UTM
UNIVERSITI TEKNOLOGI MALAYSIA

CHAPTER 5

Part 2

HYPOTHESIS TESTING
(One Sample Test)

Introduction

- In statistics, a hypothesis is a claim or statement about a property of a population.
- A hypothesis test (or test of significance) is a standard procedure for testing a claim about a property of a population.
- To conduct a Hypothesis Testing, we must establish the following :
  1) Hypothesis statement
  2) Test Statistic
  3) Significance Level @ Level of confidence
  4) Conclusion/Decision rule

Examples:

- **Genetics:** The Genetics & IVF Institute claims that its XSORT method allows couples to increase the probability of having a baby girl.

- **Business:** A newspaper headline makes the claim that most workers get their jobs through networking.

- **Medicine:** Medical researchers claim that when people with colds are treated with echinacea, the treatment has no effect.

- **Aircraft Safety:** The Federal Aviation Administration claims that the mean weight of an airline passenger (including carry-on baggage) is greater than 185 lb, which it was 20 years ago.

- **Quality Control:** When new equipment is used to manufacture aircraft altimeters, the new altimeters are better because the variation in the errors is reduced so that the readings are more consistent. (In many industries, the quality of goods and services can often be improved by reducing variation.)

Steps To Conduct Hypothesis Test:

1) State the components of a formal Hypothesis Test

Null Hypothesis: H_0

Alternative Hypothesis: H_1

Null Hypothesis: H_0

- The null hypothesis (denoted by H_0) is a statement that the value of a population parameter (such as proportion, mean, or standard deviation) is equal to some claimed value.

- We test the null hypothesis directly either reject H_0 or fail to reject H_0.

- The symbolic form of the null hypothesis is: $=$

Alternative Hypothesis: H_1

- The alternative hypothesis (denoted by H_1 or H_a or H_A) is the statement that the parameter has a value that somehow differs from the null hypothesis.

- The symbolic form of the alternative hypothesis must use one of these symbols: $\neq$, $<$, $>$.

Types of Hypothesis Test: Two-tailed, Left-tailed, Right-tailed

- Determinations of *P*-values and critical values are affected by whether a critical region is in two tails, the left tail, or the right tail. It therefore becomes important to correctly characterize a hypothesis test as two-tailed, left-tailed, or right-tailed.

- The tails in a distribution are the extreme regions bounded by critical values.

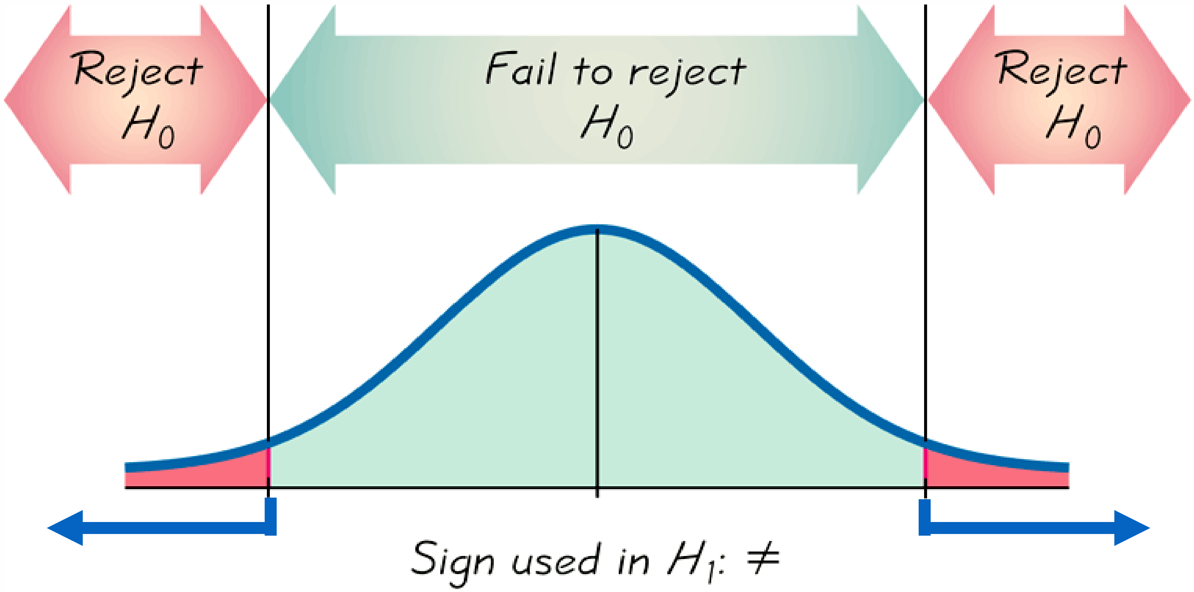
Two-tailed Test

H_0: =

H_1: ≠

α is divided equally between the two tails of the critical region

Means less than or greater than



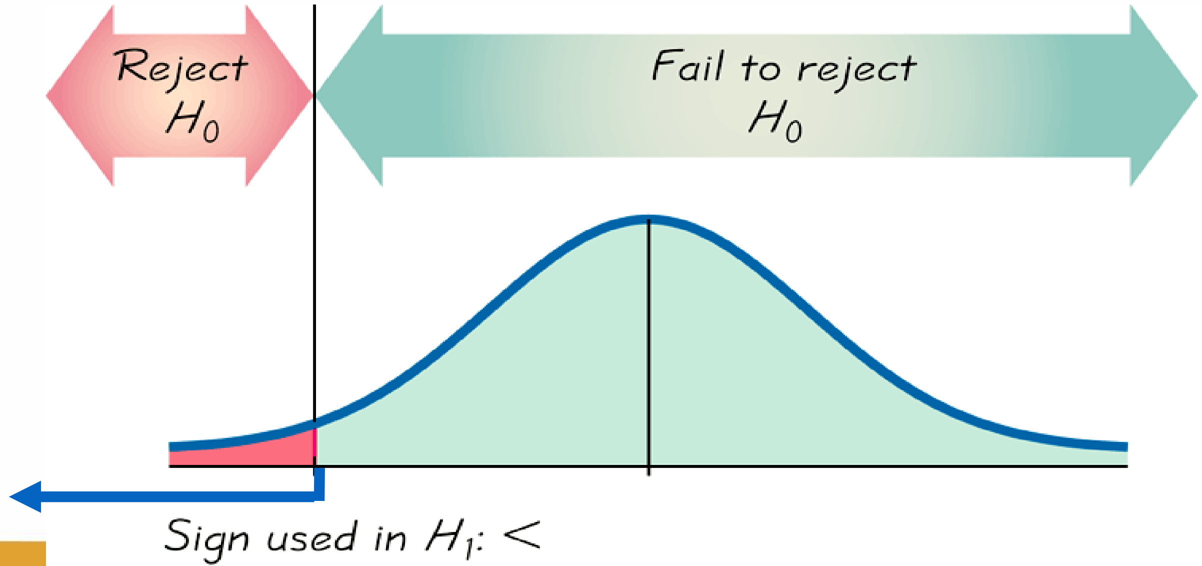
Left-tailed Test

H_0: =

H_1: <

α the left tail

Points Left



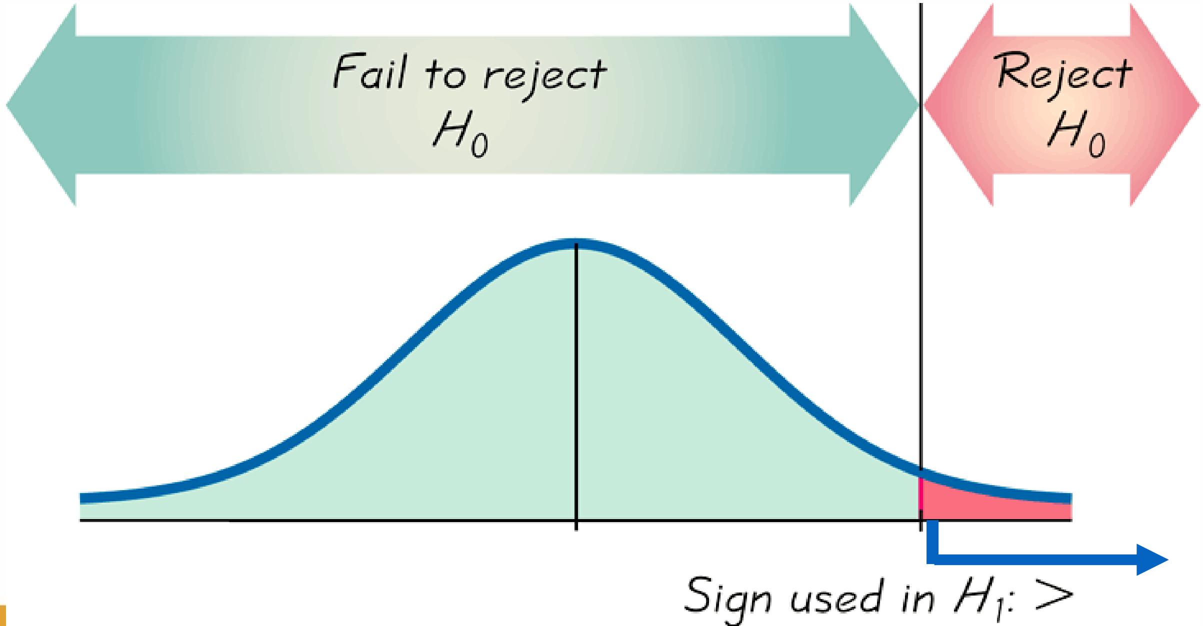
Right-tailed Test

H_0: =

H_1: >

α the right tail

Points Right



How to state the hypothesis statement?

Example # 1:

Sharing prescription drugs with others can be dangerous. A survey of a representative sample of 592 U.S. teens age 12 to 17 reported that 118 of those surveyed admitted to having shared a prescription drug with a friend. Is this sufficient evidence that **more than 10%** of teens have shared prescription medication with friends? State the component of a hypotheses test.

$$H_0: p = 0.1$$

$$H_1: p > 0.1$$

Example # 2:

Compact florescent (cfl) light bulbs are much more energy efficient than regular incandescent light bulbs. Eco bulb brand 60-watt cfl light bulbs state on the package “Average life 8000 hours”. People who purchase this brand would be unhappy if the bulbs lasted **less than 8000 hours**. A sample of these bulbs will be selected and tested. State the component of a hypotheses test.

H_0: $m = 8000$ The true mean (m) life of the cfl light bulbs

H_1: $m < 8000$ One sided alternative.

Example # 3:

Because in variation of the manufacturing process, tennis balls produced by a particular machine do not have the same diameters. Suppose the machine was initially calibrated to achieve the specification of $m = 3$ inches. However, the manager is now concerned that the diameters no longer conform to this specification. If the **mean diameter is not 3** inches, production will have to be halted. State the hypotheses statement.

$H_0: m = 3$

$H_1: m \neq 3$ Two sided alternative

Exercises #1:

For each pair of hypotheses, indicate which are not legitimate and explain why?

a) $H_0: \mu = 15; H_1: \mu \geq 15$

b) $H_0: \bar{x} = 4; H_1: \bar{x} < 15$

c) $H_0: p = 0.1; H_1: p \neq 0.1$

d) $H_0: \mu = 2.3; H_1: \mu > 2.3$

e) $H_0: p \neq 0.5; H_1: p = 0.5$

Exercises #2: Identify H_0 and H_1

1) The mean annual income of employees who took a statistics course is greater than $60,000.

2) The proportion of people aged 18 to 25 who currently use illicit drugs is equal to 0.20 (or 20%).

3) The standard deviation of duration times (in seconds) of the Old Faithful geyser is less than 40 sec.

2) Test of Statistical Hypothesis

- A test statistic is computed using sample data and is the value used to reach a conclusion to reject or fail to reject H_0.
- There are term/formula to calculate test statistic:

Test statistic for proportion: $z = \frac{\bar{p} - p}{\sqrt{\frac{pq}{n}}}$

Test statistic for population mean:

variance known or $n > 30$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

variance unknown and $n \leq 30$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

Test statistic for population standard deviation:

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

Example

Let's again consider the claim that the XSORT method of gender selection increases the likelihood of having a baby girl. Preliminary results from a test of the XSORT method of gender selection involved 14 couples who gave birth to 13 girls and 1 boy. The null and alternative hypotheses are H_0: p = 0.5 and H_1: p > 0.5. Calculate the value of the test statistic.

Example - Solution

- Given: H_0: $p = 0.5$ and H_1: $p > 0.5$
- The sample proportion of 13 girls in 14 births results in

$$\bar{p} = {}^{13}/_{14} = 0.929$$

- Using $p = 0.5, \bar{p} = 0.929$, and $n = 14$, we find the value of the test statistic as follows:

$$z = \frac{0.929 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{14}}} = 3.21$$

3) Significance Level

- The significance level (denoted by α) is the probability that the test statistic will fall in the critical region when the null hypothesis is actually true.

- Common choices for α are 0.05, 0.01, and 0.10.

Decision Criterion

There are two types of decision criteria that can be used:

i) *P*-value method

ii) Critical Region method

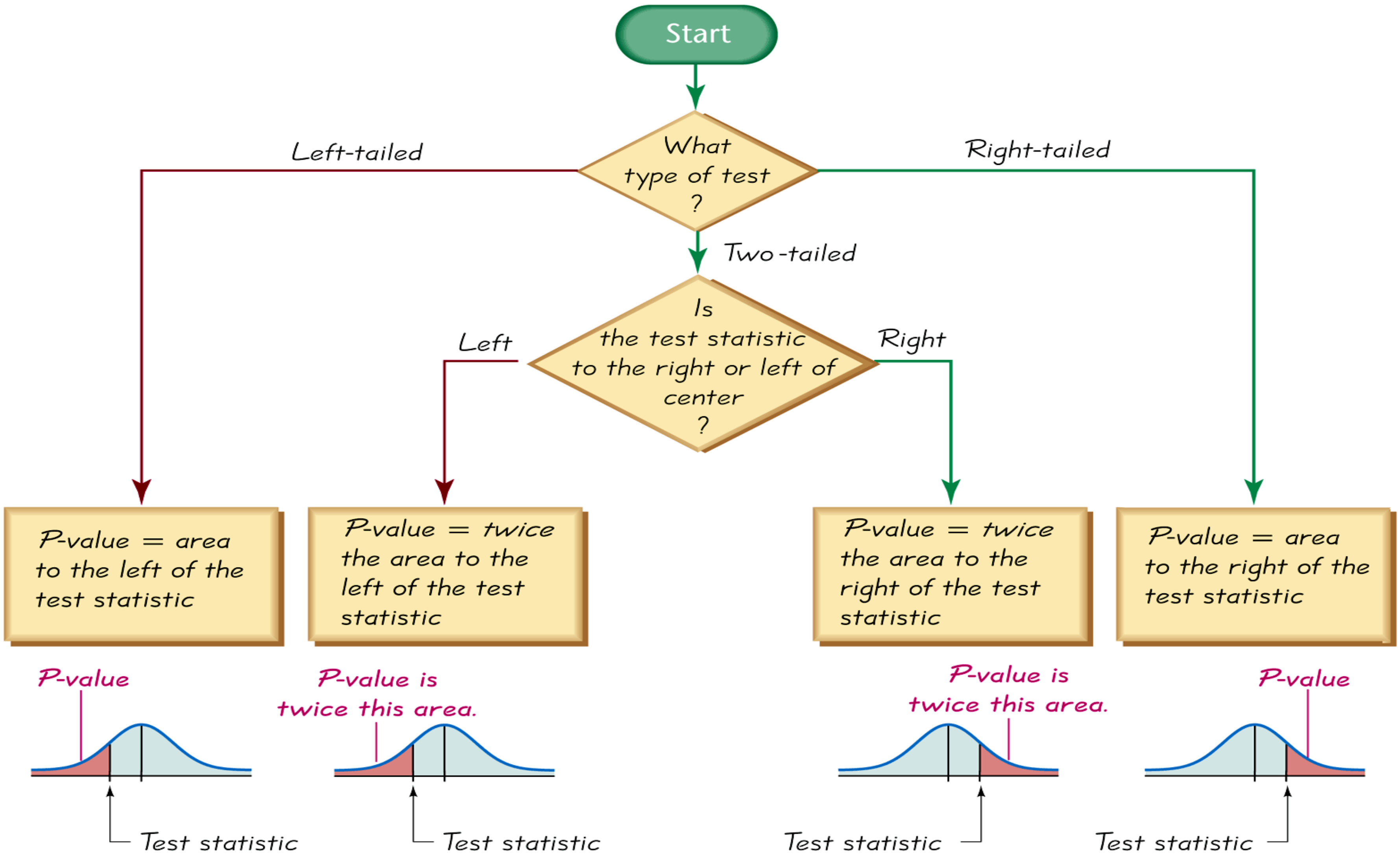
i) *P*-Value

- The *P*-value (or *p*-value or probability value) is the probability of getting a value of the test statistic that is at least as extreme as the one representing the sample data, assuming that the null hypothesis is true.

- A decision about whether to reject or fail to reject H_0 results from comparing the *P*-value to the chosen α:

H_0 should be rejected if *P*-value $\leq \alpha$,
H_0 should not be rejected if *P*-value $> \alpha$.

Procedure for Finding *P*-Value



ii) Critical Region

- The critical region (or rejection region) is the set of all values of the test statistic that cause us to reject the null hypothesis

- A critical value is any value that separates the critical region (where we reject the null hypothesis) from the values of the test statistic that do not lead to rejection of the null hypothesis.

- The critical values depend on the nature of the null hypothesis, the sampling distribution that applies, and the significance level α.

  - If the test statistic falls within the critical region, reject H_0.

  - If the test statistic does not fall within the critical region, fail to reject H_0.

Example: Find P-value

Use a 0.05 significance level and state the conclusion about the null hypothesis (reject the null hypothesis or fail to reject the null hypothesis).

i) The test statistic in a left-tailed test is $z = -1.25$.

ii) The test statistic in a right-tailed test is $z = 2.50$.

iii) The test statistic in a two-tailed test is $z = 1.75$

Example: Solution

i) The test statistic in a left-tailed test is $z = -1.25$.

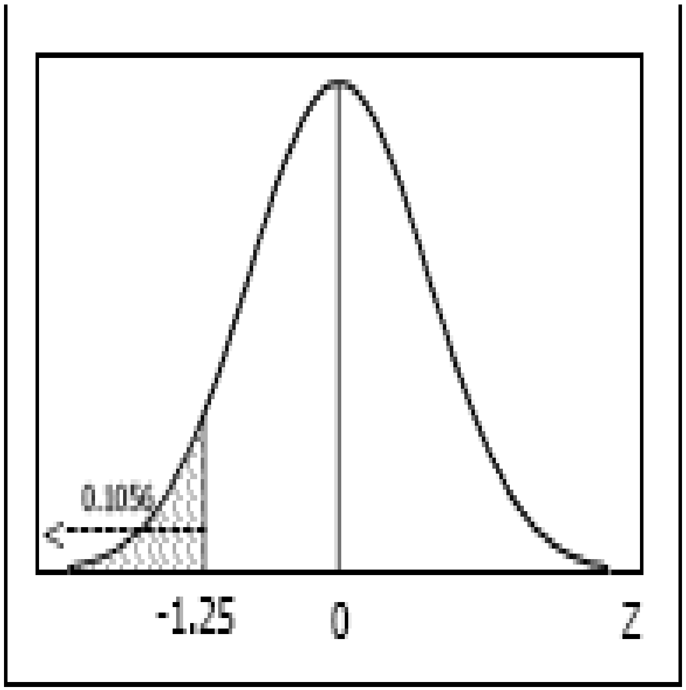
Answer:

$$P(z < -1.25) = 0.1056$$

.: p-value = 0.1056

Since 0.1056 > 0.05,

fail to reject H_0.



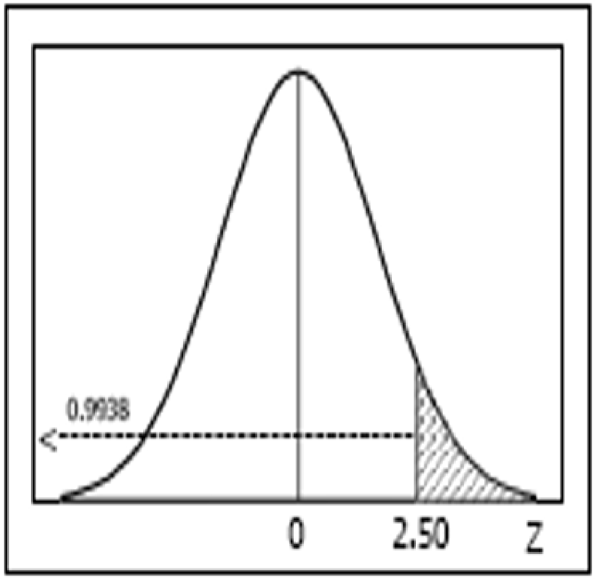
ii) The test statistic in a right-tailed test is z = 2.50.

Answer:

$$P(z > 2.5) = (1 - 0.9938) = 0.0062$$

$$\therefore P - value = 0.0062$$

$$\text{Since } 0.0062 < 0.05, \text{ reject } H_0$$



iii) The test statistic in a two-tailed test is z = 1.75.

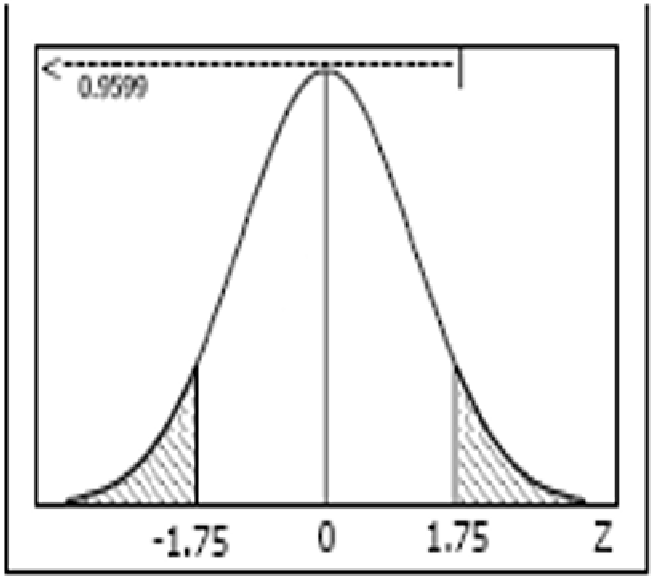
Answer:

$$P(z > 1.75) = (1 - 0.9599) = 0.0401$$

$$P(z < -1.75) = 0.0401$$

$$\therefore P - value = 2(0.0401) = 0.0802$$

Since $0.0802 > 0.05$, fail to reject H_0



Example

Consider the claim that with the XSORT method of gender selection, the likelihood of having a baby girl is different from $p = 0.5$, and use the test statistic $z = 3.21$ found from 13 girls in 14 births.

First determine whether the given conditions result in a critical region in the right tail, left tail, or two tails, then find the *P*-value.

Example - Solution

- The claim that the likelihood of having a baby girl is different from $p = 0.5$ can be expressed as $p \neq 0.5$ so the critical region is in two-tails.
- To find the *P*-value for a two-tails test, we see that the *P*-value is *twice* the area to the right of the test statistic $z = 3.21$.
- So we have:

$$P(z > 3.21) = 1 - 0.9993 = 0.0007$$

$$\therefore P\text{-value} = 2 \times 0.0007 = 0.0014 \text{ (two -tails)}$$

Error of Decision

- A Type I error is the mistake of rejecting the null hypothesis when it is actually true.
  - The symbol α (alpha) is used to represent the probability of a type I error.

- A Type II error is the mistake of failing to reject the null hypothesis when it is actually false.
  - The symbol β (beta) is used to represent the probability of a type II error.

Type I and Type II Errors

		True State of Nature	
		The null hypothesis is true	The null hypothesis is false
Decision	We decide to reject the null hypothesis	**Type I error** (rejecting a true null hypothesis) $P(\text{type I error}) = \alpha$	Correct decision
	We fail to reject the null hypothesis	Correct decision	**Type II error** (failing to reject a false null hypothesis) $P(\text{type II error}) = \beta$

Example

Assume that we are conducting a hypothesis test of the claim that a method of gender selection increases the likelihood of a baby girl, so that the probability of a baby girls is p > 0.5. Here are the null and alternative hypotheses: H_0: p = 0.5, and H_1: p > 0.5.

a) Identify a type I error.

b) Identify a type II error.

Example -Solution

a) A type I error is the mistake of rejecting a true null hypothesis, so this is a type I error: Conclude that there is sufficient evidence to support $p > 0.5$, when in reality $p = 0.5$.

b) A type II error is the mistake of failing to reject the null hypothesis when it is false, so this is a type II error: Fail to reject $p = 0.5$ (and therefore fail to support $p > 0.5$) when in reality $p > 0.5$.

Hypothesis Testing Using a Single Sample

Hypothesis Test – Single Sample

Three different types of hypothesis test- single sample:

1) Test on mean, variance known.

2) Test on mean, variance unknown.

3) Test on variance/standard deviation.

Steps in Hypothesis Testing:

1. From the problem context, identify the parameter of interest.
2. State the null hypothesis, H_0.
3. Specify an appropriate alternative hypothesis, H_1 .
4. Choose a significance level α.
5. Determine an appropriate test statistic.
6. State the rejection region for the statistic.
7. Compute any necessary sample quantities, substitute these into the equation for the test statistic, and compute that value.
8. Decide whether or not H_0 should be rejected and report that in the problem context.

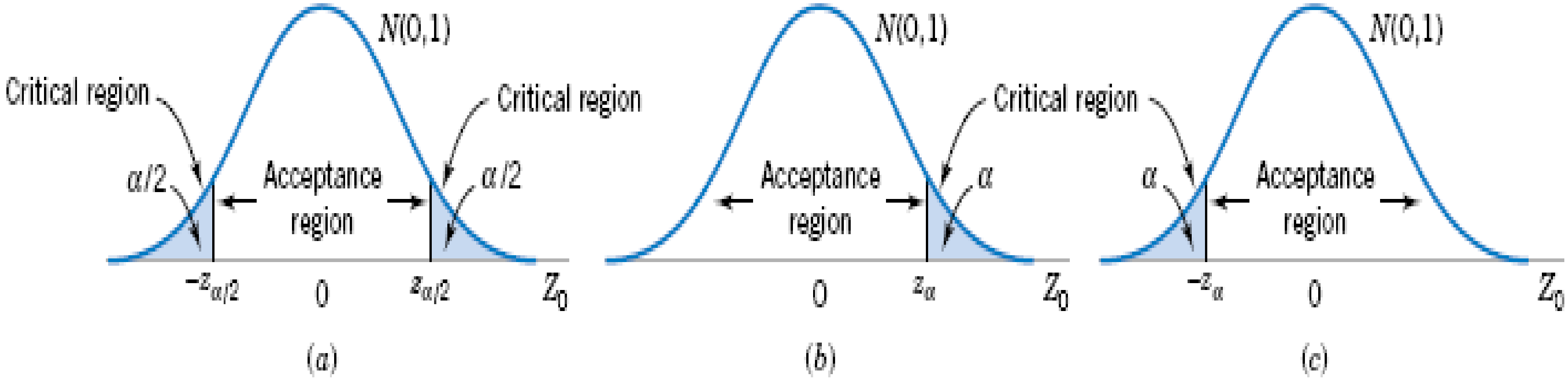
Steps 1 – 4 should be completed prior to examination of the sample data.

Test on Mean, Variance Known

When either n is large or the population distribution is approximately normal, then

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

with variance, σ^2 is known.



The distribution of Z_0 when H_0: $\mu = \mu_0$ is true, with critical region (a) the two-sided alternative H_1: $\mu \neq \mu_0$ (b) the one-sided alternative H_1: $\mu > \mu_0$ and (c) the one-sided alternative H_1: $\mu < \mu_0$.

- Null hypothesis : H_0: $\mu = \mu_o$
- Test statistic:

Alternative hypothesis	Rejection Region
$H_1: \mu \neq \mu_0$	$\lvert z \rvert \geq z_{\alpha/2}$
$H_1: \mu > \mu_0$	$z \geq z_{\alpha}$
$H_1: \mu < \mu_0$	$z \leq -z_{\alpha}$

Example

Writing a Hit Song: In the manual "How to Have a Number One the Easy Way," by KLF Publications, it is stated that a song "must be no longer than three minutes and thirty seconds" (or 210 seconds). A simple random sample of 40 current hit songs results in a mean length of 252.5 sec. Assume that the standard deviation of song lengths is 54.5 sec. Use a 0.05 significance level to test the claim that the sample is from a population of songs with a mean greater than 210 sec. What do these results suggest about the advice given in the manual?

Example - Solution

$$H_0 : \mu = 210 \sec$$

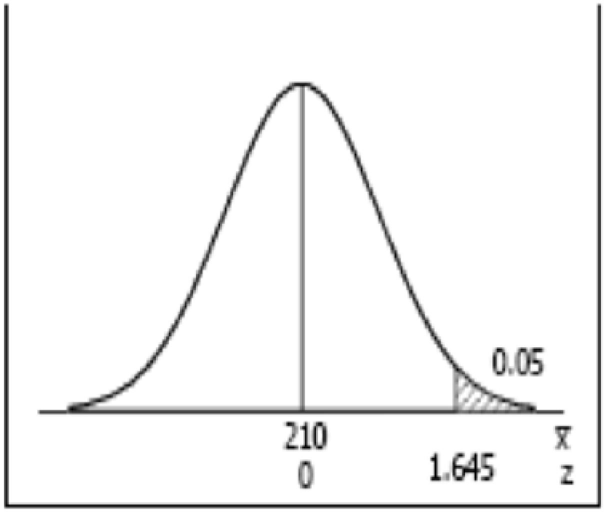
$$H_1 : \mu > 210 \sec$$

$$\alpha = 0.05$$

$$c.v = z_{0.05} = 1.645$$

$$\therefore z = \frac{\bar{x} - \mu_x}{\sigma / \sqrt{n}} = \frac{252.5 - 210}{54.5 / \sqrt{40}} = 4.93$$

$$P - value = P(z > 4.93) = 0.0001$$



Conclusion: Reject H_0 ; there is sufficient evidence to conclude that μ > 210 sec. These results suggest that the advice given in the manual is not good advice.

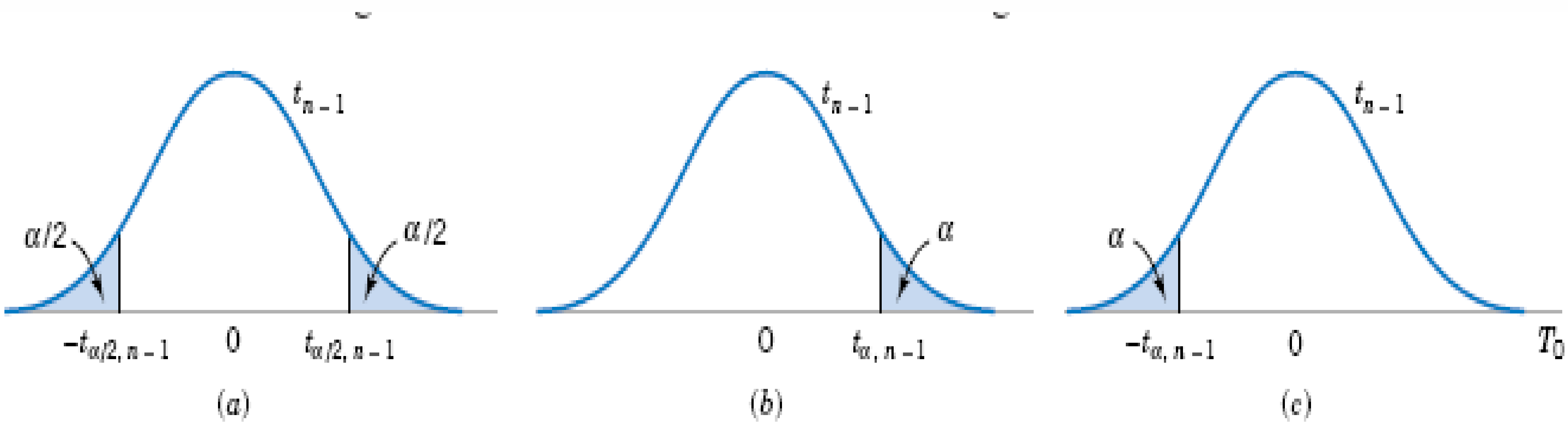
Exercise #3

Red Blood Cell Count: A simple random sample of 50 adults is obtained, and each person's red blood cell count (in cells per microliter) is measured. The sample mean is 5.23. The population standard deviation for red blood cell counts is 0.54. Use a 0.01 significance level to test the claim that the sample is from a population with a mean less than 5.4, which is a value often used for the upper limit of the range of normal values. What do the results suggest about the sample group?

Test on Mean, Variance Unknown

Test statistic concerning one mean for data that can be assumed to follow a normal distribution but variance is unknown the test statistic is:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$



The reference distribution for H_0: $\mu = \mu_0$ with critical region for (a) the two-sided alternative H_1: $\mu \neq \mu_0$ (b) the one-sided alternative H_1: $\mu > \mu_0$ and (c) the one-sided alternative H_1: $\mu < \mu_0$.

Example

The increased availability of light materials with high strength has revolutionized the design and manufacture of golf clubs, particularly drivers. Clubs with hollow heads and very thin faces can result in much longer tee shots, especially for players of modest skills. This is due partly to the “spring-like effect” that the thin face imparts to the ball. Firing a golf ball at the head of the club and measuring the ratio of the outgoing velocity of the ball to the incoming velocity can quantify this spring-like effect. The ratio of velocities is called the coefficient of restitution of the club. An experiment was performed in which 15 drivers produced by a particular club maker were selected at random and their coefficients of restitution measured. In the experiment the golf balls were fired from an air cannon so that the incoming velocity and spin rate of the ball could be precisely controlled. It is of interest to determine if there is evidence (with $\alpha = 0.05$) to support a claim that the mean coefficient of restitution exceeds 0.82. The observations follow:

0.8411	**0.8191**	**0.8182**	**0.8125**	**0.8750**	**0.8580**	**0.8532**	**0.8483**
0.8276	**0.7983**	**0.8042**	**0.8730**	**0.8282**	**0.8359**	**0.8660**	

Example - Solution

- Hypothesis:

  H_0: μ = 0.82

  H_1: μ > 0.82.
- α = 0.05.
- The test statistic is: $t = \frac{0.83724 - 0.82}{0.0237/\sqrt{15}} = 2.82$
- c.v: $t_{0.05,14} = 1.761$.
- Conclusion: Reject H_0; there is sufficient evidence that the mean coefficient of restitution exceeds 0.82.

Exercise #4

The report "Highest Paying Jobs for 2009-10 Bachelor's Degree Graduates" (National Association of Colleges and Employers, February 2010) states that the mean yearly salary offer for students graduating with a degree in accounting in 2010 is $48,722. Suppose that a random sample of 50 accounting graduates at a large university who received job offers resulted in a mean offer of $49,850 and a standard deviation of $3300. Do the sample data provide strong support for the claim that the mean salary offer for accounting graduates of this university is higher than the 2010 national average of $48,722? Test the relevant hypothesis using $\alpha = 0.05$

Test on Variance / Standard Deviation

- Test statistic: Chi-Square Distribution

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

n = sample size

s = sample standard deviation

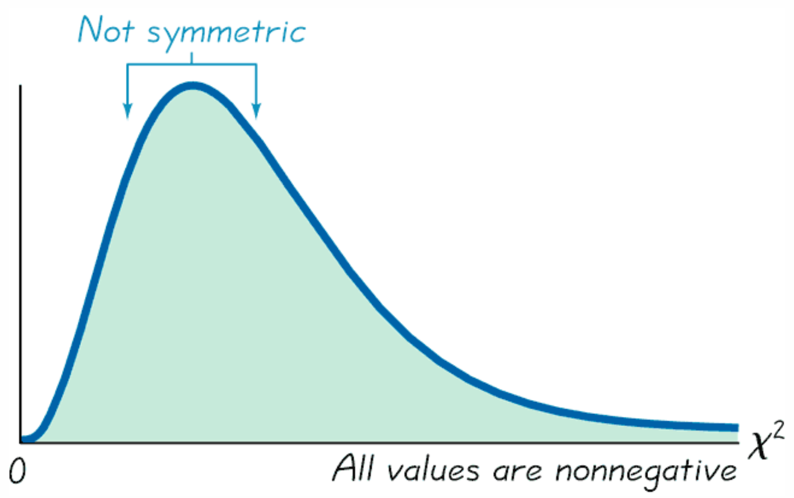
s^2 = sample variance

σ = claimed value of the population standard deviation

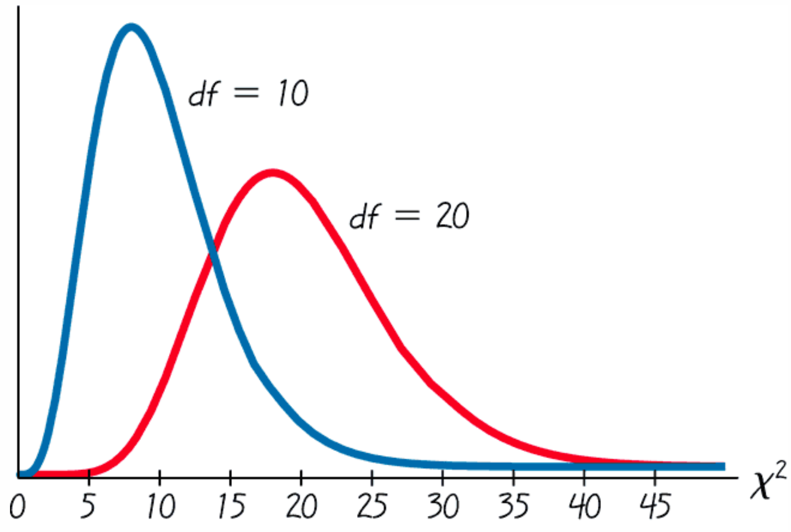
σ^2 = claimed value of the population variance

Properties of Chi-Square Distribution

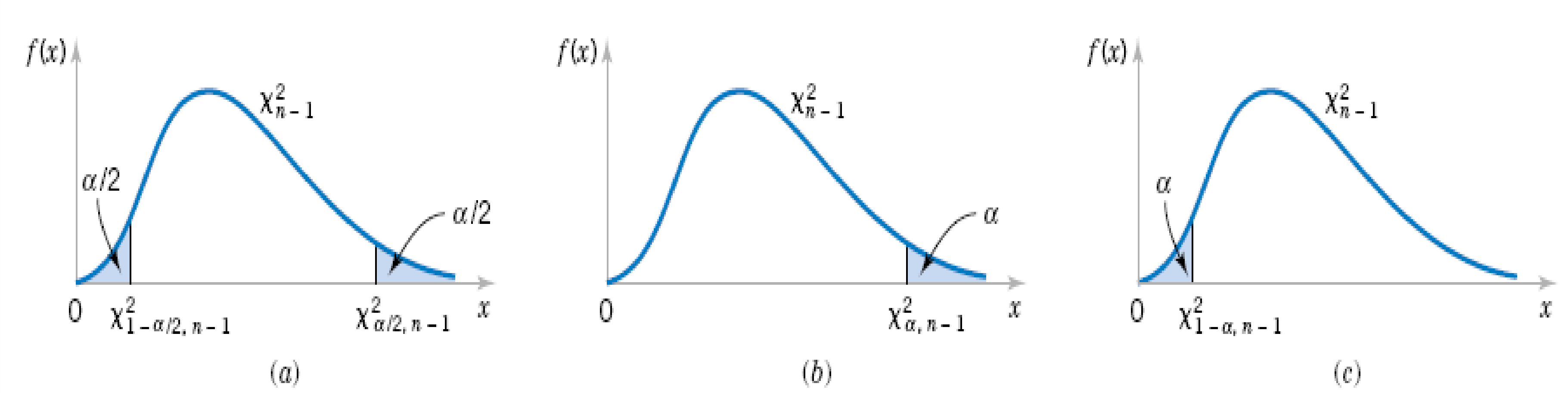
Properties of the Chi-Square Distribution



Chi-Square Distribution for 10 and 20 *df*



Different distribution for each number of *df*.



The reference distribution for H_0: $\sigma^2 = \sigma_0^2$ with critical region for (a) the two-sided alternative H_1: $\sigma^2 \neq \sigma_0^2$ (b) the one-sided(right) alternative H_1: $\sigma^2 > \sigma_0^2$ and (c) the one-sided (left) alternative H_1: $\sigma^2 < \sigma_0^2$.

Example

A police chief claims that the standard deviation in the length of response times is less than 3.7 minutes. A random sample of nine response times has a standard deviation of 3.0 minutes.

At $\alpha = 0.05$, is there are enough evidence to support the police chief's claimed? Assume the population is normally distributed.

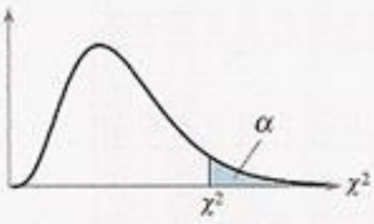
Example: Solution

- Hypothesis statement:

$$H_0: \sigma = 3.7$$
$$H_1: \sigma < 3.7$$

- Test statistic:

$$.: \chi^2 = \frac{8(3)^2}{(3.7)^2} = 5.259$$



Right tail

Degrees of freedom	α									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

Why are we using right tail?

This is the way the chi-square table is designed. So if its left-tail 5%, use the number for right-tail 95% (i.e., 1-α). Thus, $\chi^2_{0.95,8} = 2.733$

- Conclusion: Since 2.733 < 5.259, fail to reject H_0. There is sufficient evidence that support the standard deviation in the length of response times is 3.7 minutes.

Example

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 the standard deviation > 0.9 year? Use $\alpha = 0.05$ level of significance.

Example: Solution

- Hypothesis statement:

  H_0: $\sigma^2 = 0.81$

  H_1: $\sigma^2 > 0.81$

- $\alpha = 0.05$

- Test statistic: $s^2 = (1.2)^2 = 1.44$; $\sigma^2 = (0.9)^2 = 0.81$

$$\chi^2 = \frac{9(1.44)}{0.81} = 16.0$$

TABLE A-4 Chi-Square (χ^2) Distribution

Degrees of Freedom	Area to the Right of the Critical Value									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

- Critical value: $\chi^2_{0.05,9} = 16.919$
- Conclusion: Since 16.0 < 16.919, fail to reject H_0. There is **no** sufficient evidence that support the life of car batteries has standard deviation more than 0.9 years.

Exercise #5

A dairy processing company claims that the variance of the amount of fat in the whole milk processed by the company is no more than 0.25. You suspect this is wrong and find that a random sample of 41 milk containers has a variance of 0.27. At $\alpha = 0.05$, is there are enough evidence to reject the company's claimed? Assume the population is normally distributed.