TMA4267 Linear Statistical Models V2017 (L12)

Part 2: Linear regression:
Model selection [F:3.4]
Transformation and Taylor expansion
Quiz

Mette Langaas

Department of Mathematical Sciences, NTNU

To be lectured: February 24, 2017

What is the "best" model?

Acid rain in Norwegian lakes, data on $n = 26$ lakes, with

- y: measured pH in lake,
- x1: SO_4: sulfate (the salt of sulfuric acid),
- x2: $N0_3$: nitrate (the conjugate base of nitric acid),
- x3: Ca: calsium,
- x4: latent Al: aluminium,
- x5: organic substance,
- x6: area of lake,
- x7: position of lake (Telemark or Trøndelag),

Topic: choosing the "best" linear regression model!

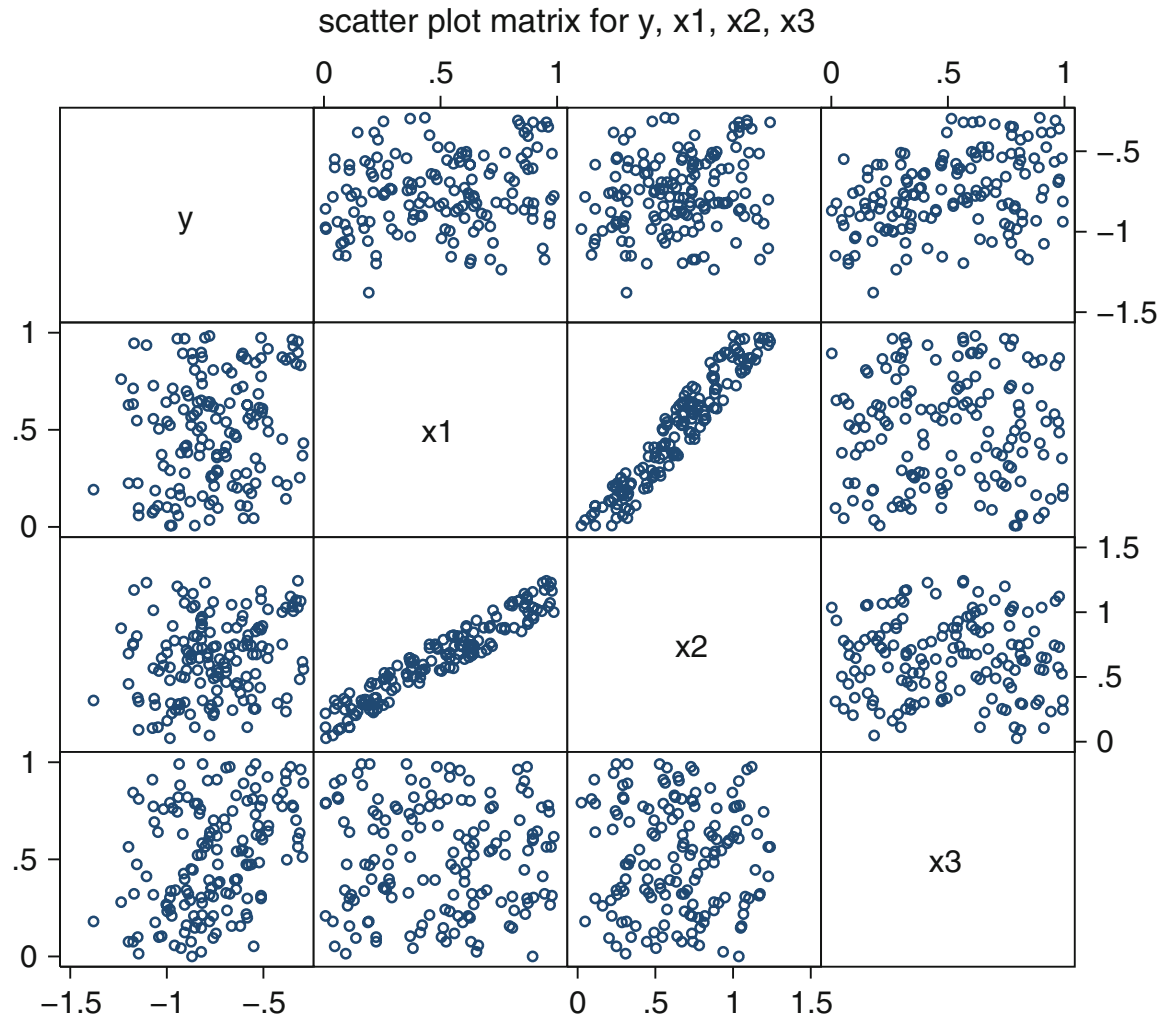
- First, debunk popular strategies (based on simulations studies were we knew the "true" model):
  - Popular 1: fit all available covariates.
    Problem: overfitting (=fitting trends and noise).
  - Popular 2: fit all available covariates, then remove the insignificant ones (=those β_j where $H_0 : \beta_j = 0$ is not rejected).

Simulated data (Fahrmeir et al: Fig 3.18, Tab3.3, Tab3.4)

True model:

$$Y \sim N(-1 + 0.3x_1 + 0.2x_3, 0.2^2)$$

where also $x_2 = x_1 + u$ is observed ($u \sim$ uniform in 0,1). The variables x_1 and x_3 are uncorrelated.



ig. 3.18 Scatter plot matrix for the variables y, x_1, x_2, and x_3

Figure from our text book: Fahrmeir et al (2013): Regression. Springer. (p.141)

Table 3.3 Results for the model based on covariates x_1, x_2, and x_3

Variable	Coefficient	Standard error	t-value	p-value	95 % Confidence interval	
intercept	−0.970	0.047	−20.46	<0.001	−1.064	−0.877
x_1	0.146	0.187	0.78	0.436	−0.224	0.516
x_2	0.027	0.177	0.15	0.880	−0.323	0.377
x_3	0.227	0.052	4.32	<0.001	0.123	0.331

Table 3.4 Results for the correctly specified model based on covariates x_1 and x_3

Variable	Coefficient	Standard error	t-value	p-value	95 % Confidence interval	
intercept	−0.967	0.039	−24.91	<0.001	−1.042	−0.889
x_1	0.173	0.055	3.17	0.002	0.065	0.281
x_3	0.226	0.052	4.33	<0.001	0.123	0.330

Table from our text book: Fahrmeir et al (2013): Regression. Springer. (p.142)

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 β_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.*

All models are wrong?

A model is a simplification or approximation of reality and hence will not reflect all of reality.

George Box noted that "all models are wrong, but some are useful". While a model can never be "truth"a model might be ranked from very useful, to useful, to somewhat useful to, finally, essentially useless.

Burnham, K. P.; Anderson, D. R. (2002), Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.

Expected squared prediction error (SPSE)

Possible criterion we want to minimize: SPSE.
Definition (j, M, ... given in classnotes)

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

can be written as:

$$\text{SPSE} = \sum_{j=1}^{J} \mathrm{E}((Y_j - \hat{Y}_{jM})^2) = n\sigma^2 + |M|\sigma^2 + \sum_{j=1}^{J}(\mu_{jM} - \mu_j)^2$$

Problem: Not useful on practise since μ_j and σ^2 are unknown.
Plan: Find a way to estimate SPSE and then choose the model M with the minimum SPSE!

How to estimate SPSE?

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

Assume we have fitted a model M with $|M|$ regression parameters.

1. Use new (independent) data – if available (seldom the case):

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

2. Cross-validation: mimic new data by dividing data into k folds (popular is $k = n$ and $k = 10$). In a for-loop let $j = 1, \ldots, k$, and use all folds except fold j to estimate regression parameter, and use the jth fold to calculated the $\widehat{SPSE}$. Sum across folds.

Choose the model M that minimizes the $\widehat{SPSE}$.

Cross-validation (5-fold)

Will be taught in TMA4300 Computational statistics and will be a backbone in TMA4268 Statistical Learning.

`http://blog-test.goldenhelix.com/wp-content/uploads/2015/04/B-fig-1.jpg`

How to estimate SPSE?

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

Assume we have fitted a model M with $|M|$ regression parameters.

3. Use existing data (only): It can be shown that $\mathrm{E}(\widehat{SPSE}) = SPSE - 2 \mid M \mid \sigma^2$ when used on the same data that was used to make the prediction, so a better estimate for existing data is

$$\widehat{SPSE} = \sum_{i=1}^{n} (Y_i - \hat{Y}_{iM})^2 + 2|M|\hat{\sigma}^2 = SSE + 2|M|\hat{\sigma}^2$$

   where $\hat{\sigma}^2$ is the same for all models M, and is often estimated using the most complex model under study.
4. Other criteria: all have the same form; a first term based on SSE (or R^2) for model M, and a second term penalizing the model complexity.

Choose the model M that minimizes the $\widehat{SPSE}$.

For models with the same model complexity – easy solution: SSE

Estimators for SPSE to be used on the same data as to be used for estimating the model parameters have the same form; a first term based on SSE (or R^2) for model M, and a second term penalizing the model complexity.
If we consider two models with the same model complexity then SSE can be used to choose between these models.

Acid rain (1). Best subset

For $1, \ldots, 7$ covariates: fit all possible models, and report the model with the smallest SSE (given below) for each value for the model complexity. Explain what you see! How many models have been searched for each model complexity?

```
regfit.full=regsubsets(y~.,data=ds)
sumreg <- summary(regfit.full)
Subset selection object
Call: regsubsets.formula(y ~ ., data = ds)
Selection Algorithm: exhaustive
         x1  x2  x3  x4  x5  x6  x7
1  ( 1 ) " " " " " " "*" " " " " " "
2  ( 1 ) "*" " " "*" " " " " " " " "
3  ( 1 ) "*" "*" "*" " " " " " " " "
4  ( 1 ) "*" "*" "*" " " "*" " " " "
5  ( 1 ) "*" "*" "*" " " "*" " " "*"
6  ( 1 ) "*" "*" "*" "*" "*" " " "*"
7  ( 1 ) "*" "*" "*" "*" "*" "*" "*"
```

Names: x1: SO_4, x2: $N0_3$, x3: Ca, x4: latent Al, x5: organic substance, x6: area of lake, x7: position of lake (Telemark or Trøndelag).

Popular model choice criteria

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.

R^2 adjusted (corrected)

$\hat{Y}_i$ is from fitting the regression model M.

Remember, for a regression model (with intercept) we have the SST=SSR+SSE.

$$\text{SST} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$\text{SSE} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$R^2 = 1 - \frac{SSE}{SST}$$

$$R^2_{\text{adj}} = 1 - \frac{\frac{SSE}{n-p}}{\frac{SST}{n-1}} = 1 - \frac{n-1}{n-p}(1 - R^2)$$

Choose the model with the *largest* R^2_{adj}.

"All" statistical software outputs this automatically! However, Fahrmeir et al (2013) believes that the penalty $n - p$ is too small.

Happiness ($n = 39$)

Are love and work the important factors determining happiness?

- y, `happiness`. 10-point scale, with 1 representing a suicidal state, 5 representing a feeling of «just muddling along», and 10 representing a euphoric state.
- x_1, `money`. Annual family income in thousands of dollars.
- x_2, `sex`. Sex was measured as the values 0 or 1, with 1 indicating a satisfactory level of sexual activity.
- x_3, `love`. 3-point scale, with 1 representing loneliness and isolation, 2 representing a set of secure relationships, and 3 representing a deep feeling of belonging and caring in the context of some family or community.
- x_4, `work`. 5-point scale, with 1 indicating that an individual is seeking other employment, 3 indicating the job is OK, and 5 indicating that the job is enjoyable.

Data taken from library faraway, data set happy.

Happy

```
> allreg=regsubsets(happy~.,data=happy)
> sumreg <- summary(allreg)
> sumreg
Subset selection object
Call: regsubsets.formula(happy ~ ., data = happy)
1 subsets of each size up to 4
Selection Algorithm: exhaustive
         money sex love work
1  ( 1 ) " "   " " "*"  " "
2  ( 1 ) " "   " " "*"  "*"
3  ( 1 ) "*"   " " "*"  "*"
4  ( 1 ) "*"   "*" "*"  "*"
```

	money	sex	love	work	N	p	R^2	R^2_{adj}
1	0.014				1	0.000747	7.3	4.8
2		−0.130			1	1	0.1	−2.6
3			2.270		1	8.35e-24	61.5	60.5
4				0.990	1	1.36e-13	29.1	27.2
5	0.016	−0.508			2	0.0504	8.8	3.8
6	0.009		2.206		2	8.77e-19	64.5	62.5
7	0.012			0.961	2	3.68e-10	34.6	31.0
8		−0.277	2.279		2	5.55e-18	62.0	59.9
9		0.610		1.079	2	3.48e-09	31.2	27.4
10			1.959	0.511	2	5.75e-20	68.1	66.3
11	0.011	−0.536	2.209		3	9.49e-16	66.2	63.3
12	0.011	0.305		1.009	3	1.84e-07	35.1	29.5
13	0.009		1.902	0.504	3	2.63e-17	70.9	68.4
14		0.108	1.944	0.530	3	2.22e-16	68.1	65.4
15	0.010	−0.149	1.919	0.476	4	9.89e-15	71.0	67.6

Intercept included, $N = p - 1$, p-value for significance of regression.

$R^2 = 1 - \frac{SSE}{SST}$, $R^2_{\text{adj}} = 1 - \frac{\frac{SSE}{n-p}}{\frac{SST}{n-\mathbf{1}}}$. Which model to prefer?

Mallows' C_p

$\hat{Y}_i$ is from fitting regression model M.
Mallows is the name of a person.

$$C_p = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\hat{\sigma}^2} - n + 2|M|$$

Minimizing Cp gives the same optimal model as minimizing $\widehat{SPSE}$.

See Exam V2015 Problem 3 for an in depth explanation of the theory behind Mallow's Cp.

AIC

Akaike information criterion – one of the most widely used.
Designed for likelihood-based inference.

For a normal regression model:

$$\text{AIC} = n \ln(\hat{\sigma}^2) + 2(|M| + 1)$$

Choose the model with the minimum AIC.

BIC

Bayesian information criterion.

For a normal regression model:

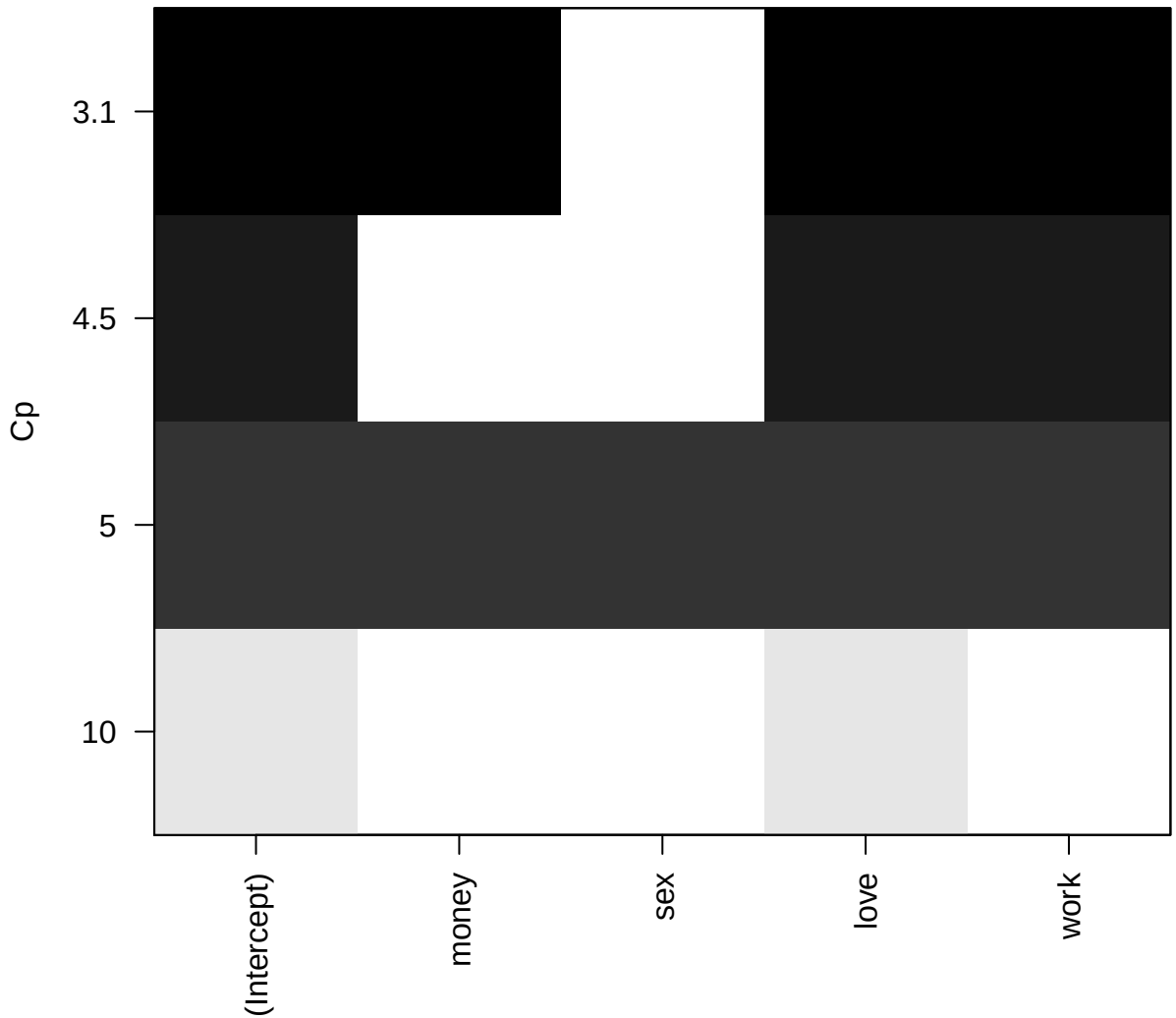
$$\text{BIC} = n \ln(\hat{\sigma}^2) + \ln(n)(|M| + 1)$$

Choose the model with the minimum BIC.

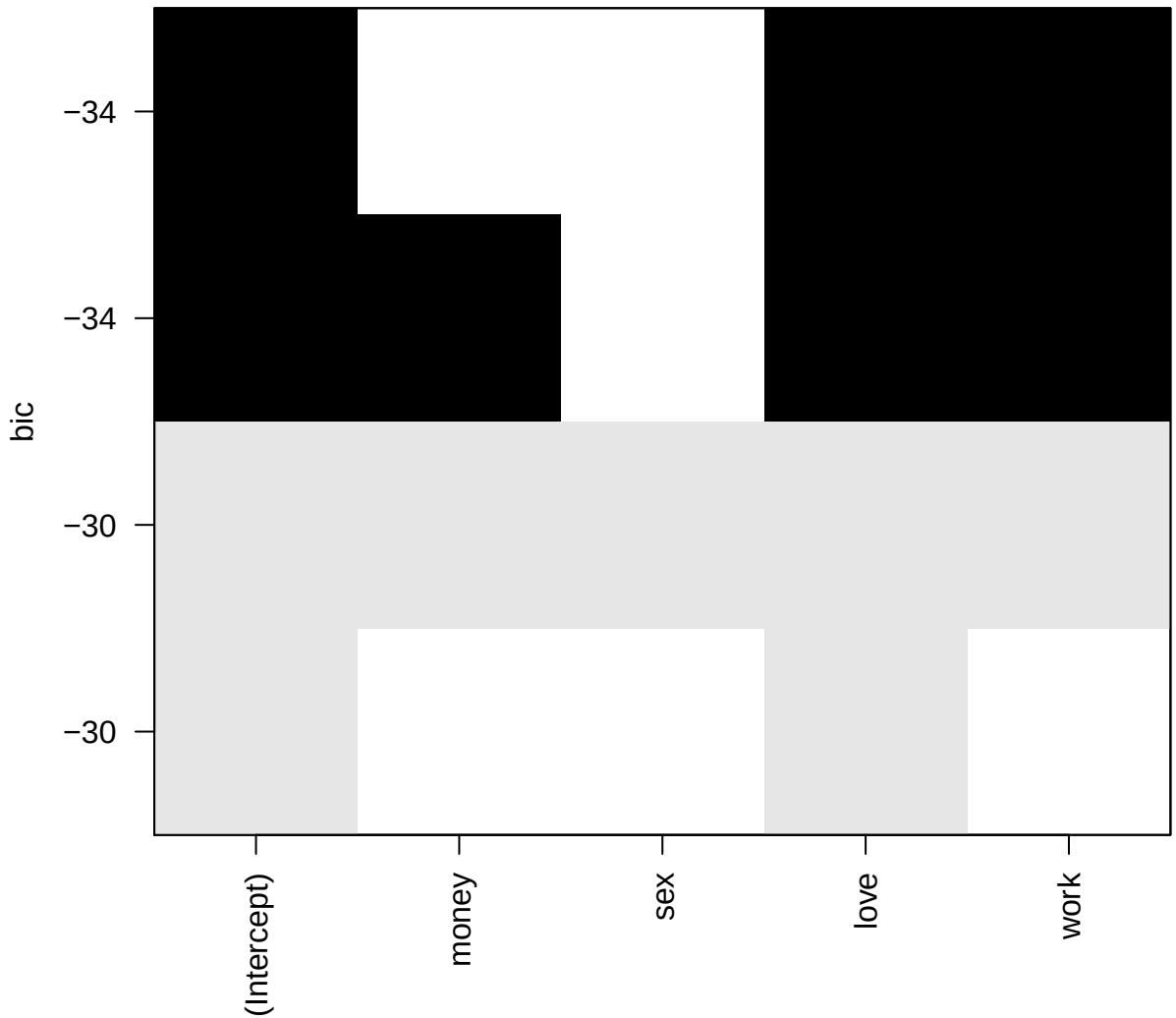
AIC and BIC are motivated in very different ways, but the final result for the normal regression model is very similar.

BIC has a larger penalty than AIC ($\log(n) vs. 2$), and will often give a smaller model (=more parsimonious models) than AIC.

Happy: Mallows' C_p



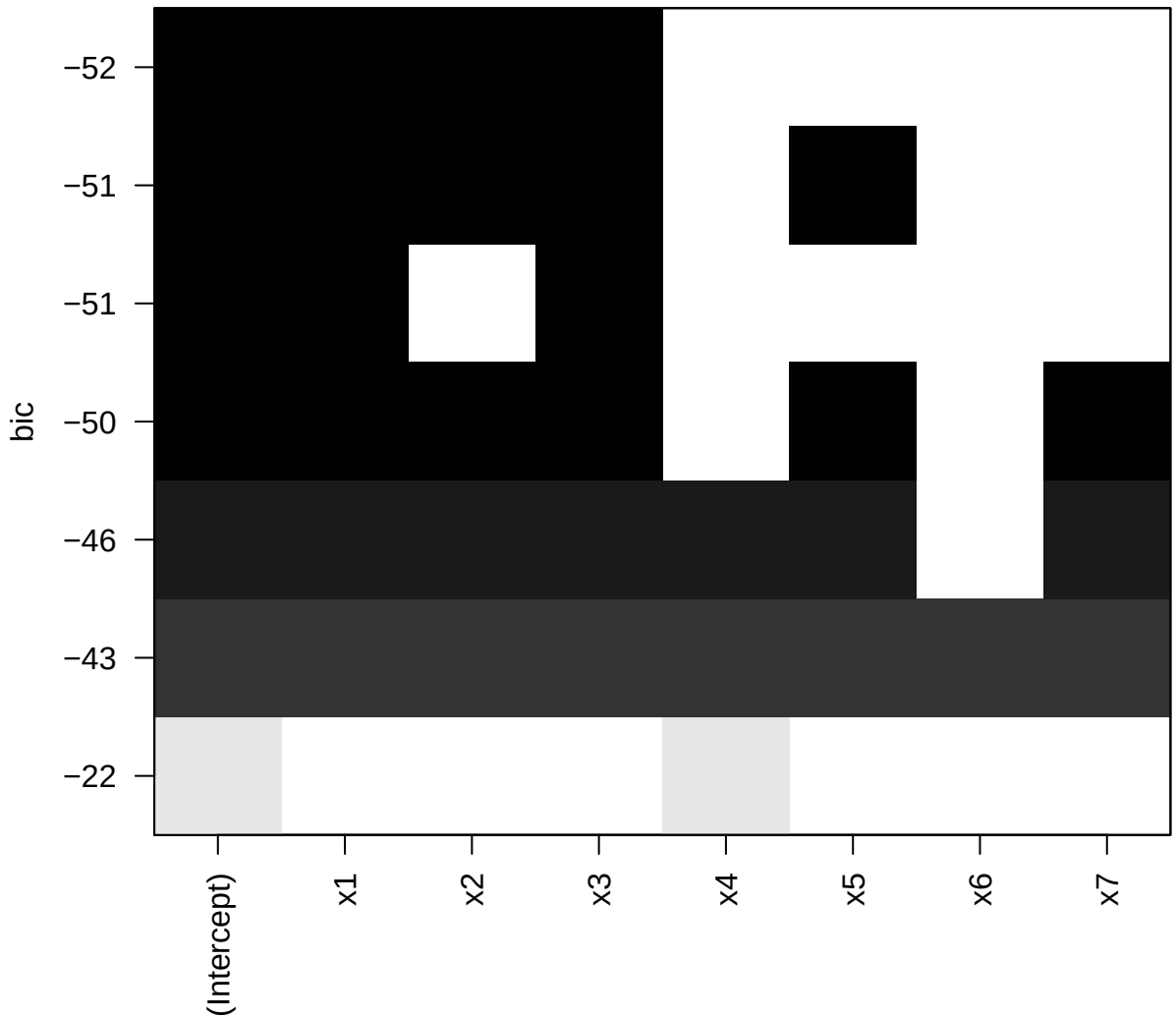
Happy: BIC



Acid rain (2)

```
Call: regsubsets.formula(y ~ ., data = ds)
1 subsets of each size up to 7
Selection Algorithm: exhaustive
         x1  x2  x3  x4  x5  x6  x7
1  ( 1 ) " " " " " " "*" " " " " " "
2  ( 1 ) "*" " " "*" " " " " " " " "
3  ( 1 ) "*" "*" "*" " " " " " " " "
4  ( 1 ) "*" "*" "*" " " "*" " " " "
5  ( 1 ) "*" "*" "*" " " "*" " " "*"
6  ( 1 ) "*" "*" "*" "*" "*" " " "*"
7  ( 1 ) "*" "*" "*" "*" "*" "*" "*"
# to mimic test set:
which.max(sumreg$adjr2) #5
which.min(sumreg$cp) #3
which.min(sumreg$bic) #3
# so, model 3 or 5 is suggested for us
# model 3: x1+x2+x3
# model 5: x1+x2+x3+x5+x7
```

Acid rain, BIC,



Practical use of the model criteria

- All subset selection: use smart "leaps and bounds" algorithm, works fine for number of covariates in the order of 40.
- Forward selection: choose starting model (only intercept), then add one new variable at each step - selected to make the best improvement in the model selection criteria. End when no improvement is made.
- Backward elimination: : choose starting model (full model), then remove one new variable at each step - selected to make the best improvement in the model selection criteria. End when no improvement is made.
- Stepwise selection: combine forward and backward.

Acid rain (3): stepAIC

```
> all=lm(happy~.,data=happy)
> stepAIC(all)
Start:  AIC=9.08
happy ~ money + sex + love + work

        Df Sum of Sq    RSS    AIC
- sex    1     0.142 38.229  7.221
<none>               38.087  9.076
- money  1     3.782 41.869 10.768
- work   1     6.386 44.473 13.122
- love   1    47.272 85.359 38.549

Step:  AIC=7.22
happy ~ money + love + work

        Df Sum of Sq    RSS    AIC
<none>               38.229  7.221
- money  1     3.723 41.952  8.846
- work   1     8.410 46.639 12.976
- love   1    47.742 85.971 36.828

Call:
lm(formula = happy ~ money + love + work, data = happy)

Coefficients:
(Intercept)        money         love         work
  -0.185936     0.008959     1.901709     0.503602
```

Acid rain (4): Forward

```
regfitF=regsubsets(y~.,data=ds,method="forward")
sumregF <- summary(regfitF)
Selection Algorithm: forward
         x1  x2  x3  x4  x5  x6  x7
1  ( 1 ) " " " " " " "*" " " " " " "
2  ( 1 ) " " " " "*" "*" " " " " " "
3  ( 1 ) "*" "*" "*" " " " " " " " "
4  ( 1 ) "*" "*" "*" "*" " " " " " "
5  ( 1 ) "*" "*" "*" "*" "*" " " " "
6  ( 1 ) "*" "*" "*" "*" "*" " " "*"
7  ( 1 ) "*" "*" "*" "*" "*" "*" "*"
which.max(sumregF$adjr2)#5
which.min(sumregF$cp) #3
which.min(sumregF$bic) #3
```

Acid rain (5): Backward

```
regfitB=regsubsets(y~.,data=ds,method="backward")
sumregB <- summary(regfitB)
Selection Algorithm: backward
         x1  x2  x3  x4  x5  x6  x7
1  ( 1 ) " " " " "*" " " " " " " " " "
2  ( 1 ) "*" " " "*" " " " " " " " " "
3  ( 1 ) "*" "*" "*" " " " " " " " " "
4  ( 1 ) "*" "*" "*" " " "*" " " " " "
5  ( 1 ) "*" "*" "*" " " "*" " " " " "*"
6  ( 1 ) "*" "*" "*" "*" "*" " " " " "*"
7  ( 1 ) "*" "*" "*" "*" "*" "*" "*" "*"
which.max(sumregB$adjr)#5
# backward finds same as best subset
which.min(sumregB$cp) #3
```

Model diagnosis

- Influential observations and outliers: impact of specific observations on model fit.
- Collinearity analysis: Highly correlated variables cause imprecise estimation of the regression parameters. (Why? Look at diagonal elements of $\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$, and look back to Problem 2 in the start of this lecture.)
- Examination of model assumptions: residual plots!

Influential observations– and outliers

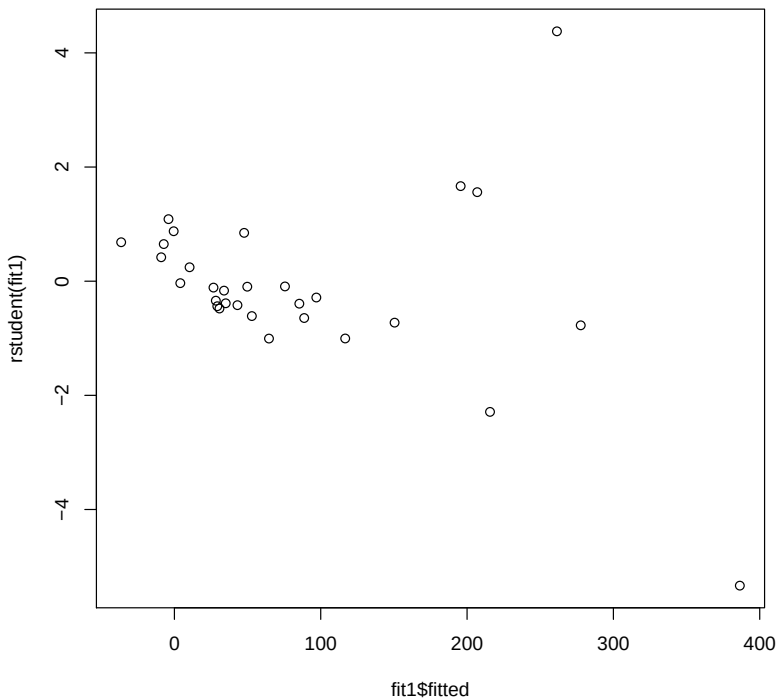
- Observations that significantly affect inferences drawn from the data are said to be influential.
- The leverage, h_{ii}, associated with the ith datapoint measures "how far the ith observation is from the other $n-1$ observations".
- Methods for assessing influential observations may be be based on change in $\boldsymbol{\beta}$ estimate when observations are deleted.
- Always investigate possible causes of an influential observation (if possible).
- Cook's distance can be used to identify influential observations.
- Robust methods (median,quantile regression) can be useful.

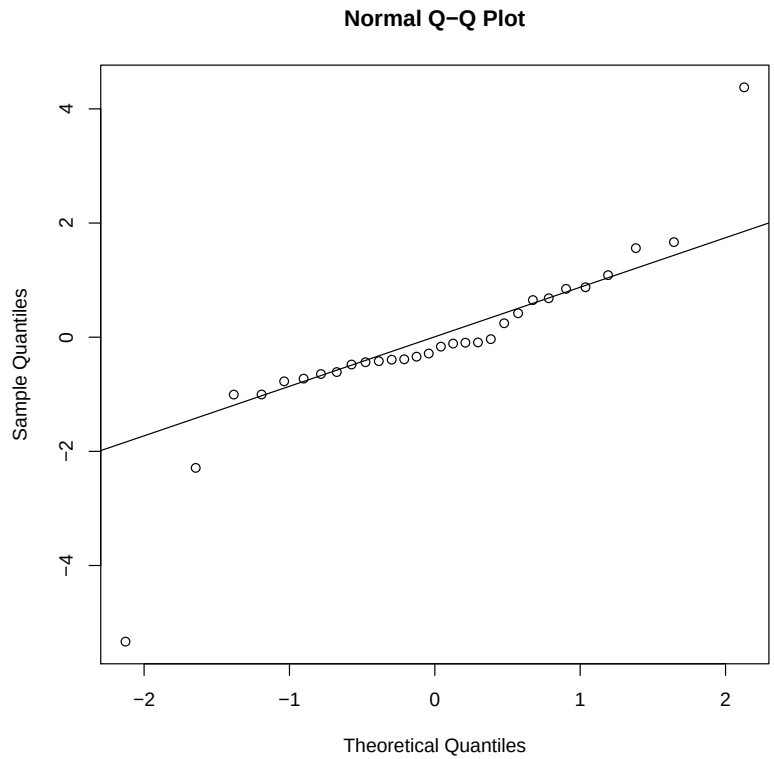
Want to understand more? Read for yourself in Fahrmeir et al (2013): p 160-166.

Transformations

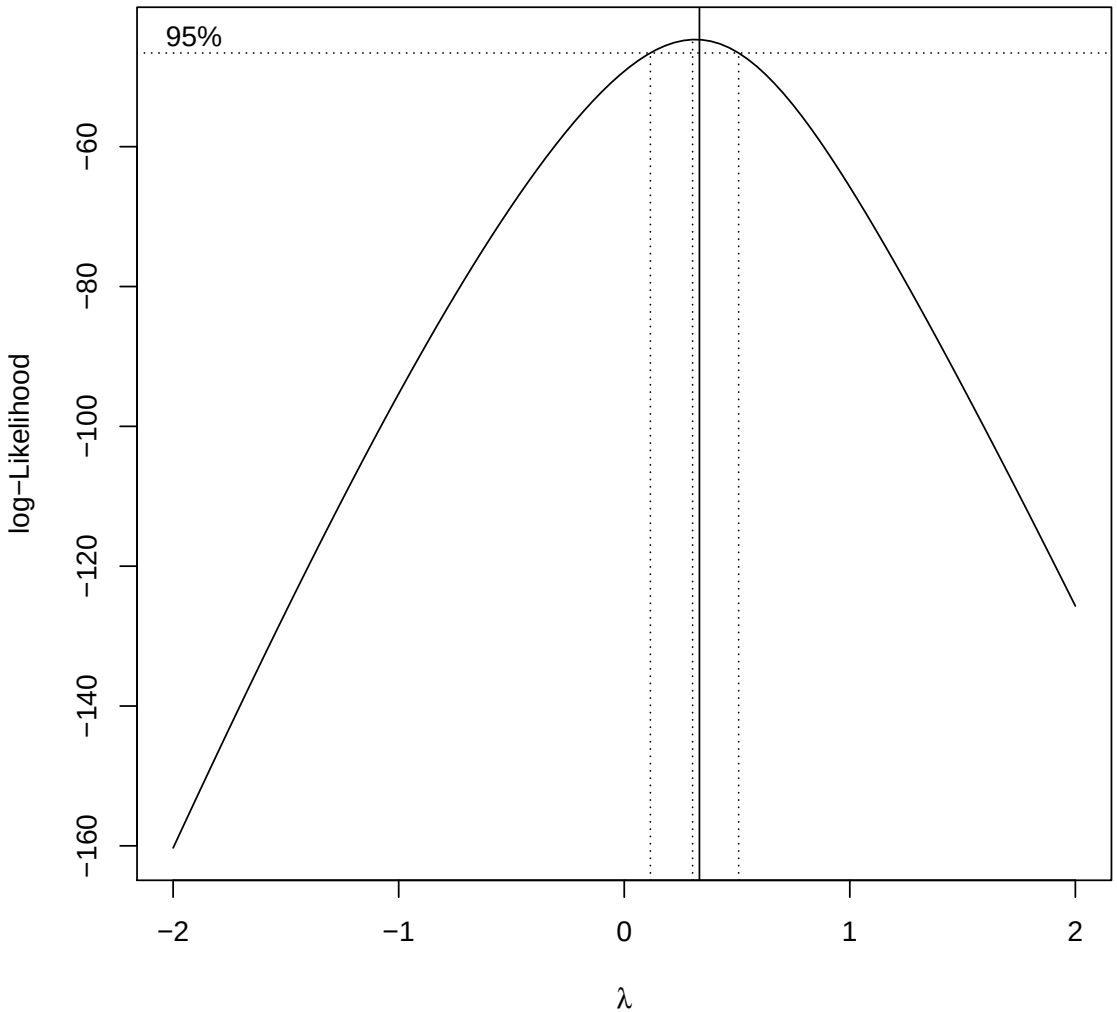
- Multiplicative or additive model?
- Box–Cox transform with profile likelihood.
- Stabilizing the variance.

Galapagos islands, Model A, Exam V2014 Problem 2



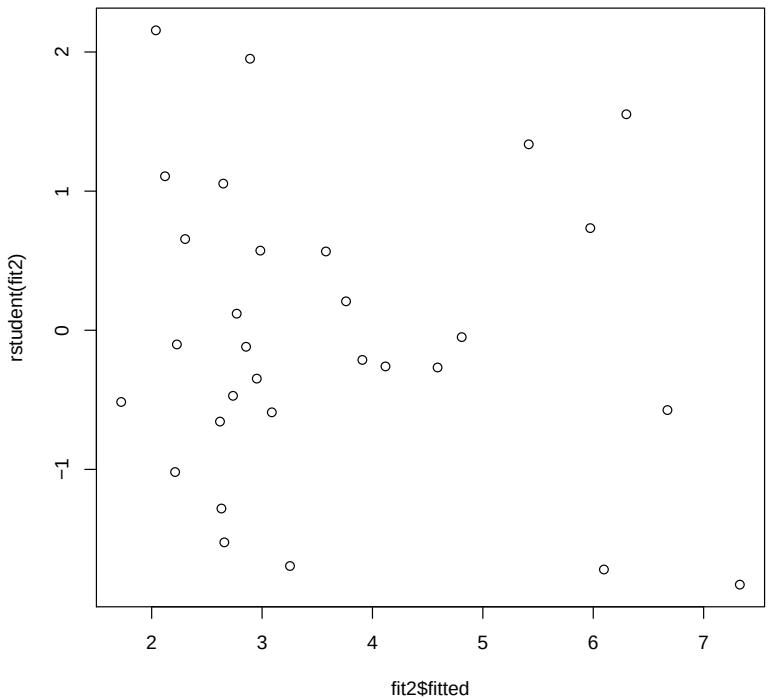


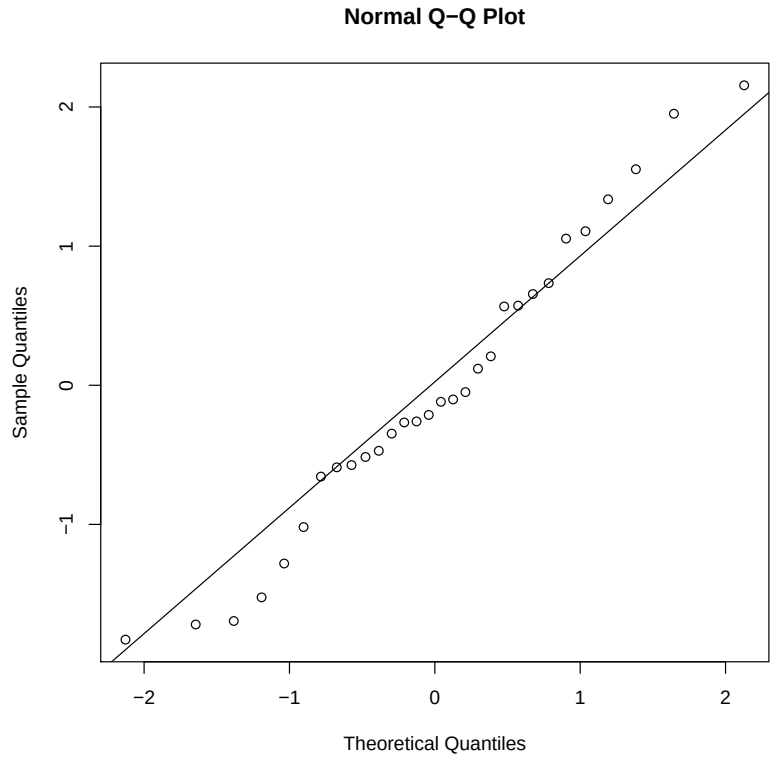
Box–Cox plot



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

Galapagos islands, Model B, Exam V2014 Problem 2





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 $g(X_1, X_2)$.

Example ln of BMI

Looking at residual plots from a regression model the conclusion was to analyse data of *BMI* on the natural logarithmic scale. After a regression model was fitted the predicted value for the ln(BMI) for a specific combination of the covariates was found to be 3.2151 with an estimated standard deviation of 0.1656. Use approximate methods to arrive at an estimate of the predicted value and estimated standard deviation on the original scale, kg/m^2, and not on the logarithmic scale.

$\mathrm{E}(g(X)$ and $\mathrm{Var}(g(X))$

- 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?
  - We can calculate $\mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$ when X is continuous, or a version thereof in the discrete case,
  - or 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:

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

Treatment of tennis elbow (exam TMA4255 V2012, 3b)

The term *tennis elbow* is used to describe a state of inflammation in the elbow, causing pain. This injury is common in people who play racquet sports, however, any activity that involves repetitive twisting of the wrist (like using a screwdriver) can lead to this condition. The condition may also be due to constant computer keyboard and mouse use.
In a randomized clinical study the aim was to compare three different methods for treatment of tennis elbow,

- A: physiotherapy intervention,
- B: corticosteroid injections and
- C: wait-and-see (the patients in the wait-and-see group did not get any treatment but was told to use the elbow as little as possible).

Treatment of tennis elbow (cont.)

We will look at the short-term effect of treatment by studying measurements at 6 weeks. All patients participating in the study only had one affected arm.
We will look at the outcome measure called *pain-free grip force*. This was measured by a digital grip dynamometer and normalized to the grip force of the unaffected arm. A pain-free grip force of 100 would mean that the affected and the unaffected arm performed equally good.

Summary statistics for each of the treatment groups.

Treatment	Sample size	Average	Standard deviation
A (physiotherapy)	63	70.2	25.4
B (injection)	65	83.6	22.9
C (wait-and-see)	60	51.8	23.0
Total	188	69.0	

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 $\mu_1 = \mathrm{E}(X_1)$ 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:

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

Bivariate function: first order Taylor

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

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

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

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

Today

- Choosing between models of *equal model complexity*: choose the model with the minimum SSE.
- Choosing between models of *different model complexity*: Model selection based on penalized criteria (Mallows Cp, R^2_{adj},AIC and BIC). Try out on RecEx4 and Compulsory Exercise 2.
- BoxCox transformation: see RecEx4.
- Work for for yourself: Taylor solution to E and Var of nonlinear function, useful when you want to look at transformations of the data or functions of parameter estimates.

Summary of Part 2 in Kahoot!