Design and analysis of experiments
Lecture 7

Jakob G. Rasmussen
Department of Mathematics
Aalborg University
Denmark

Regression

- Last time we learnt how to model a dataset using a regression model and estimate the parameters in this model, fx. slope and intercept in the case of simple regression - but there is much more to be done.
- From last time we got the following:
  - Parameter estimates: $\hat{\beta} = (X^\top X)^{-1} X^\top y$
  - Fitted values: $\hat{y} = X(X^\top X)^{-1} X^\top y$
- Thus we have estimated $\beta = (\beta_0, \ldots, \beta_k)^\top$, but we still have to estimate σ^2.

Estimate of σ^2

- Error sum of squares:

$$SS_E = |r|^2 = |y - \hat{y}|^2 = \sum_i (y_i - \hat{y}_i)^2 \sim \sigma^2 \chi^2_{n-p}$$

  where $p = k + 1$ is the number of parameters not counting σ^2, fx. $p = 2$ for simple regression.

- Estimate of σ^2:

$$\hat{\sigma}^2_E = MS_E = \frac{SS_E}{\nu_E} \sim \sigma^2 \frac{\chi^2_{\nu_E}}{\nu_E}, \quad \nu_E = n - p$$

- This is an unbiased estimate:

$$\mathbb{E}[\hat{\sigma}^2_E] = \sigma^2$$

Tests in regression models

- Consider a regression model where the response variable depends on a number of explanatory variables, e.g. multiple regression $\mu_i = \beta_0 + \beta_1 x_{1i} + \ldots \beta_k x_{ki}$ - have we included irrelevant explanatory variables?
- General hypothesis:
  H_0 : the model contains a subset of the p terms (p_0 terms)
  H_1 : the model contains all p terms
- Example:

$$H_0 : \mu = \beta_0 + \beta_1 x_1$$
$$H_1 : \mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Test statistic

- Sum of squares:

$$SS_E = |y - \hat{y}|^2 \sim \sigma^2 \chi^2_{\nu_E}, \quad \nu_E = n - p$$
$$SS_R = |\hat{y} - \hat{y}_0|^2 \sim \sigma^2 \chi^2_{\nu_R}, \quad \nu_R = p - p_0$$

($\hat{y}_0$ is the fitted values for the model under H_0)

- Test statistic:

$$F_0 = \frac{MS_R}{MS_E} = \frac{SS_R/\nu_R}{SS_E/\nu_E} \sim \frac{\chi^2_{\nu_R}/\nu_R}{\chi^2_{\nu_E}/\nu_E} = F_{\nu_R,\nu_E}$$

- We reject the hypothesis if $F_0 > F_{\nu_R,\nu_E;\alpha}$, i.e. we cannot make the hypothesized simplification of the model.
- This test can be used to simplify the model by removing one, some or all of the explanatory variables.

Test for removing just one variable

- As a special case we can test the significance of just one variable.
- Hypothesis:

$$H_0 : \beta_i = 0$$
$$H_1 : \beta_i \neq 0$$

- We reject the hypothesis if $F_0 > F_{1,\nu_E;\alpha}$
- Or (equivalently) we can make a t-test - this is included in the standard output in R, but we will get back to the details a bit later.

The all-or-nothing test

- Example:

$$H_0 : \mu = \beta_0, \qquad p_0 = 1$$
$$H_1 : \mu = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k, \qquad p = k + 1$$

- This test is used to see if there is anything useful in the model (a bit like the F-test used in ANOVA)
- The related sums of squares can be really useful in characterising the explanatory value of the model:

$$SS_E = \sum_i (y_i - \hat{y}_i)^2 \sim \sigma^2 \chi^2_{\nu_E}, \quad \nu_E = n - p = n - k - 1$$
$$SS_R = \sum_i (\hat{y}_i - \bar{y})^2 \sim \sigma^2 \chi^2_{\nu_R}, \quad \nu_R = p - p_0 = k$$
$$SS_{Tot} = \sum_i (y_i - \bar{y})^2 = SS_E + SS_R \quad \nu_{Tot} = n - 1$$

Coefficient of determination

- Coefficient of determination:

$$R^2 = \frac{SS_R}{SS_{Tot}} = 1 - \frac{SS_E}{SS_{Tot}}$$

- About R^2:
  - $0 \leq R^2 \leq 1$
  - High R^2 means the model does well in explaining the variation in the data, while low R^2 suggests important explanatory variables are missing (but the ones included may still be ok).
  - Including extra variables will never decrease R^2, so R^2 always favors complicated models.
- Adjusted coefficient of determination:

$$R^2_{\text{adj}} = 1 - \frac{MS_E}{MS_{Tot}} = 1 - \frac{n-1}{n-p}(1 - R^2)$$

Complicated models (i.e. large p) will reduce R^2_{adj}, so it will only favor more complicated models that add something significant to the explanantory value of the model.

R

- R-demo, part 1
- Exercise 1

A bit about random vectors

- Consider a 2-dimensional stochastic vector $Y = (Y_1, Y_2)^\top$ - everything generalises easily to n dimensions, but for ease of presentation we just look at the 2 dimensional case.
- The equivalent of the mean value is the mean vector:

$$\mathbb{E}[Y] = \begin{bmatrix} \mathbb{E}[Y_1] \\ \mathbb{E}[Y_2] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \mu$$

- The equivalent of the variance is the covariance matrix - first we need the 1-dimensional covariance:

$$\mathrm{Cov}(Y_i, Y_j) = \mathbb{E}[(Y_i - \mu_i)(Y_j - \mu_j)]$$

- The variance is a special case of the covariance:

$$\mathrm{Cov}(Y_i, Y_i) = \mathbb{E}[(Y_i - \mu_i)(Y_i - \mu_i)] = \mathbb{E}[(Y_i - \mu_i)^2] = \mathrm{Var}(Y_i)$$

The covariance matrix

- Covariance matrix:

$$
\begin{aligned}
\text{Cov}(Y) &= \mathbb{E}[(Y-\mu)(Y-\mu)^\top] \\
&= \mathbb{E}\left[\begin{bmatrix} Y_1-\mu_1 \\ Y_2-\mu_2 \end{bmatrix}\begin{bmatrix} Y_1-\mu_1 & Y_2-\mu_2 \end{bmatrix}\right] \\
&= \begin{bmatrix} \text{Cov}(Y_1,Y_1) & \text{Cov}(Y_1,Y_2) \\ \text{Cov}(Y_2,Y_1) & \text{Cov}(Y_2,Y_2) \end{bmatrix} \\
&= \begin{bmatrix} \text{Var}(Y_1) & \text{Cov}(Y_1,Y_2) \\ \text{Cov}(Y_2,Y_1) & \text{Var}(Y_2) \end{bmatrix}
\end{aligned}
$$

- Note:
  - The elements in the diagonal $\text{Cov}(Y_i, Y_i) = \text{Var}(Y_i)$ are variances, the rest are covariances.
  - The covariance matrix is symmetric since $\text{Cov}(Y_i, Y_j) = \text{Cov}(Y_j, Y_i)$.

Linear transformations

- Transformation: $Z = AY$ for some constant matrix of appropriate dimensions.
- Mean:
$$\mathbb{E}[Z] = A\mathbb{E}[Y]$$
Compare with $\mathbb{E}[aX] = a\mathbb{E}[X]$ for a stochastic variable X.
- Covariance:
$$\mathrm{Cov}(Z) = A\mathrm{Cov}(Y)A^\top$$
Compare with $\mathrm{Var}(X) = a^2\mathrm{Var}(X)$.

Distribution of parameter estimator

- We have found the estimate $\hat{\beta} = (X^\top X)^{-1} X^\top y$ for β, but we still need the properties of these estimates.
- Mean:

$$\begin{aligned} \mathbb{E}[\hat{\beta}] &= \mathbb{E}[(X^\top X)^{-1} X^\top y] = (X^\top X)^{-1} X^\top \mathbb{E}[y] \\ &= (X^\top X)^{-1} X^\top X \beta = \beta \end{aligned}$$

  This is an unbiased estimator.
- Covariance:

$$\begin{aligned} \mathrm{Cov}(\hat{\beta}) &= \mathrm{Cov}[(X^\top X)^{-1} X^\top y] = (X^\top X)^{-1} X^\top \mathrm{Cov}[y] X (X^\top X)^{-1} \\ &= \sigma^2 (X^\top X)^{-1} = \sigma^2 C \end{aligned}$$

  where $C = (X^\top X)^{-1}$
- This implies that $\mathbb{E}[\hat{\beta}_i] = \beta_i$ and $\mathrm{Var}(\hat{\beta}_i) = \sigma^2 C_{ii}$.

Confidence intervals & hypothesis test

- It can also be proven that $\hat{\beta}_i$ is normally distributed:

$$\hat{\beta}_i \sim N(\beta_i, \sigma^2 C_{ii}),$$

- Since we don't know σ^2 we replace it with the estimate $\hat{\sigma}_E^2 = MS_E$ - as usual this turns the normal distribution into a t-distribution.
- Confidence interval for β_i:

$$\hat{\beta}_i \pm t_{\nu_E;\alpha/2} \hat{\sigma}_E \sqrt{C_{ii}}, \qquad \nu_E = n - p$$

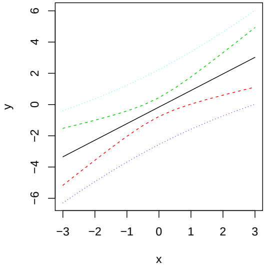
- Test statistic for testing $H_0 : \beta_i = 0$:

$$t_0 = \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_E \sqrt{C_{ii}}} = \frac{\hat{\beta}_i}{\hat{\sigma}_E \sqrt{C_{ii}}} \sim t_{\nu_E}$$

- Accept H_0 if $t_0 \in [-t_{\nu_E;\alpha/2}, t_{\nu_E;\alpha/2}]$

Mean response and prediction

- The fitted curve is where we expect the "true" curve to be.
- Inner bounds: due to errors on the estimates of β, the actual curve is located between the confidence bounds with confidence $1 - \alpha$.
- Outer bounds: if we predict a new value y_0 given som value x_0, this will be within the outer bounds with probability $1 - \alpha$.



Mean response

- x_0 is a new vector of values of the explanatory variables $x_0^\top = (1, x_{10}, x_{20}, \ldots, x_{k0})$
- Mean: $\mu_0 = x_0^\top \beta$
- Estimated mean: $\hat{\mu}_0 = x_0^\top \hat{\beta} = \hat{y}(x_0)$
- Mean and variance of $\hat{\mu}_0$:

$$\mathbb{E}[\hat{\mu}_0] = x_0^\top \mathbb{E}[\hat{\beta}] = x_0^\top \beta = \mu_0$$
$$\mathrm{Var}(\hat{\mu}_0) = x_0^\top \mathrm{Cov}(\hat{\beta}) x_0 = \sigma^2 x_0^\top C x_0$$

- Confidence interval for μ_0:

$$\hat{\mu}_0 \pm t_{\nu_E;\alpha/2} \hat{\sigma}_E \sqrt{x_0^\top C x_0}$$

Predicted value

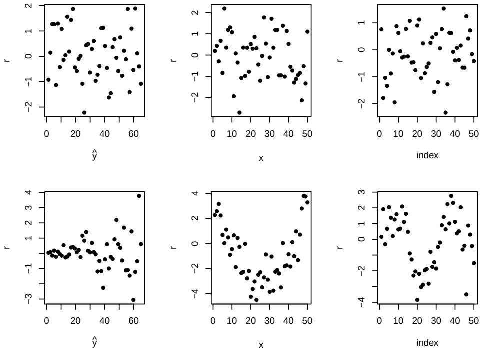
- x_0, μ_0 as before, y_0 is a new observation
- $y_0 \sim N(\mu_0, \sigma^2)$, $\hat{\mu}_0 \sim N(\mu_0, \sigma^2 x_0^\top C x_0)$ independent
- $y_0 - \hat{\mu}_0 \sim N(0, \sigma^2(1 + x_0^\top C x_0))$
- Prediction interval for y_0:

$$\hat{\mu}_0 \pm t_{\nu_E;\alpha/2} \hat{\sigma}_E \sqrt{1 + x_0^\top C x_0}$$

- Note that the only difference between the confidence interval and the prediction interval is the "$1+$" in the square root - this represents the additional uncertainty imposed by the new y_0.

Residual analysis

- We use residual analysis for checking the fit of the model.
- Useful residual-plots: r vs $\hat{y}$, r vs x_i, r vs index/time



- Never plot r vs y!

R

- R-demo, part 2
- Exercise 2 and 3