NTNU
Norwegian University of
Science and Technology

TMA4267 Linear Statistical Models V2014 (18)

Model transformations and Taylor expansion [7.2-7.4]
Design of experiments (note): full 2^k experiment

Mette Langaas

To be lectured: March 4, 2014
wiki.math.ntnu.no/emner/tma4267/2014v/start/

Transformations [7.2-7.3]

- Multiplicative or additive model?
- BoxCox transform with profile likelihood.
- Stabilizing the variance.

Transformations [7.2 - 7.3] ← the response

a) Yesterday : multiplicative vs additive model

$Y \sim N(\mu, \sigma^2)$: $Y = \mu + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$ ← additive

$Y = \mu \cdot \varepsilon$ ← multiplicative

$\log(Y) = \log \mu + \log \varepsilon$

b) the BoxCox transform (for strictly positive responses) [7.2]

$y \rightarrow g_\lambda(y)$ where

$$g_\lambda(y) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln y & \lambda = 0 \end{cases}$$

see book for $\lambda \to 0$ connection

The "best" value of λ may be chosen based on maximum likelihood theory using the profile log-likelihood (errors are normal)

$$L(\lambda) = -\frac{n}{2} \log\left(\frac{SSE_\lambda}{n} \right) + (\lambda - 1) \sum \log y_i$$

where SSE_λ is the SSE when $g_\lambda(y)$ is the response.

Prediction: may use any $\hat{\lambda}$

Interpretation: round $\hat{\lambda}$ to the nearest interpretable value

$(0.46 \to 0.5 \Rightarrow \sqrt{y})$

We don't want to transform if we don't need to, because interpretation will be difficult.

Pay attention to the CI of λ when choosing $\hat{\lambda}$ for use.

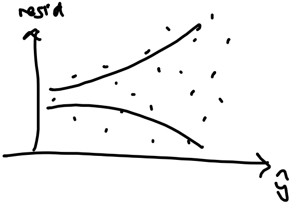
R: library(MASS), boxcox ← NB: boxcox is sensitive to outliers

λ = seq(-2, 2)

Popular: $\lambda = 0, 0.5, -1$

c) Variance stabilizing transformations [7.3]

→ Choose a transformation of y that makes $Var(y)$ constant.

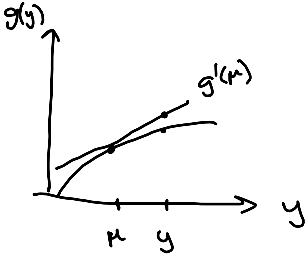


i) We observe that $Var(y) \propto E(y)$

(e.g. $Y \sim Poisson(\mu)$, $E(Y)=\mu$, $Var(Y)=\mu$).

Which $g(y)$ should we choose?

Trick: 1th order Taylor expansion of Y around μ, $E(Y)=\mu$.



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

$g'(\mu) = \left.\frac{dg}{dy}\right|_{y=\mu}$

$$E(g(Y)) \approx g(\mu) + g'(\mu) \cdot (\underbrace{E(Y)-\mu}_{0})$$

$$\approx g(\mu)$$

$$Var(g(Y)) \approx 0 + [g'(\mu)]^2 \underbrace{Var(Y-\mu)}_{Var(Y)}$$

$$\approx [g'(\mu)]^2 Var(Y)$$

How to use this? We have (in general) $Var(Y) = H(\mu)$

and we want to find $g(Y)$ so that $Var(g(Y)) = \sigma^2$ ← a constant.

$$Var(g(Y)) = \sigma^2 \approx [g'(\mu)]^2 \cdot \underbrace{Var(Y)}_{H(\mu)}$$

↗ solve for g

Let $H(\mu) = \mu$

$$\sigma^2 = [g'(\mu)]^2 \cdot \mu$$

$$g'(\mu) = \sqrt{\frac{\sigma^2}{\mu}}$$

$$g(y) \propto \int \frac{1}{\sqrt{y}} \, dy = \sqrt{y}$$

(ii) $Var(Y) \propto E(Y)^2 \Rightarrow g(y) = \ln y$

(iii) $Var(Y) \propto E(Y)^4 \Rightarrow g(y) = \frac{1}{y}$

Approximation of E and Var for nonlinear functions

- Have RV X, with mean $\mathrm{E}(X) = \mu$ and some variance $\mathrm{Var}(X)$.
- Want to look at a nonlinear function of X, called $g(X)$.
- Aim: find an approximation to $\mathrm{E}(g(X))$ and $\mathrm{Var}(g(X))$.
- And, the same for two RVs X_1 and X_2 with $g(X_1, X_2)$.
- Solution: first order Taylor approximation.

Example 1: Exam TMA4255 V2011 1d (ln of BMI)

Looking at residual plots from a one-way ANOVA the conclusion was to analyse data of *BMI* vs genotype (three groups) on the natural logaritmic scale.
In the genotype group 2 the average $ln(BMI)$ was 3.2151 and the empirical standard deviation was 0.1656.
Use approximate methods to arrive at an estimate of the mean and standard deviation for the *BMI* (that is, on the orginal scale, kg/m^2, and not on the logarithmic scale).

$\mathrm{E}(g(X))$ and $\mathrm{Var}(g(X))$: from earlier courses

— Let $g(X)$ be a general function. When is $\mathrm{E}(g(X)) = g(\mathrm{E}(X))$?
  - When $g(X)$ is a linear function of X.

— What can we do if this is not the case?
  - If g is monotone we can use the transformations formula to find the distribution of $Y = g(X)$ and then calculate $\mathrm{E}(Y)$ and $\mathrm{Var}(Y)$, if possible.

— What if we only know $\mathrm{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$ and not $f(x)$?
  - Use a Taylor series approximation of $g(X)$ around $g(\mu)$. g need to be differentiable.

Univariate function

First order Taylor approximation of $g(X)$ around μ.

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

This leads to the following approximations:

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

Example 2: Exam TMA4255 V2012 3d (fraction)

Let μ_A be the expected pain-free grip force for a population where the physiotherapy intervention treatment is used to treat tennis elbow, and μ_C be the expected pain-free grip force for a population where the wait-and-see treatment is used. Define the relative difference between these two expected values as

$$\gamma = \frac{\mu_A - \mu_C}{\mu_C}.$$

This can be interpreted as the expected relative gain by using physiotherapy instead of wait-and-see. Based on two independent random samples of size n_A and n_C from the physiotherapy and wait-and-see treatment groups, respectively, suggest an estimator, $\hat{\gamma}$, for γ.

Use approximate methods to find the expected value and variance of this estimator, that is, $\mathrm{E}(\hat{\gamma})$ and $\mathrm{Var}(\hat{\gamma})$.

Bivariate function: first order Taylor

X_1 is a RV with $\mu = \mathrm{E}(X_2)$ 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

$$g_1'(\mu_1, \mu_2) = \frac{\partial g(x_1, x_2)}{\partial x_1} \mid_{x_1=\mu_1, x_2=\mu_2}$$
$$g_2'(\mu_1, \mu_2) = \frac{\partial g(x_1, x_2)}{\partial x_2} \mid_{x_1=\mu-1, x_2=\mu_2}$$

First order Taylor approximation:

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

Bivariate function: first order Taylor

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

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

Multivariate version

From *Tabeller og formler i statistikk*.

Rekkeutvikling

En første ordens Taylorutvikling av funksjonen $g(X_1, \ldots, X_n)$ omkring $g(\mu_1, \ldots, \mu_n)$, der $\mathrm{E}(X_i) = \mu_i$, $i = 1, \ldots, n$, gir approksimasjonene

$$\mathrm{E}[g(X_1, \ldots, X_n)] \approx g(\mu_1, \ldots, \mu_n),$$

$$\mathrm{Var}[g(X_1, \ldots, X_n)] \approx \sum_{i=1}^{n} \left(\frac{\partial g(\mu_1, \ldots, \mu_n)}{\partial \mu_i} \right)^2 \mathrm{Var}(X_i) + 2 \sum_{i>j} \frac{\partial g}{\partial \mu_i} \frac{\partial g}{\partial \mu_j} \mathrm{Cov}(X_i, X_j).$$

Orthogonality

Mathematically: we look at $\boldsymbol{X}^T\boldsymbol{X}$, and remember that for the LS-regression $\mathrm{Cov}(\hat{\beta}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

- If two regressors have values independent of eachother they have zero correlation,
- and are said to be orthogonal. $\boldsymbol{x}_j^T\boldsymbol{x}_l = 0$.
- Then $\boldsymbol{X}^T\boldsymbol{X}$ will be a diagonal matrix and the regression coefficients are independent of eachother.

$$(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{b} = \boldsymbol{X}^T\boldsymbol{y}$$

$$\mathbf{A} = \mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} & \cdots & \sum_{i=1}^n x_{ki} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{1i}x_{2i} & \cdots & \sum_{i=1}^n x_{1i}x_{ki} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^n x_{ki} & \sum_{i=1}^n x_{ki}x_{1i} & \sum_{i=1}^n x_{ki}x_{2i} & \cdots & \sum_{i=1}^n x_{ki}^2 \end{bmatrix} \quad \text{and} \quad \mathbf{g} = \mathbf{X}'\mathbf{y} = \begin{bmatrix} g_0 = \sum_{i=1}^n y_i \\ g_1 = \sum_{i=1}^n x_{1i}y_i \\ \vdots \\ g_k = \sum_{i=1}^n x_{ki}y_i \end{bmatrix}$$

Orthogonality

— The normal equations are then not coupled!

$$\hat{\beta}_1 = \sum_{i=1}^{n} y_i/n$$
$$\hat{\beta}_k = (\sum_{i=1}^{n} x_{ki} y_i)/(\sum_{i=1}^{n} x_{ki}^2)$$

when we for simplicity assume that all covariates are centered (mean is zero).

— The estimate of β_2 for x_2 will not change if x_3 is also included into the model. Interpretation is easy! Fitting is easy! Testing is easy!

Multicollinearity

$\hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$, and $\text{Cov}(\hat{\beta}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

- If one covariate is correlated with another covariate then we have collinearity. (Not linearity - but a tendency of linear dependence.)
- With several correlated covariates we call this multicollinearity.
- This will make it difficult to know which variable to include in the model (several variables give much of the same information)
- and the covariance of $\hat{\beta}$ may be large since $\boldsymbol{X}^T\boldsymbol{X}$ may be nearly singular.
- And, the estimate of β_2 in a model with only x_2 will change if x_3 is also included into the model.
- This will also make prediction difficult since the prediction error will explode.
- But, this is real life (unless you do DOE using an orthogonal design).

Part 6: Design of Experiments (DOE) with 2^k factorial design

MLR: $Y = X\beta + \varepsilon \qquad \varepsilon \sim N_n(0, \sigma^2 I)$ and

$\hat{\beta} = (X^T X)^{-1} X^T Y \sim N_p(\beta, \sigma^2 (X^T X)^{-1})$

What would we want to make optimal if we could choose X?

- minimize $Var(\hat{\beta}) = tr(\sigma^2 (X^T X)^{-1})$
- minimize $det(Cov(\hat{\beta}))$
- choose X with orthogonal colums to have maximum interpretability
-
⋮

↖ all of this can be done within the research topic DOE

Our focus: 2^k factorial designs

Here the entries of the design matrix X ($n \times \underbrace{(k+1)}_{p}$) are chosen to have two possible values $\{-1, 1\}$ and each column is chosen orthogonal to the other columns.

Outline DOE

- The full 2^k experiment.
  - Coding, standard order.
  - Main and interaction effects.
  - Simple formulas for effects and SSR (due to orthogonality).
  - Lenths method, and other strategies for estimating σ^2.
  - External effect present when performing repetitions?
- Blocking in full 2^k experiments.
- Fractions of 2^k experiments.

What does it mean to choose $x_{ij} \in \{-1, 1\}$?

If I start with data e.g. on chemical yield (y) and want to study the effect of temperature (160°C, 180°C) and of chemical consentration (20%, 40%), I may recode temp & conc. to ease the mathematical presentation. Here I let

temp 160 ⇒ −1   cons 20 ⇒ −1
180 ⇒ 1   40 ⇒ 1

This will give me a regression model in the recoded variables. I will always be able to transform the regression model back to orignal units.

Important remark

- We will here denote the intercept by β_0.
- We will look at k dichotomous covariates, so we estimate $p = k + 1$ regression parameters.

The pilot plant example - Version 1

At a pilot plant a chemical process is investigated.

— The outcome of the process is measured as chemical yield (in grams).

— Two quantitative variables (factors) were investigated:

- Factor A: Temperature (in degrees C).
- Factor B: Concentration (in percentage).

Experiment no.	Temperature	Concentration	Yield
1	160	20	60
2	180	20	72
3	160	40	54
4	180	40	68
	x_1	x_2	y

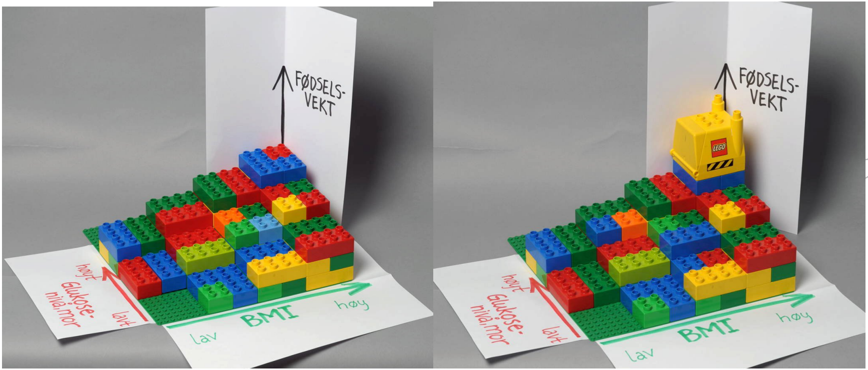


Photo from Kathrine Frey Frøslie, http://www.facebook.com/photo.php?fbid=1775971247383

Mette.Langaas@math.ntnu.no, TMA4267V2014

MLR with pilot plant data V1

```
   x1 x2  y
1 160 20 60
2 180 20 72
3 160 40 54
4 180 40 68

>lm(formula = y ~ x1 + x2 + x1*x2, data = ds)

            Estimate
(Intercept)  -14.000
x1             0.500
x2            -1.100
x1:x2          0.005
```

MLR with pilot plant V1: coded variables

```
   x1 x2  y z1 z2
1 160 20 60 -1 -1
2 180 20 72  1 -1
3 160 40 54 -1  1
4 180 40 68  1  1

lm(formula = y ~ z1 + z2 + z1 * z2, data = ds)
            Estimate
(Intercept)     63.5
z1               6.5
z2              -2.5
z1:z2            0.5
```

MLR with original and coded factors

Original variables, x_1 and x_2, gave estimated regression equation

$$\hat{y} = -14 + 0.5x_1 - 1.1x_2 + 0.005x_1 \cdot x_2$$

Coded variables, $z_1 = (x_1 - 170)/10$ and $z_2 = (x_2 - 30)/10$, gave estimated regression equation

$$\hat{y} = 63.5 + 6.5z_1 - 2.5z_2 + 0.5z_1 \cdot z_2$$

Can you compare these two results?

MLR with original and coded factors

Substitute $z_1 = (x_1 - 170)/10$ and $z_2 = (x_2 - 30)/10$ into the equation to get a estimated regression equation based on x_1 and x_2.

$$
\begin{aligned}
\hat{y} &= 63.5 + 6.5z_1 - 2.5z_2 + 0.5z_1 \cdot z_2 \\
&= 63.5 + 6.5\frac{x_1 - 170}{10} - 2.5\frac{x_2 - 30}{10} + 0.5\frac{x_1 - 170}{10} \cdot \frac{x_2 - 30}{10} \\
&= 63.5 - 6.5\frac{170}{10} + 2.5\frac{30}{10} + 0.5\frac{170 \cdot 30}{10 \cdot 10} \\
&+ x_1(6.5\frac{1}{10} - 0.5\frac{1}{10}\frac{30}{10}) + x_2(-2.5\frac{1}{10} - 0.5\frac{1}{10}\frac{170}{10}) \\
&+ 0.5\frac{1}{10}\frac{1}{10}x_1 \cdot x_2 \\
&= -14 + 0.5x_1 - 1.1x_2 + 0.005x_1 \cdot x_2
\end{aligned}
$$

Design of experiments (DOE) terminology

— Variables are called factors, and denoted A, B, C, ...
— We will only look at factors with two levels
- high, coded as $+1$ or just $+$, and
- low, coded as -1 or just $-$.

— In the pilot plant example we had two factors with two levels, thus $2 \cdot 2 = 4$ possible combinations. In general k factors with two levels gives 2^k possible combinations.

Standard notation for 2^2 experiment:

Experiment no.	A	B	AB	Level code	Response
1	-1	-1	1	1	y_1
2	1	-1	-1	a	y_2
3	-1	1	-1	b	y_3
4	1	1	1	ab	y_4
	z_1	z_2	z_{12}		y

General DOE 2^k set-up:

- factors (covariates) are called A, B, C, ...
- each factor has two levels, coded as -1 and 1

In full 2^k experiments we will conduct all possible experiments, $n = 2^k$ experiments.

All possible experiments are written in so-called standard order:

A	B	C
-1	-1	-1
1	-1	-1
-1	1	-1
1	1	-1
-1	-1	1
1	-1	1
-1	1	1
1	1	1