TMA4267 Linear Statistical Models V2017 (L10)

Part 2: Linear regression: Parameter estimation [F:3.2], Properties of residuals and distribution of estimator for error variance Confidence interval and hypothesis for one regression coefficient

Mette Langaas

Department of Mathematical Sciences, NTNU

To be lectured: February 17, 2017

Today

1. Properties for residuals (from the hat matrix), leading to properties for $\hat{\sigma}^2$,
2. Then, confidence interval and hypothesis test for regression coefficient.

The classical linear model

The model

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

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

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

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

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

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

Results so far

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

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

with mean $\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and $\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.
- Restricted maximum likelihood estimator for $\boldsymbol{\sigma}^2$:

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

- Projection matrices: idempotent, symmetric/orthogonal:

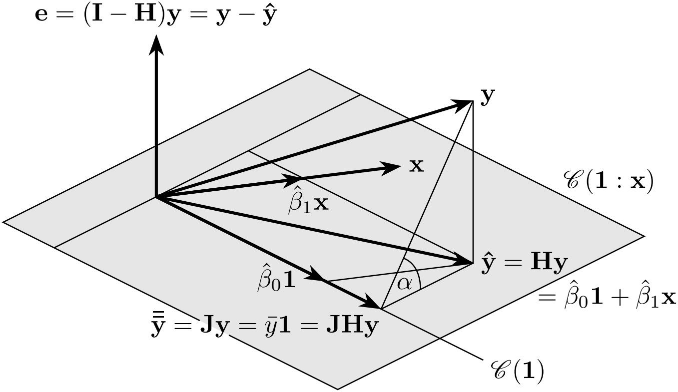
$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$$

projects onto column space of $\boldsymbol{X}$

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

projects onto space orthogonal to column space of $\boldsymbol{X}$

with important connection: predictions $\hat{\boldsymbol{Y}} = \boldsymbol{H}\boldsymbol{Y}$ and residuals $\hat{\varepsilon} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$



Putanen, Styan and Isotalo: Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty, Figure 8.3.

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 $\boldsymbol{R}\boldsymbol{S} = \boldsymbol{0}$.

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

Properties: $\hat{\beta}$ and $\hat{\sigma}^2$

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

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

has mean $\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and $\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

- In addition $\hat{\boldsymbol{\beta}}$ is best linear unbiased estimator (BLUE), that is, among all unbiased estimator it has minimum variance in each component. (More in TMA4295 Statistical Inference.)
- For the normal model: $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$.
- Restricted maximum likelihood estimator for $\boldsymbol{\sigma}^2$:

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

- For the normal model

$$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}$$

Acid rain in Norwegian lakes

Measured pH in Norwegian lakes explained by content of

- 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),

Random sample of $n = 26$ lakes.

Output from fitting the full model in R

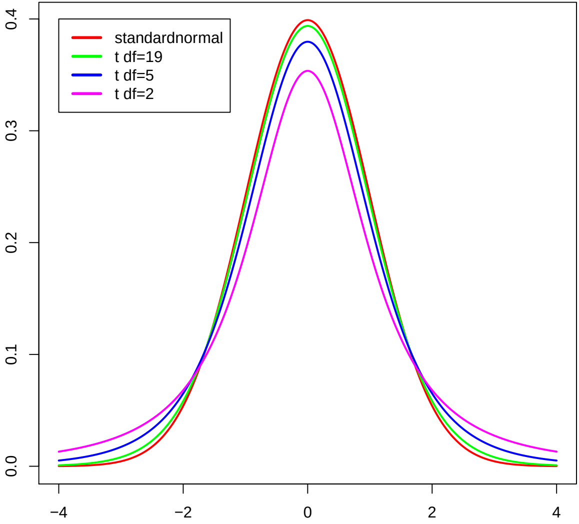
```
> fit=lm(y~.,data=ds)
> summary(fit)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.6764334  0.1389162  40.862  < 2e-16 ***
x1          -0.3150444  0.0587512  -5.362 4.27e-05 ***
x2          -0.0018533  0.0012587  -1.472    0.158
x3           0.9751745  0.1449075   6.730 2.62e-06 ***
x4          -0.0002268  0.0010038  -0.226    0.824
x5          -0.0334242  0.0225009  -1.485    0.155
x6          -0.0039399  0.0724339  -0.054    0.957
x7           0.0888722  0.1025724   0.866    0.398
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1165 on 18 degrees of freedom
Multiple R-squared:   0.93,Adjusted R-squared:  0.9027
F-statistic: 34.15 on 7 and 18 DF,  p-value: 3.904e-09
```

W. S. Gosset alias Student

Historical: Student-t fordelingen

- W.S. Gosset (1876-1937) was employed by the Guinness Brewing Company of Dublin.
- Sample sizes available for experimentation in brewing were necessarily small, and Gosset knew that a correct way of dealing with small samples was needed.
- He consulted Karl Pearson (1857-1936) of Universiy College in London about the problem. Pearson told him the current state of knowledge was unsatisfactory.
- The following year Gosset undertook a course of study under Pearson. An outcome of his study was the publication in 1908 of Gosset's paper on "The Probable Error of a Mean," which introduced a form of what later became known as Student's t-distribution.
- Gosset's paper was published under the pseudonym "Student."
- The modern form of Student's t-distribution was derived by R.A. Fisher and first published in 1925.

t-distribution



DEF: t-distribution

Let Z be a standard normal random variable and V a chi-squared random variable with parameter ν (degrees of freedom). If Z and V are independent, the distribution of the random variable T

$$T = \frac{Z}{\sqrt{V/\nu}}$$

has probability density function

$$h(t) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)\sqrt{\pi\nu}}(1+\frac{t^2}{\nu})^{-(\nu+1)/2}$$

for $-\infty < t < \infty$. This distribution is called the (Student) $t-$distribution with ν degrees of freedom.

- $\mathrm{E}(T) = 0$ if $\nu \geq 2$.
- $\mathrm{Var}(T) = \frac{\nu}{\nu-2}$ if $\nu \geq 3$.

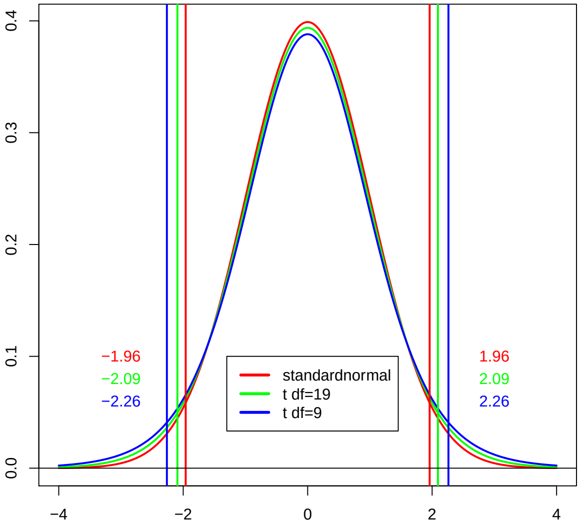
Are $\hat{\beta}$ and SSE are independent?

Independence – from Part 1:
Let $\boldsymbol{X}_{(p\times 1)}$ be a random vector from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. 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 $\boldsymbol{A}\sigma^2\boldsymbol{IB}^T = \sigma^2\boldsymbol{AB}^T = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{0}$
- 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.

Quantiles and critical values: N og t: $\alpha/2 = 0.025$



Kritiske verdier i t-fordelingen

$$P(T > t_{\alpha,\nu}) = \alpha$$

$\nu \backslash \alpha$	.150	.100	.075	.050	.025	.010	.005	.001	.0005
1	1.963	3.078	4.165	6.314	12.706	31.821	63.657	318.309	636.619
2	1.386	1.886	2.282	2.920	4.303	6.965	9.925	22.327	31.599
3	1.250	1.638	1.924	2.353	3.182	4.541	5.841	10.215	12.924
4	1.190	1.533	1.778	2.132	2.776	3.747	4.604	7.173	8.610
5	1.156	1.476	1.699	2.015	2.571	3.365	4.032	5.893	6.869
6	1.134	1.440	1.650	1.943	2.447	3.143	3.707	5.208	5.959
7	1.119	1.415	1.617	1.895	2.365	2.998	3.499	4.785	5.408
8	1.108	1.397	1.592	1.860	2.306	2.896	3.355	4.501	5.041
9	1.100	1.383	1.574	1.833	2.262	2.821	3.250	4.297	4.781
10	1.093	1.372	1.559	1.812	2.228	2.764	3.169	4.144	4.587
11	1.088	1.363	1.548	1.796	2.201	2.718	3.106	4.025	4.437
12	1.083	1.356	1.538	1.782	2.179	2.681	3.055	3.930	4.318
13	1.079	1.350	1.530	1.771	2.160	2.650	3.012	3.852	4.221
14	1.076	1.345	1.523	1.761	2.145	2.624	2.977	3.787	4.140
15	1.074	1.341	1.517	1.753	2.131	2.602	2.947	3.733	4.073
16	1.071	1.337	1.512	1.746	2.120	2.583	2.921	3.686	4.015
17	1.069	1.333	1.508	1.740	2.110	2.567	2.898	3.646	3.965
18	1.067	1.330	1.504	1.734	2.101	2.552	2.878	3.610	3.922
19	1.066	1.328	1.500	1.729	2.093	2.539	2.861	3.579	3.883
20	1.064	1.325	1.497	1.725	2.086	2.528	2.845	3.552	3.850
21	1.063	1.323	1.494	1.721	2.080	2.518	2.831	3.527	3.819
22	1.061	1.321	1.492	1.717	2.074	2.508	2.819	3.505	3.792
23	1.060	1.319	1.489	1.714	2.069	2.500	2.807	3.485	3.768
24	1.059	1.318	1.487	1.711	2.064	2.492	2.797	3.467	3.745
25	1.058	1.316	1.485	1.708	2.060	2.485	2.787	3.450	3.725
26	1.058	1.315	1.483	1.706	2.056	2.479	2.779	3.435	3.707
27	1.057	1.314	1.482	1.703	2.052	2.473	2.771	3.421	3.690
28	1.056	1.313	1.480	1.701	2.048	2.467	2.763	3.408	3.674
29	1.055	1.311	1.479	1.699	2.045	2.462	2.756	3.396	3.659
30	1.055	1.310	1.477	1.697	2.042	2.457	2.750	3.385	3.646
35	1.052	1.306	1.472	1.690	2.030	2.438	2.724	3.340	3.591
40	1.050	1.303	1.468	1.684	2.021	2.423	2.704	3.307	3.551
50	1.047	1.299	1.462	1.676	2.009	2.403	2.678	3.261	3.496
60	1.045	1.296	1.458	1.671	2.000	2.390	2.660	3.232	3.460
80	1.043	1.292	1.453	1.664	1.990	2.374	2.639	3.195	3.416
100	1.042	1.290	1.451	1.660	1.984	2.364	2.626	3.174	3.390
120	1.041	1.289	1.449	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.036	1.282	1.440	1.645	1.960	2.326	2.576	3.090	3.291

Acid rain in R

```
ds=read.table("https://www.math.ntnu.no/emner/
TMA4267/2017v/acidrain.txt",header=TRUE)
fit=lm(y~.,data=ds)
> confint(fit)
                   2.5 %       97.5 %
(Intercept)  5.384581378  5.9682854281
x1          -0.438476153 -0.1916126966
x2          -0.004497716  0.0007911594
x3           0.670735075  1.2796138706
x4          -0.002335625  0.0018820903
x5          -0.080696921  0.0138484550
x6          -0.156117992  0.1482381575
x7          -0.126624544  0.3043688780
```

P-values: `http://www.statistrikk.no/wp-content/uploads/2017/02/nerdekort.jpg`

Today

- Distribution of SSE/σ^2 is chisquared $(n-p)$.
- Independence of $\hat{\boldsymbol{\beta}}$ and SSE.
- Inference about $\boldsymbol{\beta}$ components can be performed using the t-distribution