Statistics

Chapter 9

PART 2

Agenda

- Hypothesis Testing
- Type-1 and Type –II errors
- P-Value
- Chi-square Distribution
- F-Distribution

Hypothesis Testing

- A vital role of Statistics is in verifying statements, claims, conjectures, and in general testing hypotheses. Based on a random sample, we can use Statistics to verify whether
- – a system has not been infected,
- – a hardware upgrade was efficient,
- – the average number of concurrent users increased by 2000 this year,
- – the average connection speed is 54 Mbps, as claimed by the internet service provider,
- – the proportion of defective products is at most 3%, as promised by the manufacturer,
- – service times have Gamma distribution,
- – the number of errors in software is independent of the manager’s experience,
- – etc.

Source: Michael Baron’s Probability and Statistics for Computer Scientists

Hypothesis Testing Contd…

- Testing statistical hypotheses has wide applications far beyond Computer Science.
- These methods are used to
  - prove efficiency of a new medical treatment
  - safety of a new automobile brand,
  - innocence of a defendant,
  - authorship of a document;
  - to establish cause-and-effect relationships;
  - to identify factors that can significantly improve the response
  - to fit stochastic models
  - to detect information leaks
  - and so forth.

Source: Michael Baron’s Probability and Statistics for Computer Scientists

Hypothesis Contd..

$$\underline{\text{NOTATION}} \quad \left\| \begin{array}{lcl} H_0 & = & \text{hypothesis (the null hypothesis)} \\ H_A & = & \text{alternative (the alternative hypothesis)} \end{array} \right\|$$

Source: Michael Baron’s Probability and Statistics for Computer Scientists

Example 9.22 Review

- Example 9.22. To verify that the the average connection speed is 54 Mbps, we test the hypothesis $H_0 : \mu = 54$ against the two-sided alternative $H_A : \mu = 54$, where μ is the average speed of all connections.
- However, if we worry about a low connection speed only, we can conduct a one-sided test of $H_0 : \mu = 54$ vs $H_A : \mu < 54$.
- In this case, we only measure the amount of evidence supporting the one-sided alternative $H_A : \mu < 54$. In the absence of such evidence, we gladly accept the null hypothesis.

Source: Michael Baron’s Probability and Statistics for Computer Scientists

Hypothesis Contd..

- DEFINITION 9.6 Alternative of the type $H_A : \mu \neq \mu_0$ covering regions on both sides of the hypothesis ($H_0 : \mu = \mu_0$) is a two-sided alternative.

- Alternative $H_A : \mu < \mu_0$ covering the region to the left of H_0 is one-sided, left-tail.

- Alternative $H_A : \mu > \mu_0$ covering the region to the right of H_0 is one-sided, right-tail.

Source: Michael Baron's Probability and Statistics for Computer Scientists

Hypothesis Test Contd…

DEFINITION 9.6

Alternative of the type $H_A : \mu \neq \mu_0$ covering regions on both sides of the hypothesis $(H_0 : \mu = \mu_0)$ is a **two-sided alternative**.

Alternative $H_A : \mu < \mu_0$ covering the region to the left of H_0 is **one-sided, left-tail**.

Alternative $H_A : \mu > \mu_0$ covering the region to the right of H_0 is **one-sided, right-tail**.

Source: Michael Baron's Probability and Statistics for Computer Scientists

Type 1 and Type II errors

9.4.2 Type I and Type II errors: level of significance

When testing hypotheses, we realize that all we see is a random sample. Therefore, with all the best statistics skills, our decision to accept or to reject H_0 may still be wrong. That would be a *sampling error* (Section 8.1).

Four situations are possible,

	Result of the test	
	Reject H_0	**Accept** H_0
H_0 **is true**	*Type I error*	correct
H_0 **is false**	correct	*Type II error*

DEFINITION 9.7

A **type I error** occurs when we reject the true null hypothesis.

A **type II error** occurs when we accept the false null hypothesis.

Source: Michael Baron's Probability and Statistics for Computer Scientists

Acceptance and Rejection Regions

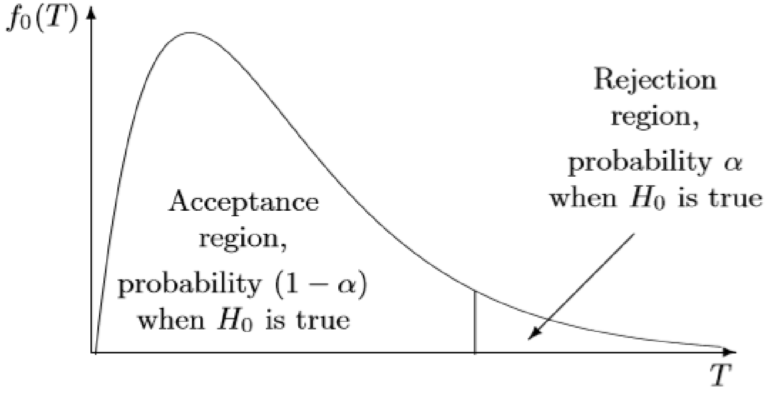


FIGURE 9.6: *Acceptance and rejection regions.*

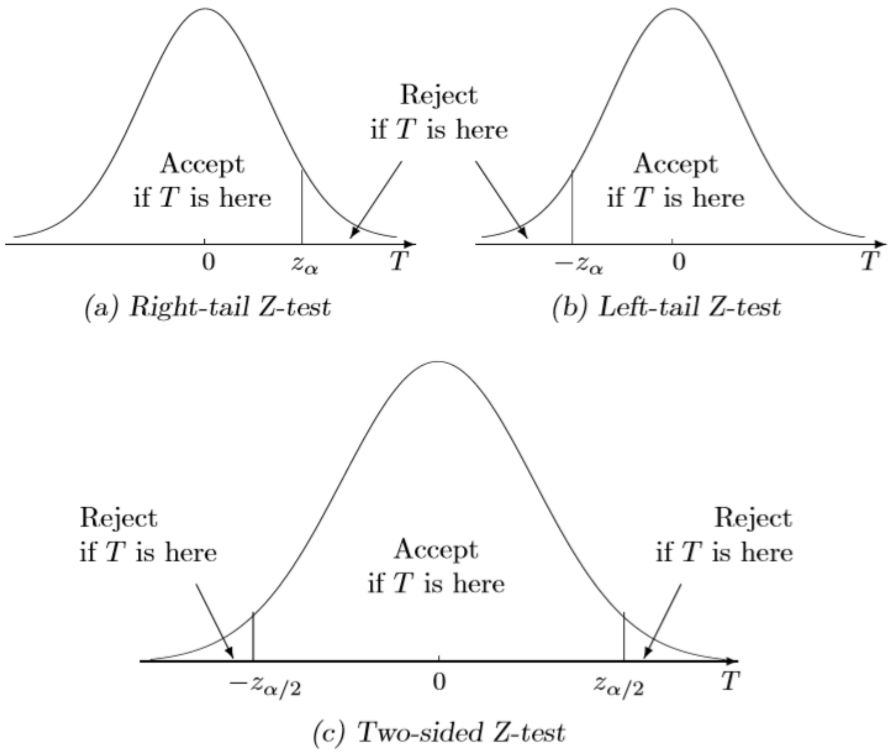


FIGURE 9.7: *Acceptance and rejection regions for a Z-test with (a) a one-sided right-tail alternative; (b) a one-sided left-tail alternative; (c) a two-sided alternative.*

Source: Michael Baron's Probability and Statistics for Computer Scientists

Z-Tests for Means and Proportions

9.4.6 Z-tests for means and proportions

As we already know,

– sample means have Normal distribution when the distribution of data is Normal;

– sample means have approximately Normal distribution when they are computed from large samples (the distribution of data can be arbitrary);

– sample proportions have approximately Normal distribution when they are computed from large samples;

– this extends to differences between means and between proportions

Source: Michael Baron's Probability and Statistics for Computer Scientists

Example 9.25 Review

Example 9.25 (Z-TEST ABOUT A POPULATION MEAN). The number of concurrent users for some internet service provider has always averaged 5000 with a standard deviation of 800. After an equipment upgrade, the average number of users at 100 randomly selected moments of time is 5200. Does it indicate, at a 5% level of significance, that the mean number of concurrent users has increased? Assume that the standard deviation of the number of concurrent users has not changed.

Source: Michael Baron’s Probability and Statistics for Computer Scientists

Null hypothesis H_0	Parameter, estimator $\theta, \hat{\theta}$	If H_0 is true: $\mathbf{E}(\hat{\theta})$	If H_0 is true: $\mathrm{Var}(\hat{\theta})$	Test statistic $Z = \dfrac{\hat{\theta} - \theta_0}{\sqrt{\mathrm{Var}(\hat{\theta})}}$
One-sample Z-tests for means and proportions, based on a sample of size n				
$\mu = \mu_0$	$\mu, \bar{X}$	μ_0	$\dfrac{\sigma^2}{n}$	$\dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$
$p = p_0$	$p, \hat{p}$	p_0	$\dfrac{p_0(1-p_0)}{n}$	$\dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
Two-sample Z-tests comparing means and proportions of two populations, based on independent samples of size n and m				
$\mu_X - \mu_Y = D$	$\mu_X - \mu_Y$, $\bar{X} - \bar{Y}$	D	$\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}$	$\dfrac{\bar{X} - \bar{Y} - D}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$
$p_1 - p_2 = D$	$p_1 - p_2$, $\hat{p}_1 - \hat{p}_2$	D	$\dfrac{p_1(1-p_1)}{n} + \dfrac{p_2(1-p_2)}{m}$	$\dfrac{\hat{p}_1 - \hat{p}_2 - D}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}}$
$p_1 = p_2$	$p_1 - p_2$, $\hat{p}_1 - \hat{p}_2$	0	$p(1-p)\left(\dfrac{1}{n} + \dfrac{1}{m}\right)$, where $p = p_1 = p_2$	$\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}}$ where $\hat{p} = \dfrac{n\hat{p}_1 + m\hat{p}_2}{n+m}$

TABLE 9.1: *Summary of Z-tests.*

Source: Michael Baron’s Probability and Statistics for Computer Scientists

Unknown Pop Std Dev: T-Tests

As we decided in Section 9.3, when we don't know the population standard deviation, we estimate it. The resulting *T-statistic* has the form

$$t = \frac{\hat{\theta} - \mathbf{E}(\hat{\theta})}{s(\hat{\theta})} = \frac{\hat{\theta} - \mathbf{E}(\hat{\theta})}{\sqrt{\widehat{\mathrm{Var}}(\hat{\theta})}}.$$

In the case *when the distribution of* $\hat{\theta}$ *is Normal*, the test is based on *Student's T-distribution* with acceptance and rejection regions according to the direction of H_A:

(a) For a **right-tail alternative**,

$$\begin{cases} \text{reject } H_0 & \text{if} \quad t \geq t_\alpha \\ \text{accept } H_0 & \text{if} \quad t < t_\alpha \end{cases} \tag{9.18}$$

(b) For a **left-tail alternative**,

$$\begin{cases} \text{reject } H_0 & \text{if} \quad t \leq -t_\alpha \\ \text{accept } H_0 & \text{if} \quad t > -t_\alpha \end{cases} \tag{9.19}$$

(c) For a **two-sided alternative**,

$$\begin{cases} \text{reject } H_0 & \text{if} \quad |t| \geq t_{\alpha/2} \\ \text{accept } H_0 & \text{if} \quad |t| < t_{\alpha/2} \end{cases} \tag{9.20}$$

Source: Michael Baron's Probability and Statistics for Computer Scientists

Example 9.28 (UNAUTHORIZED USE OF A COMPUTER ACCOUNT, CONTINUED). A long-time authorized user of the account makes 0.2 seconds between keystrokes. One day, the data in Example 9.19 on p. 260 are recorded as someone typed the correct username and password. At a 5% level of significance, is this an evidence of an unauthorized attempt?

Source: Michael Baron’s Probability and Statistics for Computer Scientists

Example 9.29 (CD WRITER AND BATTERY LIFE). Does a CD writer consume extra energy, and therefore, does it reduce the battery life on a laptop?

Example 9.20 on p. 262 provides data on battery lives for laptops with a CD writer (sample $\boldsymbol{X}$) and without a CD writer (sample $\boldsymbol{Y}$):

$$n = 12,\ \bar{X} = 4.8,\ s_X = 1.6;\ m = 18,\ \bar{Y} = 5.3,\ s_Y = 1.4;\ s_p = 1.4818.$$

Testing

$$H_0 : \mu_X = \mu_Y \quad \text{vs} \quad H_A : \mu_X < \mu_Y$$

at $\alpha = 0.05$, we obtain

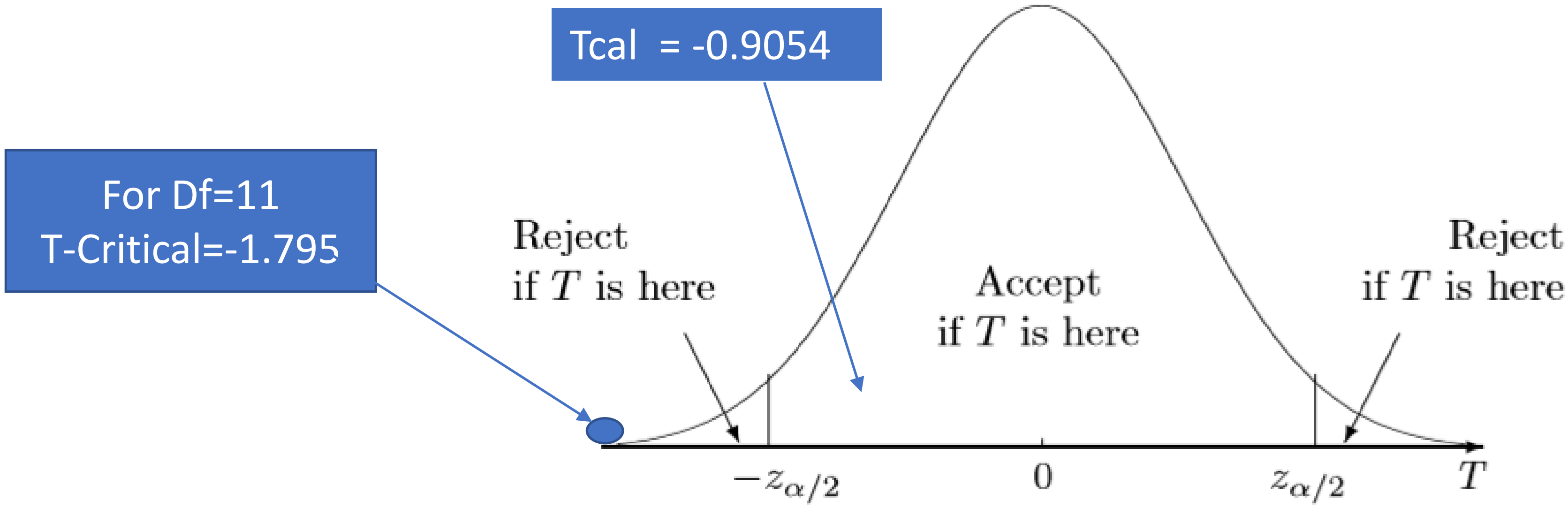
$$t = \frac{\bar{X} - \bar{Y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{4.8 - 5.3}{(1.4818)\sqrt{\frac{1}{18} + \frac{1}{12}}} = -0.9054.$$

For Df=11
T-Critical=-1.795

The rejection region for this left-tail test is $(-\infty, -z_\alpha] = (-\infty, -1.645]$. Since $t \notin \mathcal{R}$, we accept H_0 concluding that there is *no evidence that laptops with a CD writer have a shorter battery life.* ◊

Source: Michael Baron's Probability and Statistics for Computer Scientists

Problem 9.29 Contd…



(c) Two-sided Z-test

What about P-Value

- Review of P-Value

DEFINITION 9.9

P-value is the lowest significance level α that forces rejection of the null hypothesis.

P-value is also the highest significance level α that forces acceptance of the null hypothesis.

Source: Michael Baron's Probability and Statistics for Computer Scientists

P-Values for Z-Test

Understanding P-values

Looking at Tables 9.3 and 9.4, we see that *P-value* is the probability of observing a test statistic *at least as extreme as* Z_{obs} or t_{obs}. Being "extreme" is determined by the alternative. For a right-tail alternative, large numbers are extreme; for a left-tail alternative, small

Hypothesis H_0	Alternative H_A	P-value	Computation
$\theta = \theta_0$	right-tail $\theta > \theta_0$	$\boldsymbol{P}\{Z \geq Z_{\text{obs}}\}$	$1 - \Phi(Z_{\text{obs}})$
	left-tail $\theta < \theta_0$	$\boldsymbol{P}\{Z \leq Z_{\text{obs}}\}$	$\Phi(Z_{\text{obs}})$
	two-sided $\theta \neq \theta_0$	$\boldsymbol{P}\{\lvert Z\rvert \geq \lvert Z_{\text{obs}}\rvert\}$	$2(1 - \Phi(\lvert Z_{\text{obs}}\rvert))$

TABLE 9.3: *P-values for Z-tests.*

Source: Michael Baron's Probability and Statistics for Computer Scientists

P-Values for T-Test

Hypothesis H_0	Alternative H_A	P-value	Computation
$\theta = \theta_0$	right-tail $\theta > \theta_0$	$\boldsymbol{P}\{t \geq t_{\text{obs}}\}$	$1 - F_\nu(t_{\text{obs}})$
	left-tail $\theta < \theta_0$	$\boldsymbol{P}\{t \leq t_{\text{obs}}\}$	$F_\nu(t_{\text{obs}})$
	two-sided $\theta \neq \theta_0$	$\boldsymbol{P}\{\lvert t \rvert \geq \lvert t_{\text{obs}} \rvert\}$	$2(1 - F_\nu(\lvert t_{\text{obs}} \rvert))$

TABLE 9.4: *P-values for T-tests (F_ν is the cdf of T-distribution with the suitable number ν of degrees of freedom).*

Source: Michael Baron’s Probability and Statistics for Computer Scientists

P-Values for a Z-Test

Example 10.3: A random sample of 100 recorded deaths in the United States during the past year showed an average life span of 71.8 years. Assuming a population standard deviation of 8.9 years, does this seem to indicate that the mean life span today is greater than 70 years? Use a 0.05 level of significance.

Source: Probability and Statistics for Engineers and Scientists by Walopole et al

P-

Example 10.5: The Edison Electric Institute has published figures on the number of kilowatt hours used annually by various home appliances. It is claimed that a vacuum cleaner uses an average of 46 kilowatt hours per year. If a random sample of 12 homes included in a planned study indicates that vacuum cleaners use an average of 42 kilowatt hours per year with a standard deviation of 11.9 kilowatt hours, does this suggest at the 0.05 level of significance that vacuum cleaners use, on average, less than 46 kilowatt hours annually? Assume the population of kilowatt hours to be normal.

Solution:

1. H_0: $\mu = 46$ kilowatt hours.
2. H_1: $\mu < 46$ kilowatt hours.
3. $\alpha = 0.05$.
4. Critical region: $t < -1.796$, where $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ with 11 degrees of freedom.
5. Computations: $\bar{x} = 42$ kilowatt hours, $s = 11.9$ kilowatt hours, and $n = 12$. Hence,

$$t = \frac{42 - 46}{11.9/\sqrt{12}} = -1.16, \qquad P = P(T < -1.16) \approx 0.135.$$

6. Decision: Do not reject H_0 and conclude that the average number of kilowatt hours used annually by home vacuum cleaners is not significantly less than 46.

Source: Probability and Statistics for Engineers and Scientists by Walopole et al

Chi-Square Distribution

9.5.1 Variance estimator and Chi-square distribution

We start by estimating the population variance $\sigma^2 = \text{Var}(X)$ from an observed sample $\boldsymbol{X} = (X_1, \ldots, X_n)$. Recall from Section 8.2.4 that σ^2 is estimated *unbiasedly* and *consistently* by the sample variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2.$$

The summands $\left(X_i - \bar{X}\right)^2$ are not quite independent, as the Central Limit Theorem on p. 93 requires, because they all depend on $\bar{X}$. Nevertheless, the distribution of s^2 is approximately Normal, under mild conditions, when the sample is large.

For small to moderate samples, the distribution of s^2 is not Normal at all. It is not even symmetric. Indeed, why should it be symmetric if s^2 is always non-negative!

Distribution of the sample variance

When observations $X_1, \ldots, X_n$ are independent and Normal with $\text{Var}(X_i) = \sigma^2$, the distribution of

$$\frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma}\right)^2$$

is *Chi-square with* $(n-1)$ *degrees of freedom*

Chi-square distribution, or χ^2, is a continuous distribution with density

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, \quad x > 0,$$

Source: Michael Baron's Probability and Statistics for Computer Scientists

Chi-Squre Distribution Contd..

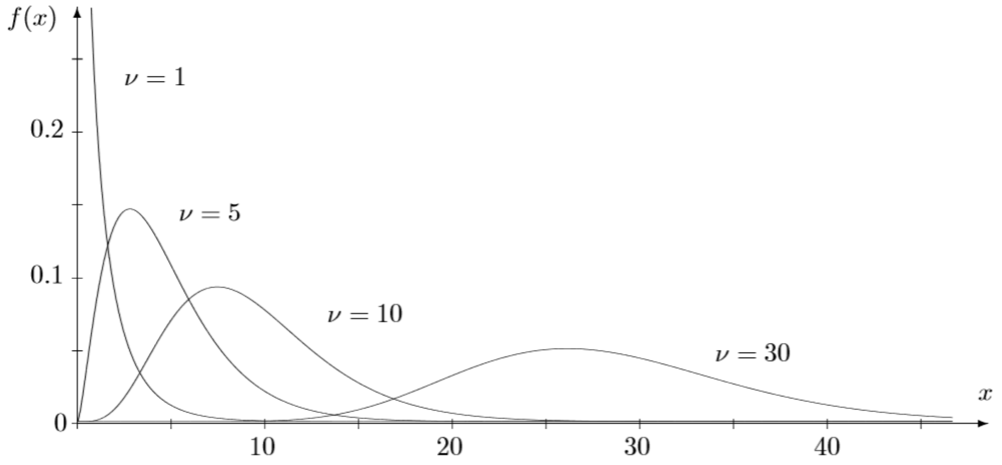


FIGURE 9.12: *Chi-square densities with ν = 1, 5, 10, and 30 degrees of freedom. Each distribution is right-skewed. For large ν, it is approximately Normal.*

Chi-square distribution (χ^2)

$$\begin{array}{lcl} \nu & = & \text{degrees of freedom} \\ f(x) & = & \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)}\, x^{\nu/2-1} e^{-x/2}, \quad x > 0 \\ \mathbf{E}(X) & = & \nu \\ \mathrm{Var}(X) & = & 2\nu \end{array} \qquad (9.23)$$

Source: Michael Baron’s Probability and Statistics for Computer Scientists

Confidence Interval for the population Variance

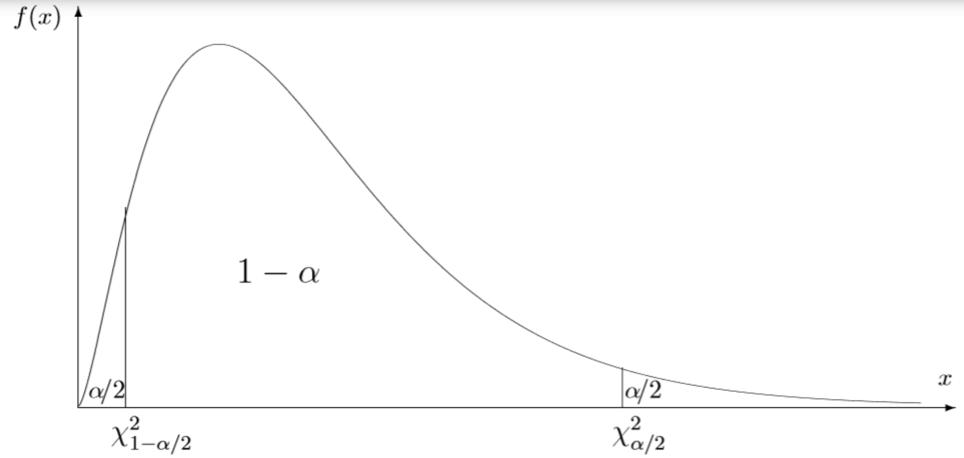


FIGURE 9.13: *Critical values of the Chi-square distribution.*

Confidence interval for the standard deviation

$$\left[\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}},\ \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}\right]$$

Source: Michael Baron’s Probability and Statistics for Computer Scientists

Example 9.40. In Example 9.31 on p. 278, we relied on the reported parameters of the measurement device and assumed the known standard deviation $\sigma = 2.2$. Let us now rely on the data only and construct a 90% confidence interval for the standard deviation. The sample contained $n = 6$ measurements, 2.5, 7.4, 8.0, 4.5, 7.4, and 9.2.

Solution. Compute the sample mean and then the sample variance,

$$\bar{X} = \frac{1}{6}(2.5 + \ldots + 9.2) = 6.5;$$

$$s^2 = \frac{1}{6-1}\left\{(2.5-6.5)^2 + \ldots + (9.2-6.5)^2\right\} = \frac{31.16}{5} = 6.232.$$

(actually, we only need $(n-1)s^2 = 31.16$).

From Table A6 of Chi-square distribution with $\nu = n - 1 = 5$ degrees of freedom, we find the critical values $\chi^2_{1-\alpha/2} = \chi^2_{0.95} = 1.15$ and $\chi^2_{\alpha/2} = \chi^2_{0.05} = 11.1$. Then,

$$\left[\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}},\ \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}\right] = \left[\sqrt{\frac{31.16}{11.1}},\ \sqrt{\frac{31.16}{1.15}}\right] = \underline{[1.68,\ 5.21]}.$$

is a 90% confidence interval for the population standard deviation (and by the way, $[1.68^2, 5.21^2] = [2.82, 27.14]$ is a 90% confidence interval for the variance). ◊

Source: Michael Baron's Probability and Statistics for Computer Scientists

Chi-Square Distribution: Testing variance

9.5.3 Testing variance

Suppose now that we need *to test* the population variance, for example, to make sure that the actual variability, uncertainty, volatility, or risk does not exceed the promised value. We'll derive a level α test based on the Chi-square distribution of the rescaled sample variance.

Level α test

Let $X_1, \ldots, X_n$ be a sample from the Normal distribution with the unknown population variance σ^2. For testing the null hypothesis

$$H_0 : \sigma^2 = \sigma_0^2,$$

compute the χ^2-*statistic*

$$\chi^2_{\text{obs}} = \frac{(n-1)s^2}{\sigma_0^2}.$$

Source: Michael Baron's Probability and Statistics for Computer Scientists

Chi-Square Distribution: Testing variance Contd…

Testing against the *right-tail* alternative $H_A : \sigma^2 > \sigma_0^2$, reject H_0 if $\chi^2_{\text{obs}} \geq \chi^2_\alpha$.

Testing against the *left-tail* alternative $H_A : \sigma^2 < \sigma_0^2$, reject H_0 if $\chi^2_{\text{obs}} \leq \chi^2_{1-\alpha}$.

Testing against the *two-sided* alternative $H_A : \sigma^2 \neq \sigma_0^2$, reject H_0 if either $\chi^2_{\text{obs}} \geq \chi^2_{\alpha/2}$ or $\chi^2_{\text{obs}} \leq \chi^2_{1-\alpha/2}$.

As an exercise, please verify that in each case, the probability of type I error is exactly α.

Null Hypothesis	Alternative Hypothesis	Test statistic	Rejection region	P-value
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\dfrac{(n-1)s^2}{\sigma_0^2}$	$\chi^2_{\text{obs}} > \chi^2_\alpha$	$\boldsymbol{P}\left\{\chi^2 \geq \chi^2_{\text{obs}}\right\}$
	$\sigma^2 < \sigma_0^2$		$\chi^2_{\text{obs}} < \chi^2_\alpha$	$\boldsymbol{P}\left\{\chi^2 \leq \chi^2_{\text{obs}}\right\}$
	$\sigma^2 \neq \sigma_0^2$		$\chi^2_{\text{obs}} \geq \chi^2_{\alpha/2}$ or $\chi^2_{\text{obs}} \leq \chi^2_{1-\alpha/2}$	$2\min\left(\boldsymbol{P}\left\{\chi^2 \geq \chi^2_{\text{obs}}\right\}, \boldsymbol{P}\left\{\chi^2 \leq \chi^2_{\text{obs}}\right\}\right)$

TABLE 9.5: χ^2-*tests for the population variance*

Source: Michael Baron’s Probability and Statistics for Computer Scientists

F-Distribution

9.5.4 Comparison of two variances. F-distribution.

In this section, we deal with two populations whose variances need to be compared. Such inference is used for the comparison of accuracy, stability, uncertainty, or risks arising in two populations.

Distribution of the ratio of sample variances

For independent samples $X_1, \ldots, X_n$ from Normal (μ_X, σ_X) and $Y_1, \ldots, Y_m$ from Normal (μ_Y, σ_Y), the standardized ratio of variances

$$F = \frac{s_X^2/\sigma_X^2}{s_Y^2/\sigma_Y^2} = \frac{\sum(X_i - \bar{X})^2/\sigma_X^2/(n-1)}{\sum(Y_i - \bar{Y})^2/\sigma_Y^2/(m-1)}$$

has *F-distribution with* $(n-1)$ and $(m-1)$ *degrees of freedom.*

(9.29)

Source: Michael Baron's Probability and Statistics for Computer Scientists

F-Distribution

Reciprocal property of F-distribution

The critical values of $F(\nu_1, \nu_2)$ and $F(\nu_2, \nu_1)$ distributions are related as follows,

$$F_{1-\alpha}(\nu_1, \nu_2) = \frac{1}{F_\alpha(\nu_2, \nu_1)} \tag{9.32}$$

Confidence interval for the ratio of variances

$$\left[\frac{s_X^2}{s_Y^2 F_{\alpha/2}(n-1, m-1)},\ \frac{s_X^2 F_{\alpha/2}(m-1, n-1)}{s_Y^2} \right] \tag{9.33}$$

Source: Michael Baron’s Probability and Statistics for Computer Scientists

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.

***Solution*:** Let σ_1^2 and σ_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.

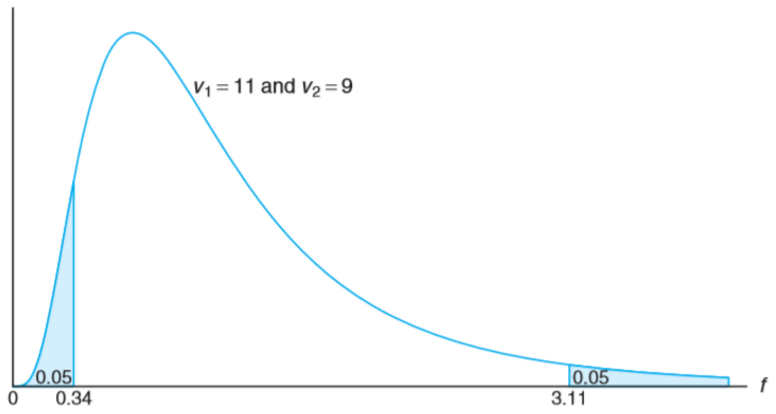


Figure 10.20: Critical region for the alternative hypothesis $\sigma_1^2 \neq \sigma_2^2$.

Source: Probability and Statistics for Engineers and Scientists by Walopole et al

9.5.6 F-tests comparing two variances

In this section, we test the null hypothesis about a *ratio of variances*

$$H_0 : \frac{\sigma_X^2}{\sigma_Y^2} = \theta_0 \tag{9.34}$$

against a one-sided or a two-sided alternative. Often we only need to know if two variances are equal, then we choose $\theta_0 = 1$. F-distribution is used to compare variances, so this test is called the **F-test**.

The test statistic for (9.34) is

$$F = \frac{s_X^2}{s_Y^2} / \theta_0,$$

which under the null hypothesis equals

$$F = \frac{s_X^2 / \sigma_X^2}{s_Y^2 / \sigma_Y^2}.$$

Source: Michael Baron's Probability and Statistics for Computer Scientists

F-Distribution Contd..

Null Hypothesis $H_0 : \frac{\sigma_X^2}{\sigma_Y^2} = \theta_0$		Test statistic $F_{\text{obs}} = \frac{s_X^2}{s_Y^2} / \theta_0$
Alternative Hypothesis	Rejection region	P-value Use F$(n-1, m-1)$ distribution
$\frac{\sigma_X^2}{\sigma_Y^2} > \theta_0$	$F_{\text{obs}} \geq F_\alpha(n-1, m-1)$	$\boldsymbol{P}\{F \geq F_{\text{obs}}\}$
$\frac{\sigma_X^2}{\sigma_Y^2} < \theta_0$	$F_{\text{obs}} \leq F_\alpha(n-1, m-1)$	$\boldsymbol{P}\{F \leq F_{\text{obs}}\}$
$\frac{\sigma_X^2}{\sigma_Y^2} \neq \theta_0$	$F_{\text{obs}} \geq F_{\alpha/2}(n-1, m-1)$ or $F_{\text{obs}} < 1/F_{\alpha/2}(m-1, n-1)$	$2 \min\left(\boldsymbol{P}\{F \geq F_{\text{obs}}\}, \boldsymbol{P}\{F \leq F_{\text{obs}}\}\right)$

TABLE 9.6: *Summary of F-tests for the ratio of population variances*

Source: Michael Baron’s Probability and Statistics for Computer Scientists