Lecture 16: Testing for equality of variances: the F-test

Announcements:

- Midterm exam 2 has been graded and results will be posted to the website.

Announcements:

- Midterm exam 2 has been graded and results will be posted to the website.
- Problem set 8 due now.

Announcements:

- Midterm exam 2 has been graded and results will be posted to the website.
- Problem set 8 due now.
- Problem set 9 due next Tuesday – Croatian translation will be uploaded to website today.

Lecture 16: Testing for equality of variances: the F-test

Announcements:

- Midterm exam 2 has been graded and results will be posted to the website.
- Problem set 8 due now.
- Problem set 9 due next Tuesday – Croatian translation will be uploaded to website today.
- Today: Testing for equality of variances with an F test.

Measures of variability

Testing for equality of variances: the F-test

t - test for diffs in μ

z - test " " "

F - test for diff in σ^2

Measures of variability

- Variance $= Var(x) = \sigma_x^2$

Measures of variability

- Variance $= Var(x) = \sigma_x^2$
- Standard deviation $= \sqrt{Var(x)} = \sigma_x$

Variance of a distribution

Definition

Variance of a distribution

Definition

The variance of a discrete probability distribution is

Variance of a distribution

Definition

The variance of a discrete probability distribution is

$$Var(x) = E(x-\mu)^2 = \sum_i (x_i - \mu)^2 P(x_i). = \sigma^2$$

Variance of a distribution

Definition

The variance of a discrete probability distribution is

$$Var(x) = E(x-\mu)^2 = \sum_i (x_i - \mu)^2 P(x_i).$$

The variance of a continuous probability distribution is

Variance of a distribution

Definition

The variance of a discrete probability distribution is

$$Var(x) = E(x-\mu)^2 = \sum_i (x_i - \mu)^2 P(x_i). \quad = \sigma^2$$

The variance of a continuous probability distribution is

$$Var(x) = E(x-\mu)^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx. \quad = \sigma^2$$

Calculation of the sample variance

Calculation of the sample variance

Sample variance

Calculation of the sample variance

$$\text{Sample variance} = s^2 =$$

Calculation of the sample variance

$$\text{Sample variance} = s^2 = \frac{\sum_1^n (x_i - \overline{x})^2}{n-1}$$

Calculation of the sample variance

estimate of σ^2

$$\text{Sample variance} = s^2 = \frac{\sum_1^n (x_i - \overline{x})^2}{n-1}$$

Sample standard deviation

Calculation of the sample variance

$$\text{Sample variance} = s^2 = \frac{\sum_1^n (x_i - \overline{x})^2}{n-1}$$

$$\text{Sample standard deviation} = s =$$

Calculation of the sample variance

$$\text{Sample variance} = s^2 = \frac{\sum_1^n (x_i - \overline{x})^2}{n-1}$$

$$\text{Sample standard deviation} = s = \sqrt{s^2}$$

Calculation of the sample variance

$$\text{Sample variance} = s^2 = \frac{\sum_1^n (x_i - \overline{x})^2}{n-1}$$

$$\text{Sample standard deviation} = s = \sqrt{s^2}$$

$n - 1$ is called v or the "degrees of freedom"

Calculation of the sample variance

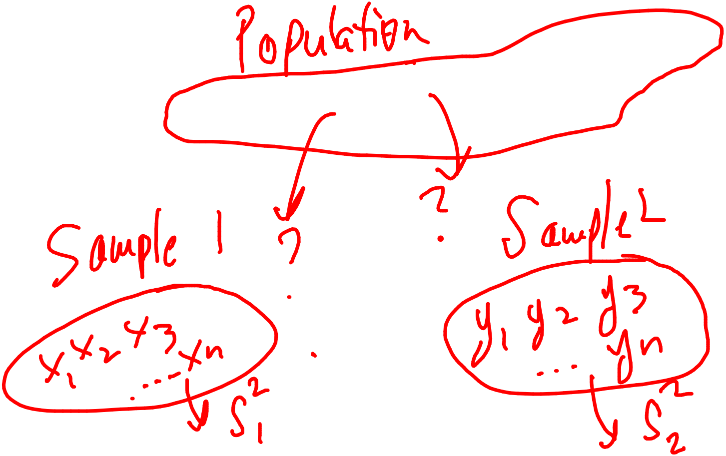
$$\text{Sample variance} = s^2 = \frac{\sum_1^n (x_i - \overline{x})^2}{n-1}$$

$$\text{Sample standard deviation} = s = \sqrt{s^2}$$

$n-1$ is called v, or the "degrees of freedom" used in the calculation of s.

F-test steps

1. We want to know if the two population variances are equal. Null hypothesis is that $\sigma_1^2 = \sigma_2^2$.



F-test steps

1. We want to know if the two population variances are equal. Null hypothesis is that $\sigma_1^2 = \sigma_2^2$.
2. Calculate the sample variance in the two samples, s_1^2 and s_2^2.

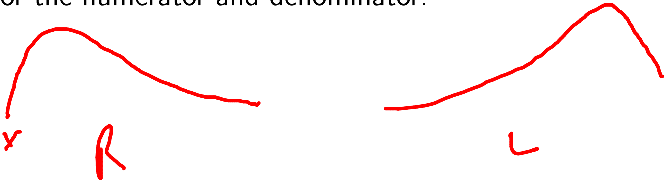
F-test steps

1. We want to know if the two population variances are equal. Null hypothesis is that $\sigma_1^2 = \sigma_2^2$.
2. Calculate the sample variance in the two samples, s_1^2 and s_2^2.
3. Form the ratio of the larger over the smaller.

F-test steps

1. We want to know if the two population variances are equal. Null hypothesis is that $\sigma_1^2 = \sigma_2^2$.
2. Calculate the sample variance in the two samples, s_1^2 and s_2^2.
3. Form the ratio of the larger over the smaller.
4. This ratio has an F-distribution with degrees of freedom for the numerator and denominator.

$\frac{s_1^2}{s_2^2}$



F-test steps

1. We want to know if the two population variances are equal. Null hypothesis is that $\sigma_1^2 = \sigma_2^2$.
2. Calculate the sample variance in the two samples, s_1^2 and s_2^2.
3. Form the ratio of the larger over the smaller. = F
4. This ratio has an F-distribution with degrees of freedom for the numerator and denominator.
5. Compare the observed F with F_{crit}, which is qf(1 - alpha/2, dfn, dfd). Note that we divide alpha by 2 for a two-tailed test.

F-test steps

1. We want to know if the two population variances are equal. Null hypothesis is that $\sigma_1^2 = \sigma_2^2$.
2. Calculate the sample variance in the two samples, s_1^2 and s_2^2.
3. Form the ratio of the larger over the smaller.
4. This ratio has an F-distribution with degrees of freedom for the numerator and denominator.
5. Compare the observed F with F_{crit}, which is qf(1 - alpha/2, dfn, dfd). Note that we divide alpha by 2 for a two-tailed test.
6. Calculate the probability of the observed F as 2(1 - pf(F, dfn, dfd)). Note that we multiply by 2 if this is a two-tailed test.

Bumpus data

$$F = \frac{0.434}{0.176} = 2.47$$

$$p = 0.04$$

In females, the variance of the humerus length of survivors was 0.176, and the variance for the non-survivors was 0.434. The number of survivors was 21, and the number of non-survivors was 28. Question: is the population variance of survivors different from the population variance of non-survivors?

$\therefore$ reject H_0. Conclude

$$\sigma_s^2 \neq \sigma_{ns}^2$$

$$\sigma_{ns}^2 > \sigma_s^2$$

Bumpus data

In females, the variance of the humerus length of survivors was 0.176, and the variance for the non-survivors was 0.434. The number of survivors was 21, and the number of non-survivors was 28. Question: is the population variance of survivors different from the population variance of non-survivors?

Exam scores, men and women

In our first midterm exam, the sample variance for women was 59.2, and the sample variance for men was 68.9. There were 14 women and 13 men taking the exam. Is the true population variance for women different from that for men?

Confidence interval for the population variance

Confidence interval for the population variance

$$\chi^2 = \frac{vs^2}{\sigma^2},$$

Confidence interval for the population variance

$$\chi^2 = \frac{vs^2}{\sigma^2},$$

v is the degrees of freedom; so:

Confidence interval for the population variance

$$\chi^2 = \frac{vs^2}{\sigma^2},$$

v is the degrees of freedom; so:

$$P(\chi^2_{0.025} < \frac{vs^2}{\sigma^2} < \chi^2_{0.975}) = 0.95$$

Confidence interval for the population variance

$$\chi^2 = \frac{vs^2}{\sigma^2},$$

v is the degrees of freedom; so:

$$P(\chi^2_{0.025} < \frac{vs^2}{\sigma^2} < \chi^2_{0.975}) = 0.95$$

$$P(\frac{vs^2}{\chi^2_{0.975}} < \sigma^2 < \frac{vs^2}{\chi^2_{0.025}}) = 0.95$$

Confidence interval for the population variance

$$\chi^2 = \frac{vs^2}{\sigma^2},$$

v is the degrees of freedom; so:

$$P(\chi^2_{0.025} < \frac{vs^2}{\sigma^2} < \chi^2_{0.975}) = 0.95$$

$$P(\frac{vs^2}{\chi^2_{0.975}} < \sigma^2 < \frac{vs^2}{\chi^2_{0.025}}) = 0.95$$

High end of 95% confidence interval:

Confidence interval for the population variance

$$\chi^2 = \frac{vs^2}{\sigma^2},$$

v is the degrees of freedom; so:

$$P(\chi^2_{0.025} < \frac{vs^2}{\sigma^2} < \chi^2_{0.975}) = 0.95$$

$$P(\frac{vs^2}{\chi^2_{0.975}} < \sigma^2 < \frac{vs^2}{\chi^2_{0.025}}) = 0.95$$

High end of 95% confidence interval: $\frac{vs^2}{\chi^2_{0.025}}$

Confidence interval for the population variance

$$\chi^2 = \frac{vs^2}{\sigma^2},$$

v is the degrees of freedom; so:

$$P(\chi^2_{0.025} < \frac{vs^2}{\sigma^2} < \chi^2_{0.975}) = 0.95$$

$$P(\frac{vs^2}{\chi^2_{0.975}} < \sigma^2 < \frac{vs^2}{\chi^2_{0.025}}) = 0.95$$

High end of 95% confidence interval: $\frac{vs^2}{\chi^2_{0.025}}$

Low end of 95% confidence interval:

Confidence interval for the population variance

$$\chi^2 = \frac{vs^2}{\sigma^2},$$

v is the degrees of freedom; so:

$$P(\chi^2_{0.025} < \frac{vs^2}{\sigma^2} < \chi^2_{0.975}) = 0.95$$

$$P(\frac{vs^2}{\chi^2_{0.975}} < \sigma^2 < \frac{vs^2}{\chi^2_{0.025}}) = 0.95$$

High end of 95% confidence interval: $\frac{vs^2}{\chi^2_{0.025}}$

Low end of 95% confidence interval: $\frac{vs^2}{\chi^2_{0.975}}$