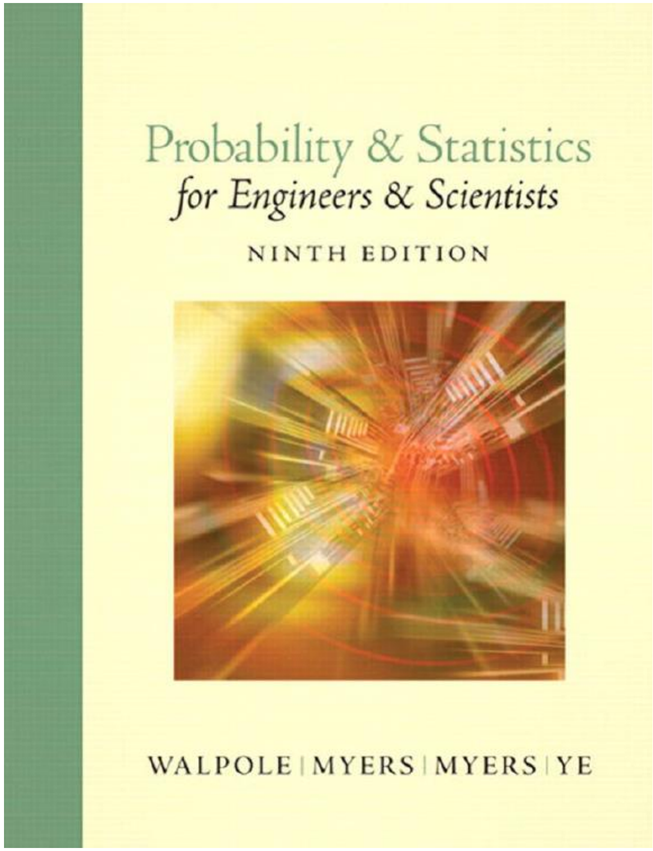
# Statistical Analysis

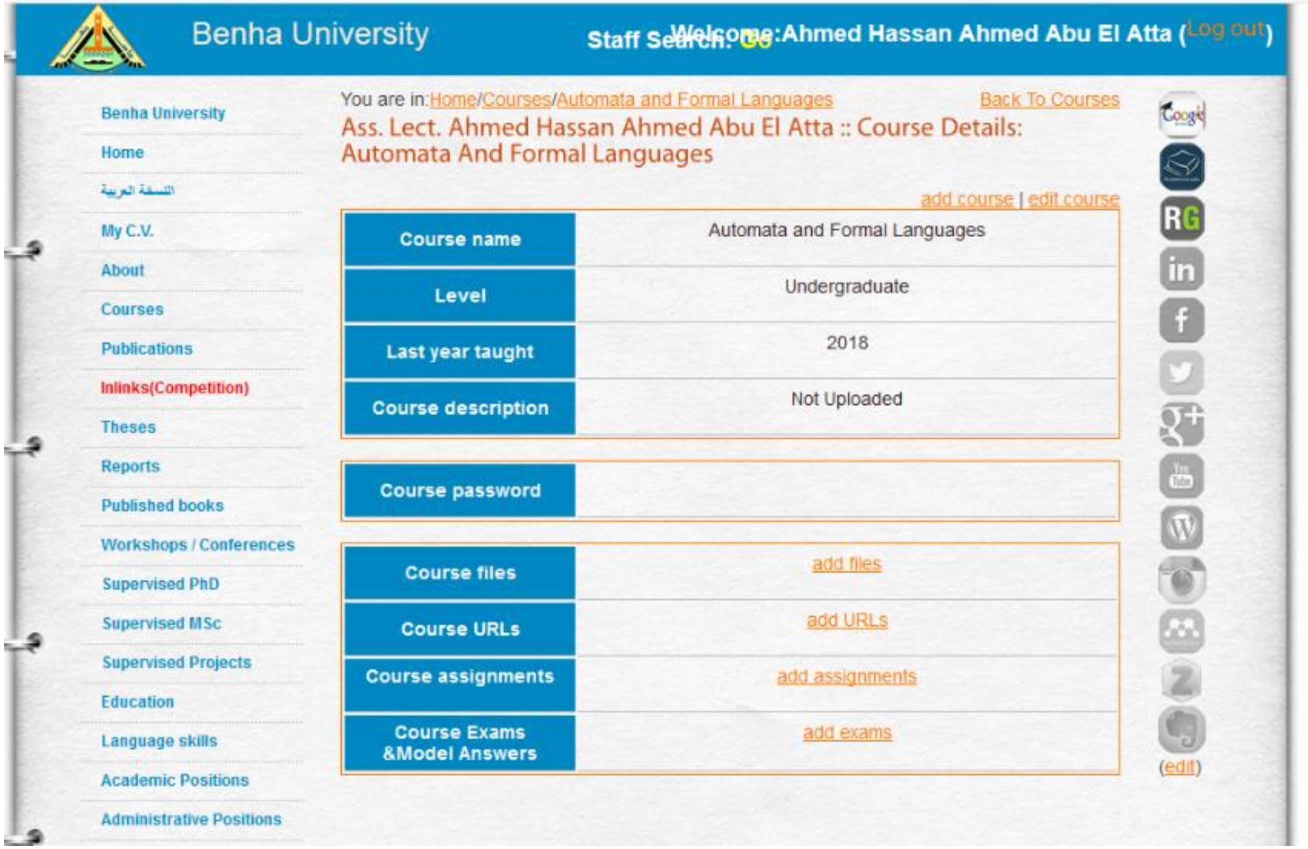
Lecture 08

# 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 Tests of Hypotheses

CHAPTER 10

# 10.10 One- and Two-Sample Tests Concerning Variances

# One Sample Test Concerning Variance

Let us first consider the problem of testing the null hypothesis $H_0$ that the population variance $\sigma^2$ equals a specified value $\sigma_0^2$ against one of the usual alternatives $\sigma^2 < \sigma_0^2$, $\sigma^2 > \sigma_0^2$, or $\sigma^2 \neq \sigma_0^2$. The appropriate statistic on which to base our decision is the chi-squared statistic of Theorem 8.4, which was used in Chapter 9 to construct a confidence interval for $\sigma^2$. Therefore, if we assume that the distribution of the population being sampled is normal, the chi-squared value for testing $\sigma^2 = \sigma_0^2$ is given by

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2},$$

of significance, the critical region is $\chi^2 < \chi^2_{1-\alpha/2}$ or $\chi^2 > \chi^2_{\alpha/2}$. For the one-sided alternative $\sigma^2 < \sigma_0^2$, the critical region is $\chi^2 < \chi^2_{1-\alpha}$, and for the one-sided alternative $\sigma^2 > \sigma_0^2$, the critical region is $\chi^2 > \chi^2_{\alpha}$.

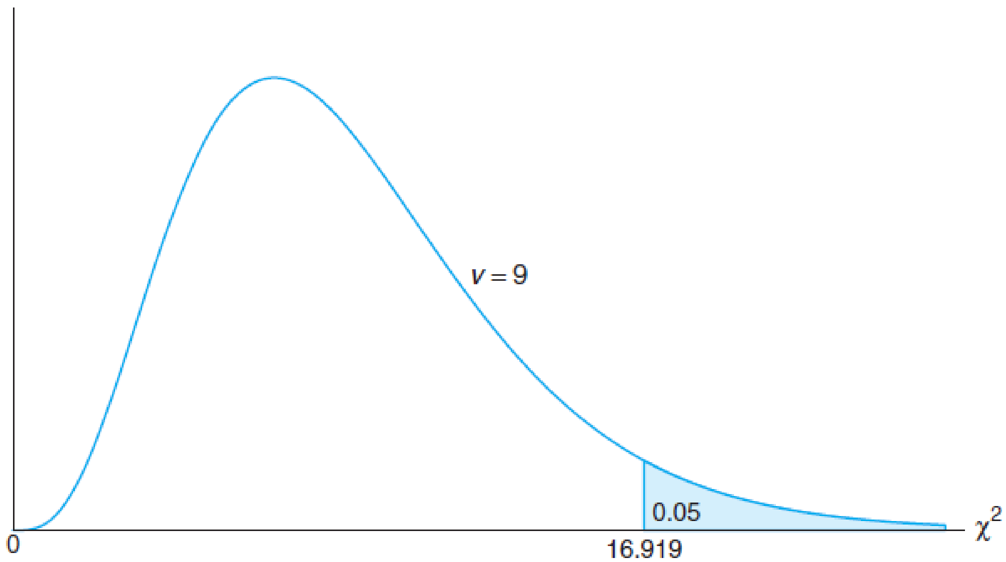
# Example 10.12:

A manufacturer of car batteries claims that the life of the company's batteries is approximately normally distributed with a standard deviation equal to 0.9 year.

If a random sample of 10 of these batteries has a standard deviation of 1.2 years, do you think that $\sigma > 0.9$ year? Use a 0.05 level of significance.

# Solution

1. $H_0$: $\sigma^2 = 0.81$.
2. $H_1$: $\sigma^2 > 0.81$.
3. $\alpha = 0.05$.
4. Critical region: From Figure 10.19 we see that the null hypothesis is rejected when $\chi^2 > 16.919$, where $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$, with $v = 9$ degrees of freedom.




5. Computations: $s^2 = 1.44$, $n = 10$, and

$$\chi^2 = \frac{(9)(1.44)}{0.81} = 16.0, \qquad P \approx 0.07.$$

6. Decision: The $\chi^2$-statistic is not significant at the 0.05 level.

# Two Samples Test Concerning Variances

Now let us consider the problem of testing the equality of the variances $\sigma_1^2$ and $\sigma_2^2$ of two populations. That is, we shall test the null hypothesis $H_0$ that $\sigma_1^2 = \sigma_2^2$ against one of the usual alternatives

$$\sigma_1^2 < \sigma_2^2, \quad \sigma_1^2 > \sigma_2^2, \quad \text{or} \quad \sigma_1^2 \neq \sigma_2^2.$$

For independent random samples of sizes $n_1$ and $n_2$, respectively, from the two populations, the **f-value for testing $\boldsymbol{\sigma_1^2 = \sigma_2^2}$** is the ratio

$$f = \frac{s_1^2}{s_2^2},$$

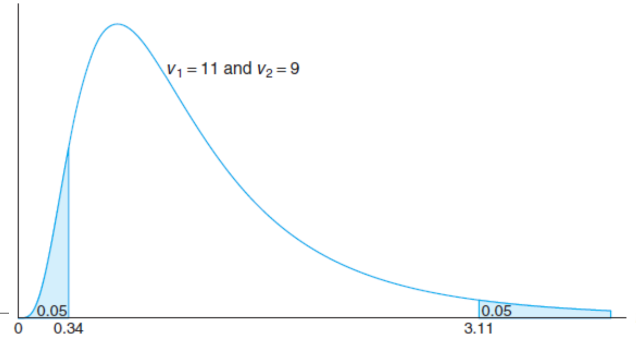
of size $\alpha$ corresponding to the one-sided alternatives $\sigma_1^2 < \sigma_2^2$ and $\sigma_1^2 > \sigma_2^2$ are, respectively, $f < f_{1-\alpha}(v_1, v_2)$ and $f > f_\alpha(v_1, v_2)$. For the two-sided alternative $\sigma_1^2 \neq \sigma_2^2$, the critical region is $f < f_{1-\alpha/2}(v_1, v_2)$ or $f > f_{\alpha/2}(v_1, v_2)$.

# Example 10.13:

In testing for the difference in the abrasive wear of the two materials in Example 10.6, we assumed that the two unknown population variances were equal. Were we justified in making this assumption? Use a 0.10 level of significance.

An experiment was performed to compare the abrasive wear of two different laminated materials. Twelve pieces of material 1 were tested by exposing each piece to a machine measuring wear. Ten pieces of material 2 were similarly tested. In each case, the depth of wear was observed. The samples of material 1 gave an average (coded) wear of 85 units with a sample standard deviation of 4, while the samples of material 2 gave an average of 81 with a sample standard deviation of 5. Can we conclude at the 0.05 level of significance that the abrasive wear of material 1 exceeds that of material 2 by more than 2 units? Assume the populations to be approximately normal with equal variances.

# Solution :



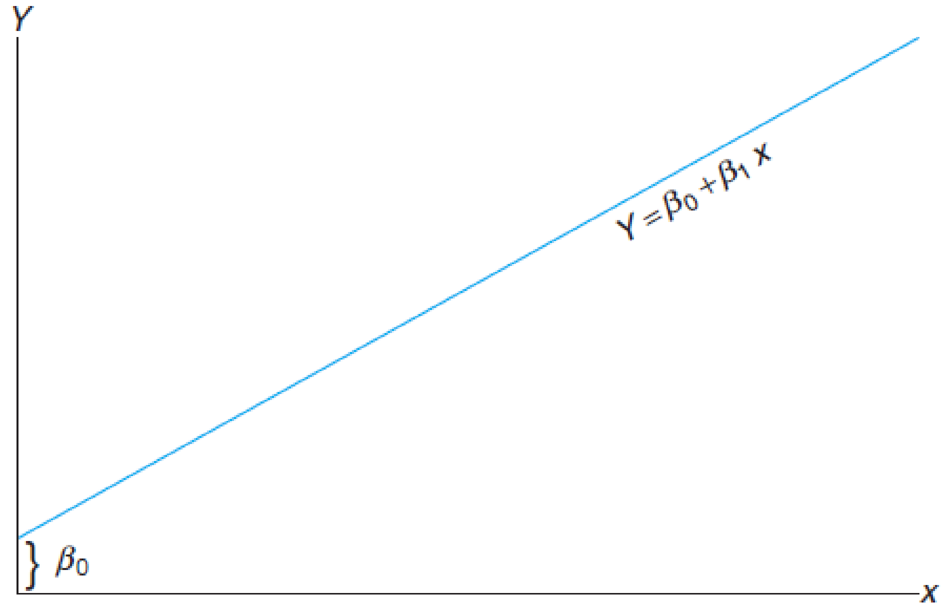

Let $\sigma_1^2$ and $\sigma_2^2$ be the population variances for the abrasive wear of material 1 and material 2, respectively.

1. $H_0$: $\sigma_1^2 = \sigma_2^2$.
2. $H_1$: $\sigma_1^2 \neq \sigma_2^2$.
3. $\alpha = 0.10$.
4. Critical region: From Figure 10.20, we see that $f_{0.05}(11, 9) = 3.11$, and, by using Theorem 8.7, we find

$$f_{0.95}(11, 9) = \frac{1}{f_{0.05}(9, 11)} = 0.34.$$

   Therefore, the null hypothesis is rejected when $f < 0.34$ or $f > 3.11$, where $f = s_1^2/s_2^2$ with $v_1 = 11$ and $v_2 = 9$ degrees of freedom.
5. Computations: $s_1^2 = 16$, $s_2^2 = 25$, and hence $f = \frac{16}{25} = 0.64$.
6. Decision: Do not reject $H_0$. Conclude that there is insufficient evidence that the variances differ.

# Linear Regression



# Estimating the Regression Coefficients

Given the sample $\{(x_i, y_i);\ i = 1, 2, \ldots, n\}$, the least squares estimates $b_0$ and $b_1$ of the regression coefficients $\beta_0$ and $\beta_1$ are computed from the formulas

$$b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \text{ and}$$

$$b_0 = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n} = \bar{y} - b_1 \bar{x}.$$

Table 11.1: Measures of Reduction in Solids and Oxygen Demand

| Solids Reduction, $x$ (%) | Oxygen Demand Reduction, $y$ (%) | Solids Reduction, $x$ (%) | Oxygen Demand Reduction, $y$ (%) |
|---|---|---|---|
| 3 | 5 | 36 | 34 |
| 7 | 11 | 37 | 36 |
| 11 | 21 | 38 | 38 |
| 15 | 16 | 39 | 37 |
| 18 | 16 | 39 | 36 |
| 27 | 28 | 39 | 45 |
| 29 | 27 | 40 | 39 |
| 30 | 25 | 41 | 41 |
| 30 | 35 | 42 | 40 |
| 31 | 30 | 42 | 44 |
| 31 | 40 | 43 | 37 |
| 32 | 32 | 44 | 44 |
| 33 | 34 | 45 | 46 |
| 33 | 32 | 46 | 46 |
| 34 | 34 | 47 | 49 |
| 36 | 37 | 50 | 51 |
| 36 | 38 | | |

Estimate the regression line for the pollution data of Table 11.1.

$$\sum_{i=1}^{33} x_i = 1104, \ \sum_{i=1}^{33} y_i = 1124, \ \sum_{i=1}^{33} x_i y_i = 41{,}355, \ \sum_{i=1}^{33} x_i^2 = 41{,}086$$

Therefore,

$$b_1 = \frac{(33)(41{,}355) - (1104)(1124)}{(33)(41{,}086) - (1104)^2} = 0.903643 \text{ and}$$

$$b_0 = \frac{1124 - (0.903643)(1104)}{33} = 3.829633.$$

Thus, the estimated regression line is given by

$$\hat{y} = 3.8296 + 0.9036x.$$