TMA4267 Linear statistical models

Part 3: Hypothesis testing and ANOVA

March 14, 2017

Happiness

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.072081	0.852543	-0.085	0.9331
money	0.009578	0.005213	1.837	0.0749
sex	-0.149008	0.418525	-0.356	0.7240
love	1.919279	0.295451	6.496	1.97e-07
work	0.476079	0.199389	2.388	0.0227

For which covariates would we reject the null hypothesis $\beta = 0$ at significance level 1%?

A money

B sex

C love

D work

Type I errors

What is a commonly used name for the type I errors?

A true positives

B false positives

C false negatives

D true negatives

Linear hypotheses

$H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}$ in a regression model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.
n=number of observations,
p = number of estimated regression coefficients
r=number of linear hypotheses (rank of $\boldsymbol{C}$).

What is the distribution of F_{obs}
$= \frac{1}{r}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d})^T(\hat{\sigma}^2\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{C}^T)^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d})$?

A $F_{r,n-p}$

B $F_{p,n-r}$

C $N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$

D $N(0, \sigma^2\boldsymbol{I})$

ANOVA

Which type of covariate coding is used in the one-way ANOVA model with design matrix given as:

```
1    1    0    0    0
1    0    1    0    0
1    0    1    0    0
1    0    0    1    0
1    0    0    0    1
1    0    0    0    1
1   -1   -1   -1   -1
1   -1   -1   -1   -1
```

A Continuous

B Effect coding

C Dummy variable coding

D Categorical

ANOVA

Is the interaction term significant at significance level 0.01?

```
> res <- lm(Words~Age*Process)
> anova(res)
             Df  Sum Sq Mean Sq F value    Pr(>F)
Age           1  240.25  240.25 29.9356 3.981e-07 ***
Process       4 1514.94  378.74 47.1911 < 2.2e-16 ***
Age:Process   4  190.30   47.58  5.9279 0.0002793 ***
Residuals    90  722.30    8.03
```

A Yes

B Not enough information to decide

C No

p-value from true null hypothesis

For a continuous test statistic that gives an exact p-value, what is the distribution the p-value when the null hypothesis is true?

A Normal

B Chisquared

C Exponential

D Uniform

FWER

V=number of false positives and

R=number of rejections.

The familywise error rate FWER is

A $E(V/R)$

B $E(V)$

C $P(V/R > 0.05)$

D $P(V > 0)$

Bonferroni

α=level for control of FWER.

α_{loc}=cut-off on p-value

m =number of tests.

What is the Bonferroni rule?

A $\alpha_{\text{loc}} = m\alpha$

B $\alpha_{\text{loc}} = \frac{\alpha}{m}$

C $\alpha_{\text{loc}} = \alpha^m$

D $\alpha_{\text{loc}} = (1-\alpha)^{1/m}$

Correct?

Are you sure you want to read the correct answers? Maybe try first? The answers are explained on the next two slides.

Answers

1. C: only love is significant on level 1%, since this is the only p-value below 0.01 (last column).
2. B: type I errors are called false positive findings
3. A: linear hypotheses with $F_{r,n-p}$-distributed statistic.
4. B: Effect coding is used in ANOVA.

Answers

5. A: Interaction term has p-value below 0.01.
6. D: p-values from true nulls are uniform.
7. D: FWER is the probability of one or more false positives.
8. B: Bonferroni rule is α/m.