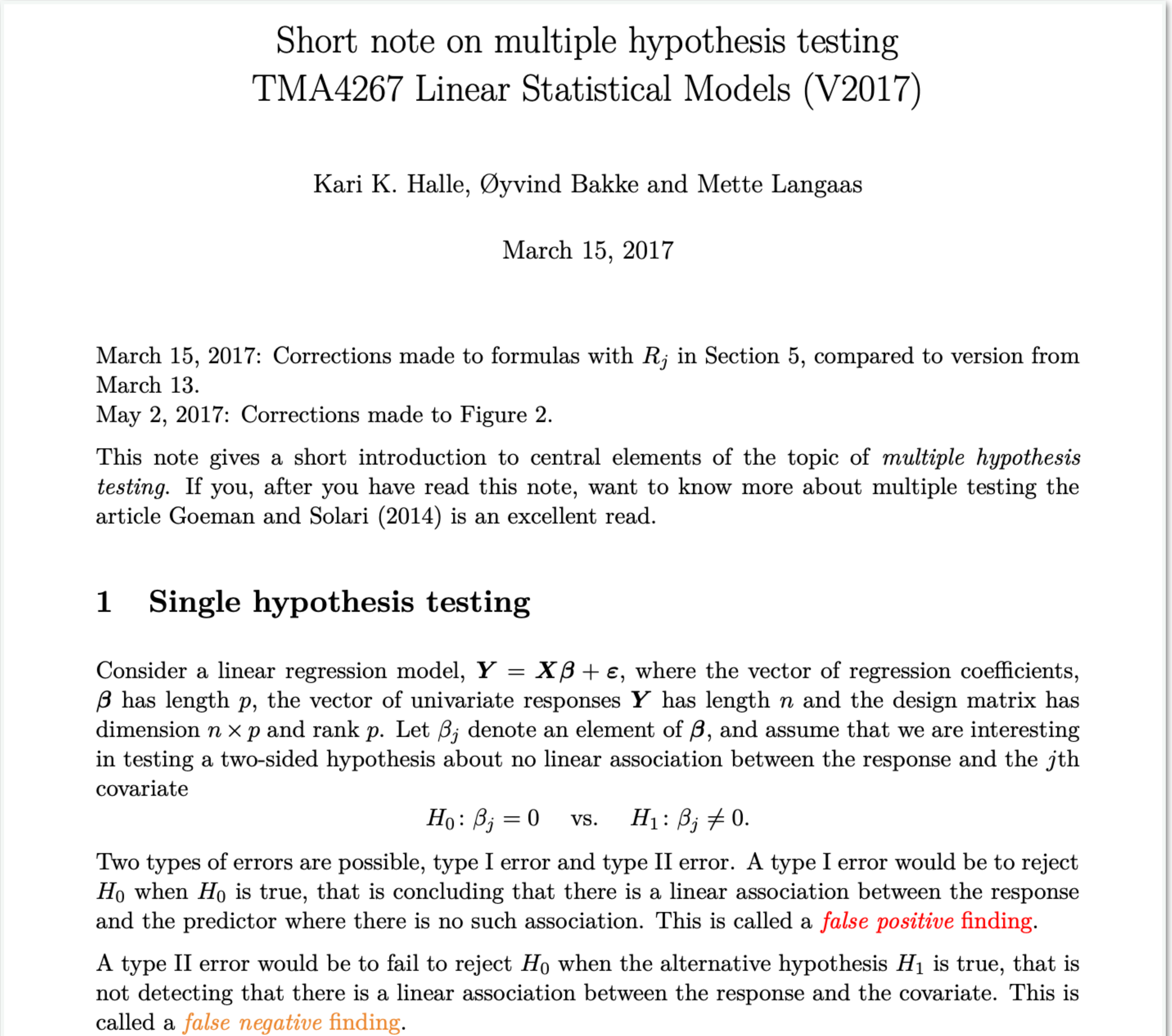
Evaluation	Weighting	Duration	Examination aids
School exam	100/100	4 hours	C

Course content

Random vectors. Multivariate normal distribution. Multiple linear regression. Analysis of variance. Multiple hypothesis testing. Design of experiments.

Multiple testing

https://www.math.ntnu.no/emner/TMA4267/2025v/files/multtest.pdf



Short note on multiple hypothesis testing
TMA4267 Linear Statistical Models (V2017)

Kari K. Halle, Øyvind Bakke and Mette Langaas

March 15, 2017

March 15, 2017: Corrections made to formulas with R_j in Section 5, compared to version from March 13.
May 2, 2017: Corrections made to Figure 2.

This note gives a short introduction to central elements of the topic of *multiple hypothesis testing*. If you, after you have read this note, want to know more about multiple testing the article Goeman and Solari (2014) is an excellent read.

1 Single hypothesis testing

Consider a linear regression model, $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where the vector of regression coefficients, $\boldsymbol{\beta}$ has length p, the vector of univariate responses $\boldsymbol{Y}$ has length n and the design matrix has dimension $n \times p$ and rank p. Let β_j denote an element of $\boldsymbol{\beta}$, and assume that we are interesting in testing a two-sided hypothesis about no linear association between the response and the jth covariate

$$H_0\colon \beta_j = 0 \quad \text{vs.} \quad H_1\colon \beta_j \neq 0.$$

Two types of errors are possible, type I error and type II error. A type I error would be to reject H_0 when H_0 is true, that is concluding that there is a linear association between the response and the predictor where there is no such association. This is called a *false positive* finding.

A type II error would be to fail to reject H_0 when the alternative hypothesis H_1 is true, that is not detecting that there is a linear association between the response and the covariate. This is called a *false negative* finding.

Repetition

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_i \qquad Y = X\beta + \varepsilon$$

$$H_0 : C\beta = 0$$

$$H_1 : C\beta \neq 0$$

```
Call:
lm(formula = y ~ x1 + x2)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.9807 -5.1992 -0.0145  3.4825  9.9923 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   0.8972     3.5586   0.252 0.803971    
x1            1.9854     0.4060   4.890 0.000138 ***
x2           -3.9486     2.0082  -1.966 0.065821 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.938 on 17 degrees of freedom
Multiple R-squared:  0.6491,	Adjusted R-squared:  0.6078 
F-statistic: 15.72 on 2 and 17 DF,  p-value: 0.0001363
```

$$C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Repetition

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_i \qquad Y = X\beta + \varepsilon$$

$$H_0 : C\beta = 0$$

$$H_1 : C\beta \neq 0$$

```
Call:
lm(formula = y ~ x1 + x2)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.9807 -5.1992 -0.0145  3.4825  9.9923 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   0.8972     3.5586   0.252 0.803971    
x1            1.9854     0.4060   4.890 0.000138 ***
x2           -3.9486     2.0082  -1.966 0.065821 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.938 on 17 degrees of freedom
Multiple R-squared:  0.6491,	Adjusted R-squared:  0.6078 
F-statistic: 15.72 on 2 and 17 DF,  p-value: 0.0001363
```

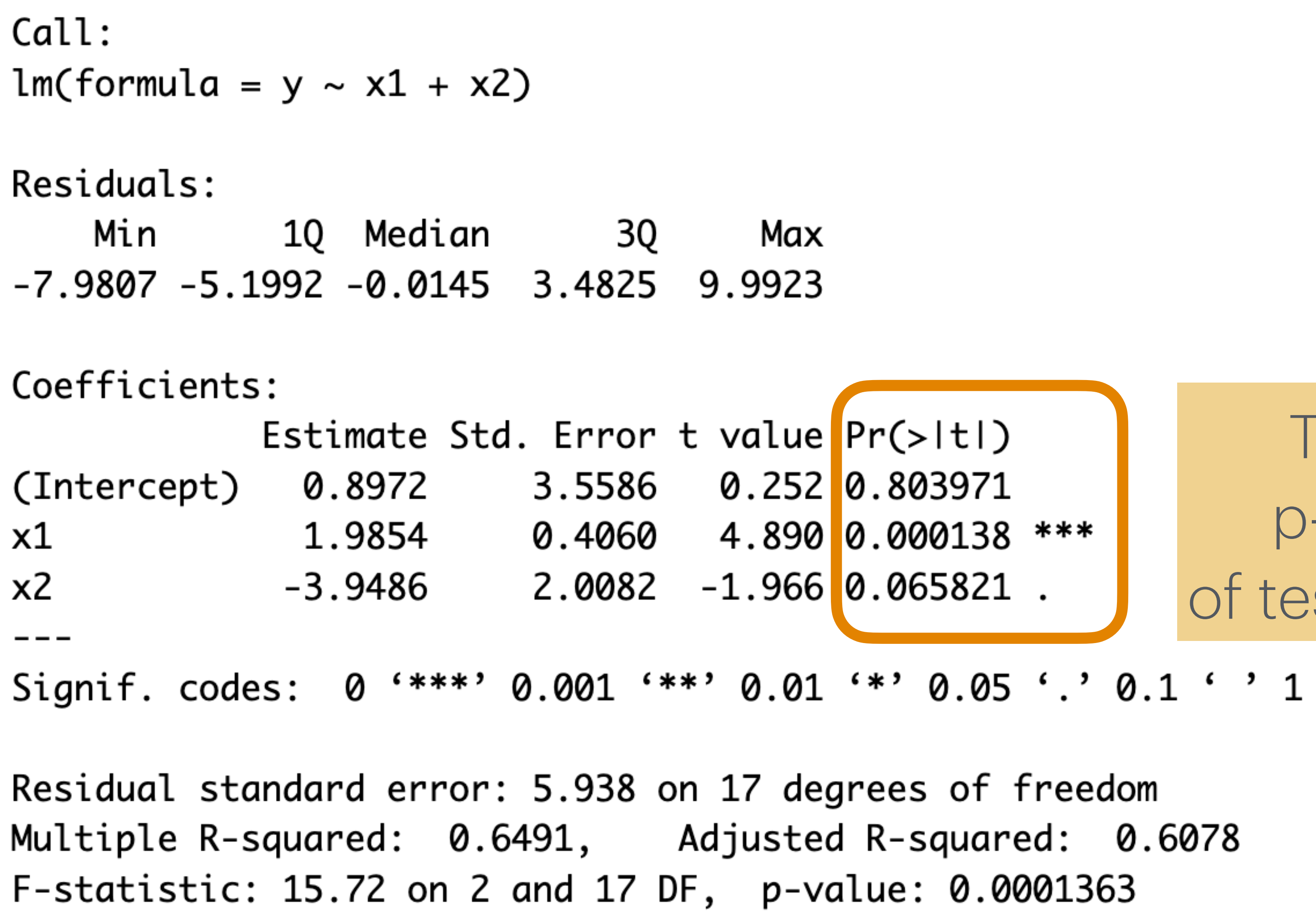
$$C = [0 \quad 1 \quad 0]$$

Repetition

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_i \qquad Y = X\beta + \varepsilon$$

$$H_0 : C\beta = 0$$

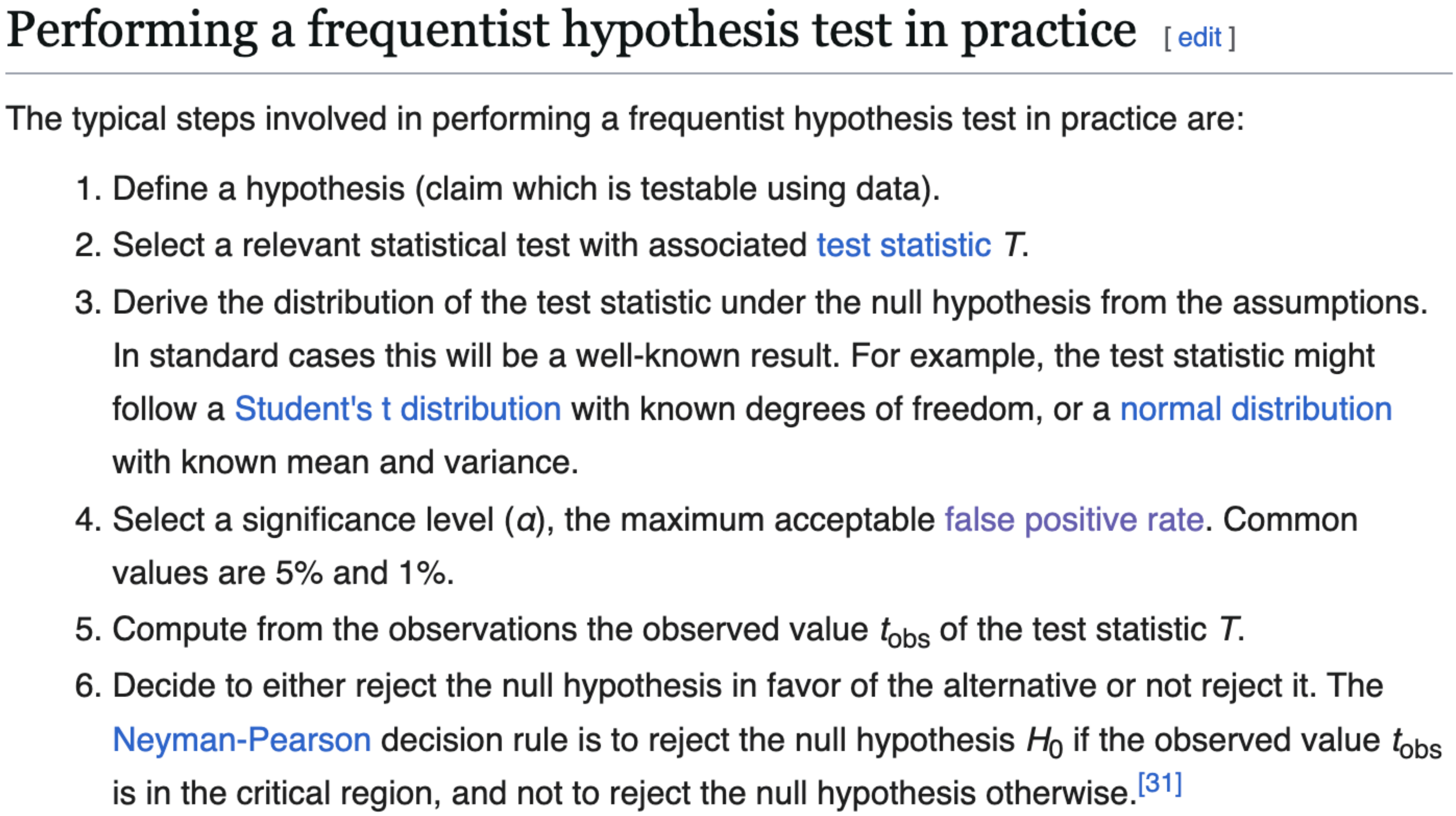
$$H_1 : C\beta \neq 0$$



```
Call:
lm(formula = y ~ x1 + x2)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.9807 -5.1992 -0.0145  3.4825  9.9923 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   0.8972     3.5586   0.252 0.803971    
x1            1.9854     0.4060   4.890 0.000138 ***
x2           -3.9486     2.0082  -1.966 0.065821 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.938 on 17 degrees of freedom
Multiple R-squared:  0.6491,	Adjusted R-squared:  0.6078 
F-statistic: 15.72 on 2 and 17 DF,  p-value: 0.0001363
```

Today we focus on p-values, regardless of test statistic (Z, T, F, etc)

Hypothesis testing: some basics



Performing a frequentist hypothesis test in practice

The typical steps involved in performing a frequentist hypothesis test in practice are:

1. Define a hypothesis (claim which is testable using data).
2. Select a relevant statistical test with associated test statistic T.
3. Derive the distribution of the test statistic under the null hypothesis from the assumptions. In standard cases this will be a well-known result. For example, the test statistic might follow a Student's t distribution with known degrees of freedom, or a normal distribution with known mean and variance.
4. Select a significance level (α), the maximum acceptable false positive rate. Common values are 5% and 1%.
5. Compute from the observations the observed value t_{obs} of the test statistic T.
6. Decide to either reject the null hypothesis in favor of the alternative or not reject it. The Neyman-Pearson decision rule is to reject the null hypothesis H_0 if the observed value t_{obs} is in the critical region, and not to reject the null hypothesis otherwise.[31]

https://en.wikipedia.org/wiki/Statistical_hypothesis_test

Hypothesis testing: some basics

The null hypothesis H_0 is never proven, we search for evidence of H_1

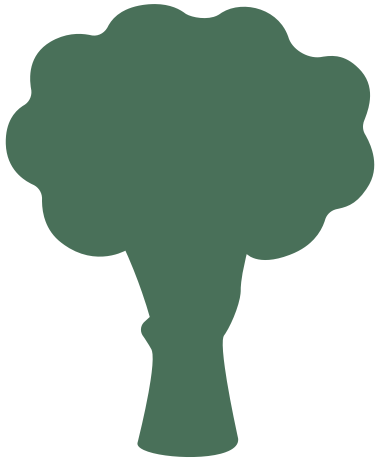
"By chance" often refers to events that can occur if H_0 is true

"Evidence" is data that are sufficiently unlikely to occur under H_0,
and points towards H_1

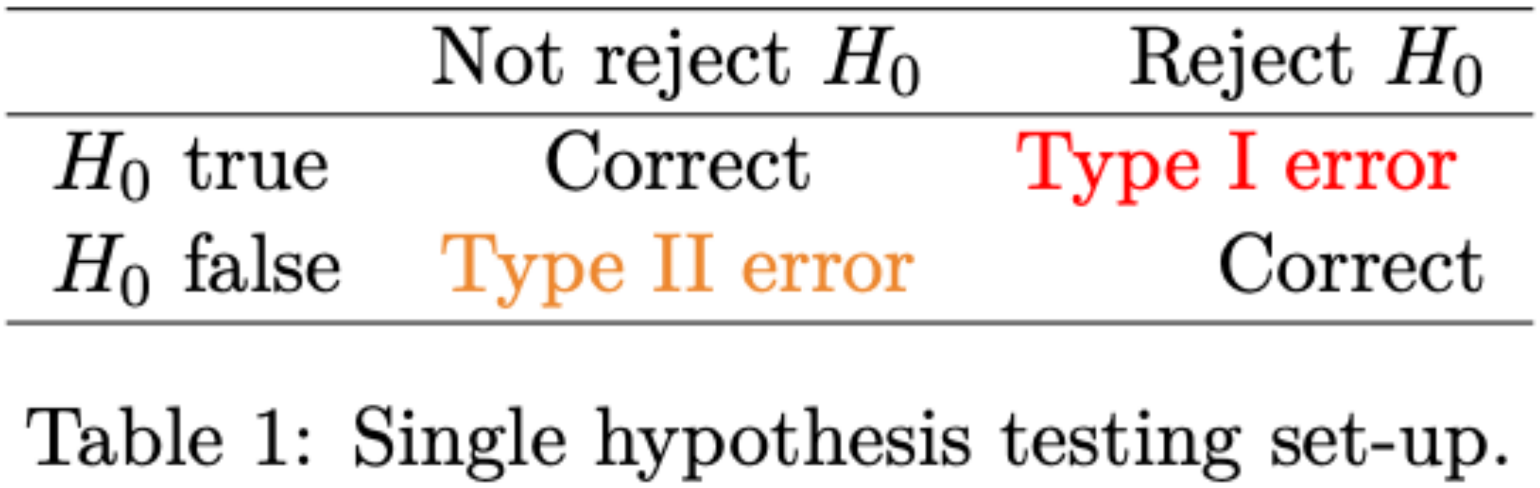
Statistical evidence is not necessarily the same as truth/causality,
but it can help us (researchers, industry, etc.) to learn or take actions

Example/repetition: Simple hypothesis

Broccoli consumption



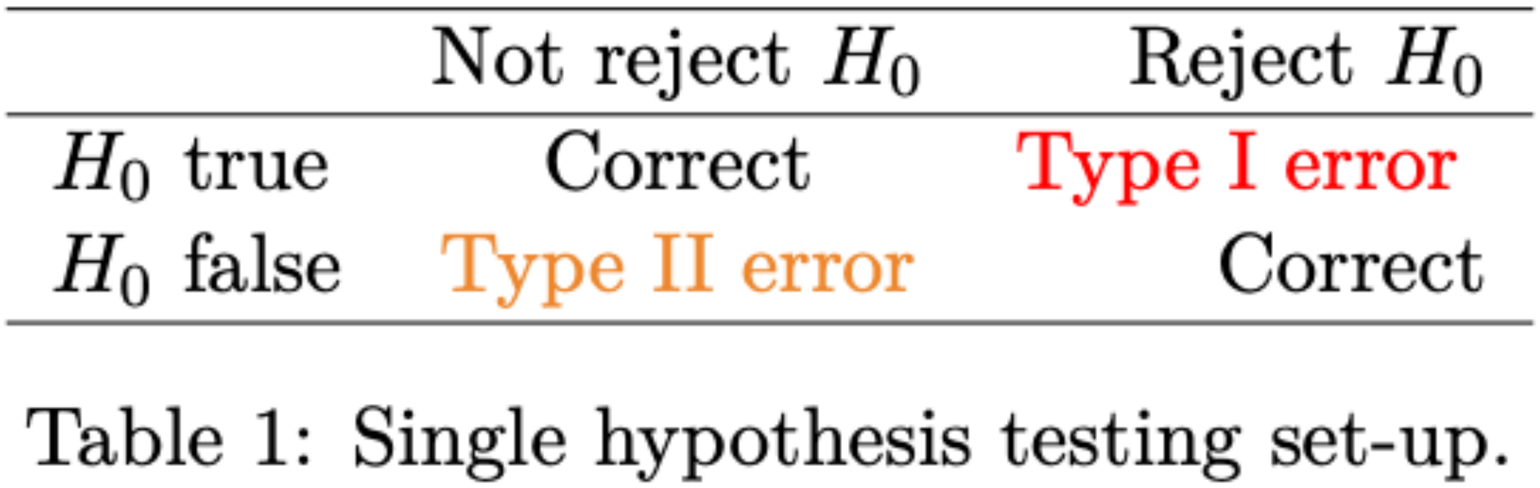
What is good enough evidence for rejecting H_0?



	Not reject H_0	Reject H_0
H_0 true	Correct	Type I error
H_0 false	Type II error	Correct

Table 1: Single hypothesis testing set-up.

What is good enough evidence for rejecting H_0?



	Not reject H_0	Reject H_0
H_0 true	Correct	Type I error
H_0 false	Type II error	Correct

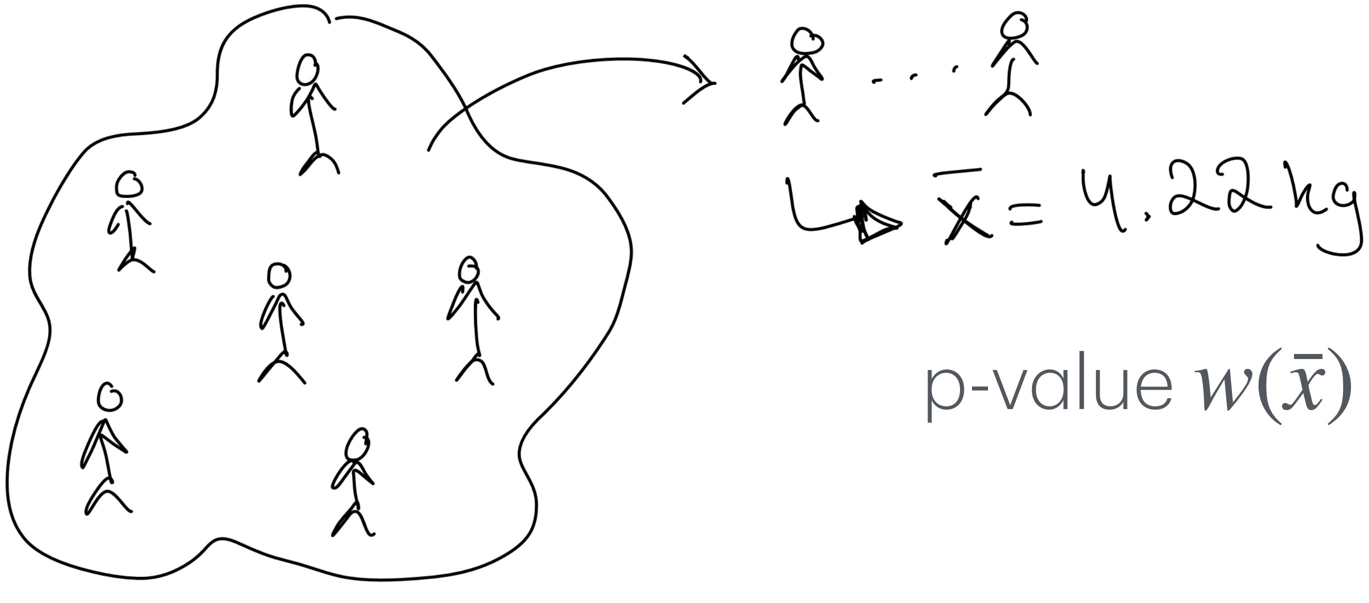
Table 1: Single hypothesis testing set-up.

If we want to control the type-1 error to some level α (e.g. 0.05), then we can only reject H_0 if we obtain a p-value less than α.

Equivalently, we can find a critical value to compare a test statistic (such as an average) to.

Broccoli example: Reject H_0 if we observe $\bar{x} > k$
where k is such that $P(\bar{X} > k \mid H_0) = \alpha$

Example: Broccoli consumption

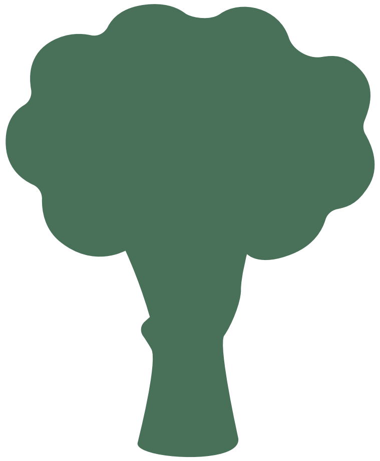


$\bar{x} = 4.22$ kg

p-value $w(\bar{x}) = 0.41$

What is the distribution of the p-value when H_0 is true?

Make a sketch!



Summing up first part

The p-value is a function of the data (sample) and will therefore vary every time we repeat an experiment.

Small p-values are generally taken as evidence of the alternative hypothesis

We control the type-1 error rate at some level α by rejecting H_0 only when we observe $p < \alpha$

Multiple testing

We accept a 5% chance of false positive findings.

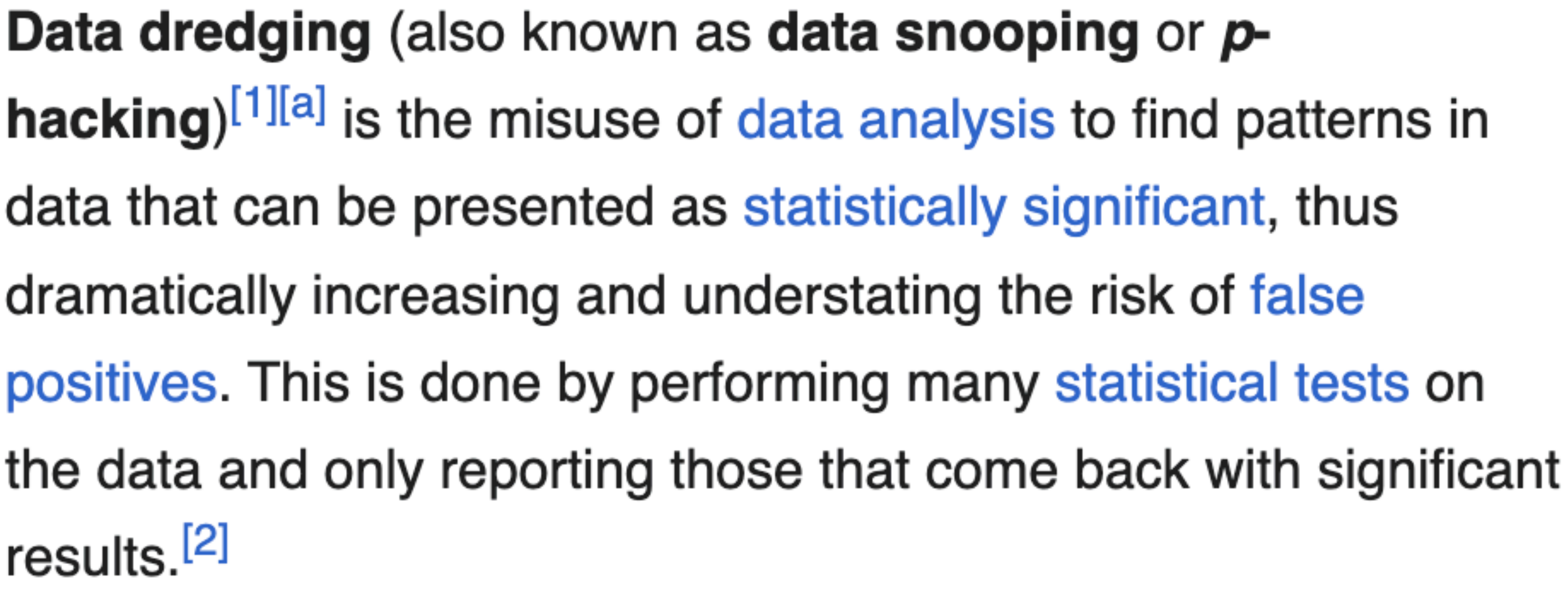
Should all 5 p-values be compared to 0.05?

If the p-values can be assumed independent,
and all 5 null hypotheses are true,
then we have a 23% chance of at least one
false positive if all p-values are compared to 0.05

Multiple testing

If an experiment is repeated many times, there is a large chance of (at least once) observing $p \leq \alpha$

This 'finding' should not be taken as evidence of anything at all



Data dredging (also known as **data snooping** or ***p*-hacking**)[1][a] is the misuse of data analysis to find patterns in data that can be presented as statistically significant, thus dramatically increasing and understating the risk of false positives. This is done by performing many statistical tests on the data and only reporting those that come back with significant results.[2]

https://en.wikipedia.org/wiki/Data_dredging

What if we need to do many tests?

We accept a 5% chance of false positive findings.

a) Adjust the p-values so that they can be compared to 0.05

b) Find a "local" level that we compare the individual p-values to

Bonferroni method

When we perform m tests, we control the FWER at the level α, by comparing each p-value to the local significance level $\alpha_{loc} = \alpha / m$

The Bonferroni method is often conservative (a larger value for α_{loc} could have been used)

Summary

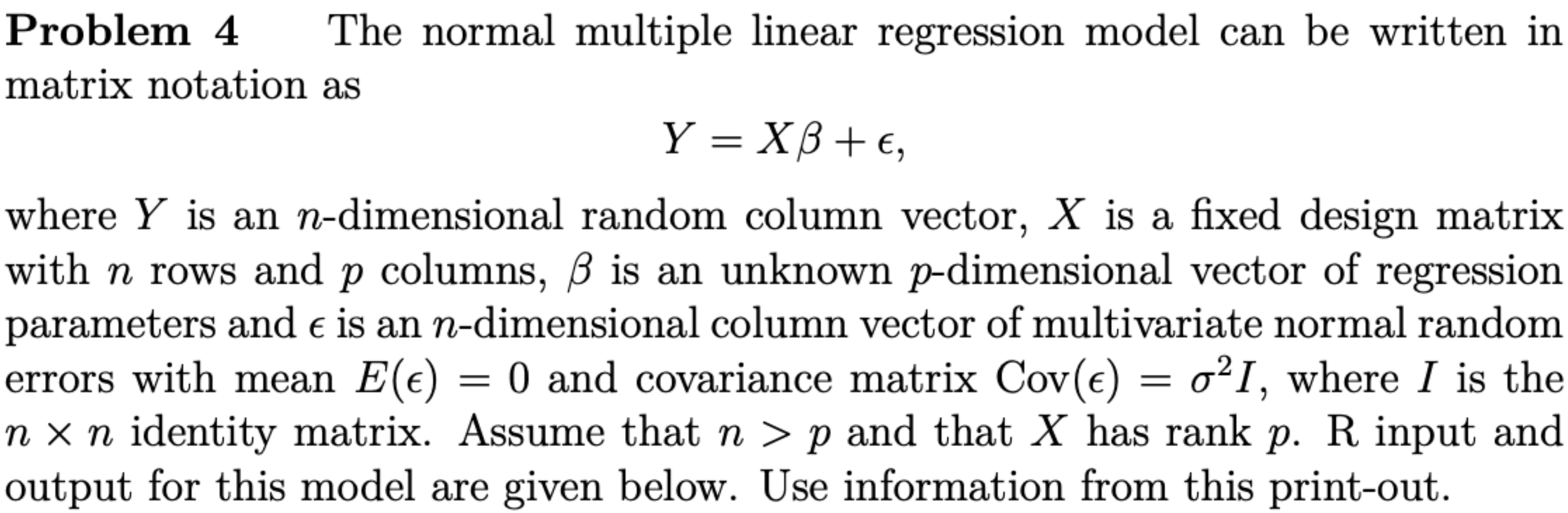
Family-wise error rate (FWER) is the probability of at least one false positive when m tests are performed

The Bonferroni method:
Control the FWER at level α by comparing each of the m p-values to $\alpha_{loc} = \alpha/m$

The Sidak method:
If p-values can be assumed independent*,
control the FWER at level α by comparing each of the m p-values to $\alpha_{loc} = 1 - (1 - \alpha)^{1/m}$

* satisfies Sidak equations

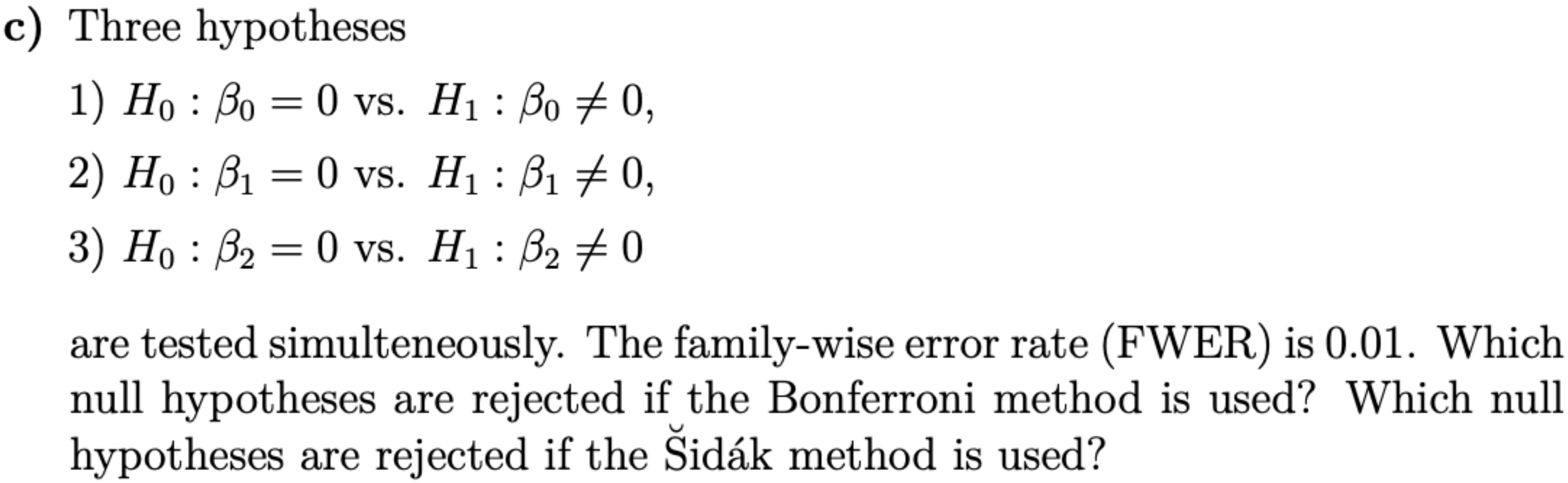
Exam 2024 (summer)



Problem 4 The normal multiple linear regression model can be written in matrix notation as

$$Y = X\beta + \epsilon,$$

where Y is an n-dimensional random column vector, X is a fixed design matrix with n rows and p columns, β is an unknown p-dimensional vector of regression parameters and ϵ is an n-dimensional column vector of multivariate normal random errors with mean $E(\epsilon) = 0$ and covariance matrix $\text{Cov}(\epsilon) = \sigma^2 I$, where I is the $n \times n$ identity matrix. Assume that $n > p$ and that X has rank p. R input and output for this model are given below. Use information from this print-out.



c) Three hypotheses

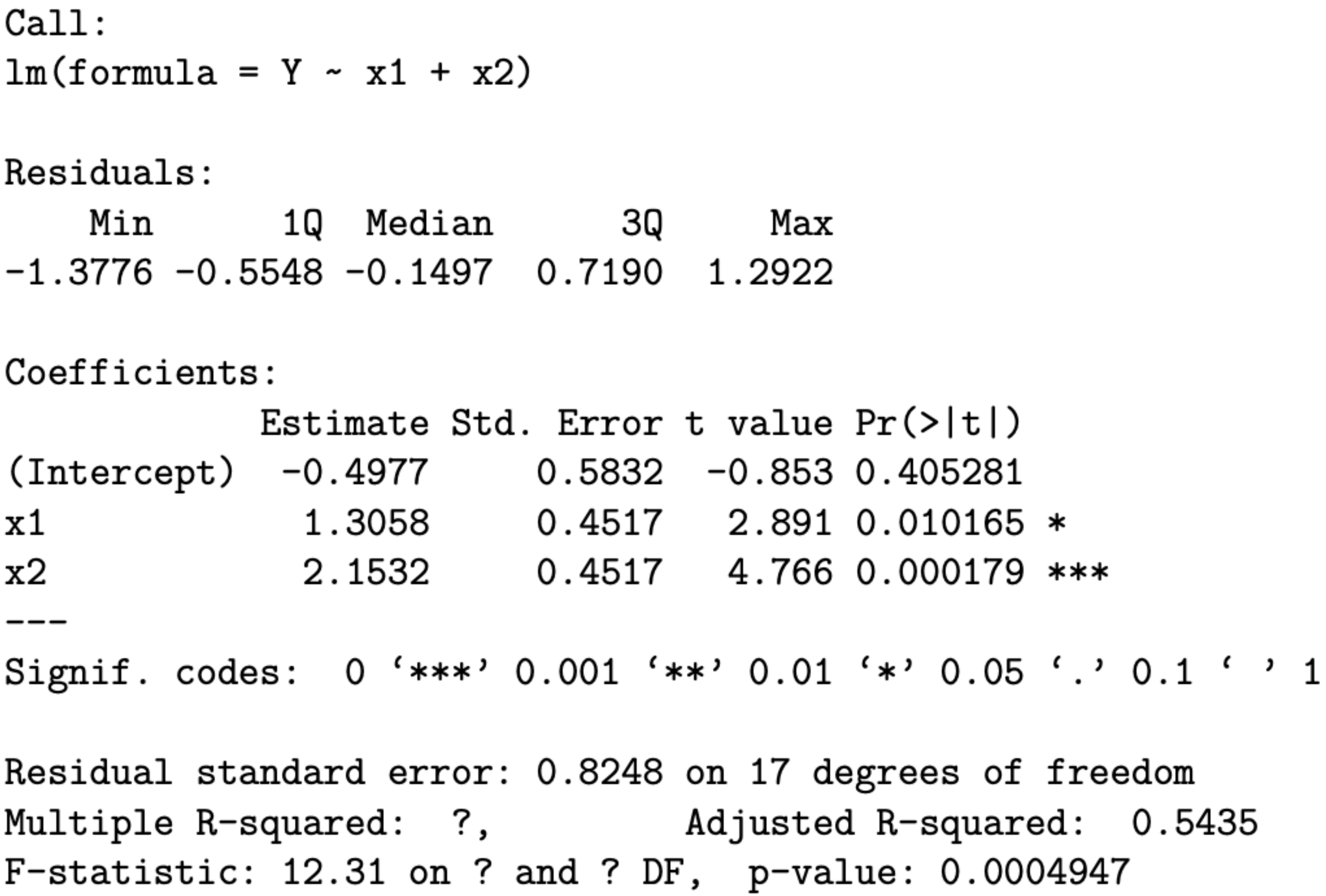
1) $H_0 : \beta_0 = 0$ vs. $H_1 : \beta_0 \neq 0$,

2) $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$,

3) $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$

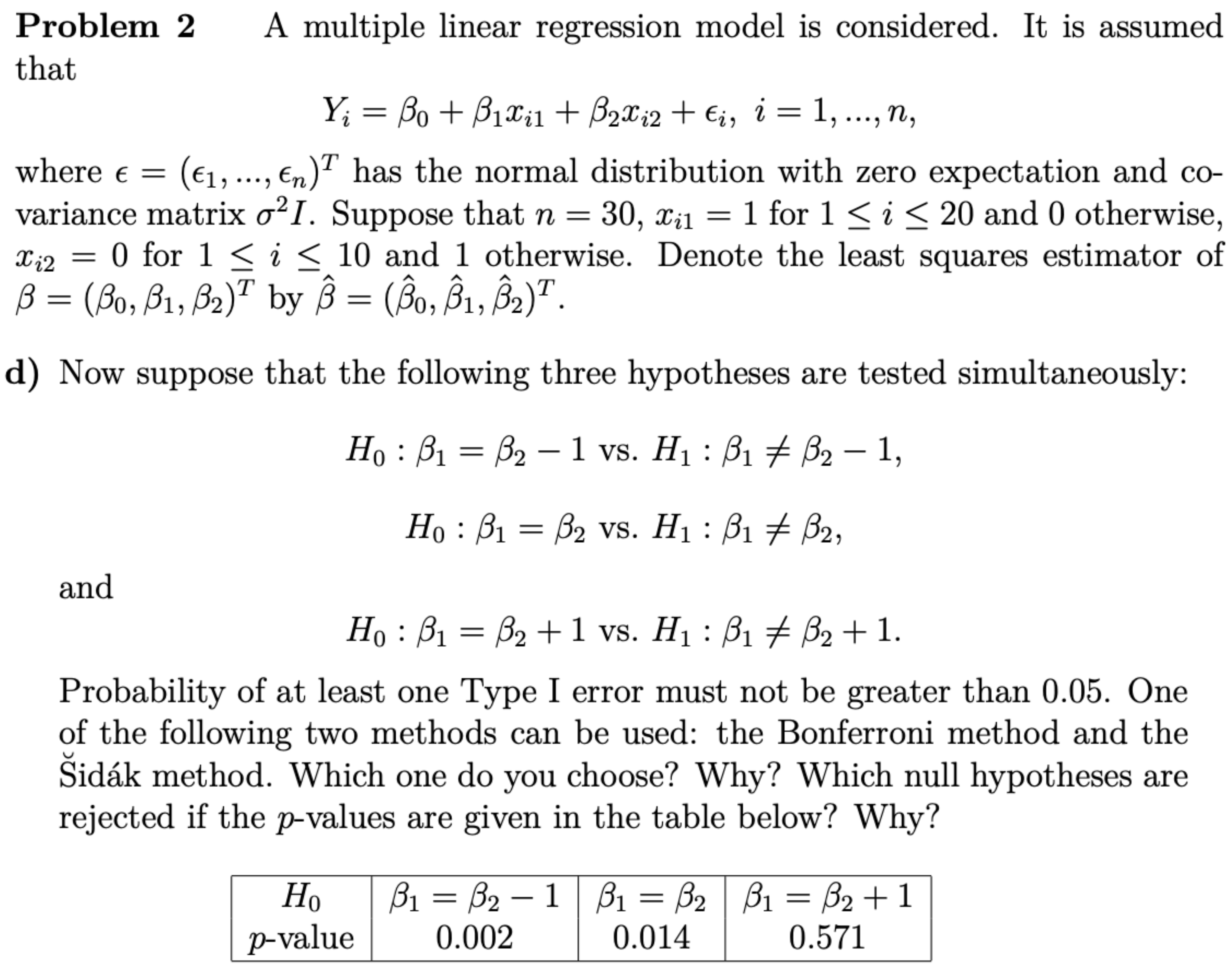
are tested simulteneously. The family-wise error rate (FWER) is 0.01. Which null hypotheses are rejected if the Bonferroni method is used? Which null hypotheses are rejected if the Šidák method is used?

Exam 2024 (summer)



```
Call:
lm(formula = Y ~ x1 + x2)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.3776 -0.5548 -0.1497  0.7190  1.2922 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -0.4977     0.5832  -0.853 0.405281    
x1            1.3058     0.4517   2.891 0.010165 *  
x2            2.1532     0.4517   4.766 0.000179 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8248 on 17 degrees of freedom
Multiple R-squared:  ?,         Adjusted R-squared:  0.5435 
F-statistic: 12.31 on ? and ? DF,  p-value: 0.0004947
```

Exam 2023



Problem 2 A multiple linear regression model is considered. It is assumed that

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i, \ i = 1, ..., n,$$

where $\epsilon = (\epsilon_1, ..., \epsilon_n)^T$ has the normal distribution with zero expectation and covariance matrix $\sigma^2 I$. Suppose that $n = 30$, $x_{i1} = 1$ for $1 \le i \le 20$ and 0 otherwise, $x_{i2} = 0$ for $1 \le i \le 10$ and 1 otherwise. Denote the least squares estimator of $\beta = (\beta_0, \beta_1, \beta_2)^T$ by $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)^T$.

d) Now suppose that the following three hypotheses are tested simultaneously:

$$H_0 : \beta_1 = \beta_2 - 1 \text{ vs. } H_1 : \beta_1 \neq \beta_2 - 1,$$

$$H_0 : \beta_1 = \beta_2 \text{ vs. } H_1 : \beta_1 \neq \beta_2,$$

and

$$H_0 : \beta_1 = \beta_2 + 1 \text{ vs. } H_1 : \beta_1 \neq \beta_2 + 1.$$

Probability of at least one Type I error must not be greater than 0.05. One of the following two methods can be used: the Bonferroni method and the Šidák method. Which one do you choose? Why? Which null hypotheses are rejected if the p-values are given in the table below? Why?

H_0	$\beta_1 = \beta_2 - 1$	$\beta_1 = \beta_2$	$\beta_1 = \beta_2 + 1$
p-value	0.002	0.014	0.571

You should now be able to discuss this:

When working with observations in a large data-set,
100 (exploratory) comparisons were made. One comparison (difference between lifestyle A and lifestyle B on body mass index) gave a p-value less than the chosen significance level alpha = 0.05; p = 0.037.

A study was published and reported a significant difference between lifestyle A and lifestyle B (p = 0.037).

What is the problem here?