Statistical Data Analysis

Dr. Jana de Wiljes

14. Dezember 2021

Universität Potsdam

Hypothesis testing for Regression parameters

eg $R_1 = (0, 1, 0 \dots 0)$ $\quad R_1 \cdot \beta = \begin{pmatrix} 0 \\ \beta_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

$\beta_2 = 0 \quad \beta_3 = 0$

$x_1 \quad \cancel{x_2} \quad \cancel{x_3} \quad x_4 \quad x_5$

$r = 0$

In general, the following hypotheses can be tested:

1. $H_0 : \mathbf{R}_1\beta = r$ vs. $H_1 : \mathbf{R}_1\beta \neq r$
2. $H_0 : \mathbf{R}_1\beta \geq r$ vs. $H_1 : \mathbf{R}_1\beta < r$
3. $H_0 : \mathbf{R}_1\beta \leq r$ vs. $H_1 : \mathbf{R}_1\beta > r$

Under H_0:

$$\mathbf{R}_1\hat{\beta} \overset{H_0}{\sim} \mathcal{N}(r, \sigma^2 \mathbf{R}_1(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{R}_1^\top) \tag{1}$$

holds. For unknown σ^2, a reasonable test-statistic is

grade

$$T = \frac{\mathbf{R}_1\hat{\beta} - r}{\hat{\sigma}\sqrt{\mathbf{R}_1(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{R}_1^\top}} \sim t_{n-p-1} \tag{2}$$

quantile grade

α

$\cancel{\beta \leq 1}$

$R_1\beta \geq 1$

The corresponding rejection areas are:

1. $|T| > t_{1-\alpha/2, n-p-1}$
2. $T < t_{1-\alpha, n-p-1}$
3. $T > t_{1-\alpha, n-p-1}$

α 0.01 0.05

$(1-\alpha)$-confidence intervals for $\mathbf{R}_1\hat{\beta}$ are:

$$\mathbf{R}_1\hat{\beta} \pm t_{n-p-1, 1-\alpha/2}\,\hat{\sigma}\sqrt{\mathbf{R}_1(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{R}_1^\top} \tag{3}$$

z_i)

Hypothesis for general parameter identification problems

Setting:

1. Let $(\mathbb{P})_{\theta \in \Theta}$ a family of probability measures on the sample space $(\mathcal{X}, \mathcal{A})$.
2. Find disjunct subsets Θ_1 and Θ_2 of parameter space $\boxed{\Theta = \Theta_1 \cup \Theta_2}$ and $\boxed{\Theta_1 \cap \Theta_2 = \emptyset}$
3. θ unknown

$\boxed{1 \quad 2}$

$\boxed{0 \quad 1}$

Hypothesis:

θ_0

1. Null hypothesis H_0: $\theta \in \Theta_0$ 1
2. alternative hypothesis H_1: $\theta \in \Theta_1$ 2

$\{3\} \cup \{\Theta \setminus \{3\}\}$

Example

Flipping a coin

$\text{Bin}(n, \theta)$, θ unknown

Question: Is the coin fair?

$\Theta = (0, 1)$

$\rightarrow$ $\boxed{p = 0.5}$ for "heads"?

$\Theta = \Theta_0 \cup \Theta_1 = \{\frac{1}{2}\} \cup \Theta \setminus \{1/2\}$

$\boxed{n = 200}$ $\rightsquigarrow$ sample set: S random variable (How many times we observe heads)

Null hypothesis H_0: Coin is fair: $\theta = \frac{1}{2}$

Alternative hypothesis H_1: Coin <u>not</u> fair d.h. $\theta \neq 1/2$

$$\mathbb{E}_{H_0}[S] = 200 \cdot \frac{1}{2} = 100$$

Idea: · $|S - 100|$ is large $\rightsquigarrow$ maybe our Null Hypothesis should be rejected

· Need to choose a constant $c \in \{0, 1, \dots\}$ so that we reject H_0 if

$\boxed{|S - 100| > c}$

0.01

$\boxed{0.05}$

Note: There are two types of errors:

$\rightarrow$ Error 1: H_0 is rejected, despite it being true $\Leftarrow$

$\rightarrow$ Error 2: H_0 is <u>not</u> rejected, but H_0 is <u>not</u> true

$$\boxed{\mathbb{P}_{H_0}[|S - 100| > c]} = 2 \cdot \mathbb{P}_{H_0}[S > 100 + c] = \boxed{2 \sum_{k=100+c+1}^{200} \binom{200}{k} \boxed{\frac{1}{2^{200}}}}$$

α

$S \sim \text{Bin}(200, 1/2)$ (under H_0 is true)

Example

$$P_{H_0}[|S-100|>c] = \begin{cases} \boxed{0.05896} & \boxed{c=13} \\ \boxed{0.04003} & \boxed{c=14} \end{cases}$$

$\rightsquigarrow$ we reject H_0 if $\boxed{|S-100|}>14$

$$\boxed{\theta = 1/2 + 10^{-10}}$$

$$\boxed{\mathbb{P}_{H_0}[|S-100| \geq |\underbrace{150}_{}-100|]} = 2\mathbb{P}_{H_0}[S \geq 150] = 2\sum_{k=150}^{200}\binom{200}{k}\frac{1}{2^{200}}$$

$$\approx \boxed{8.393 \cdot 10^{-13}}$$

$$\boxed{p = 0.01}$$

Example

$\boxed{X_1, \ldots, X_n \sim N(\mu, \sigma^2)}$

case: want to test for μ with σ^2 is known (Gauß-Test)

Hypothesis: H_0 $\mu = \mu_0$ vs H_1: $\mu \neq \mu_0$

H_0 $\mu \geq \mu_0$ vs H_1: $\mu < \mu_0$

H_0 $\mu \leq \mu_0$ vs H_1: $\mu > \mu_0$

mean

$$T := \sqrt{n}\, \frac{\overline{X}_n - \mu_0}{\sigma_0}$$

$\boxed{\mu = \mu_0}$ (H_0 is true) $T \sim N(0,1)$

H_0 is rejected if $|T| > \boxed{z_{1-\frac{\alpha}{2}}}$

Quantile of Gauss density

choose: $\alpha = c$

$\boxed{\alpha = 0.01}$

Test for μ with σ^2 unknown

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

$$\boxed{T := \sqrt{n}\, \frac{\overline{X}_n - \mu_0}{S_n} \sim t_{n-1}}$$

Example

case: $\sigma_1^2 = \sigma_2^2$ and $\boxed{\mu_1 \text{ and } \mu_2 \text{ are unknown}}$

S_X^2, S_Y^2

$X_1, \dots X_n \sim N(\mu_1, \sigma_1^2)$

$Y_1, \dots, Y_n \sim N(\mu_2, \sigma_2^2)$

$\rightsquigarrow$ F - test

$$T := \frac{S_X^2}{S_Y^2} \sim F_{n-1, m-1}$$

$\boxed{H_0 : \sigma_1^2 = \sigma_2^2}$

$$\boxed{\begin{array}{l} T < F_{n-1, m-1, \frac{\alpha}{2}} \\ T < F_{n-1, m-1, 1-\frac{\alpha}{2}} \end{array}}$$

Neyman-Pearson-Theory

Setting: $\Theta_0 = \{\theta_0\}$, $\Theta_1 = \{\theta_1\}$, $\Theta = \{\Theta_0, \Theta_1\}$

Assumption: The associate probability measures $\mathbb{P}_{\theta_0}$ and $\mathbb{P}_{\theta_1}$ have densities h_0 and h_1 for a measure λ on $(\mathcal{X}, \mathcal{A})$

Def: Let $k \in [0, \infty]$ and $\gamma \in [0, 1]$. A likelihood-quotient-test (LQ-test) is of the form

$$\phi(x) = \begin{cases} 1, & \text{if } \frac{h_1(x)}{h_0(x)} > k \\ 1, & \text{if } \frac{h_1(x)}{h_0(x)} < k \\ \gamma, & \text{if } \frac{h_1(x)}{h_0(x)} = k. \end{cases} \tag{4}$$

Neyman-Pearson Lemma

Lemma: Let ϕ be a LQ-test with $\mathbb{E}_{\theta_0}[\phi(X)] = \alpha$. Then

$$\mathbb{E}_{\theta_1}[\phi(X)] = \sup_{\psi:\mathbb{E}_{\theta_0}[\psi(X)]\leq\alpha} \mathbb{E}_{\theta_1}[\psi(X)] \tag{5}$$

Further for every $\alpha \in (0, \infty)$ it is possible to find $k \in [0, \infty]$ and $\gamma \in [0, 1]$ so that for a predefined Test ϕ

$$\mathbb{E}_{\theta_0}[\phi(X)] = \alpha \tag{6}$$

Proof/Example

Proof/Example

Proof/Example

Proof/Example

Regularization

Ridge Regularization (L_2)

$$\hat{\beta}^{Ridge} = \underset{\beta \in \mathbb{R}^p}{\arg\min} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_2^2 \tag{7}$$

- decreases variance but increases bias (for increasing λ)
- Can improve predictive performance
-

$$\hat{\beta}^{Ridge} = (\mathbf{X}^\top \mathbf{X} + \lambda I_p)^{-1} \mathbf{X}^\top \mathbf{y} \tag{8}$$

p+1

hyper parameter

Identity $\mathbb{R}^{p \times p}$

$\beta \in \mathbb{R}^p$

$\beta \in \mathbb{R}^{p+1}$

β_0
$\vdots$
β_{p-1}

Lasso Regularization (L_1)

likelihood prior

$$\hat{\beta}^{Lasso} = \arg\min_{\beta\mathbb{R}} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda\|\beta\|_1 \tag{9}$$

- LASSO=Least Absolute Shrinkage and Selection Operator
- This penalty allows coefficients to shrink towards exactly zero
- LASSO usually leads to sparse models, that are easier to interpret

$Y = x_1 \quad x_2 \quad x_3 \quad x_4 \quad \ldots$

Y_{snow} temperature