# TMA4267 Linear Statistical Models V2017 (L21+L22)

## Summing up

Mette Langaas

Department of Mathematical Sciences, NTNU

To be lectured: April 25 and 28, 2017

# Outline: Summing up (today and on Friday)

- Learning outcomes
- Overview, important concepts and exam problems V2016:
  - Part 1: Multivariate RVs and the multivariate normal distribution
  - Part 2: Linear regression
  - Part 3: Hypothesis testing and analysis of variance
  - Part 4: Design of experiments
- Final reading list
- Exam
- Activities before the exam

# TMA4267 Linear statistical models
# Learning outcome, Knowledge

- The student has strong theoretical knowledge about the most popular statistical models and methods that are used in science and technology, with emphasis on regression-type statistical models.

# TMA4267 Linear statistical models
## Learning outcome, Knowledge

- The student has strong theoretical knowledge about the most popular statistical models and methods that are used in science and technology, with emphasis on regression-type statistical models.
- The statistical properties of the multivariate normal distribution are well known to the student, and the student is familiar with the role of the multivariate normal distribution within linear statistical models.

# Linear Statistical Models (L1)

Simple linear regression (height of child explained by mid-parent height):

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

Multiple linear regression (also include other explanatory variables):

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon$$

The multiple linear regression model is our linear statistical model! So, why is this course not called "Regression"?

# Linear Statistical Models (L1)

Simple linear regression (height of child explained by mid-parent height):

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

Multiple linear regression (also include other explanatory variables):

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon$$

The multiple linear regression model is our linear statistical model!
So, why is this course not called "Regression"?
We include theory that focus on mathematical understanding: multivariate random variables, the multivariate normal distribution, projections, idempotent matrices, multiple hypothesis testing, design of experiments, ....

# TMA4267 Linear statistical models
## Learning outcome, Skills

- The student knows how to design an experiment and
- how to collect informative data of high quality to study a phenomenon of interest.
- Subsequently, the student is able to choose a suitable statistical model,
- apply sound statistical methods, and
- perform the analyses using statistical software.
- The student knows how to present the results from the statistical analyses, and how to draw conclusions about the phenomenon under study.

# TMA4267: Parts

- Part 1: Multivariate RVs and the multivariate normal distribution [L1-L6, CompEx1, RecEx1-2]
  - Data consists of simultaneous measurements on many variables: we work with *random vectors and random matrices*.
  - There is a strong connection between the *multivariate normal distribution* and the classical linear model.

# TMA4267: Parts

- Part 1: Multivariate RVs and the multivariate normal distribution [L1-L6, CompEx1, RecEx1-2]
  - Data consists of simultaneous measurements on many variables: we work with *random vectors and random matrices*.
  - There is a strong connection between the *multivariate normal distribution* and the classical linear model.
- Part 2: Linear regression [L7-L12, CompEx2, RecEx3-4]
  - We want to understand the relationship between many variables: with focus on linear relationships through the *classical linear model* (multiple linear regression).

# TMA4267: Parts

- Part 1: Multivariate RVs and the multivariate normal distribution [L1-L6, CompEx1, RecEx1-2]
  - Data consists of simultaneous measurements on many variables: we work with *random vectors and random matrices*.
  - There is a strong connection between the *multivariate normal distribution* and the classical linear model.
- Part 2: Linear regression [L7-L12, CompEx2, RecEx3-4]
  - We want to understand the relationship between many variables: with focus on linear relationships through the *classical linear model* (multiple linear regression).
- Part 3: Hypothesis testing and analysis of variance [L13-16, CompEx3, RecEx5]
  - We need to know how the scientific process - which often lead to performing hypotheses test (liner hypotheses), and to know about issues with reproducibility and multiple tests.

# TMA4267: Parts

- Part 1: Multivariate RVs and the multivariate normal distribution [L1-L6, CompEx1, RecEx1-2]
  - Data consists of simultaneous measurements on many variables: we work with *random vectors and random matrices*.
  - There is a strong connection between the *multivariate normal distribution* and the classical linear model.
- Part 2: Linear regression [L7-L12, CompEx2, RecEx3-4]
  - We want to understand the relationship between many variables: with focus on linear relationships through the *classical linear model* (multiple linear regression).
- Part 3: Hypothesis testing and analysis of variance [L13-16, CompEx3, RecEx5]
  - We need to know how the scientific process - which often lead to performing hypotheses test (liner hypotheses), and to know about issues with reproducibility and multiple tests.
- Part 4: Design of experiments [L17-20, CompEx4, RecEx6]
  - If we want to collect data, we need to do know how to *design and perform an experiment*, that we analyse using the methods of Part 2.

# Final grade in TMA4267

- 20% of final grade from the 4 compulsory exercises,
- and the remaining 80% on the 4hrs written exam.
- Written exam:
  - mostly focussed on the "knowledge learning outcome"
  - 8 "questions" each with maximum 10 points score
  - the plan is 3*Easy+3*Medium+2*Hard
  - the plan is 3 from Part 1, 3 from Part 2, 1 from Part 3 and 1 from Part 4.
- Remember that all answers must be *justified* and *correct notation and vocabulary* used to score well.
- The written exam must give at least 41% score (that is at least 32-33 out of 80 points) for a passing grade.

# Part 1: Multivariate RVs and the multivariate normal distribution [L1-L6, CompEx1, RecEx1-2]

Curriculum:

- Härdle, Simar (2015): Applied Multivariate Statistical Analysis, Fourth edition, Springer.
  - Chapter 2
  - Chapter 3.3 (89-93)
  - Chapter 4.1-4.5 (117-149)
  - Chapter 5.1 (p.181-190)
  - Chapter 11.1-11.3 (p. 319-331)
- Fahrmeir, Kneib, Lang and Marx (2013): Regression, Springer.
  - Appendix B
- Slides and handouts ($\sim$ 200 pages): `https://www.math.ntnu.no/emner/TMA4267/2017v/TMA4267V2017Part1.pdf`

# First half of Part 1: Working with random vectors and matrices

- $\boldsymbol{X}$ $p$-dimensional random vector, characterized by
- pdf $f$, and/or cdf $F$, and/or momentgenerating function $M_X(t) = \mathrm{E}(e^{\boldsymbol{t}^T \boldsymbol{X}})$.
- Moments: E and Cov, and rules for linear and quadratic forms (here properties of idempotent matrices comes in).
  - $\boldsymbol{X}$ has mean $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu}$ and covariance matrix $\mathrm{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}$, and $\boldsymbol{C}$ is a constant matrix. Then $\boldsymbol{CX}$ has mean $\mathrm{E}(\boldsymbol{CX}) = \boldsymbol{C\mu}$ and $\mathrm{Cov}(\boldsymbol{CX}) = \boldsymbol{C\Sigma C}^T$.
  - The "trace-formula": $\mathrm{E}(\boldsymbol{X}^T \boldsymbol{AX}) = \mathrm{tr}(\boldsymbol{A\Sigma}) - \boldsymbol{\mu}^T \boldsymbol{A\mu}$.
- Understanding the covariance matrix $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{X}) = E((\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T)$, and spectral decomposition for positive definite covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{P\Lambda P}^T$, principal components.

# Second half of Part 1: The multivariate normal distribution

- Derivation, pdf $f$ and mgf $M_X(t)$.
- Properties galore!
- Connections to chi-square, t and F distributions.
- Connected to regression: distribution of errors!
- Connection to quadratic forms and idempotent matrices, used in proofs in Part 2.

# Properties galore

Let $\boldsymbol{X}_{(p\times 1)}$ be a random vector from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

1. The grapical contours of the mvN are ellipsoids (shown using spectral decomposition). [CompEx1.1b]
2. Linear combinations of components of $\boldsymbol{X}$ are (multivariate) normal (proof using MGF). [CompEx1.1a]
3. All subsets of the components of $\boldsymbol{X}$ are (multivariate) normal (special case of the above).
4. Zero covariance implies that the corresponding components are independently distributed (proof using MGF). [CompEx1.1a]
5. $\boldsymbol{A\Sigma B}^T = \boldsymbol{0} \Leftrightarrow \boldsymbol{AX}$ and $\boldsymbol{BX}$ are independent (will be very important in Part 2). [CompEx1.2b]
6. The conditional distributions of the components are (multivariate) normal. $\boldsymbol{X}_2 \mid (\boldsymbol{X}_1 = \boldsymbol{x}_1) \sim N_{p2}(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$.

# Properties of symmetric idempotent matrices

A symmetric matrix $\boldsymbol{A}$ is idempotent, $\boldsymbol{A}^2 = \boldsymbol{A}$, and has the following properties (RecEx1.P7).

1. The eigenvalues are 0 and 1.
2. The rank of a symmetric matrix (actually: a diagonalizable quadratic matrix) equals the number of nonero eigenvaluse of the matrix. Should be known from previous courses.
3. (Combining 1+2). If a $(n \times n)$ symmetric idempotent matrix $\boldsymbol{A}$ has rank $r$ then $r$ eigenvalues are 1 and $n - r$ are 0.
4. The trace and rank of a symmetric projection matrix are equal: $tr(\boldsymbol{A}) = \text{rank}(\boldsymbol{A})$.
5. The matrix $\boldsymbol{I} - \boldsymbol{A}$ is also idempotent, and $\boldsymbol{A}(\boldsymbol{I} - \boldsymbol{A}) = 0$.

# Quadratic forms [F:B3.3, Theorem B.2]

Random vector $\boldsymbol{X}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, symmetric constant matrix $\boldsymbol{A}$.

- Quadratic form: $\boldsymbol{X}^T \boldsymbol{A} \boldsymbol{X}$.
- The "trace-formula": $\mathrm{E}(\boldsymbol{X}^T \boldsymbol{A} \boldsymbol{X}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) - \boldsymbol{\mu}^T \boldsymbol{A} \boldsymbol{\mu}$.

Then, let $\boldsymbol{X} \sim N_p(\boldsymbol{0}, \boldsymbol{I})$, and $\boldsymbol{R}$ is a symmetric and idempotent matrix with rank $r$.

$$\boldsymbol{X}^T \boldsymbol{R} \boldsymbol{X} \sim \chi^2_r$$

Now, also $\boldsymbol{S}$ is a symmetric and idempotent matrix with rank $s$, and $\boldsymbol{R}\boldsymbol{S} = \boldsymbol{0}$.

$$\frac{s\boldsymbol{X}^T \boldsymbol{R} \boldsymbol{X}}{r\boldsymbol{X}^T \boldsymbol{S} \boldsymbol{X}} \sim F_{r,s}$$

# V2016: Problem 1

**Independent random variables**

Assume that $\boldsymbol{X}$ is a bivariate normal random variable and that $\mathrm{E}(\boldsymbol{X}) = \begin{pmatrix} 5 \\ 3 \end{pmatrix}$ and $\mathrm{Cov}(\boldsymbol{X}) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. Let $\boldsymbol{Y} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \boldsymbol{X}$.

Find the distribution of $\boldsymbol{Y}$.

Specify $a$, $b$ such that $\boldsymbol{Y}$ and $\begin{pmatrix} 2 & a \\ b & 1 \end{pmatrix} \boldsymbol{X}$ are independent random variables. Justify your answer.

# Grading V2016

Results for 53 students:

| Grade | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| Frequency % | 28 | 28 | 23 | 13 | 2 | 6 |

| Item | 1a | 2a | 2b | 2c | 2d | 2e | 3a | 3b |
|---|---|---|---|---|---|---|---|---|
| Average score | 9.25 | 8.47 | 7.42 | 4.37 | 5.97 | 4.68 | 8.82 | 6.18 |

# Theoretical problems - derivations and proofs

- V2016 Problem 3ab: Properties of estimator for $\sigma^2$
- V2015 Problem 3abc: Problem Mallows' Cp.
- V2014 Problem 4ab: weighted regression, omitting parts of regression.
- K2014 Problem 4abc: idempotent $\boldsymbol{H}$ and $\boldsymbol{I} - \boldsymbol{H}$, distribution of SSE and independence of $\hat{\boldsymbol{\beta}}$ and SSE.

# V2016 Problem 3: Properties of estimator for $\sigma^2$

Let $\boldsymbol{Y}$ be an $n \times 1$ random vector with mean $\mu\mathbf{1}$ and covariance matrix $\sigma^2\boldsymbol{I}$, where $\mathbf{1}$ is an $n \times 1$ vector with all elements equal to 1 and $\boldsymbol{I}$ is an $n \times n$ identity matrix. Further, denote by $Y_i$ element $i$ of $\boldsymbol{Y}$, and let $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\mathbf{1}^{\mathsf{T}}\boldsymbol{Y}$.
An estimator for $\sigma^2$ is

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \frac{1}{n-1}\boldsymbol{Y}^{\mathsf{T}}\left(\boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathsf{T}}\right)\boldsymbol{Y}.$$

We give the following useful result. Let $\boldsymbol{X}$ be an $n \times 1$ random vector with mean $\boldsymbol{\eta}$ and covariance matrix $\boldsymbol{\Sigma}$, and let $\boldsymbol{C}$ be an $n \times n$ symmetric constant matrix. Then,

$$\mathrm{E}(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{C}\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{C}\boldsymbol{\Sigma}) + \boldsymbol{\eta}^{\mathsf{T}}\boldsymbol{C}\boldsymbol{\eta}. \tag{1}$$

First, write down the value of $\mathbf{1}^{\mathsf{T}}\mathbf{1}$, and the matrices $\mathbf{1}\mathbf{1}^{\mathsf{T}}$ and $\boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathsf{T}}$ for $n = 4$.
What are key characteristics of the matrix $\boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathsf{T}}$ (symmetric or not, idempotent or not, rank)?
Use Equation (1) to find $\mathrm{E}(S^2)$.

# V2016 Problem 3b

Let us now assume that $\boldsymbol{Y}$ is multivariate normally distributed with the mean and covariance given above.
Show that $\frac{1}{\sigma^2}\boldsymbol{Y}^{\mathsf{T}}(\boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathsf{T}})\boldsymbol{Y}$ follows a $\chi^2$-distribution, and also derive the number of degrees of freedom.
Use this result to find the variance of $S^2$.
Is the random variable $\frac{1}{n}\mathbf{1}^{\mathsf{T}}\boldsymbol{Y}$ and the random vector $(\boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathsf{T}})\boldsymbol{Y}$ independent? Justify your answer.
Finally, find the distribution of

$$\frac{n(\frac{1}{n}\mathbf{1}^{\mathsf{T}}\boldsymbol{Y} - \mu)^2}{\frac{1}{n-1}\boldsymbol{Y}^{\mathsf{T}}(\boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathsf{T}})\boldsymbol{Y}}.$$

Justify your answer.

# Part 2: Linear regression [L7-L12, CompEx2, RecEx3-4]

Curriculum:

- Fahrmeir, Kneib, Lang and Marx (2013): Regression, Springer.
  - Chapter 3.
  - Appendix B
- Slides and handouts: `https://www.math.ntnu.no/emner/TMA4267/2017v/TMA4267V2017Part2.pdf`

# The classical linear model

The model

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$$

is called a classical linear model if the following is true:

1. $\mathrm{E}(\varepsilon) = \mathbf{0}$.
2. $\mathrm{Cov}(\varepsilon) = \mathrm{E}(\varepsilon\varepsilon^T) = \sigma^2\boldsymbol{I}$.
3. The design matrix has full rank, $\mathrm{rank}(\boldsymbol{X}) = k + 1 = p$.

The classical *normal* linear regression model is obtained if additionally

4. $\varepsilon \sim N_n(\mathbf{0}, \sigma^2\boldsymbol{I})$

holds. For random covariates these assumptions are to be understood conditionally on $\boldsymbol{X}$.

# Conditional mean and covariance

If we believe that the vector with elements $Y$ and $\boldsymbol{X}$ are multivariate normal $N_{k+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ we may look at the partition

$$\begin{pmatrix} Y \\ \boldsymbol{X} \end{pmatrix} \sim N_{k+1}\left( \begin{pmatrix} \mu_Y \\ \boldsymbol{\mu_X} \end{pmatrix}, \begin{pmatrix} \Sigma_{YY} & \boldsymbol{\Sigma}_{YX} \\ \boldsymbol{\Sigma}_{XY} & \boldsymbol{\Sigma}_{XX} \end{pmatrix} \right)$$

The conditional distributions of the components are (multivariate) normal, with conditional mean and variance of $Y \mid \boldsymbol{X} = \boldsymbol{x}$ are

$$\begin{aligned} \mathrm{E}(Y \mid \boldsymbol{X} = \boldsymbol{x}) &= \mu_Y + \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_X) \\ \mathrm{Var}(Y \mid \boldsymbol{X} = \boldsymbol{x}) &= \Sigma_Y - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY} \end{aligned}$$

Observe: mean is linear in $\boldsymbol{x}$ and variance independent of $\boldsymbol{x}$.

# Parameter estimation

- Least squares and maximum likelihood estimator for $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

- Restricted maximum likelihood estimator for $\boldsymbol{\sigma}^2$:

$$\hat{\boldsymbol{\sigma}^2} = \frac{1}{n-p}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \frac{\text{SSE}}{n-p}$$

- Projection matrices: idempotent, symmetric/orthogonal:

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$$
$$\boldsymbol{I} - \boldsymbol{H} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$$

with important connection:

$$\hat{\boldsymbol{Y}} = \boldsymbol{H}\boldsymbol{Y}$$
$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{I} - \boldsymbol{H}\boldsymbol{Y}$$

# Alternative summery of Geometry of Least Squares

- Mean response vector: $E(\boldsymbol{Y}) = \boldsymbol{X}\beta$
- As $\beta$ varies, $\boldsymbol{X}\beta$ spans the model plane of all linear combinations. I.e. the space spanned by the columns of $\boldsymbol{X}$: the column-space of $\boldsymbol{X}$.
- Due to random error (and unobserved covariates), $\boldsymbol{Y}$ is not exactly a linear combination of the columns of $\boldsymbol{X}$.
- LS-estimation chooses $\hat{\beta}$ such that $\boldsymbol{X}\hat{\beta}$ is the point in the column-space of $\boldsymbol{X}$ that is closes to $\boldsymbol{Y}$.
- The residual vector $\hat{\varepsilon} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$ is perpendicular to the column-space of $\boldsymbol{X}$.
- Multiplication by $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ projects a vector onto the column-space of $\boldsymbol{X}$.
- Multiplication by $\boldsymbol{I} - \boldsymbol{H} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ projects a vector onto the space perpendicular to the column-space of $\boldsymbol{X}$.

# Properties for the normal linear model

- Least squares and maximum likelihood estimator for $\beta$:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

  with $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$.
- Restricted maximum likelihood estimator for $\boldsymbol{\sigma}^2$:

$$\hat{\boldsymbol{\sigma}^2} = \frac{1}{n-p}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \frac{\text{SSE}}{n-p}$$

  with $\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}$.
- Statistic for inference about $\beta_j$, $c_{jj}$ is diagonal element $j$ of $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

$$T_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{c_{jj}}\hat{\sigma}} \sim t_{n-p}$$

- Residuals (raw): $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$, $\mathrm{E}(\hat{\boldsymbol{\varepsilon}}) = \boldsymbol{0}$ and $\mathrm{Cov}(\hat{\boldsymbol{\varepsilon}}) = \sigma^2(\boldsymbol{I} - \boldsymbol{H})$ where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$.

# Are $\hat{\beta}$ and SSE are independent?

Independence – from Part 1:
Let $\boldsymbol{X}_{(p\times 1)}$ be a random vector from $N_p(\mu, \Sigma)$. Then $\boldsymbol{AX}$ and $\boldsymbol{BX}$ are independent iff $\boldsymbol{A\Sigma B}^T = \boldsymbol{0}$.

We have:

- $\boldsymbol{Y} \sim N_n(\boldsymbol{X}\beta, \sigma^2\boldsymbol{I})$
- $\boldsymbol{AY} = \hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$, and
- $\boldsymbol{BY} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$.
- Now $\boldsymbol{A}\sigma^2\boldsymbol{I}\boldsymbol{B}^T = \sigma^2\boldsymbol{A}\boldsymbol{B}^T = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{0}$
- since $\boldsymbol{X}(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{X} - \boldsymbol{HX} = \boldsymbol{X} - \boldsymbol{X} = \boldsymbol{0}$.
- We conclude that $\hat{\beta}$ is independent of $(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$,
- and, since SSE=function of $(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$: SSE=$\boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$,
- then $\hat{\beta}$ and SSE are independent.

# V2016 Problem 2: Plant stress (slightly modified to reflect changes to the reading list

At the Department of Biology at NTNU researchers use the model plant *Arabidopsis thaliana* to study the response of a plant to different sources of stress. In an experiment *Arabidopsis thaliana* seedlings were subject to a stress situation. The following factors were fitted:

- $D$ (damage): $D = 1$ means that the plant was damaged mechanically by cutting into the leaves of the plant by a pair of scissors. $D = -1$ means damage was not inflicted (no cutting).
- $F$ (flagellin): $F = 1$ means that the pathogen-derived peptide flagellin was sprayed on the leaves of the plant. $F = -1$ means water (not flagellin) was sprayed.
- $T$ (time): Plants were harvested at two different time points after the stress situation. $T = 1$ means that the plant was harvested 60 minutes after the stress situation and $T = -1$ means that the plant was harvested 30 minutes after the stress situation.

Thus, we have three factors, $D$, $F$ and $T$, each at two levels. In the study experiments for all possible combinations of the three factors were performed four times yielding 32 experiments in total.

# V2016 Problem 2: Plant stress

The response measured in the experiment, was the observed gene activity level (a continuous measurement) of each of around 40 000 genes. We will only focus on the gene activity level of one of these genes, the AT1G32920 gene, and we denote the gene activity level of this gene by $Y$. It is known that this gene is active in response to wounding.
For experiment number $i$ (where $i = 1, \ldots, 32$): $Y_i$ is the observed response, $D_i$ is chosen value of $D$, $F_i$ is chosen value of $F$, and $T_i$ is chosen value of $T$. A multiple regression model with all main effects, and two- and three-way interactions, was considered,

$$Y_i = \beta_0 + \beta_D D_i + \beta_F F_i + \beta_T T_i + \beta_{D:F} D_i F_i + \beta_{D:T} D_i T_i + \beta_{F:T} F_i T_i + \beta_{D:F:T} D_i F_i T_i + \varepsilon_i,$$

where $i = 1, \ldots, 32$, and we assume $\varepsilon_i$ independent and identically normally distributed with mean 0 and variance $\sigma^2$. We refer to this as the *full model*.
Note that the interactions are simply products of the factors. The vector of regression parameters is
$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 & \beta_D & \beta_F & \beta_T & \beta_{D:F} & \beta_{D:T} & \beta_{F:T} & \beta_{D:F:T} \end{pmatrix}^{\mathsf{T}}$, and the $i$th row of the design matrix $\boldsymbol{X}$ is $\begin{pmatrix} 1 & D_i & F_i & T_i & D_i F_i & D_i T_i & F_i T_i & D_i F_i T_i \end{pmatrix}$.

# V2016 Problem 2: Plant stress

Here you find R-commands and print-out from fitting the full model.

```
# data is in "standard order" in data frame with name "ds"
> ds %showing only rows 1-6 and 27-32 for space considerations
          Y  D  F  T
1  15.45169 -1 -1 -1
2  15.15908 -1 -1 -1
3  14.93064 -1 -1 -1
4  15.06569 -1 -1 -1
5  14.51032 -1 -1  1
6  14.76922 -1 -1  1
...
27 18.23645  1  1 -1
28 17.70327  1  1 -1
29 16.66523  1  1  1
30 16.96046  1  1  1
31 16.73133  1  1  1
32 16.57248  1  1  1
> fit=lm(Y~D*F*T,data=ds)
> model.matrix(fit)%only showing rows 1-6 and 27-32
   (Intercept)  D  F  T D:F D:T F:T D:F:T
1            1 -1 -1 -1   1   1   1    -1
2            1 -1 -1 -1   1   1   1    -1
3            1 -1 -1 -1   1   1   1    -1
4            1 -1 -1 -1   1   1   1    -1
5            1 -1 -1  1   1  -1  -1     1
6            1 -1 -1  1   1  -1  -1     1
...
27           1  1  1 -1   1  -1  -1    -1
28           1  1  1 -1   1  -1  -1    -1
29           1  1  1  1   1   1   1     1
30           1  1  1  1   1   1   1     1
31           1  1  1  1   1   1   1     1
32           1  1  1  1   1   1   1     1
```

# V2016 Problem 2a: Plant stress

In the print-out from `summary(fit)` *four* numerical values are replaced by question marks. Calculate numerical values for each of these, and explain what each of the values means.

```
> summary(fit)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.15942    0.04140  ?       < 2e-16
D            0.93739    0.04140  22.644  < 2e-16
F            0.28546    0.04140   6.896 3.93e-07
T           -0.52354    0.04140 -12.647 4.18e-12
D:F         -0.08878    0.04140  -2.145  0.04231
D:T         -0.00242    ?        -0.058  0.95386
F:T         -0.12614    0.04140  -3.047  0.00555
D:F:T        0.09099    0.04140   2.198  ?
Residual standard error: 0.2342 on 24 degrees of freedom
Multiple R-squared:      ?, Adjusted R-squared:  0.9594
F-statistic: 105.6 on 7 and 24 DF,  p-value: < 2.2e-16
```

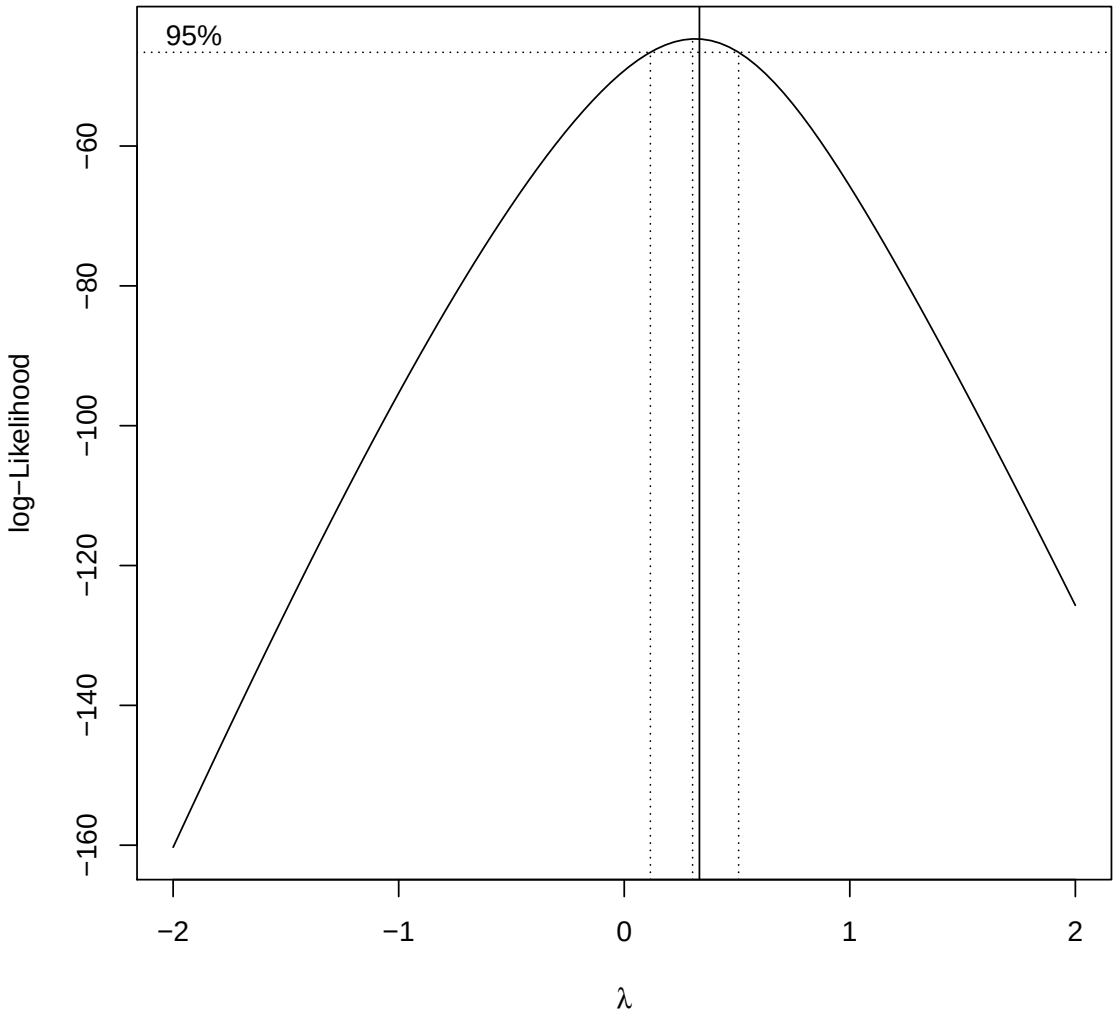
# Examination of model assumptions

1. Linearity of covariates: $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}$
2. Homoscedastic error variance: $Cov(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}$.
3. Uncorrelated errors: $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$.
4. Additivity of errors: $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}$
5. Assumption of normality: $\boldsymbol{\varepsilon} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$

# Plotting residuals

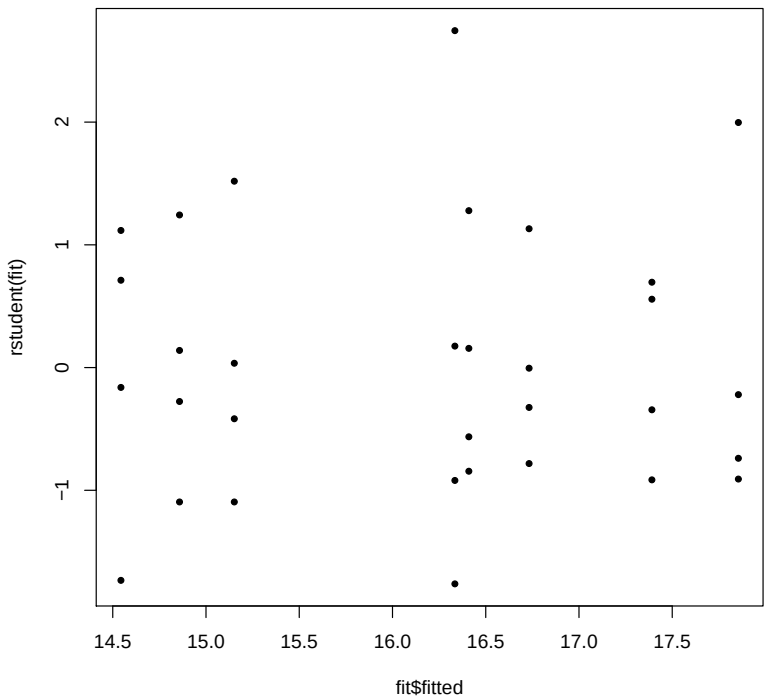
1. Plot the residuals, $r_i^*$ against the predicted values, $\hat{y}_i$.
   - Dependence of the residuals on the predicted value: wrong regression model?
   - Nonconstant variance: transformation or weighted least squares is needed?
2. Plot the residuals, $r_i^*$, against predictor variable or functions of predictor variables. Trend suggest that transformation of the predictors or more terms are needed in the regression.
3. Assessing normality of errors: QQ-plots and histograms of residuals. As an additional aid a test for normality can be used, but must be interpreted with caution since for small sample sizes the test is not very powerful and for large sample sizes even very small deviances from normality will be labelled as significant.
4. Plot the residuals, $r_i^*$, versus time or collection order (if possible). Look for dependence or autocorrelation.

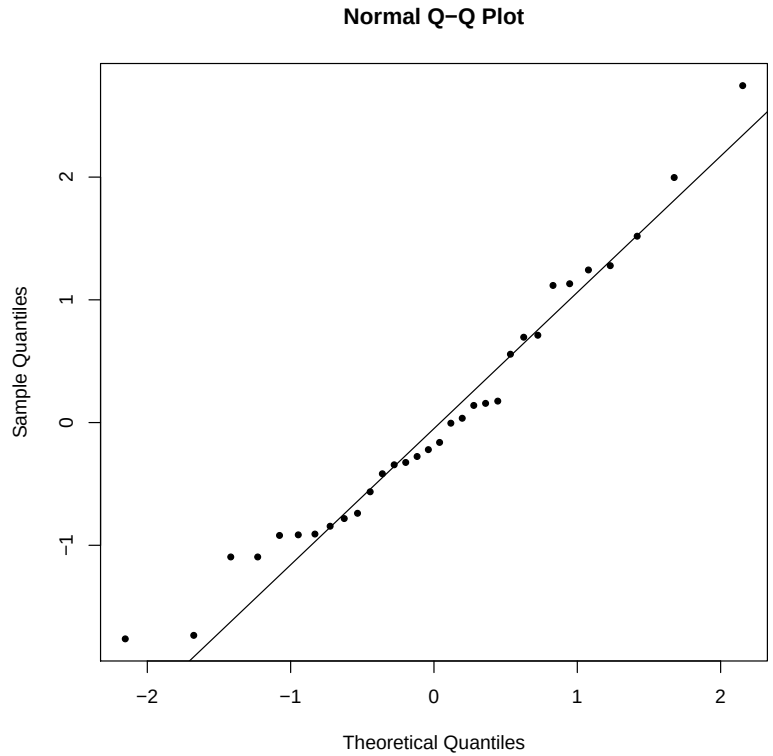
# Box–Cox plot




Box–Cox transformation plot based on Model A for the Galapagos data set, RecEx4. Line at $x = 1/3$.

# V2016 Plant stress 2b: Residual plots

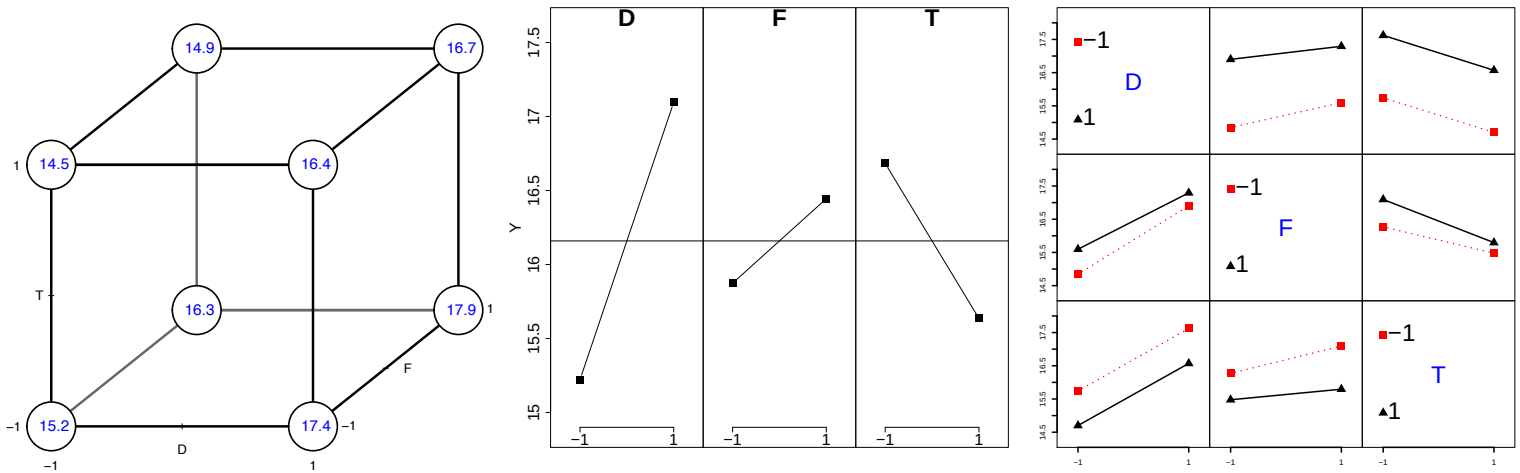


# V2016 Plant stress 2b: R-code

```
> library(FrF2)
> MEPlot(fit)
> IAPlot(fit)
> cubePlot(fit,"D","F","T",round=1,size=0.33,main="")
> plot(fit$fitted,rstudent(fit),pch=20)
> qqnorm(rstudent(fit),pch=20)
> qqline(rstudent(fit))
> ad.test(rstudent(fit))
Anderson-Darling normality test
data:  rstudent(fit)
A = 0.43191, p-value = 0.2869
```

# V2016 Plant stress 2b: Cube, main effects and interaction effects



# V2016: Plant stress 2b

How would you, based on the figures given and the R-output, evaluate the fit of the model?

How would you explain to a biologist what the estimated main effect of damage means in practice? How would you explain the estimated interaction effect between damage and flagellin?

Let $\gamma = 2^{\beta_F - \beta_D}$ be a new parameter of interest.
Suggest an estimator, $\hat{\gamma}$, for $\gamma$. Use approximate methods to find the expected value and variance of this estimator, that is, $\mathrm{E}(\hat{\gamma})$ and $\mathrm{Var}(\hat{\gamma})$. Use results in the prinout to calculate numerical value for $\hat{\gamma}$, and estimated numerical values for $\mathrm{E}(\hat{\gamma})$ and $\mathrm{Var}(\hat{\gamma})$.
Hint: You may use that $2^x = \exp(x \ln 2)$, where ln is the natural logarithm.

# First order Taylor expansion: Univariate function

$X$ is RV with $\mathrm{E}(X) = \mu$, and we look at function $g(X)$.
First order Taylor approximation of $g(X)$ around $\mu$.

$$g(X) \approx g(\mu) + g'(\mu)(X - \mu)$$

This leads to the following approximations:

$$\mathrm{E}(g(X)) \approx g(\mu)$$
$$\mathrm{Var}(g(X)) \approx [g'(\mu)]^2 \mathrm{Var}(X)$$

# First order Taylor expansion: Bivariate function

$X_1$ is a RV with $\mu_1 = \mathrm{E}(X_1)$ and $X_2$ is a RV with $\mu_2 = \mathrm{E}(X_2)$.
Let $g$ be a bivariate function of $X_1$ and $X_2$, and define

$$g_1'(\mu_1, \mu_2) = \frac{\partial g(x_1, x_2)}{\partial x_1} \Big|_{x_1=\mu_1, x_2=\mu_2}$$

$$g_2'(\mu_1, \mu_2) = \frac{\partial g(x_1, x_2)}{\partial x_2} \Big|_{x_1=\mu_1, x_2=\mu_2}$$

First order Taylor approximation:

$$g(X_1, X_2) \approx g(\mu_1, \mu_2) + g_1'(\mu_1, \mu_2)(X_1 - \mu_1) + g_2'(\mu_1, \mu_2)(X_2 - \mu_2)$$

$$\mathrm{E}(g(X_1, X_2)) \approx g(\mu_1, \mu_2)$$

$$\mathrm{Var}(g(X_1, X_2)) \approx [g_1'(\mu_1, \mu_2)]^2 \mathrm{Var}(X_1) + [g_2'(\mu_1, \mu_2)]^2 \mathrm{Var}(X_2) + 2 \cdot g_1'(\mu_1, \mu_2) \cdot g_2'(\mu_1, \mu_2) \mathrm{Cov}(X_1, X_2)$$

# Topic: choosing the "best" linear regression model!

- First, debunk popular strategies (based on simulations studies were we knew the "true" model):
  - Popular 1: fit all available covariates.
    Problem: overfitting (=fitting trends and noise).
  - Popular 2: fit all available covariates, then remove the insignificant ones (=those $\beta_j$ where $H_0 : \beta_j = 0$ is rejected).
    Problem: may also remove important covariates that are correlated with unimportant ones - but insignificant because being masked by the unimportant ones.
- Study of irrelevant and missing covariates:

  Irrelevant : variables that are included in the regression but should not have been (IQ of lumberjack)

  missing : variables that are not included, but should have been (omitting height in the tree volum example)

  Conclusion in book: the model should not contain irrelevant covariates, and we should aim for a sparse model. Take home message is the "Law of parsimony": *If two models are not very different – then always choose the simplest one.*

# Model selection

Want to choose the model that minimize the

$$\text{SPSE} = \sum_{j=1}^{J} \mathrm{E}((Y_j - \hat{Y}_{jM})^2)$$

Several solution based on test set or cross-validation. We have focused on using only the original data and a penalized criterion: a first term based on SSE (or $R^2$) for model $M$, and a second term penalizing the model complexity.

$R^2$ adjusted (corrected)

Mallows' $C_p$

Akaike Information Criterion (AIC)

Bayesian Information Criterion (BIC)

NB: there is no overall best choice for criterion - all of these are used.

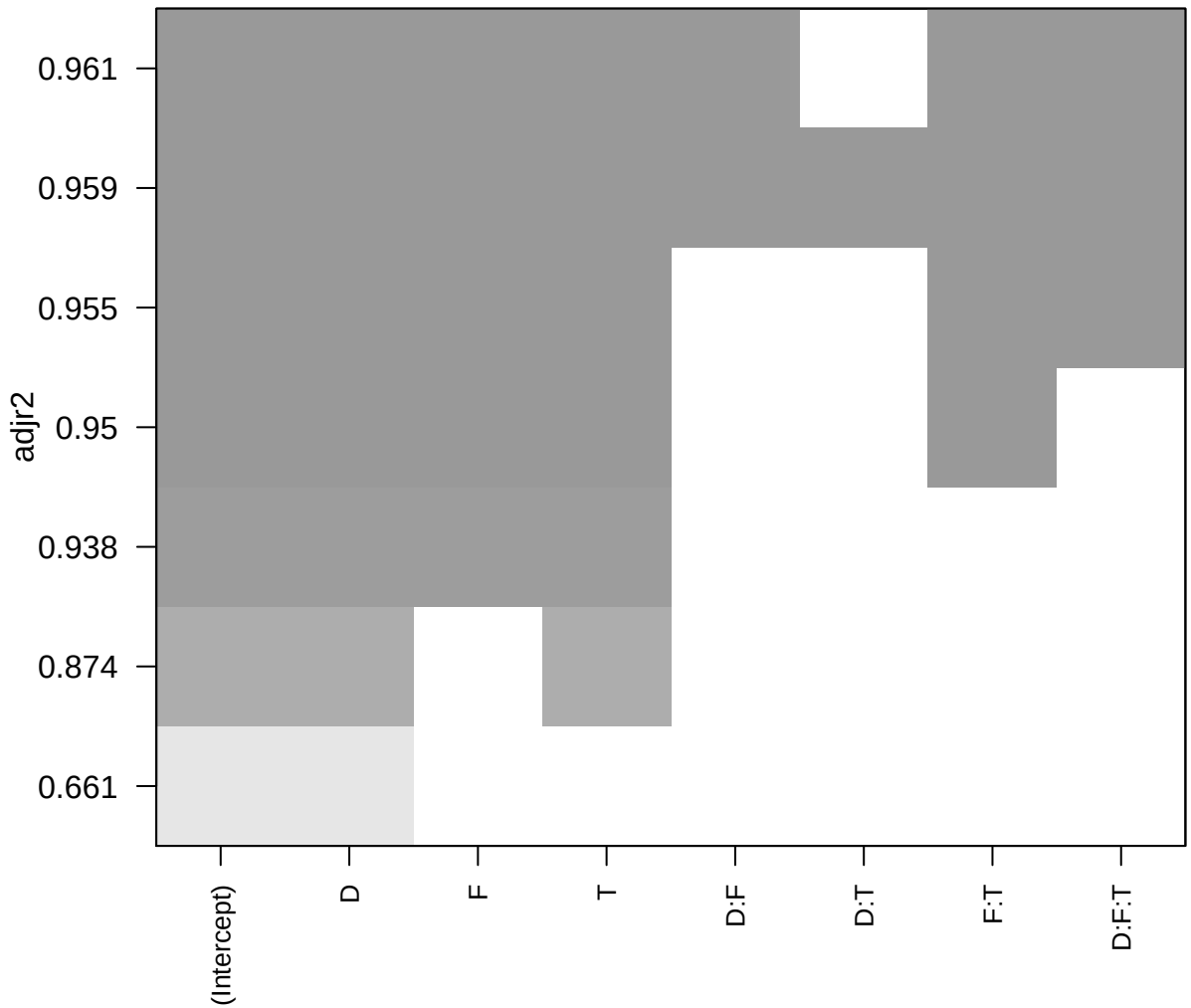
# V2016 Plant stress 2d: model selection (missing part with lasso not on reading list)

The researchers want to use the data to fit a prediction model, and want to consider reduced versions of the full model, using best subset model selection.

Explain briefly what is done in the best subset model selection, and choose a good model based on the $R^2_{\text{adj}}$-criterion. Write down the fitted regression model for the model you choose.

```
> x <- model.matrix(fit)[,-1]; dim(x)
[1] 32  7
> y <- ds$Y
> library(leaps)
> bests <- regsubsets(x,y)
> sumbests=summary(bests)
> sumbests
1 subsets of each size up to 7
Selection Algorithm: exhaustive
         D   F   T   D:F D:T F:T D:F:T
1  ( 1 ) "*" " " " " " " " " " " " "
2  ( 1 ) "*" " " "*" " " " " " " " "
3  ( 1 ) "*" "*" "*" " " " " " " " "
4  ( 1 ) "*" "*" "*" " " " " "*" " "
5  ( 1 ) "*" "*" "*" " " " " "*" "*"
6  ( 1 ) "*" "*" "*" "*" " " "*" "*"
7  ( 1 ) "*" "*" "*" "*" "*" "*" "*"
> plot(bests,scale="adjr2",col=gray(seq(0.6,0.9,length=20)))
> round(sumbests$adjr2,3)
[1] 0.661 0.874 0.938 0.950 0.955 0.961 0.959
```

# V2016: Plant stress 2d



# Part 3: Hypothesis testing and analysis of variance [L13-L16, CompEx3, RecEx5]

Curriculum:

- Fahrmeir, Kneib, Lang and Marx (2013): Regression, Springer.
  - Chapter 3.3
  - Appendix B
- Härdle, Simar (2015): Applied Multivariate Statistical Analysis, Fourth edition, Springer.
  - Chapter 8.1.1.
- Note: Multiple testing by Halle, Bakke and Langaas.
- Slides and handouts: `https://www.math.ntnu.no/emner/TMA4267/2017v/TMA4267V2017Part3.pdf`

# MORE TO COME HERE

# V2016 Problem 2: Plant stress

At the Department of Biology at NTNU researchers use the model plant *Arabidopsis thaliana* to study the response of a plant to different sources of stress. In an experiment *Arabidopsis thaliana* seedlings were subject to a stress situation. The following factors were fitted:

- $D$ (damage): $D = 1$ means that the plant was damaged mechanically by cutting into the leaves of the plant by a pair of scissors. $D = -1$ means damage was not inflicted (no cutting).
- $F$ (flagellin): $F = 1$ means that the pathogen-derived peptide flagellin was sprayed on the leaves of the plant. $F = -1$ means water (not flagellin) was sprayed.
- $T$ (time): Plants were harvested at two different time points after the stress situation. $T = 1$ means that the plant was harvested 60 minutes after the stress situation and $T = -1$ means that the plant was harvested 30 minutes after the stress situation.

Thus, we have three factors, $D$, $F$ and $T$, each at two levels. In the study experiments for all possible combinations of the three factors were performed four times yielding 32 experiments in total.

# V2016 Problem 2: Plant stress

The response measured in the experiment, was the observed gene activity level (a continuous measurement) of each of around 40 000 genes. We will only focus on the gene activity level of one of these genes, the AT1G32920 gene, and we denote the gene activity level of this gene by $Y$. It is known that this gene is active in response to wounding.
For experiment number $i$ (where $i = 1, \ldots, 32$): $Y_i$ is the observed response, $D_i$ is chosen value of $D$, $F_i$ is chosen value of $F$, and $T_i$ is chosen value of $T$. A multiple regression model with all main effects, and two- and three-way interactions, was considered,

$$Y_i = \beta_0 + \beta_D D_i + \beta_F F_i + \beta_T T_i + \beta_{D:F} D_i F_i + \beta_{D:T} D_i T_i + \beta_{F:T} F_i T_i + \beta_{D:F:T} D_i F_i T_i + \varepsilon_i,$$

where $i = 1, \ldots, 32$, and we assume $\varepsilon_i$ independent and identically normally distributed with mean 0 and variance $\sigma^2$. We refer to this as the *full model*.
Note that the interactions are simply products of the factors. The vector of regression parameters is
$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 & \beta_D & \beta_F & \beta_T & \beta_{D:F} & \beta_{D:T} & \beta_{F:T} & \beta_{D:F:T} \end{pmatrix}^{\mathsf{T}}$, and the $i$th row of the design matrix $\boldsymbol{X}$ is $\begin{pmatrix} 1 & D_i & F_i & T_i & D_i F_i & D_i T_i & F_i T_i & D_i F_i T_i \end{pmatrix}$.

# V2016 Problem 2a: Plant stress

In the print-out from `summary(fit)` *four* numerical values are replaced by question marks. Calculate numerical values for each of these, and explain what each of the values means.

```
> summary(fit)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.15942    0.04140  ?       < 2e-16
D            0.93739    0.04140  22.644  < 2e-16
F            0.28546    0.04140   6.896 3.93e-07
T           -0.52354    0.04140 -12.647 4.18e-12
D:F         -0.08878    0.04140  -2.145  0.04231
D:T         -0.00242    ?        -0.058  0.95386
F:T         -0.12614    0.04140  -3.047  0.00555
D:F:T        0.09099    0.04140   2.198  ?
Residual standard error: 0.2342 on 24 degrees of freedom
Multiple R-squared:      ?, Adjusted R-squared:  0.9594
F-statistic: 105.6 on 7 and 24 DF,  p-value: < 2.2e-16
```

# V2016: Plant stress 2c - with the full model

The researchers want to test the hypothesis

$H_0: \beta_{D:T} = \beta_{F:T} = \beta_{D:F:T} = 0$ vs.

$H_1:$ at least one of $\beta_{D:T}$, $\beta_{F:T}$, $\beta_{D:F:T}$ is different from 0.

Perform the hypothesis test at a significance level of your own choice. All the numerical values you need for the calculations are found in the R-printout.

# Part 4: Design of experiments [L17-L20, CompEx4, RecEx6]

Curriculum:

- Note on Design of experiments by Tyssedal (not p 19: partial confounding and p28 fold-over).
- Slides and handouts: `https://www.math.ntnu.no/emner/TMA4267/2017v/TMA4267V2017Part4.pdf`

# MORE TO COME HERE

# Topic: choosing the "best" linear regression model!

- First, debunk popular strategies (based on simulations studies were we knew the "true" model):
  - Popular 1: fit all available covariates.
    Problem: overfitting (=fitting trends and noise).
  - Popular 2: fit all available covariates, then remove the insignificant ones (=those $\beta_j$ where $H_0 : \beta_j = 0$ is rejected).
    Problem: may also remove important covariates that are correlated with unimportant ones - but insignificant because being masked by the unimportant ones.
- Study of irrelevant and missing covariates:

  Irrelevant : variables that are included in the regression but should not have been (IQ of lumberjack)

  missing : variables that are not included, but should have been (omitting height in the tree volum example)

  Conclusion in book: the model should not contain irrelevant covariates, and we should aim for a sparse model. Take home message is the "Law of parsimony": *If two models are not very different – then always choose the simplest one.*

# V2016: Problem 1e - full vs. reduced model

The researchers choose to use the following *reduced model* for prediction:

$$Y_i = \beta_0 + \beta_D D_i + \beta_F F_i + \beta_T T_i + \beta_{D:F} D_i F_i + \varepsilon_i,$$

where $i = 1, \ldots, 32$, and we assume $\varepsilon_i$ independent and identically normally distributed with mean 0 and variance $\sigma^2$.

Compare the estimated regression parameters and the estimated standard deviations of the estimated regression parameters for the full model and the reduced model, and explain what you observe.
Based on the reduced model, provide a prediction and a 95% prediction interval for the gene activity level for the factor combination $D = 1$, $F = 1$, $T = -1$.

Hint: In a multiple linear regression with $n \times p$ design matrix $\boldsymbol{X}$, estimated regression coefficients $\hat{\boldsymbol{\beta}}$ and unbiased estimated error variance $s^2$, a $(1-\alpha)100\%$ prediction interval at $\boldsymbol{x}_0$ is given as

$$\boldsymbol{x}_0^T \hat{\boldsymbol{\beta}} \pm t_{\frac{\alpha}{2}, n-p}\, s \sqrt{1 + \boldsymbol{x}_0^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_0},$$

where $t_{\alpha/2, n-p}$ denotes the value in the $t$-distribution with $n-p$ degrees of freedom that has area $\frac{\alpha}{2}$ to the right.

# V2016: 2d - full vs. reduced model

```
> fitRED=lm(Y~D+F+T+D:F,data=ds)
> summary(fitRED)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.15942    0.04919 328.528  < 2e-16
D            0.93739    0.04919  19.057  < 2e-16
F            0.28546    0.04919   5.804 3.56e-06
T           -0.52354    0.04919 -10.644 3.66e-11
D:F         -0.08878    0.04919  -1.805   0.0822
Residual standard error: 0.2782 on 27 degrees of freedom
Multiple R-squared:   0.95,Adjusted R-squared:  0.9426
F-statistic: 128.4 on 4 and 27 DF,  p-value: < 2.2e-16
> qt(0.025,32,lower.tail=FALSE)
[1] 2.036933
> qt(0.025,27,lower.tail=FALSE)
[1] 2.051831
> qt(0.025,24,lower.tail=FALSE)
[1] 2.063899
```

# Grading V2016

Results for 53 students:

| Grade | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| Frequency % | 28 | 28 | 23 | 13 | 2 | 6 |

| Item | 1a | 2a | 2b | 2c | 2d | 2e | 3a | 3b |
|---|---|---|---|---|---|---|---|---|
| Average score | 9.25 | 8.47 | 7.42 | 4.37 | 5.97 | 4.68 | 8.82 | 6.18 |

# Exam

- 9.00-13.00, May 19, 2017.
- Written.
- Makes up 80% of the final grade, the remaining 20 % from the four compulsory exercises.
- Permitted aids: (Code C). One yellow A5 with own handwritten notes, Rottmann: Matematisk formelsamling, Tabeller og formler i statistikk, specified calculator.

# Why one yellow A5 sheet?

- Force you to structure the course key concepts?
- Memorizing not needed?
- Security blanket.

# Final reading list

- Fairmeir et al (2013):
  Chapter 3 and Appendix B.
- Härdle et al (2015):
  Chapters 2, 3.3, 4.1-4.5, 5.1, 8.1.1 and 11.1-11.3.
- Multiple testing note by Halle, Bakke and Langaas.
- DOE-note by Tyssedal.
- BoxCox: from L12 in lecture notes/handouts (and on several exams).
- The 4 compulsory and 6 recommended exercises.

# Comparison with reading list earlier years

**Not on the reading list V2017, but on before:**
Analysis of contingency tables.
The most complex parts of Design of experiments (folding, combining blocking and fractionating).
Random effects ANOVA.
More effort made earlier with quadratic forms for ANOVA (especially with idempotent centering matrices and sums-of-squares).
Hotelling $T^2$.
Penalized regression: lasso and ridge (will be part of TMA4268 Statistical learning).

**New(ish) on the reading list:**
Replication crises, properties of $p$-values.
Multiple testing with FWER and FDR.
Testing of linear hypotheses (F:3.3) (new in 2016)
Effect coding in linear regression to see analysis of variance just as a special case of regression, and using the linear hypothesis F-test instead of much work with sums-of-squares (new in 2016)

# Activities before the exam

- The exam is Friday, May 19, 9-13.
- Exam problems from earlier year is available from the course www-page (also outside Bb).
- Supervision - and you may sit and work - Sentralbygg 2 - room 822 (booked 10-14)
  - Before May 15 - just stop by the office of Jacob or Mette (better to stop by than sending email - difficult to give good answer on email).
  - Monday May 15: 10-12
  - Tuesday May 16: 10-12
  - Thursday May 18: 10-12
- After the exam: tentative solutions posted
- and hopefully (if allowed) automatic feedback given together with exam grade.

# Statistics courses

- Autumn semester
  - TMA4295 Statistical Inference
  - TMA4285 Time series
  - TMA4315 Generalized linear models
- Spring semester
  - TMA4250 Spatial statistics
  - TMA4268 Statistical learning
  - TMA4275 Survival analysis
  - TMA4300 Computational statistics

# Future studies?

What is your current plan of topic for future studies?

- A: Statistics
- B: Mathematics
- C: Numerics
- D: Other
- E: Don't know

Use your smart phone, or other devise with internet access and go to **http://clicker.math.ntnu.no/**, and then select TMA4267 as classroom.