Development Statistics

S10 ANOVA

Fujikawa, Kiyoshi
Nagoya University, GSID

Distribution of sample mean

- The population follows a normal distribution

$$x_i \sim N(\mu, \sigma^2)$$

- Whether σ² is known or unknown

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

Normalization

- Normalization of normal distribution

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \quad \Rightarrow \quad z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

Definition of Chi-square distribution

- The population is N(0,1)
- Take sample 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* :χ^2 (*n*)

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

Sample variance

- The population is N(μ,σ^2)
- Take sample whose size is *n*

Unbiased sample Variance

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

Practically speaking, Chi square is ...

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

$$= \sum_i [(x_i - \mu) - (\bar{x} - \mu)]^2$$

$$= \sum_i (x_i - \mu)^2 - 2(\bar{x} - \mu)\sum_i (x_i - \mu) + n(\bar{x} - \mu)^2$$

$$= \sum_i (x_i - \mu)^2 - \frac{(\bar{x} - \mu)^2}{1/n}$$

Chi square distribution

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

Definition of T distribution

Standard normal **Chi square (n)**

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

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

T (n)

Practically speaking,

Standard normal **Chi square (n-1)**

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

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

T (n-1)

F distribution

- Two independent Chi distributions
- The ratio follows F distribution

$$C_1 \sim \chi^2(n_1) \qquad C_2 \sim \chi^2(n_2)$$

$$\frac{C_1 / n_1}{C_2 / n_2} \sim F(n_1, n_2)$$

Difference means among groups

- When the number of the groups = 2
  - Test of the difference of means of two populations
  - Based on T-distribution
- When the number of the groups >= 3
  - Analysis of Variance = ANOVA
  - Based on F-distribution

Data

group 1	12.2	18.8	18.2
group 2	22.2	20.5	14.6
group 3	20.8	19.5	26.3
group 4	26.4	32.6	31.1
group 5	24.5	21.2	22.4

Test of difference of the means of two populations

- Assuming the population variance is the same

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

$$S^2 = \frac{\sum_{i=1}^{n1}(X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{n2}(X_{2i} - \bar{X}_2)^2}{n_1 + n_2 - 2}$$

- You can do a statistical test based on t-distribution

Use Excel sheet !

- T-statistics

	2	3	4	5
1	-0.865	-1.957	-4.842	-2.719
2		-0.998	-3.687	-1.442
3			-2.799	-0.218
4				3.489

Assumption of ANOVA

	distribution
1	$N(\mu+\alpha_1, \sigma^2)$
2	$N(\mu+\alpha_2, \sigma^2)$
3	$N(\mu+\alpha_3, \sigma^2)$
4	$N(\mu+\alpha_4, \sigma^2)$
5	$N(\mu+\alpha_5, \sigma^2)$

Mean square Decomposition of total variation

- Total variation
  = Intra-group variation + Inter-group variation

$$S_T = S_{\text{intra}} + S_{\text{inter}}$$

$$\sum_i \sum_j (x_{ij} - \bar{x})^2$$

$$= \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 + \sum_i n_i (\bar{x}_i - \bar{x})^2$$

If $\alpha_A = \alpha_B = \alpha_C = \alpha_D = \alpha_E$

- Then, the population distribution is same

$$x_{ij} \sim N(\mu, \sigma^2)$$

- Therefore (see slide No.6)

$$\sum_j (x_{ij} - \bar{x}_i)^2 / \sigma^2 \sim \chi^2(2)$$

- Therefore

$$\sum_i \sum_j (x_{ij} - \bar{x}_i)^2 / \sigma^2 \sim \chi^2(10)$$

If $\alpha_A=\alpha_B=\alpha_C=\alpha_D=\alpha_E$

- Distribution of mean of each group

$$\bar{x}_i \sim N(\mu, \sigma^2/3)$$

- Therefore (see slide No.6)

$$\frac{\sum_i (\bar{x}_i - \bar{x})^2}{\sigma^2/3} \sim \chi^2(4)$$

Then

If $\alpha_A=\alpha_B=\alpha_C=\alpha_D=\alpha_E$

$$\frac{\dfrac{\sum_i (\bar{x}_i - \bar{x})^2}{\sigma^2/3}/4}{\dfrac{\sum_i \sum_j (x_{ij} - \bar{x}_i)^2}{\sigma^2}/10} = \frac{\sum_i 3(\bar{x}_i - \bar{x})^2/4}{\sum_i \sum_j (x_{ij} - \bar{x}_i)^2/10} \sim F(4,10)$$

One-way ANOVA

Factor	Square sum	Degree of freedom	Mean square	F statistic
Inter-group	S (intra)	Group -1	V (intra)	$\frac{V\ (intra)}{V\ (Inter)}$
Intra-group	S (Inter)	N - group	V (Inter)	

One-way ANOVA

Factor	Square sum	Degree of freedom	Mean square	F statistic
Inter-group	314.391	4	78.598	7.079
Intra-group	111.027	10	11.103	