Development Statistics

S09 Statistical Test 2

Fujikawa, Kiyoshi
Nagoya University, GSID

Today's mission

- You understand χ^2 distribution
- You understand t distribution
- You understand t test
- You understand χ^2 test

$\chi 2$ distribution

- The population is N(0,1)
- You take samples whose size is n

$$x_i \sim N(0,1)$$

- The squared sum of *n* samples follows Chi-square distribution with degree of freedom n :$\chi^2(n)$

$$C = \sum_{i=1}^{n} x_i \sim \chi^2(n)$$

Sample variance

- The population is N(μ,σ²)
- You take samples whose size is n

Unbiased sample Variance

$$s^2 = \frac{\sum_i (x_i - \overline{x})^2}{n-1}$$

Chi square distribution

$$C = (n-1)\frac{s^2}{\sigma^2} \sim \chi^2(n-1)$$

Definition of t

$$z \sim N(0,1) \qquad C \sim \chi^2(n)$$

$$t = \frac{Z}{\sqrt{\frac{C}{n}}} \sim t(n)$$

Practically speaking

$$z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1) \qquad C = (n-1)\frac{s^2}{\sigma^2} \sim \chi^2(n-1)$$

$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} \sim t(n-1)$$

Test of difference of the means of two populations 1a

- If the population variance is known

$$z_1 = \frac{\overline{X}_1 - \mu_1}{\sqrt{\frac{\sigma_1^2}{n_1}}} \sim N(0,1) \qquad z_2 = \frac{\overline{X}_2 - \mu_2}{\sqrt{\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$$z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Test of difference of the means of two populations 1b

- H0： μ1 = μ2
- H1： μ1 ≠ μ2
- If H0 is true

$$z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Test of difference of the means of two populations 2a

- If the sample size is large enough
- You can assume s ≑ σ

$$z_1 = \frac{\bar{X}_1 - \mu_1}{\sqrt{\frac{s_1^2}{n_1}}} \sim N(0,1) \qquad z_2 = \frac{\bar{X}_2 - \mu_2}{\sqrt{\frac{s_2^2}{n_2}}} \sim N(0,1)$$

$$z = \frac{(\bar{X}_1 - \bar{X}_1) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$$

Test of difference of the means of two populations 2b

- H0:　μ1 = μ2
- H1:　μ1 ≠ μ2
- If H0 is true

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$$

Test of difference of the means of two populations 3a

- If the sample size is small
- In case the population variance is the same

$$t = \frac{(\overline{X}_1 - \overline{X}_1) - (\mu_1 - \mu_2)}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t(n_1 + n_2 - 2)$$

$$s^2 = \frac{\sum_i (x_{1i} - \overline{X}_1)^2 + \sum_j (x_{2j} - \overline{X}_2)^2}{n_1 + n_2 - 2}$$

Test of difference of the means of two populations 3b

- H0： μ1 = μ2
- H1： μ1 ≠ μ2
- If H0 is true

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t(n_1 + n_2 - 2)$$

Test of difference of the ratios of two populations

$$z = \frac{\hat{p}_1 - p_1}{\sqrt{\frac{p_1(1-p_1)}{n_1}}} \sim N(0,1) \qquad z = \frac{\hat{p}_2 - p_2}{\sqrt{\frac{p_2(1-p_2)}{n_2}}} \sim N(0,1)$$

$$z = \frac{(\hat{p}_1 - \hat{p}_1) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0,1)$$

Test of difference of the ratios of two populations

$$z = \frac{(\hat{p}_1 - \hat{p}_1) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0,1)$$

- If $p_1=p_2=p$

$$z = \frac{\hat{p}_1 - \hat{p}_1}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} \sim N(0,1)$$

goodness-of-fit test

observation	O1	…	On
expectation	E1	…	En

$$C = \sum_i \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(n-1)$$

independence test

	C Ca1		C Ca m
R Ca 1	E 11	…	E 1m
…	…		…
R Ca n	E n1	…	E nm

$$C = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2[(n-1)(m-1)]$$