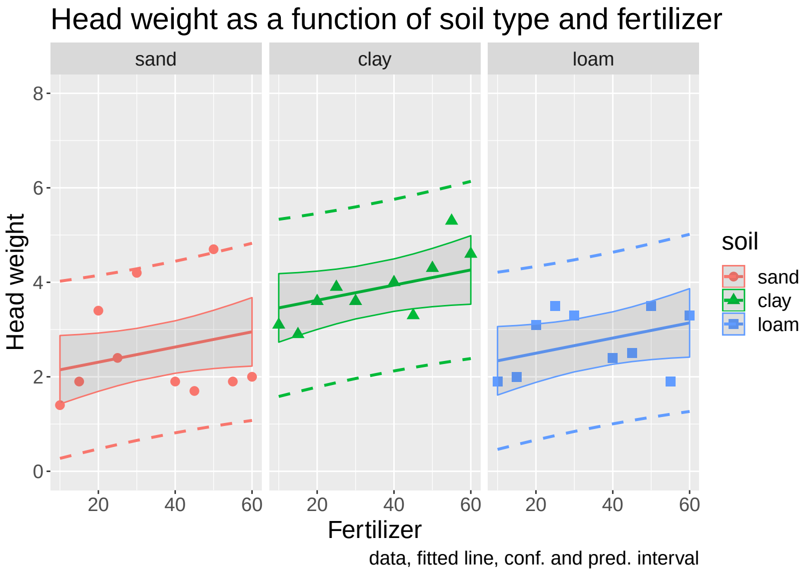
MASM22/FMSN30: Linear and Logistic Regression, 7.5 hp
FMSN40: . . . with Data Gathering, 9 hp

Lecture 4, spring 2023
Significance tests for model parameters

Mathematical Statistics / Centre for Mathematical Sciences
Lund University

29/3-23


Head weight as a function of soil type and fertilizer
sand
clay
loam
Head weight
0
2
4
6
8
20
40
60
Fertilizer
soil
sand
clay
loam
data, fitted line, conf. and pred. interval

Model: $Y_i = \beta_0 + \beta_1 x_{i2} + \beta_2 x_{i2} + \beta_3 x_3 + \epsilon_i$
where $x_1 = 1$ if clay (0 if not clay), $x_2 = 1$ if loam (0 if not loam) and $x_3 =$ fertilizer.

Questions

- Does the amount of fertilizer have a significant effect on the head weight, when we already have soil type in the model?
  Test H$_0$: $\beta_3 = 0$ against H$_1$: $\beta_3 \neq 0$. **t-test**.
- Are there significant systematic differences between any of the soil types, when we have fertilizer in the model?
  Test H$_0$: $\beta_1 = \beta_2 = 0$ against H$_1$: "at least one of β_1 and β_2 is $\neq 0$". **Partial F-test**.
- Is the model better than nothing?
  Test H$_0$: $\beta_1 = \beta_2 = \beta_3 = 0$ against H$_1$: "at least one of β_1, β_2 and β_3 is $\neq 0$". **Global F-test**.

Hypothesis testing

We want to test a null hypothesis H_0 against an alternative hypothesis H_1. We will reject H_0 in favour of H_1 only if our results are very unlikely to appear by chance if H_0 were true. Thus, we want to either

- determine if the difference between what we observed and what we would expect to observe if H_0 were true, is too large to be due to just random variablility, or, equivalently,
- determine if the chance of observing something this far away from what we expected if H_0 were true, by chance, is small.

Significance level

We define the significance level α of a test by

$$\alpha = P(\text{reject } \mathrm{H}_0 \mid \mathrm{H}_0 \text{ is true})$$

This is the probability of concluding that H_0 is wrong when it is in fact correct. False positive rate. Typically we choose $\alpha = 0.05$.

Test statistic

A test statistic is a suitable function of our observations that has a known distribution when H_0 is true, and can be expected to lie in the extreme end(s) of this distribution if H_0 is not true.

- Reject H_0 at significance level α if the test statistic lies above the α quantile (or outside the $1-\alpha/2$ and $\alpha/2$ quantiles) in its distribution. One-sided or two-sided depends on the type of statistic.

P-value

The P-value of a test is defined as the probability of observing as large a result as the one produced by your data, or an even more extreme result, when H_0 is true. If the P-value is small our result is more extreme than would be expected by chance and H_0 is unlikely to be true.

- Reject H_0 at significance level α if P-value $< \alpha$.

We can use the test statistic and its distribution to calculate the P-value.

t-test for β_j

We want to test if one x-variable, x_j, has any relevance in the model, given all the other variables.

We want to test H$_0$: $\beta_j = 0$ against H$_1$: $\beta_j \neq 0$.

Since $\hat{\beta}_j \sim N(\beta_j, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}_{jj}) = N(0, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}_{jj})$ when H$_0$ is true, a suitable test statistic is

$$t = \frac{\hat{\beta}_j - 0}{d(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{s\sqrt{(\mathbf{X}'\mathbf{X})^{-1}_{jj}}} \sim t(n-(p+1))$$

and we should reject H$_0$ at significance level α if

- $|t| > t_{\alpha/2,\, n-(p+1)}$ or
- P-value $= P(|t(n-(p+1))| > |t|) < \alpha$ or
- $0 \notin I_{\beta_j}$.

All three variants are equivalent.

Cabbage: soil and fertilizer estimates

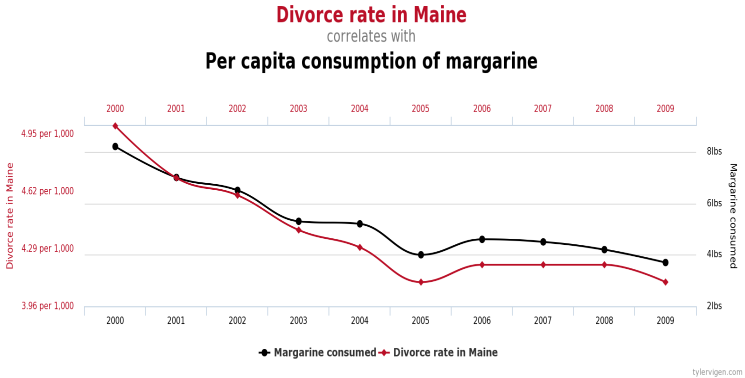
Variable	param	est	s.e.	t-value	P-value	95 % C.I.
intercept	β_0	1.99	0.42	4.74	$7 \cdot 10^{-5}$	$(1.13, 2.85)$
clay	β_1	1.31	0.38	3.48	0.002	$(0.54, 2.08)$
loam	β_2	0.19	0.38	0.50	0.62	$(-0.58, 0.96)$
fertilize	β_3	0.016	0.009	1.73	0.09	$(-0.003, 0.035)$
resid.std.dev	σ	0.84	df $= 26$			

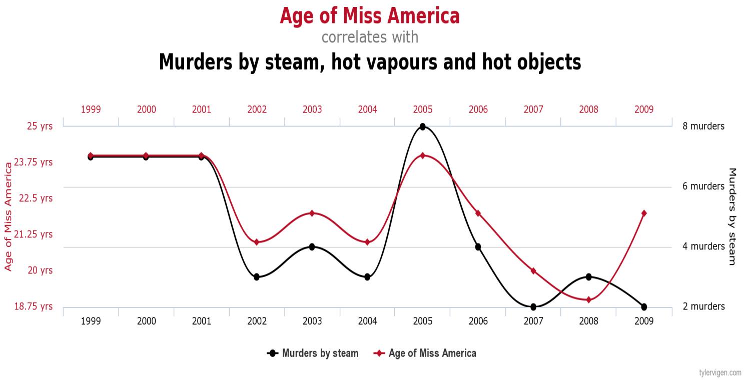
- Since $|t| = |\frac{0.016}{0.009}| = |1.73| \not> t_{0.05/2,\,26} = 2.06$ we should not reject H_0: $\beta_3 = 0$ at significance level $\alpha = 5\,\%$.
- Since P-value $= P(|t(26)| > 1.73) = 0.09 \not< 0.05$ we should not reject H_0 at significance level $\alpha = 5\,\%$
- Since the confidence interval for β_3 covers 0 we should not reject H_0 at significance level $\alpha = 5\,\%$.

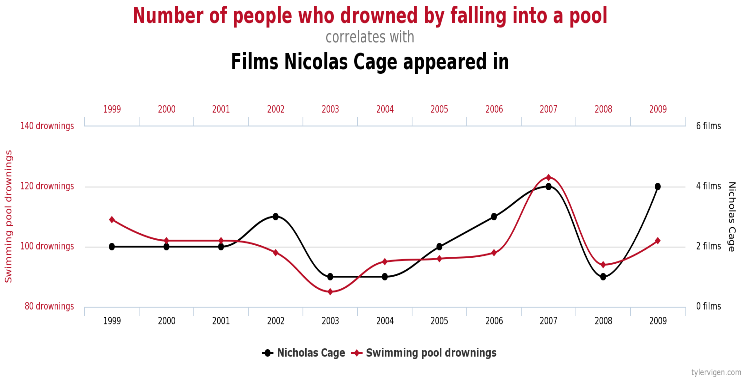
The amount of fertilizer does not have a significant effect on head weight.

R: two-sided quantile and P-value in a $t(f)$-distribution
quantile = `qt(alpha/2, f, lower.tail = FALSE)` = `qt(1 - alpha/2, f)`
P-value = `2*pt(abs(t), f, lower.tail = FALSE)`

Some relations that are absurd and irrelevant




Age of Miss America
correlates with
Murders by steam, hot vapours and hot objects
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
25 yrs
23.75 yrs
22.5 yrs
21.25 yrs
20 yrs
18.75 yrs
Age of Miss America
8 murders
6 murders
4 murders
2 murders
Murders by steam
Murders by steam
Age of Miss America
tylervigen.com


Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
140 drownings
120 drownings
100 drownings
80 drownings
Swimming pool drownings
6 films
4 films
2 films
0 films
Nicholas Cage
Nicholas Cage
Swimming pool drownings
tylervigen.com

Testing the significance of several variables at once

- When we have categorical variables with more than two categories, we replace one variable by two, or more, dummy-variables.
- How to test the significance of the original variable?
- Not t-test. That just tests one of the categories against the reference.
- Solution: Divide the variability in Y into different parts, depending on (groups of) different variables. ANOVA (ANalysis Of VAriance)
- This also works for testing several continous variables at the same time.

Variance decomposition

- Null model with only β_0: $Y_i \sim N(\beta_0, \sigma^2)$ with $\hat{Y}_i = \hat{\beta}_0 = \bar{Y}$ and

$$\hat{\sigma}^2 = s_0^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1} = \frac{\text{SS}(\text{Total}_{\text{corrected}})}{n-1}$$

- How much of this variability in Y can be explained by the linear relationship with the x-variables?
- Full model: $Y_i \sim N(\mathbf{x}_i\boldsymbol{\beta}, \sigma^2)$ with $\hat{Y}_i = \mathbf{x}_i\hat{\boldsymbol{\beta}}$
- Idea: Divide Y_i into a noice part and a signal part:

$$Y_i = \underbrace{Y_i - \hat{Y}_i}_{e_i} + \hat{Y}_i = \underbrace{e_i}_{\text{noice}} + \underbrace{\hat{Y}_i}_{\text{signal}}$$

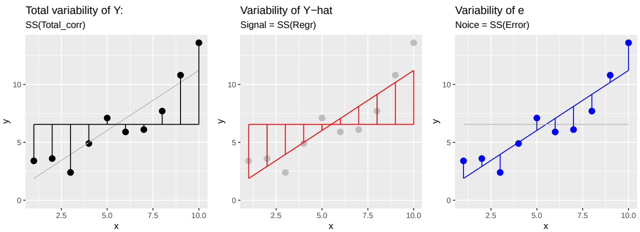
- Divide $\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ into corresponding noice and signal parts.
  If at least one $\beta_j \neq 0$, for $j = 1, \ldots, p$, the signal should be stronger than the noice.

Variance decomposition (cont.)

$$\mathrm{SS}(\text{Total}_{\text{corrected}}) = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\underbrace{e_i + \hat{Y}_i}_{Y_i} - \bar{Y})^2$$

$$= \sum_{i=1}^{n} e_i^2 + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 - \underbrace{2\sum_{i=1}^{n} e_i(\hat{Y}_i - \bar{Y})}_{=0 \text{ see next slide}}$$

$$= \underbrace{\mathrm{SS}(\text{Error})}_{\text{noice}} + \underbrace{\mathrm{SS}(\text{Regr})}_{\text{signal}}$$



(*) Proof of cross-term = 0

The Least squares estimation requires

$$\frac{\partial Q(\boldsymbol{\beta})}{\partial \beta_0} = \sum_{i=1}^{n} (Y_i - \mathbf{x}_i\boldsymbol{\beta}) = 0 \Rightarrow \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (Y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}) = 0.$$

Rewriting (see also Lecture 5)

$$\mathbf{X}\hat{\boldsymbol{\beta}} = \underbrace{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{=\mathbf{P}}\mathbf{Y} = \mathbf{P}\mathbf{Y}, \qquad \mathbf{P}' = \mathbf{P}, \qquad \mathbf{P}\mathbf{P} = \mathbf{P}$$

gives

$$\sum_{i=1}^{n} e_i\hat{Y}_i = \mathbf{e}'\hat{\mathbf{Y}} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{Y} - \mathbf{P}\mathbf{Y})'\mathbf{P}\mathbf{Y} =$$
$$= (\mathbf{Y}' - \mathbf{Y}'\mathbf{P}')\mathbf{P}\mathbf{Y} = \mathbf{Y}'\mathbf{P}\mathbf{Y} - \mathbf{Y}'\underbrace{\mathbf{P}'\mathbf{P}}_{=\mathbf{P}}\mathbf{Y} = 0$$

and thus

$$\sum_{i=1}^{n} e_i(\hat{Y}_i - \bar{Y}) = \sum_{i=1}^{n} e_i\hat{Y}_i - \bar{Y}\sum_{i=1}^{n} e_i = 0 - \bar{Y}\cdot 0 = 0.$$

If H_0 is correct

If H_0: $\beta_1 = \ldots = \beta_p = 0$ is correct then $E(Y_i) = \beta_0$ and $E(\hat{Y}_i) = E(\mathbf{x}_i\hat{\boldsymbol{\beta}}) = \mathbf{x}_i\boldsymbol{\beta} = \beta_0$.

- Using $Y_1, \ldots, Y_n$ with $\hat{\beta}_0 = \bar{Y}$ we know that

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1} = \frac{\text{SS}(\text{Total}_{\text{corrected}})}{n-1}$$

  is an unbiased estimate of σ^2 (Basic statistics course).

- Using $\hat{Y}_1, \ldots, \hat{Y}_n$ as a sample with $\hat{\beta}_0 = \bar{\hat{Y}} = \bar{Y}$ we get

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{p} = \frac{\text{SS}(\text{Regr})}{p} = \text{MS}(\text{Regr})$$

  Note: the $\hat{Y}_i$ are strongly dependent so we have to compensate by dividing by p in order to get an unbiased estimate (Advanced course in inference theory).

These two $\hat{\sigma}^2$ are only correct when H_0 is true. When H_0 is wrong they will be too large. We need another estimate that is always correct to compare with.

Full model

Using the full model with all the β-parameters we know that

$$\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-(p+1)} = \frac{\text{SS(Error)}}{n-(p+1)} = \text{MS(Error)}$$

is an unbiased estimate of σ^2 (Lecture 1). And this is true even if some (or all) of $\beta_1, \ldots, \beta_p$ happen to be 0!

Analysis of variances

The signal-to-noice comparison will use the two variance estimates $\hat{\sigma}^2 = \text{MS(Regr)} =$ Signal and $\hat{\sigma}^2 = \text{MS(Error)} =$ Noice.

Chi-squared distribution, $\chi^2(f)$

If $Z_i \sim N(0,1)$, $i = 1, \ldots, n$ are independent then

$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$$

If $Z_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, n$ are independent then

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (Z_i - \bar{Z})^2 \sim \chi^2(n-1)$$

The parameter f is called the degrees of freedom.

F-distribution, $F(f_1,\ f_2)$

If $Z_1 \sim \chi^2(f_1)$ and $Z_2 \sim \chi^2(f_2)$ are independent then

$$\frac{Z_1/f_1}{Z_2/f_2} \sim F(f_1,\ f_2)$$

Distribution of Sums of Squares

We always have that

$$\frac{1}{\sigma^2}\text{SS(Error)} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \sim \chi^2(n-(p+1))$$

If H_0 is true we also have that

$$\frac{1}{\sigma^2}\text{SS(Regr)} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \sim \chi^2(p)$$

Since they are independent it follows that, when H_0 is true,

$$\frac{\frac{1}{\sigma^2}\text{SS(Regr)}/p}{\frac{1}{\sigma^2}\text{SS(Error)}/(n-(p+1))} = \frac{\text{MS(Regr)}}{\text{MS(Error)}} \sim F(p,\, n-(p+1))$$

Global F-test

Test $H_0 : \beta_1 = ... = \beta_p = 0$ vs H_1 : at least one $\beta_j \neq 0$, $j = 1, ..., p$

- If H_0 is true then none of the x-variables are needed and the signal is just noice and $\dfrac{\text{MS(Regr)}}{\text{MS(Error)}} \approx 1$
- If H_0 is wrong a consistent part of the response variation is due to the signal and $\dfrac{\text{MS(Regr)}}{\text{MS(Error)}} \gg 1$
- If H_0 is true then:

$$F = \frac{\text{MS(Regr)}}{\text{MS(Error)}} \sim F(p,\, n-(p+1)) \qquad \text{(one-sided test)}$$

and we can reject H_0, in favour of H_1 at significance level α if

$$F > F_{\alpha,\,p,\,n-(p+1)}$$

Example: cabbage weight, soil type and fertilizer

Are any of the variables in the model necessary?
H_0: $\beta_1 = \beta_2 = \beta_3 = 0$.
In R: run the `summary()` for your model
On the last line you have
`F-statistic:  5.721 on 3 and 26 DF, p-value:  0.003814`.
This means that

- $F = \dfrac{\text{MS(Regr)}}{\text{MS(Error)}} = 5.721$
- The test statistic F follows an $F(3, 26)$-distribution if H_0 is true.
- P-value $= P(F(3,\ 26) > 5.721) = 0.003814$ if H_0 is true.

Since P-value $< 0.05 = \alpha$ we should reject H_0. Our data suggest that the model does contain at least one relevant covariate (it doesn't say which one!)

Partial F-test: testing a subset of parameters

Model: $Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$.
Question: are k specific β-parameters $= 0$
(e.g. the k last: $\beta_{p-k+1} = \ldots = \beta_p = 0$)?

Procedure:

- Estimate the full model with all $p+1$ parameters: $\mathrm{SS}(\text{Error}_{\text{full}})$ with $df = n-(p+1)$.
- Estimate the reduced model with $p+1-k$ parameters: $\mathrm{SS}(\text{Error}_{\text{reduced}})$ with $df = n-(p+1-k)$.
- Calculate the increase in $\mathrm{SS}(\text{Error})$: $Q = \mathrm{SS}(\text{Error}_{\text{reduced}}) - \mathrm{SS}(\text{Error}_{\text{full}})$ with $df = n-(p+1-k)-(n-(p+1)) = k$.
- Is this increase too large? Reject H_0 at $\alpha = 5\%$ if

$$F = \frac{Q/k}{s^2_{\text{full}}} > F_{\alpha,\, k, n-(p+1)}$$

Example: do we need the soil type(s)?

Compare the full model: $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$
against the reduced model $Y_i = \beta_0 + \beta_3 x_{i3} + \epsilon_i$.
Test H_0: $\beta_1 = \beta_2 = 0$ against H_1: "$\beta_1 \neq 0$ and/or $\beta_2 \neq 0$"
In R: `anova(model.reduced, model.full)` gives

$$F = \frac{Q/k}{s^2_{\text{full}}} = \frac{(28.427 - 18.405)/2}{18.405/26} = 7.08,$$

$$\text{P-value} = 0.0035$$

Since $F = 7.08 > F_{0.05,\,2,\,26} = 3.37$ or P-value $= 0.0035 < 0.05$ we should reject H_0. Yes, we need the soil types in the model.

Important!

- Partial F-test for comparison between a large and a reduced model can only be performed if the models are nested.
  All variables in the reduced model must be present in the full model.
- Data used must be the same for both models.

Warning: Running `anova(model)` on just one model builds the ANOVA table sequentially using Type I Sums of Squares, assigning as much of the variability as possible to the x-variable that comes first in the `lm()`-function, then it assigns as much of the remaining variability as possible to the second x-variable, etc. This is not what we want!
We should use Type II in models without interaction and Type III in models with interaction. This will automatically be the case when using `anova(model.reduced, model.full)`.

Note:

- A t-test is for the significance of a single β-parameter, given all others in the model.
- An F-test tests several β-parameters at the same time but does not say which β_j's are significantly $\neq 0$.
- An F-test is always one-sided because a larger model can never explain less than a reduced model.

R:
quantile $F_{\alpha,\, f_1,\, f_2} =$
= `qf(alpha, f1, f2, lower.tail = FALSE)` = `qf(1 - alpha, f1, f2)`,
P-value = `pf(F, f1, f2, lower.tail = FALSE)`