Statistical hypothesis tests

remember: t test to compare group means

hypothesis → $H_0: \mu_{male} = \mu_{female}$

alternative (hypothesis) → $H_1: \mu_{male} \neq \mu_{female}$

data example:

```
> t.test(wage~gender)

        Welch Two Sample t-test

data:  wage by gender
t = 4.8853, df = 530.55, p-value = 1.369e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.265164 2.966949
sample estimates:
  mean in group male mean in group female
            9.994913             7.878857
```

decision based on p value

mean for males

mean for females

here: p value = 0.00000 1369 = 0.0001369 %

< 5 %

⇒ we reject the hypothesis (at level 5 %)

(we would also reject at 1 % or 0,1 % or 10 %)

other test (with t.test in R)

$H_0: \mu = 9$ ← expected wage equals 9 US $

$H_1: \mu \neq 9$

⇒ one sample t test

```
> t.test(wage,mu=9)

        One Sample t-test

data:  wage
t = 0.1082, df = 533, p-value = 0.9139
alternative hypothesis: true mean is not equal to 9
95 percent confidence interval:
 8.587194 9.460933
sample estimates:
mean of x
 9.024064
```

$\Rightarrow$ (we accept H_0)

better: we do not reject H_0!

"political correct formulation"

(decision from data)

(truth)		H_0	H_1
	H_0	OK	α (1st type error in the decision)
	H_1	β (2nd type error in the decision)	OK

- it is complicated to minimize α and β simultaneously

$\Rightarrow$ predefine α typically 5% (or 10% or 1%)

$\Rightarrow$ if we reject then we have either an error α or everything is OK

$\Rightarrow$ if we reject H_0 then our maximal error is α

(as if we do not reject H_0 : we do not know the size of β)

Statistical tests in regression

- t tests for the coefficients β_j

Coefficient of the jth variable

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

we could also test another value than 0, but this is the most interesting case

$\Rightarrow$ in order to perform such a test we need some more assumptions on ε_i (or on our model)

$\Rightarrow$ normality assumption for ε

(or later on: large sample size)

Normality Assumption

Until now we assumed:

$$E\varepsilon = \mathbf{0}\,, \quad \mathrm{Var}(\varepsilon) = \sigma^2\mathbf{I}$$

Let us here assume that the ε_i are

$$\varepsilon_i \sim N(0, \sigma^2) \text{ iid} \quad \Longleftrightarrow \quad \varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I})$$

additional
assumption : normality of ε_i

$\Longrightarrow$ it follows (slide 39)

Distribution of the LSE under Normality

Theorem:

Under the assumption $\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ hold:

(i) $\widehat{\beta} \sim N\left(\beta, \sigma^2(\mathcal{X}^\top\mathcal{X})^{-1}\right)$

(ii) $\widehat{\beta}$ (and therefore also $\widehat{\mathbf{Y}} = \mathcal{X}\widehat{\beta}$) is independent from $\mathrm{RSS} = \mathbf{Y}^\top(\mathbf{I} - \mathbf{P})\mathbf{Y}$

(iii) $\frac{1}{\sigma^2}\mathrm{RSS} \sim \chi^2_{n-p-1}$ ← chi square distribution with $n-(p+1)$ degrees of freedom

(vi) $\frac{\widehat{\beta}_j - \beta_j}{\widehat{\sigma}\sqrt{c_{jj}}} \sim t_{n-p-1}$ for each $j = 0, \ldots, p$

(where c_{jj} ist the jth diagonal element of the matrix $(\mathcal{X}^\top\mathcal{X})^{-1}$)

now $\hat{\beta}$ has a normal distribution and we need that distribution to establish a test for $\hat{\beta}$ or for each single $\hat{\beta}_j$

this is the test statistic that we are going to use for our t tests now

→ t test because a t_{n-p-1} distribution is used

→ we see that we also have to estimate σ (or σ^2)

Going back to our test problem:

$$H_0 : \beta_j = 0$$ hypothesis (null hypothesis)

$$H_1 : \beta_j \neq 0$$ alternative (alternative hypothesis)

$\Rightarrow$ Why is this test interesting for us?

remember the model (multiple linear)

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_j X_j + \dots + \beta_p X_p + \varepsilon$$

if $\beta_j = 0$

$\Rightarrow$ the term $\beta_j X_j$ is not relevant

$\Rightarrow$ no need to consider variable X_j

test statistic

known (estimated)

absolute value

if $H_0 : \beta_j = 0$

$$\left| \frac{\hat{\beta}_j - 0}{\hat{\sigma} \underbrace{\sqrt{c_{jj}}}_{\text{known (see slide 39)}}} \right| = \left| \frac{\hat{\beta}_j}{\hat{\sigma} \sqrt{c_{jj}}} \right|$$

to be defined: estimated via RSS

We reject the hypothesis H_0

if $$\underbrace{\left| \frac{\hat{\beta}_j}{\hat{\sigma} \sqrt{c_{jj}}} \right|}_{\text{reject if this is "to large"}} > c_\alpha$$

critical value determined by the $t_{n-(p+1)} = t_{n-p-1}$ distribution estimation of $\hat{\sigma}$:

$$\hat{\sigma}^2 = \frac{RSS}{n-\underbrace{(p+1)}} = \frac{RSS}{n-p-1}$$

Sample size

no. of all Coefficients (incl. intercept)

$\Longrightarrow$ please see Slides 32 - 34

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$

hat matrix :

$$P = X\,(X^T X)^{-1} X^T$$

called P because it is a projection matrix : $P^2 = P$

$(\Rightarrow P^k = P)$

$\Longrightarrow (I-P)^2 = I-P$

$(\Rightarrow (I-P)^k = I-P)$

t test in practice

```
> summary(lm1)
```

(see : CPS1985_ttests.R)

```
Call:
lm(formula = log(wage) ~ education + experience + I(experience^2))

Residuals:
     Min       1Q   Median       3Q      Max 
-2.12709 -0.31543  0.00671  0.31170  1.98418 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      0.5203218  0.1236163   4.209 3.01e-05 ***
education        0.0897561  0.0083205  10.787  < 2e-16 ***
experience       0.0349403  0.0056492   6.185 1.24e-09 ***
I(experience^2) -0.0005362  0.0001245  -4.307 1.97e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4619 on 530 degrees of freedom
Multiple R-squared:  0.2382,	Adjusted R-squared:  0.2339 
F-statistic: 55.23 on 3 and 530 DF,  p-value: < 2.2e-16
```

4 t tests

→ one t test for each coefficient :

(1) $H_0 : \beta_0 = 0$ vs. $H_1 : \beta_0 \neq 0$

⋮

(4) $H_0 : \beta_3 = 0$ vs. $H_1 : \beta_3 \neq 0$

⇒ all p values are very small

⇒ we reject all 4 hypotheses at 5% (also at 1% or 10% ...)

in our model : $p+1 = 4 \Rightarrow p = 3$

$n = 534 \Rightarrow n-(p+1) = 534 - 4 = 530$

→ all coefficients $\beta_0, \beta_1, \beta_2, \beta_3$ are significantly different from 0 at 5% (1%, 10%)

recall $H_0 : \beta_j = 0$

$\Rightarrow$ if H_0 is not rejected:

β_j is not significantly different from 0

$\Rightarrow$ variable X_j may not be not be relevant in the model

if H_0 is rejected:

(we are interested in rejecting H_0!)

β_j is significantly different from 0

$\Rightarrow$ variable X_j is relevant!

Note: the numerical value of $\hat{\beta}_j$ does not matter in general, for its significance it is the p value to look at