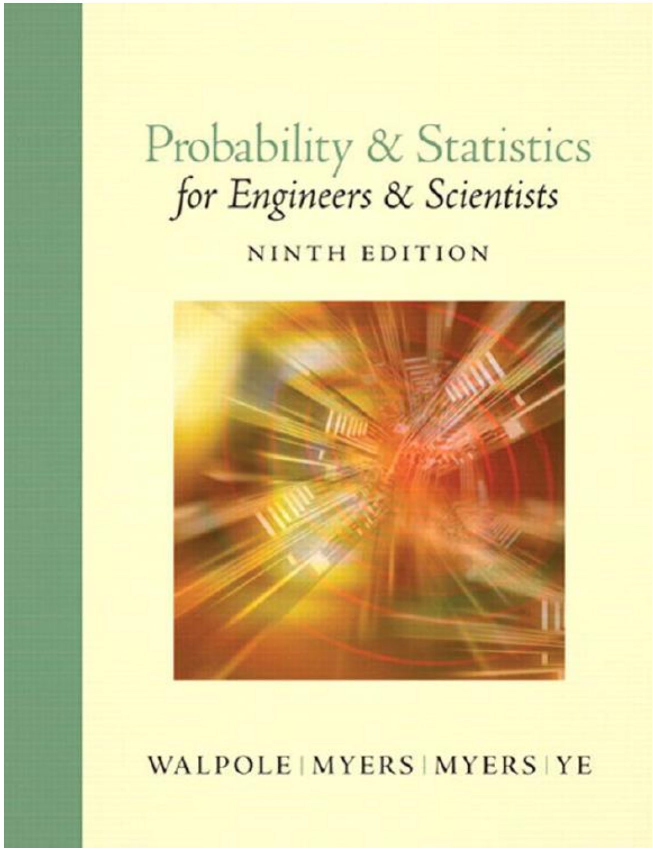
# Statistical Analysis

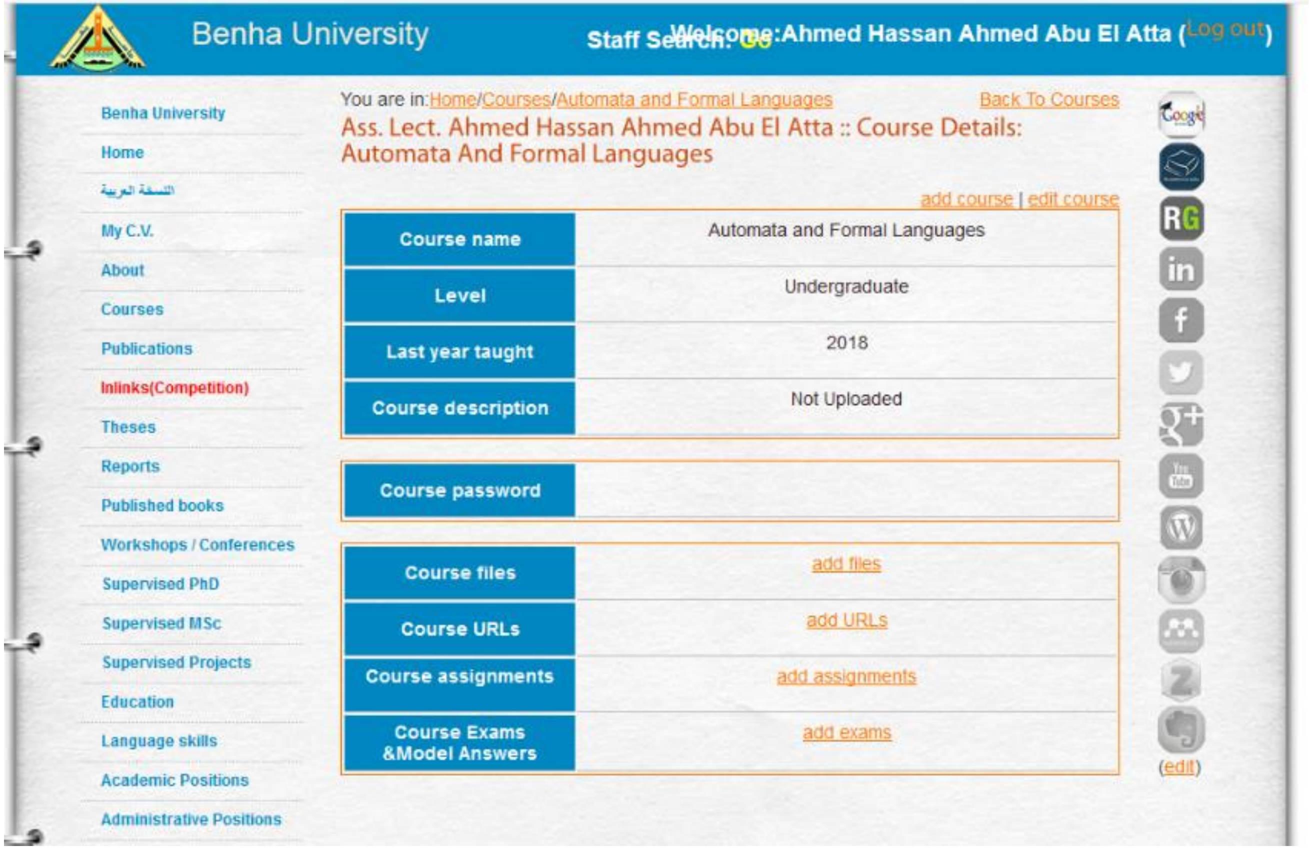
Lecture 06

# Books


Probability & Statistics
for Engineers & Scientists
NINTH EDITION
WALPOLE | MYERS | MYERS | YE

# PowerPoint

http://www.bu.edu.eg/staff/ahmedaboalatah14-courses/14767



# One- and Two-Sample Estimation Problems

CHAPTER 9

# 9.12 Single Sample: Estimating the Variance

# Point Estimating for Variance

If a sample of size $n$ is drawn from a normal population with variance $\sigma^2$ and the sample variance $s^2$ is computed, we obtain a value of the statistic $S^2$. This computed sample variance is used as a point estimate of $\sigma^2$. Hence, the statistic $S^2$ is called an estimator of $\sigma^2$.

# Interval Estimating of σ²

An interval estimate of $\sigma^2$ can be established by using the statistic

$$X^2 = \frac{(n-1)S^2}{\sigma^2}.$$

According to Theorem 8.4, the statistic $X^2$ has a chi-squared distribution with $n-1$ degrees of freedom when samples are chosen from a normal population. We may write (see Figure

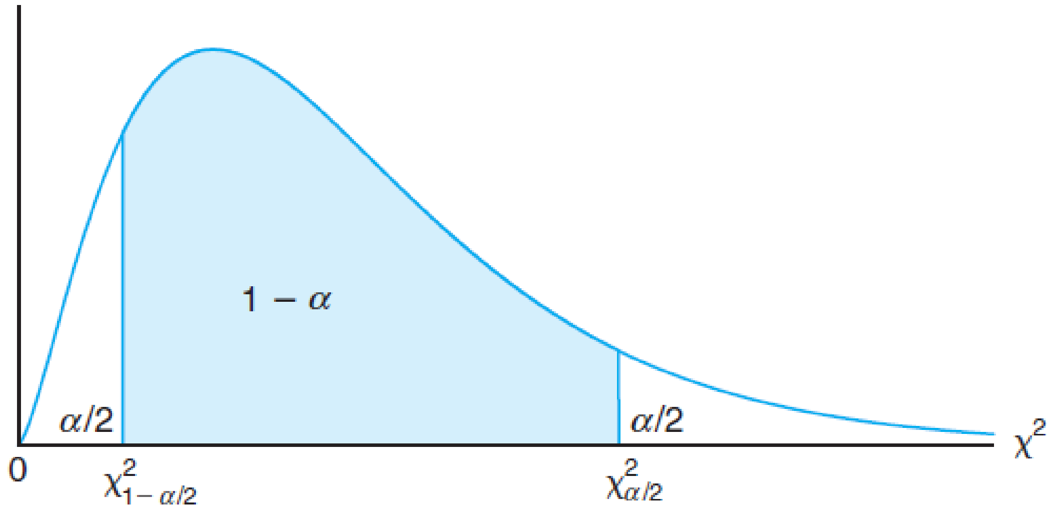



Figure 9.7: $P(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2}) = 1 - \alpha$.

# Interval Estimating of σ2

Dividing each term in the inequality by $(n-1)S^2$ and then inverting each term (thereby changing the sense of the inequalities), we obtain

$$P\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right] = 1-\alpha.$$

For a random sample of size $n$ from a normal population, the sample variance $s^2$ is computed, and the following $100(1-\alpha)\%$ confidence interval for $\sigma^2$ is obtained.

# Confidence Interval for $\sigma^2$

If $s^2$ is the variance of a random sample of size $n$ from a normal population, a $100(1-\alpha)\%$ confidence interval for $\sigma^2$ is

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}},$$

where $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ are $\chi^2$-values with $v = n-1$ degrees of freedom, leaving areas of $\alpha/2$ and $1-\alpha/2$, respectively, to the right.

An approximate $100(1-\alpha)\%$ confidence interval for $\sigma$ is obtained by taking the square root of each endpoint of the interval for $\sigma^2$.

# Example 9.18:

The following are the weights, in decagrams, of 10 packages of grass seed distributed by a certain company: 46.4, 46.1, 45.8, 47.0, 46.1, 45.9, 45.8, 46.9, 45.2, and 46.0. Find a 95% confidence interval for the variance of the weights of all such packages of grass seed distributed by this company, assuming a normal population.

# Example 9.18: *Solution* :

First we find

$$s^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$$

$$= \frac{(10)(21,273.12) - (461.2)^2}{(10)(9)} = 0.286.$$

To obtain a 95% confidence interval, we choose $\alpha = 0.05$. Then, using Table A.5 with $v = 9$ degrees of freedom, we find $\chi^2_{0.025} = 19.023$ and $\chi^2_{0.975} = 2.700$. Therefore, the 95% confidence interval for $\sigma^2$ is

$$\frac{(9)(0.286)}{19.023} < \sigma^2 < \frac{(9)(0.286)}{2.700},$$

or simply $0.135 < \sigma^2 < 0.953$.

# 9.13 Two Samples: Estimating the Ratio of Two Variances

# Point Estimating of the Ratio of Two Population Variances

A point estimate of the ratio of two population variances $\sigma_1^2/\sigma_2^2$ is given by the ratio $s_1^2/s_2^2$ of the sample variances. Hence, the statistic $S_1^2/S_2^2$ is called an estimator of $\sigma_1^2/\sigma_2^2$.

# Interval Estimating of the Ratio of Two Population Variances

If $\sigma_1^2$ and $\sigma_2^2$ are the variances of normal populations, we can establish an interval estimate of $\sigma_1^2/\sigma_2^2$ by using the statistic

$$F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}.$$

According to Theorem 8.8, the random variable $F$ has an $F$-distribution with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ degrees of freedom. Therefore, we may write (see Figure 9.8)

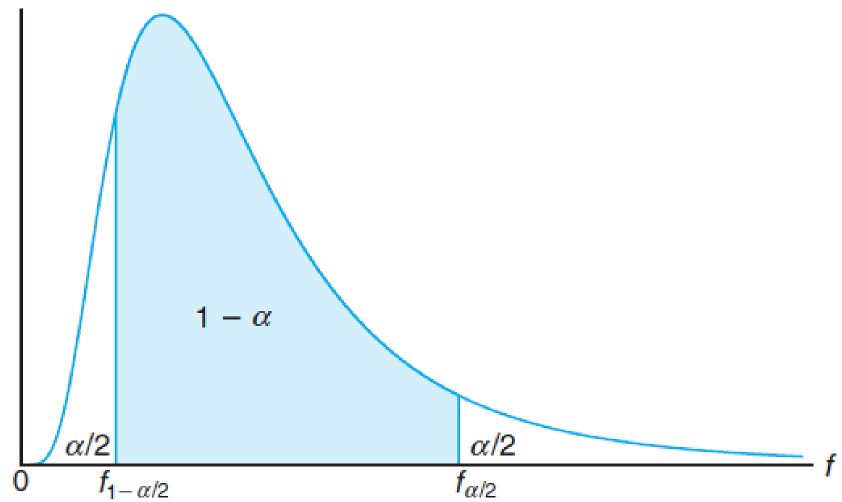



Figure 9.8: $P[f_{1-\alpha/2}(v_1, v_2) < F < f_{\alpha/2}(v_1, v_2)] = 1 - \alpha$.

# Interval Estimating of the Ratio of Two Population Variances

$$P[f_{1-\alpha/2}(v_1, v_2) < F < f_{\alpha/2}(v_1, v_2)] = 1 - \alpha,$$

where $f_{1-\alpha/2}(v_1, v_2)$ and $f_{\alpha/2}(v_1, v_2)$ are the values of the $F$-distribution with $v_1$ and $v_2$ degrees of freedom, leaving areas of $1 - \alpha/2$ and $\alpha/2$, respectively, to the right.

Substituting for $F$, we write

$$P\left[f_{1-\alpha/2}(v_1, v_2) < \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} < f_{\alpha/2}(v_1, v_2)\right] = 1 - \alpha.$$

Multiplying each term in the inequality by $S_2^2/S_1^2$ and then inverting each term, we obtain

$$P\left[\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\frac{1}{f_{1-\alpha/2}(v_1, v_2)}\right] = 1 - \alpha.$$

# Interval Estimating of the Ratio of Two Population Variances

The results of Theorem 8.7 enable us to replace the quantity $f_{1-\alpha/2}(v_1, v_2)$ by $1/f_{\alpha/2}(v_2, v_1)$. Therefore,

$$P\left[\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha/2}(v_2, v_1)\right] = 1 - \alpha.$$

For any two independent random samples of sizes $n_1$ and $n_2$ selected from two normal populations, the ratio of the sample variances $s_1^2/s_2^2$ is computed, and the following $100(1-\alpha)\%$ confidence interval for $\sigma_1^2/\sigma_2^2$ is obtained.

# Confidence Interval for $\sigma^2_1/\sigma^2_2$

If $s_1^2$ and $s_2^2$ are the variances of independent samples of sizes $n_1$ and $n_2$, respectively, from normal populations, then a $100(1-\alpha)\%$ confidence interval for $\sigma_1^2/\sigma_2^2$ is

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{\alpha/2}(v_2, v_1),$$

where $f_{\alpha/2}(v_1, v_2)$ is an $f$-value with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right, and $f_{\alpha/2}(v_2, v_1)$ is a similar $f$-value with $v_2 = n_2 - 1$ and $v_1 = n_1 - 1$ degrees of freedom.

As in Section 9.12, an approximate $100(1-\alpha)\%$ confidence interval for $\sigma_1/\sigma_2$ is obtained by taking the square root of each endpoint of the interval for $\sigma_1^2/\sigma_2^2$.

# Example 9.19:

A confidence interval for the difference in the mean orthophosphorus contents, measured in milligrams per liter, at two stations on the James River was constructed in Example 9.12 on page 290 by assuming the normal population variance to be unequal. Justify this assumption by constructing 98% confidence intervals for $\sigma_1^2/\sigma_2^2$ and for $\sigma_1/\sigma_2$, where $\sigma_1^2$ and $\sigma_2^2$ are the variances of the populations of orthophosphorus contents at station 1 and station 2, respectively.

# Example 9.19: Solution :

From Example 9.12, we have $n_1 = 15$, $n_2 = 12$, $s_1 = 3.07$, and $s_2 = 0.80$. For a 98% confidence interval, $\alpha = 0.02$. Interpolating in Table A.6, we find $f_{0.01}(14, 11) \approx 4.30$ and $f_{0.01}(11, 14) \approx 3.87$. Therefore, the 98% confidence interval for $\sigma_1^2/\sigma_2^2$ is

$$\left(\frac{3.07^2}{0.80^2}\right)\left(\frac{1}{4.30}\right) < \frac{\sigma_1^2}{\sigma_2^2} < \left(\frac{3.07^2}{0.80^2}\right)(3.87),$$

which simplifies to $3.425 < \frac{\sigma_1^2}{\sigma_2^2} < 56.991$. Taking square roots of the confidence limits, we find that a 98% confidence interval for $\sigma_1/\sigma_2$ is

$$1.851 < \frac{\sigma_1}{\sigma_2} < 7.549.$$

Since this interval does not allow for the possibility of $\sigma_1/\sigma_2$ being equal to 1, we were correct in assuming that $\sigma_1 \neq \sigma_2$ or $\sigma_1^2 \neq \sigma_2^2$ in Example 9.12.