Hypothesis testing

A particular machine is expected to produce metal rods of length 3 m. After a few years of operation, it is suspected that the lengths of rods produced have decreased and the machine needs to be serviced. To determine whether there is sufficient evidence to support the suspicion, we can conduct a **hypothesis test** using the average length of a random sample of 500 (for instance) rods produced.

In hypothesis testing, we begin with an assumption (known as the **null hypothesis** which is denoted by H_0) and use the sample data and our knowledge of statistical theory to decide whether the sample data support the null hypothesis.

[Remind students: when you draw conclusions about hypothesis testing, make sure you say "reject H_0" or "do not reject H_0" - AVOID saying "accept H_0"]

In hypothesis testing, a hypothesis may be accepted or rejected not with certainty but with confidence that the likelihood of error in making the decision is small.

Null hypothesis (H_0) is the statistical hypothesis about the value of the population mean μ (remind students that null hypothesis MUST be stated with the Greek symbol to represent population parameter) which we wish to test.

$$Eg. \qquad H_0 : \mu = 2$$

$$(it\ is\ wrong\ to\ write \quad H_0 :\ \bar{x} = 2)$$

Alternative hypothesis (denoted by H_a) is the statistical hypothesis that we wish to test against the null hypothesis. When the null hypothesis is rejected, we are in favor of the alternative hypothesis. In general, **it is the alternative hypothesis which the researcher hopes to "prove"' is true**.

In the above example about metal rods, the hypothesis test is carried out on the mean length of rods produced by the machine.

Let μ be the average length (in m) of rods produced by the machine.

$$H_0: \ \mu = 2$$
$$H_a: \ \underline{\mu < 2}$$

Note: we are interested in testing if population mean has decreased, hence we use $<$.

Exercise: In the following examples, formulate the null and alternative hypotheses:

(a) A random sample of students was each given an IQ test to decide whether they have a mean IQ of 100 or whether their IQ is above 100.

Let μ be the mean IQ score of the population (of students).

$$H_0: \ \mu = 100$$
$$H_a: \ \mu > 100$$

(b) Suppose that according to national standards, 5-year old children of heights between 86.5 cm to 91.5 cm have a mean weight of 13 kg. The weights of a random sample of 5-year old children of heights 86.5 cm- 91.5 cm are collected. Are these children below average weight for their height?

Let μ be the mean weight of all 5-year old children with heights between 86.5 cm to 91.5 cm.

$$H_0: \ \mu = 13$$
$$H_a: \ \mu < 13$$

(c) A machine produces components of mean length 10 cm if it is operating correctly. A random sample of components is produced by the machine is taken and each of their lengths measured. Is the machine operating correctly?

Let μ be the mean length of components for the population.

$$H_0: \ \mu = 10$$
$$H_a: \ \mu \neq 10$$

(d) The time for a worker to repair an electrical instrument is normally distributed. The repair times for 10 such instruments chosen at random are taken. Suppose the worker claims that his average repair time for the instrument is not more than 200 hours. Test if there is sufficient evidence to dispute the worker's claim.

Let μ be the average repair tie for the instrument.

$$H_0: \ \mu = 200$$
$$H_a: \ \mu > 200$$

Note! You want to "show" his hours are > 200, disputing his claim

Two-sided tests

You can tell by looking at the alternative hypotheses. If it is of the form with an inequality $\neq$ sign, then it is a two sided test.

$$Eg. \quad H_a: \ \mu \neq 100 \leftarrow 2 \ \textit{sided test}$$

One-sided tests

You can tell by looking at the alternative hypotheses. If H_a is of the form $\mu < \mu_0$, or H_a is of the form $\mu > \mu_0$, it is a one sided test.

$$Eg. \quad H_a: \ \mu \ > 5 \leftarrow 1 \ \textit{sided test}$$

Significance level of a test

The level of significance is denoted by α (eg, α=0.05). It is the probability of rejecting H_0 when it is actually true (that is, in probability notation, you can think of it as P(rejecting $H_0|H_0$ is true)=α).

This level of significance refers to how much error we allow for wrongly rejecting H_0.

[Explain to class that testing a hypothesis can result in two possible errors: rejecting a null hypothesis that is true (α, committing such an error is called Type I error), or not rejecting null hypothesis when it is false (we commit a Type II error in this case)].

		Truth about population	
		H_0 true	H_a true
Decision based on sample	reject H_0	Type I error	correct decision
	do not reject H_0	correct decision	Type II error

In many introductory Statistics course, we learn testing hypothesis using 3 different methods:

- Using confidence intervals to make inferences about hypothesis tests
- Using rejection regions to perform hypothesis testing
- Using p-values to perform hypothesis testing.

However, for STAT 241/251, we learn only the first two methods (the first method using confidence intervals to make inferences about hypothesis testing is found in free coursenotes; the 2^{nd} method using rejection region is not found in your free coursenotes but I will be teaching it in lectures and you can find this in any standard textbook on Statistics).

The next section is very important (page 134 in your course notes):

<u>How to perform 2-sided tests</u>

Method 1 (Using confidence intervals):

Step 1: Construct a (1-α)100% confidence interval for μ

Step 2: Reject $H_0 : \mu = \mu_0$ if μ_0 lies outside the interval.

If you've the confidence interval handy, you can use it to make inferences regarding hypothesis test.

Method 2 (Using rejection region):

See procedure (last 2 pages of this lecture). Here is an example using t-test (but z is also very similar)

<u>Example</u>

$$\text{Assumptions: } \begin{cases} (1)\ \textit{Simple random sample} \\ (2)\ \textit{Normal distribution} \\ (3)\ \textit{Population variance unknown} \quad \leftarrow \textit{so we use } t-\textit{test} \end{cases}$$

Let μ be ...

$H_0 : \mu = 8.3$

μ_0

$H_a : \mu \neq 8.3$

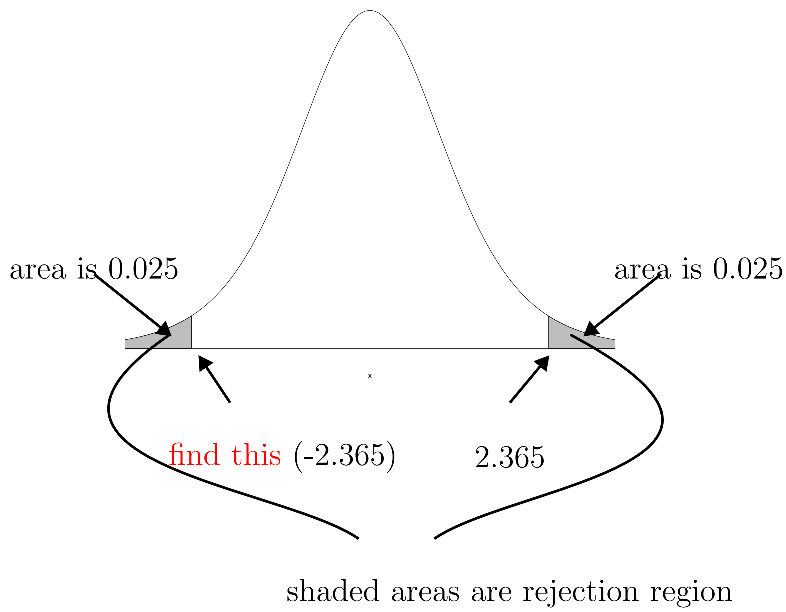
[Let's say question tells to perform hypothesis test at 5% significance level, so α=0.05]

Under H_0, our test statistic is $T = \frac{\bar{x}-\mu_0}{s/\sqrt{n}},\ df = n-1$

[Let's say $\bar{x} = 5.0, s = 3.63, n = 8$]

$$t_{obs} = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{5.0-8.3}{3.63/\sqrt{8}} = -2.57$$

[Draw a picture to guide yourself where the rejection region is]



We look up $t_{\alpha/2,n-1} = t_{0.025,7}$. We then check if our t_{obs} falls inside the rejection region.If t_{obs} is inside rejection region, we reject H_0 (otherwise, we do not reject H_0).

In this example, t_{obs} is inside rejection region $\Rightarrow$ we reject H_0.

Conclusion: At the 5% significance level, the data provides sufficient evidence that the mean diameter is not equal to 8.3.

How to perform 1-sided tests:

Method 1 (Using confidence intervals):

Your free course-notes use this method.

Step 1: Construct a $(1-2\alpha)100\%$ **confidence interval for** μ

Step 2: For $H_a < \mu_0$

Reject H_0 **if** μ_0 **is larger than the upper end of the interval**

For $H_a > \mu_0$

Reject H_0 **if** μ_0 **is smaller than the lower end of the interval**

(See Pg 135 of free course-notes for an example using this method to perform 1-sided test).

Method 2 (Using rejection region):

Personally, I find it easier to use rejection region method for 1-sided tests.

Assumptions: ① Simple random sample

② Population is normally distributed

③ σ is unknown ← so we use t-test

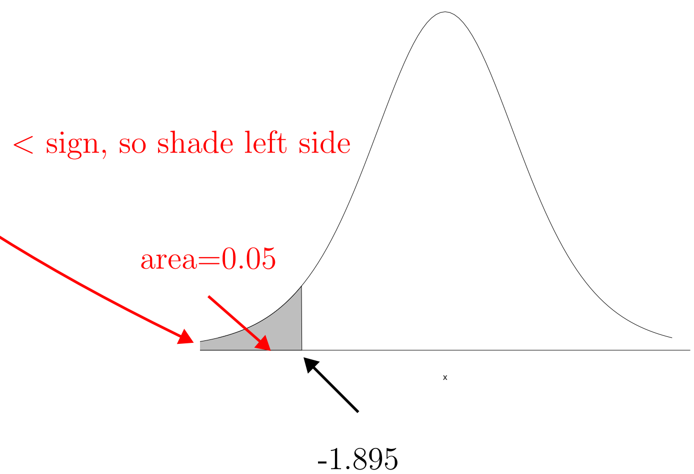
$H_0 : \mu = 8.3$

$H_a : \mu < 8.3$

[say $\alpha = 0.05, \bar{x} = 5.0, s = 3.63, n = 8$]

Under H_0, our test statistic is $T = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$

$$t_{obs} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = -2.57$$



We look up $t_{0.05,7}$ (remember to use $-t_{0.05,7}$ if rejection region is on the left side).

Since $t_{obs} = -2.57$ falls within the rejection region, we reject H_0.

At the 5% significance level, the data provides evidence that the mean is less than 8.3.

Example 8.1+8.2+8.3 in course notes:

A scientist wishes to detect small amounts of contamination in the environment. To test her measurement procedure, she spiked 12 specimens with a known concentration (2.5 ug/l of lead). The readings for the 12 specimens are

1.9 2.4 2.2 2.1 2.4 1.5 2.3 1.7 1.9 1.9 1.5 2.0

$\bar{y} = 1.9833$ $s^2 = 0.09787879$

Make sure you know how to calculate the sample mean and the sample variance for exams. Also, if your scientific calculator has such functions, make sure you learn how to verify your answers! (NOTE: graphing calculators such as TI-83 an TI-84 are not allowed in exams!!! Penalty applies when caught as such calculators are capable of performing integration and linear regression - Ch 11)

(a) Test at level $\alpha = 0.05$

$$H_0 : \mu = 2.5$$

$$H_a : \mu \neq 2.5$$

(b) Test at level $\alpha = 0.05$

$$H_0 : \mu = 2.3$$

$$H_a : \mu < 2.3$$

(a) Note: This is a two-sided test

Step1: Construct a $(1-\alpha)100\%$ CI for μ.

Since $\alpha = 0.05$, we construct a 95% CI for μ.

Let μ be the <u>true mean</u> of the scientist's measurement.
population mean

95% CI for μ is:

n-1

$$\bar{y} \pm t_{0.025,11} \cdot se(\bar{y})$$

$$1.9833 \pm 2.201 \times 0.09031371$$

$$=(1.785, 2.182)$$

$H_0 : \mu = 2.5$

$H_a : \mu \neq 2.5$

Step2: Since 2.5 is not in the interval (1.785, 2.182), we reject H_0. There is statistical evidence that the population mean is not equal to 2.5.

(b) This is a one-sided test

$$H_0 : \mu = 2.3$$

$$H_a : \mu < 2.3$$

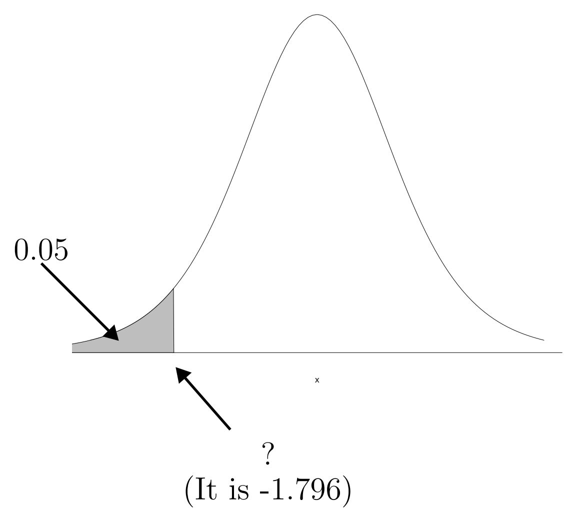
Say we wish to test at $\alpha = 0.05$

Test statistic:

$$t_{obs} = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{1.9833 - 2.3}{\sqrt{0.09787879/12}} = -3.51$$

We look up $t_{11,0.05} = 1.796$, and since $H_a : \mu < 2.3$, our critical region is the area in the left of $t = -1.796$

Since $t_{obs} = -3.51$ is under critical region, we reject H_0 and conclude there is statistical evidence that the population mean is less than 2.3 ug/l.



Procedure for one-sample z-test

State Assumptions: 1. Simple random sample

2. Normal population

3. σ known $\leftarrow$ when σ known or large samples, we perform z-test

Step 1: (define μ and write out H_0 and H_a)

Let μ be ...

always write = sign for H_0

$H_0 : \mu = \mu_0$ $\leftarrow$ fill in an appropriate number for H_0 based on question

H_a : you've to decide among one of 3 choices

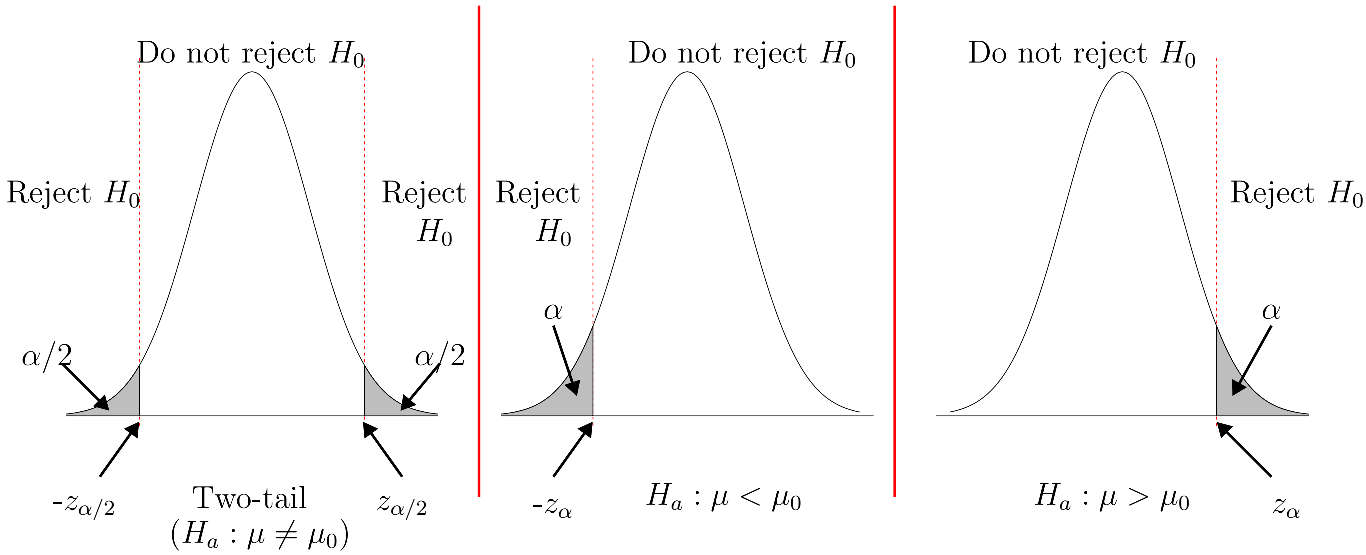
$H_a : \mu \neq \mu_0$	OR	$H_a : \mu < \mu_0$	OR	$H_a : \mu > \mu_0$
$(2 - tail\ test)$		$(1 - tail\ test)$		$(1 - tail\ test)$

Step 2: Decide on significance level, α. Usually α is given in question (eg, $\alpha = 0.05$ or $\alpha = 0.01$). If not given, use $\alpha = 0.05$ which is commonly used.

Step 3: Under H_0, our test statistic is $Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$. Calculate $z_{obs} = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$=some number.

observed

Step 4: Identify Rejection Region



If z_{obs} falls within the rejection region, reject H_0; otherwise, we do not reject H_0.

for instance

Step 5: Draw conclusion. Eg, we reject H_0 and conclude that at the 5% significance level, the data provide sufficient evidence to conclude that ....

eg the mean diameter is greater than 5mm

Procedure for one-sample t-test

State assumptions: 1. Simple random sample

2. Normal population

3. σ unknown $\leftarrow$ use t-test when σ unknown and small sample size

Step 1: (define μ and state H_0 and H_a)

Let μ be ...

always write = for H_0

$H_0 : \mu = \mu_0$ $\leftarrow$ fill in appropriate number from question

H_a: one of 3 choices

$$H_a : \mu \neq \mu_0 \qquad OR \qquad H_a : \mu < \mu_0 \qquad OR \qquad H_a : \mu > \mu_0$$

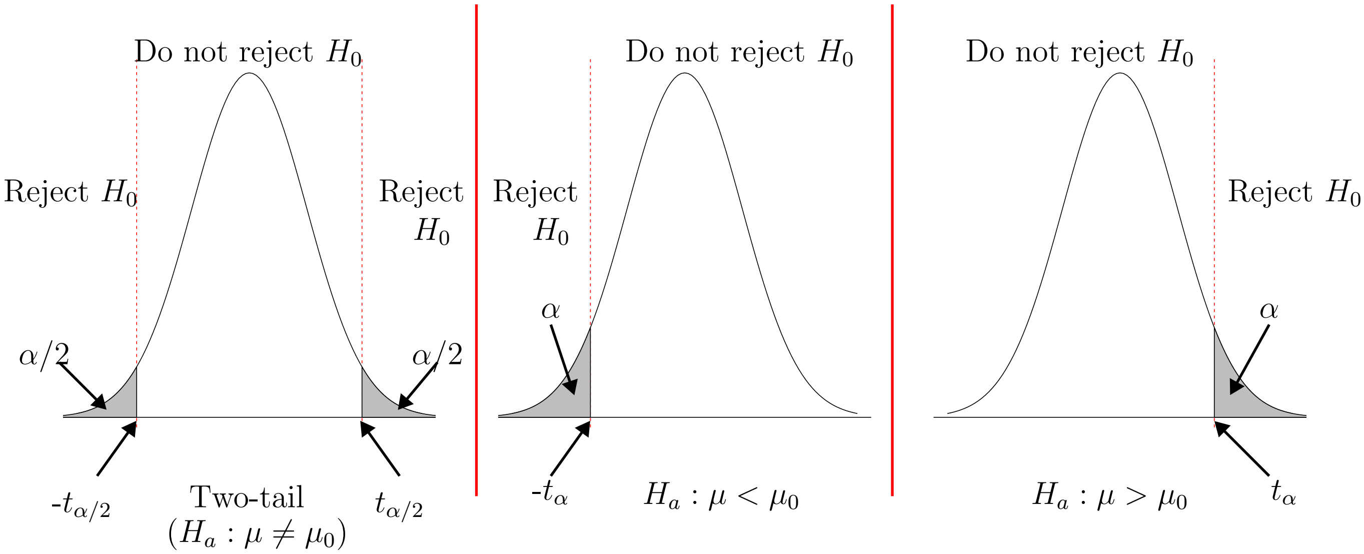
$$(2 - tail\ test) \qquad (1 - tail\ test) \qquad (1 - tail\ test)$$

Step 2: Decide on significance level, α. Usually α is given (eg $\alpha = 0.05$ or $\alpha = 0.01$). If not given, assume $\alpha = 0.05$ in question.

Step 3: Under H_0, our test statistic is $T = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}, df = n - 1$. Calculate $t_{obs} = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = some\ number$.

observed

Step 4: Identify rejection region



If t_{obs} falls inside the rejection region, reject H_0. Otherwise, we do not reject H_0.

for instance

Step 5: Draw conclusion. Eg, we reject H_0 and conclude that at the 5% significance level, the data provide sufficient evidence to conclude that .....

eg the mean diameter is greater than 5mm