REGIONAL TRAINING COURSE ON APPLIED ECONOMETRIC ANALYSIS (SUMMER SCHOOL) FOR YOUNG ECONOMISTS / RESEARCHERS ORGANIZED BY WIUT AND IFPRI
JUNE 4-15, 2018

Introduction to statistical inference

Farrukh Ataev
Lecturer, WIUT
6 June 2018

Outline

- **Session 5 Introduction to statistical inference**
  - (Tue 6 June at 13:30-15:00)
- **Session 6 Introduction to statistical inference**
  - (Tue 6 June at 15:30-17:00)

Session 6 outline

- ANOVA F-test (for means, equal population variances)
- Chi-squared test (for independence)

ANOVA F-test to compare k treatment means

- $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$
- H_a: at least two treatment means differ
- Test statistic: $F = \frac{MST}{MSE}$
- Rejection rule: $F > F_\alpha$ (or $p < \alpha$),

where F_α is based on $(k-1)$ numerator degree of freedom and $(n-k)$ denominator degree of freedom.

Conditions required for a valid ANOVA F-test:

1) The samples are randomly selected independently from k treatment populations.

2) All k sampled populations are approximately normal.

3) The population variances are equal (i.e. $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$)

Example

Case 1: A training course manager compared the performance of three different groups on a test. Determine if there is a significant difference among the groups using a 5% level ($\alpha = 0.05$). The results of the test are in the following table:

	Group 1	Group 2	Group 3
	61	66	48
	75	82	78
	50	44	63
$\bar{x}$	62	64	63
s^2	157	364	225

$$n = n_1 + n_2 + n_3 = 9 \qquad \bar{\bar{x}} = \frac{\overline{x_1} + \overline{x_2} + \overline{x_3}}{3} = 63$$

$$MST = \frac{SST}{k-1} = \frac{\sum_{j=1}^{3} n_j\left(\bar{x}_j + \bar{\bar{x}}\right)}{3-1} =$$

$$= \frac{3(62-63)^2 + 3(64-63)^2 + 3(63-63)^2}{3-1} = 3$$

$$MSE = \frac{SSE}{n-k} = \frac{\sum_{j=1}^{3}(n_j - 1)s_j^2}{9-3} =$$

$$= \frac{2 \cdot 157 + 2 \cdot 364 + 2 \cdot 225}{6} = 249$$

$$F = \frac{MST}{MSE} = \frac{3}{249} = 0.01$$

Reject H_0 if $F > F_\alpha$

$$F_{\alpha;(k-1;n-k)} = F_{0.05;(2,6)} = 5.14$$

Do not reject H_0 because $0.01 \ngtr 5.14$

F-test for equal population variances

One-tail:

- $H_0: \sigma_1^2 = \sigma_2^2$
- $H_a: \sigma_1^2 < \sigma_2^2$ (or $\sigma_1^2 > \sigma_2^2$)
- Test statistic: $F = \frac{s_2^2}{s_1^2}$ (or $F = \frac{s_1^2}{s_2^2}$)
- Rejection rule: $F > F_\alpha$ (or $p < \alpha$),

where F_α is based on $(n_1 - 1)$ numerator degree of freedom and $(n_2 - 1)$ denominator degree of freedom.

Two-tail:

- $H_0: \sigma_1^2 = \sigma_2^2$
- $H_a: \sigma_1^2 \neq \sigma_2^2$
- Test statistic: $F = \frac{s_2^2}{s_1^2}$ (when $s_2^2 > s_1^2$)
- Rejection rule: $F > F_{\alpha/2}$ (or $p < \alpha/2$),

where $F_{\alpha/2}$ is based on $(n_1 - 1)$ numerator degree of freedom and $(n_2 - 1)$ denominator degree of freedom.

Example

Case 2: A manufacturer of paper products wants to compare the variation in daily production levels at two paper mills. Independent random samples of days are selected from each mill and the production levels (in units) recorded. The data are shown below:

Mill 1	**34**	**18**	**28**	**21**	**32**	**40**	**22**	**23**	**22**	**29**	**25**	**10**	**38**					
Mill 2	31	13	27	19	22	18	23	22	21	13	18	15	24	13	19	18	19	23

Do these data provide sufficient evidence to indicate a difference in the variability of production levels at the two paper mills? Use $\alpha = 0.1$.

$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$

$H_a: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$

$\overline{x_1} = 26.3, \overline{x_2} = 19.3, s_1 = 8.36, s_2 = 4.85, n_1 - 1 = 12, n_2 - 1 = 17.$

Reject H_0 if $F > F_{\alpha/2}$

Critical value: $F_{0.05;(12;17)} = 2.38$

Test statistic: $F = \frac{s_1^2}{s_2^2} = 2.97$

Reject H_0 because $2.97 > 2.38$

Miltinomial experiment

- Binomial experiment: qualitative data that fall in two categories
- Multinomial experiment: qualitative data that fall in more than two categories.

- Confident interval for p_k in Multinomial experiment:

$$p_k = \widehat{p_k} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p_k}(1-\widehat{p_k})}{n}}$$

Hypothesis test for multinomial experiment (one-way)

$H_0: p_1 = p_{1,0}; p_2 = p_{2,0}; \dots; p_k = p_{k,0}$

H_a: at least one does not equal to its hypothesized value.

Test statistic: $\chi^2 = \sum \frac{(n_i - E_i)^2}{E_i} = \sum \frac{\left(n_i - np_{i,0}\right)^2}{np_{i,0}}$, where $E_i = np_{i,0}$ is expected cell count

Rejection rule: $\chi^2 > \chi^2_\alpha$ (or $p > \alpha$), where χ^2_α has $(k-1)$ degrees of freedom.

Example

Case 3: A sample of 200 consumers provided data on preference among 5 different brands of coffee:

Brand	1	2	3	4	5
Consumers	18	65	55	30	32

a) Construct a 90% confidence interval for the proportion of cell 3.

b) Test the claim that the probabilities show no preference. Use $\alpha = 0.05$.

a) $p_3 = \widehat{p_3} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p_3}(1-\widehat{p_3})}{n}} = \frac{55}{200} \pm z_{0.05}\sqrt{\frac{0.275 \cdot 0.725}{200}} = 0.275 \pm 0.05.$

b) $H_0: p_1 = p_2 = p_3 = p_4 = p_5 = \frac{1}{5}$

Reject H_0 if $\chi^2 > \chi^2_\alpha$

H_a: at least one of them is not equal to $\frac{1}{5}$

Critical value: $\chi^2_{0.05;5-1} = 9.48773$

Test statistic: $\chi^2 = \sum \frac{(n_i - E_i)^2}{E_i} = \sum \frac{(n_i - np_{i,0})^2}{np_{i,0}} = \frac{\left(18 - 200 \cdot \frac{1}{5}\right)^2}{200 \cdot \frac{1}{5}} + \cdots + \frac{\left(32 - 200 \cdot \frac{1}{5}\right)^2}{200 \cdot \frac{1}{5}} = 37.45$

Reject H_0 because $37.45 > 9.48773$.

Reading

1) Murray R. Spiegel, *Schaum's outline of Theory and Problems of Probability and Statistics*, McGraw-Hill, 23 edition, 1998.

2) Nitis Mukhopadhyay, Probability and Statistical Inference, Marcel Dekker, Inc. 2000.