Formulas - U2

29th September 2023 ①

Normal Probability function

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$$

$$\mu = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx - \mu^2$$

$$\left.\begin{aligned} &* P(\mu-\sigma, \mu+\sigma) = 68\% \\ &* P(\mu-2\sigma, \mu+2\sigma) = 95\% \\ &* P(\mu-3\sigma, \mu+3\sigma) = 99.7\% \end{aligned}\right\} \text{areas}$$

$$z = \frac{x-\mu}{\sigma}$$

$* \; X \sim N(\mu_1, \sigma_1^2) \quad Y \sim N(\mu_2, \sigma_2^2)$

$$X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$$

$$X-Y \sim N(\mu_1-\mu_2, \sigma_1^2+\sigma_2^2)$$

$$aX+bY \sim N(a\mu_1+b\mu_2, a^2\sigma_1^2+b^2\sigma_2^2)$$

→ Chebyshev's Inequality

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

$$P(\mu-k\sigma < X < \mu+k\sigma) \geq 1-\frac{1}{k^2}$$

→

$$\text{Bias } \hat{\theta} = \mu_{\hat{\theta}} - \theta$$

$$MSE_{\hat{\theta}} = Var(\hat{\theta}) + (\text{Bias of } \hat{\theta})^2$$

$$MSE_{\hat{\theta}} = \sigma_{\hat{\theta}}^2 + (\mu_{\hat{\theta}} - \theta)^2$$

$$MSE_{\hat{\theta}} = E(\hat{\theta}-\theta)^2 = \mu(\hat{\theta}-\theta)^2$$

- Discrete Random Variable

$$\mu_x = \sum x P(X=x)$$

$$\sigma_x^2 = \sum x^2 P(X=x) - \mu_x^2 = \sum P \cdot (x-\bar{x})^2$$

- Probability Density Function

$$f(x) = \text{pdf}$$

$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

$$P(a \leq x \leq b) = P(a < x \leq b) = P(a \leq x < b) = P(a < x < b) = \int_a^b f(x)\, dx$$

- CDF

$$F(x) = \int \underset{\text{pdf}}{f(t)}\, dt$$

- Continuous Random Variable

$$\mu_x = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

$$\sigma_x^2 = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu_x^2$$

uncertainty $\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

→ CLT

can be applied for Binomial random variable

sample proportion $\hat{p} = \frac{x}{n}$

can be normal - rescaled μ, σ

n is large, p is not too close to 0 or 1

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$S_n \sim N(n\mu, n\sigma^2)$

$\mu = p \qquad \sigma = \sqrt{\frac{pq}{n}}$

$$z = \frac{x-\mu}{\sigma} = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$$

→ MLE

Bernoulli $\quad \hat{p} = \bar{x}$; $L = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$; $L = p^{\sum x_i}(1-p)^{n-\sum x_i}$

Binomial $\quad \hat{p} = \frac{x}{n}$; $L = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$

Poisson (λ) $\quad \hat{\lambda} = \bar{x}$; $L = \frac{\lambda^{\sum x_i} e^{-n\lambda}}{\prod x_i!}$

Normal $\quad \hat{\mu} = \bar{x} = \frac{\sum x_i}{n}$

$\hat{\sigma}^2 = \sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$; $L = \prod \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$$L = \sigma^{-n}(2\pi)^{-n/2} \cdot \exp\left[\frac{-1}{2\sigma^2}\sum(x-\mu)^2\right]$$

28th Nov, 2023 ①

U3

CI :

$\bar{x} \pm MOE$

$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$, $\bar{x} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$

$\bar{x} - \bar{y} \pm Z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$

$90\% \rightarrow |Z_{0.05}| = 1.645$

$95\% \rightarrow |Z_{0.025}| = 1.96$

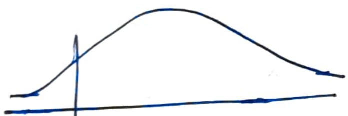
$99\% \rightarrow |Z_{0.005}| = 2.58$

· Agresti Coull : $\tilde{n} = n + 4$ $\tilde{p} = \frac{X+2}{\tilde{n}}$

CI : $\tilde{p} \pm Z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$

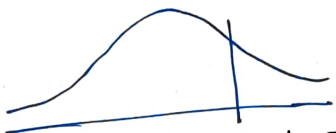
One Sided CI

LCB



$\bar{x} - Z_\alpha \frac{\sigma}{\sqrt{n}}$

UCB



$\bar{x} + Z_\alpha \frac{\sigma}{\sqrt{n}}$

Hypothesis Testing

	H_1 (alt)	H_0 (null)
left-tailed	$<$	$\geq$
right-tailed	$>$	$\leq$
Two-tailed	$\neq$	$=$

$z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

Wilcoxon signed-rank test

for $n > 20$

$$Z = \frac{S_{+} - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Wilcoxon rank-sum test

if $m, n > 8$

$$Z = \frac{W - m(m+n+1)/2}{\sqrt{mn(m+n+1)/12}}$$

Chi-squared Test

$$H_0 : p_1 = p_{01}, p_2 = p_{02}, \ldots\ldots p_k = p_{0k}$$

chi-squared statistic $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$

observed

expected

for r × c data $\cancel{\chi^2}$

$$E_{ij} = \frac{R_{ij} \times C_{ij}}{N}$$

row sum

col sum

row sum + col sum

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Fixed Level Testing

Basically if $P > \alpha \Rightarrow$ accept H_0

$P < \alpha \Rightarrow$ reject H_0

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

to find rejection region do $\bar{x} = z\left(\frac{s}{\sqrt{n}}\right) + \mu$

Errors

Basically

	H_0 is True	H_0 is false
accepted H_0	✓	Type 2
rejected H_0	Type 1	✓ $(1-\beta)$ Power

* H_0 rejection $\Rightarrow$ Type 1

Co-relation coefficient

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}$$

Least Squares line

$$\hat{y}_i = \overset{A}{\hat{\beta}_0} + \overset{B}{\hat{\beta}_1} x_i$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_1 = r\frac{s_y}{s_x}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Co-efficient of determination (r^2)

Total sum of squares — Error sum of squares

$$r^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2 - \sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Total sum of squares = Regression sum of sq + Error sum of sq

$$r^2 = \frac{\text{Regression sum of sq}}{\text{Total sum of sq}}$$

Fitted value $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

residual $e = y - \hat{y}$

13th December, 2023

Formulas U-4

$$r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

→ eqn. of least squares line : $y_i = \hat{\beta}_0 + \hat{\beta}_1 x$

Fitted value $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x$

residual point $e_i = y_i - \hat{y}_i$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} , \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = r \frac{s_y}{s_x}$$

→ Goodness of fit

Total sum of squares = Regression sum of squares + Error sum of sq.

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \qquad \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$r^2 = \frac{\text{regression sum of squares}}{\text{Total sum of squares}}$$

coefficient of determination

→ Estimate of error variance

$$s^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2} = \frac{(1 - r^2) \sum_{i=1}^{n} (y_i - \bar{y}_i)^2}{n-2}$$

$$s_{\hat{\beta}_0} = s \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \qquad s_{\hat{\beta}_1} = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}}$$