MSP_8_cviko

neděle 3. listopadu 2024 10:05

Assume that the observations
11,5; 6,7; 7,8; 7,1; 10,3; 12,2; 11,4; 12,6; 9,8; 10,9 follow a normal distribution with unknown parameters μ, σ.

- Compute 95% CI's for both μ and σ.
- Using $\alpha = 5\%$ test a hypothesis that $H_0 : \mu = 11$ against $H_1 : \mu \neq 11$.
- Derive a test for $H_0 : \sigma^2 = \sigma_0^2$.
- Test $H_0 : \sigma^2 = 2$ against $H_1 : \sigma^2 \neq 2$ using $\alpha = 1\%$

$\bar{X} = 10{,}03 \quad s^2(x) = 4{,}542$

$s(x) = \sqrt{\frac{\wedge}{n-1} \dots} \doteq 2{,}131$

$\mu \in \langle 8{,}505 ; 11{,}555 \rangle$

$\sigma^2 \in \langle 2{,}149 ; 15{,}140 \rangle$

$\sigma \in \langle 1{,}466 ; 3{,}891 \rangle$

$t = -1{,}44 \quad t_{0{,}975}(9) \doteq 2{,}262$

$t \in \overline{W}_\alpha \quad \overline{W}_\alpha = \langle -2{,}262 ; 2{,}262 \rangle$

$\Rightarrow$ NEZAMÍTÁM

1) Rozptyl nezávisí na μ

2) pro $N(\mu;1)$ platí $\frac{(n-1)\,S^2(X)}{1} \sim \chi^2(n-1)$

3) pro $N(\mu;\sigma_0^2) \rightarrow \frac{(n-1)\cdot S^2(X)}{\sigma_0^2} \sim \chi^2(n-1)$

$\Rightarrow$ KDYŽ PLATÍ H_0 $\quad t = \frac{(n-1)S^2(X)}{\sigma_0^2} \sim \chi^2(n-1)$

$t = 20{,}44 \quad \overline{W}_{0{,}01} = \langle 1{,}735 , 23{,}589 \rangle$

$t \in \overline{W} \Rightarrow$ NEZ.

You are a doctor and want to determine if a new medication reduces blood pressure in patients. You have a group of patients whose blood pressure was measured before and after taking the medication. You want to compare their results before and after the treatment to see if there was a reduction in blood pressure. For any hypothesis testing use $\alpha = 5\%$

ID	1	2	3	4	5	6	7	8	9	10
Before	150	160	145	155	148	152	149	153	157	151
After	147	157	145	154	147	150	148	151	155	149

Table: Blood Pressure Measurements Before and After Treatment

Compute descriptive statistics for both datasets.

$\bar{B} = 152 \quad \bar{A} = 150{,}3$

$S^2(B) = 19{,}78 \quad S^2(A) = 15{,}34$

$S(B) = 4{,}45 \quad S(A) = 3{,}92$

$\bar{D} = 1{,}7$

$S(D) = 0{,}949$

$H_0 : \mu_d = 0 \qquad D = B - A$

$H_1 : \mu_d > 0$

$t = 5{,}67 \qquad \overline{W}_\alpha = (-\infty ; 1{,}833 \rangle$

Use Central limit theorem (wiki, Python illustration) to derive a test of a hypothesis about equality of probability parameters in 2 different sets of Bernoulli trials ($H_0 : \pi_x = \pi_y$, use joint standard deviation). Imagine that the probability of Bernoulli trials corresponds to the proportion of defective parts manufactured on 2 different manufacturing stations. You observed 11 defects out of 211 at station x and 23 out of 303 at station y.

- Test a hypothesis $H_0 : \pi_x = \pi_y$ against $H_1 : \pi_x \neq \pi_y$ for $\alpha = 5\%$
- Derive a likelihood ratio test for the same H_0

1) CLT and $H_0 \rightarrow \sum$ of Bernoulli trials with π

$$\frac{\sum x_i}{n_x} - \frac{\sum y_i}{n_y} = \hat{\pi}_x - \hat{\pi}_y \rightarrow 0 \quad (\text{podle } H_0)$$

$E((X - E(X))^2) = \pi(1-\pi) \rightarrow$ Bernoulli trials Variance

$$\frac{\sum x_i + \sum y_i}{n_x + n_y} = f = \hat{\pi}$$

Linear model $\mathbb{X} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \quad b = \begin{pmatrix} \hat{\pi}_x \\ \hat{\pi}_y \end{pmatrix} \quad \mathbb{V} = \sigma^2 \mathbb{I}$

with known σ^2 $\quad \dfrac{\frac{\sum x_i}{n_x} - \frac{\sum y_i}{n_y}}{\sigma \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \overset{as.}{\sim} N(0;1)$; for Bernoulli trials $\sigma^2 = \pi(1-\pi)$

$\rightarrow$ use joint estimate f

po úpravě:

$$t = \frac{\hat{\pi}_x - \hat{\pi}_y}{\sqrt{f(1-f)}} \cdot \sqrt{\frac{n_x \cdot n_y}{n_x + n_y}} = -1{,}07$$

$t \in \overline{W} \Rightarrow$ NEZ.

$\overline{W}_{0{,}05} = \langle -1{,}96 ; 1{,}96 \rangle$

$$l(\pi) = \sum x_i \ln(\pi) + \sum (1 - x_i) \ln(1-\pi)$$

$$LR = 2\Big[\Big(\sum_i x_i \ln(\bar{x}) + (n_x - \sum_i x_i)\ln(1-\bar{x}) + \sum_j y_j \ln(\bar{y}) + (n_y - \sum y_i)\ln(1-\bar{y})\Big) -$$

$$- \Big(\sum_i x_i + \sum_j y_j\Big)\ln(f) + \Big((n_x - \sum_i x_i) + (n_y - \sum_j y_j)\Big)\ln(1-f)\Big) = \dots$$

Find if there is a statistically significant difference ($\alpha = 5\%$) among mean test scores of high school graduation exams in different regions of Czechia in 2023. (data are randomly generated). You only obtained self reported subset of scores from each region.

region	JM	O	V	Z
no. of obs.	199	129	60	~~113~~ 115
mean score	74, 81	72, 11	71, 00	72, 61
sum fo squares	1123833, 76	677710, 65	305452, 98	600399, 20

Assume that all necessary assumption for ANOVA hold. Arange your results into ANOVA table.

$n = 503 \quad \bar{x} = 73{,}1623 \quad \sum_{i=1}^{n} x_i^2 = 2707395{,}61$

$S_{JM} = 540{,}27$

$S_O = 142{,}85$

$S_V = 280{,}53$

$S_Z = 34{,}47$

	S	df	MS	F-value
Region	998,72	3	332,907	11,74
ERROR	14140,594	499	28,338	
TOTAL	15139,317	502	30,158	

$\overline{W}_\alpha = \langle 0 ; F_{0,95}(3 ; 499) \rangle = \langle 0 ; 2{,}6227 \rangle \qquad F \notin \overline{W}_\alpha \Rightarrow H_0$ ZAM.