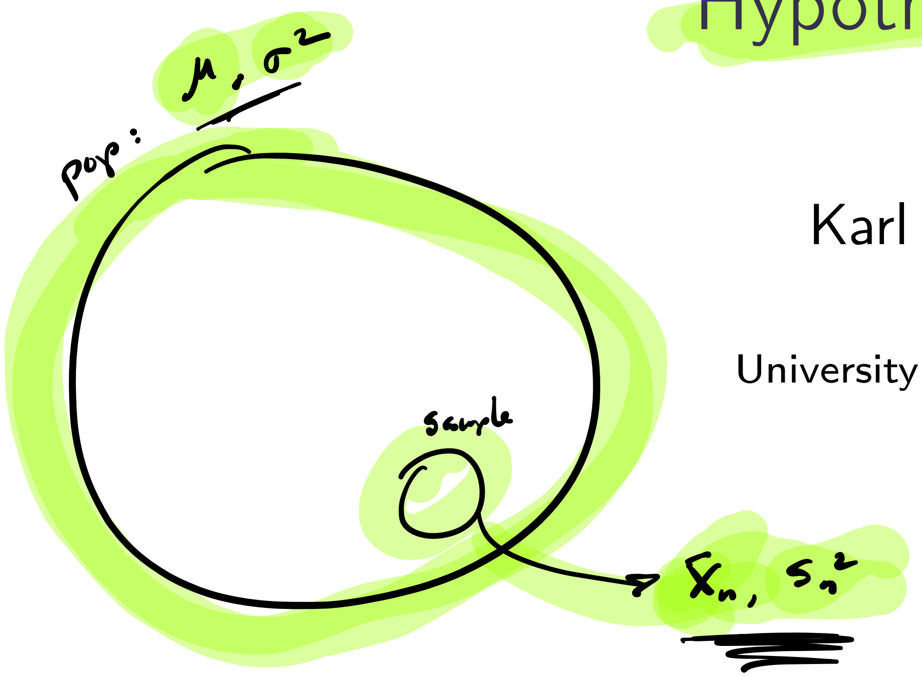
STAT 515 fa 2023 Lec 14 slides

Hypothesis testing



Karl B. Gregory

University of South Carolina

C.I.

$$\bar{X}_n \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\bar{X}_n \pm t_{n-1, \alpha/2} \frac{s_n}{\sqrt{n}}$$

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

informed guess

A *statistical inference* is a conclusion about a pop. parameter based on a rs.

Specifically, a decision concerning contradictory statements about the parameter:

- The *null hypothesis* H_0. ← represents "status quo", no finding.
- The *alternate hypothesis* H_1. ← represents some new "finding"

The decision is whether to

1. Reject H_0, thereby concluding that H_1 is true.
2. Not reject H_0, thereby not concluding anything.

A *test of hypotheses* is a rule for when to reject H_0 based on the data.

Exercise: We want to know whether a coin is unfair. Let p be the prob. of heads.

We want to test

$$H_0\colon p = 1/2 \text{ versus } H_1\colon p \neq 1/2.$$

① Reject H_0, conclude H_1

② Fail to reject H_0, conclude nothing.

Suppose we toss the coin 100 times. Discuss the following:

LARGE prob of Type II error

$\boxed{\# \text{ heads} \geq 79 \text{ or } \leq 1}$

$p = .6$

1. Reject or fail to reject H_0 if 51 heads observed?
2. Reject or fail to reject H_0 if 60 heads observed?
3. Reject or fail to reject H_0 if 90 heads observed?
4. Reject or fail to reject H_0 if 50 heads observed?
5. What possible evidence could convince us that $p = 1/2$?
6. If the coin is fair, find prob. of observing a # of heads ≥ 60 or ≤ 40.

$X \sim \text{Binomial}(n=100, p=1/2)$ $P(X \geq 60) = 1 - P(X \leq 59) = 1 - \text{pbinom}(59, 100, 1/2) = 0.028$

$P(X \leq 40) = \text{pbinom}(40, 100, 1/2) = 0.028$

D = After − Before

$\boxed{\mu < 0}$

Exercise: Is a treatment effective in lowering cholesterol levels? Let μ represent the average difference (after-minus-before treatment) in cholesterol levels.

We want to test

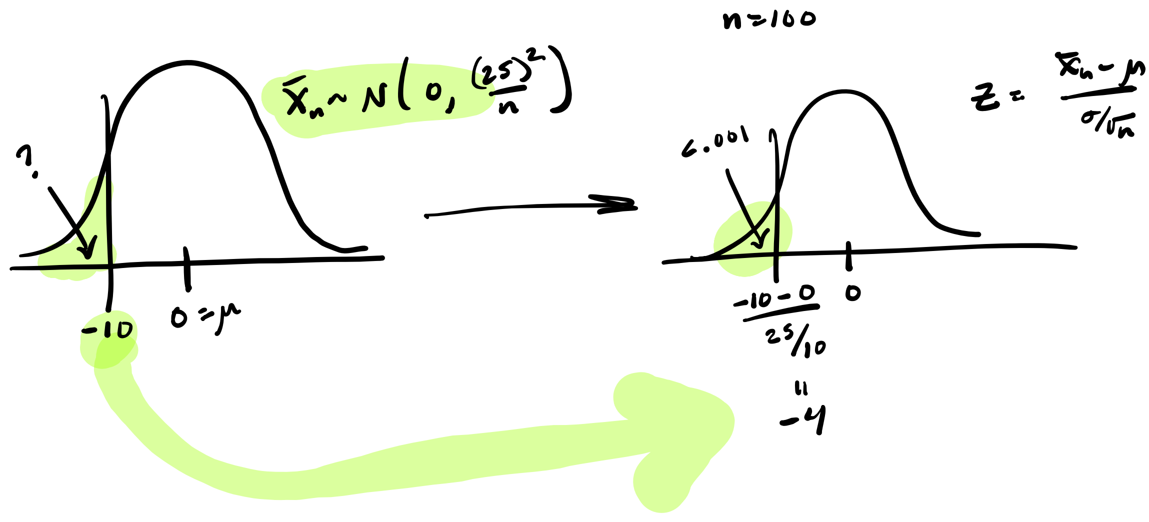
μ_0 μ_0

$$H_0\colon \mu \geq 0 \text{ versus } \boxed{H_1\colon \mu < 0.}$$

Suppose we obtain $\bar{X}_n$ from $n = 100$ subjects. Discuss the following:

average difference After − Before.

1. Reject or fail to reject H_0 if $\bar{X}_n = 10$? ← Not reject Ho. No evidence against Ho
2. Reject or fail to reject H_0 if $\bar{X}_n = -10$? ← We don't know until we account for the variance.
3. If the changes in chol. level are $N(\mu = 0, \sigma^2 = (25)^2)$, find $P(\bar{X}_n < -10)$.



$\Rightarrow$ reject H_0.

Trade-off between making Type I error prob. small and
Type II error prob small.

	H_0 true	H_0 false
reject H_0	Type I error	correct decision
Fail to reject H_0	correct decision	Type II error

Our data may lead us to an incorrect decision about H_0 and H_1:

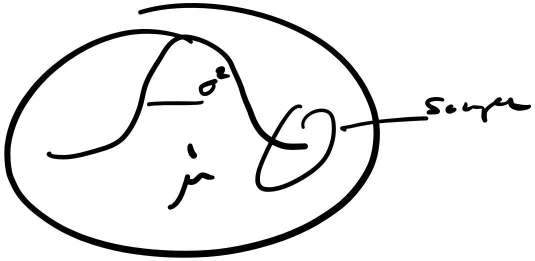
- A *Type I error* is rejecting H_0 when H_0 is true.
- A *Type II error* is failing to reject H_0 when H_0 is false.

Make table summarizing possible outcomes of inference.

We like to calibrate our tests of hypotheses such that $P(\text{Type I error}) \leq \alpha$.

Then we call α the *significance level* of the test.

Suppose $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, with σ known.



We will consider null and alternate hypotheses of the form

$$\begin{matrix} H_0\colon \mu \geq \mu_0 \\ H_1\colon \mu < \mu_0 \end{matrix} \quad \text{or} \quad \begin{matrix} H_0\colon \mu = \mu_0 \\ H_1\colon \mu \neq \mu_0 \end{matrix} \quad \text{or} \quad \begin{matrix} H_0\colon \mu \leq \mu_0 \\ H_1\colon \mu > \mu_0. \end{matrix}$$

Here μ_0 is a value specified by the researcher called the *null value*.

Exercise: For each set of hypotheses, find a test based on the *test statistic*

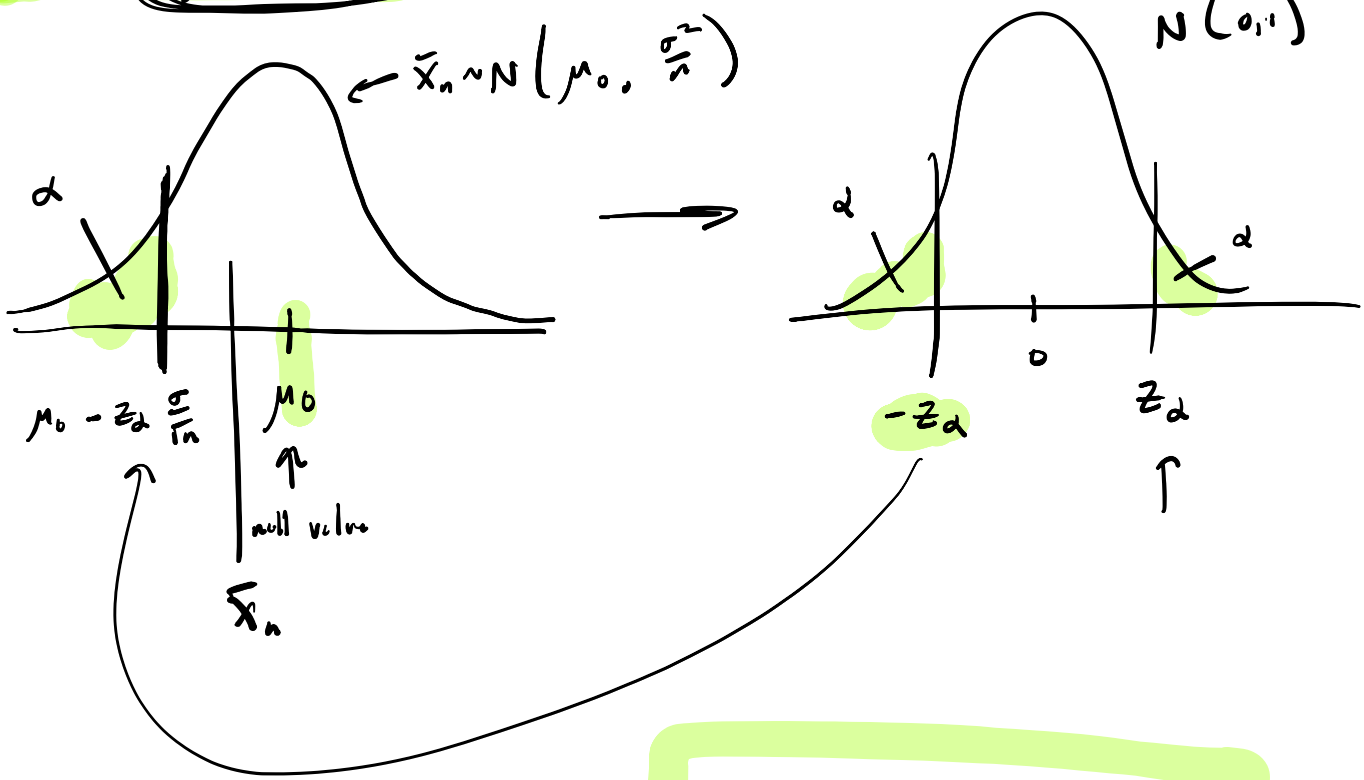
$$\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$$

with $P(\text{Type I error}) \leq \alpha$.

$H_0: \mu \geq \mu_0$

$H_1: \mu < \mu_0$

$z = \dfrac{\bar{x}_n - \mu}{\sigma/\sqrt{n}}$ $\qquad$ $\bar{x}_n = \dfrac{\sigma}{\sqrt{n}} z + \mu$

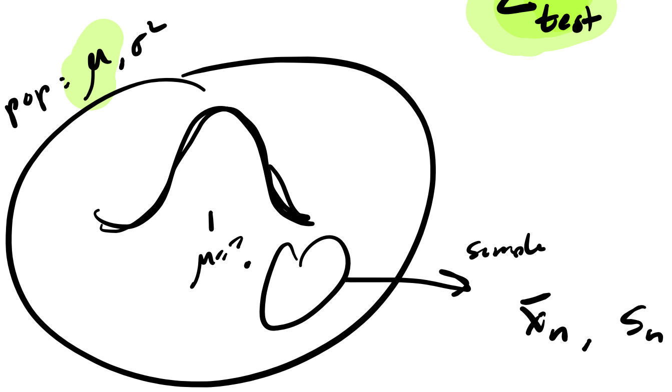


Reject H_0 when $\boxed{\bar{x}_n < \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}}$

Equivalently

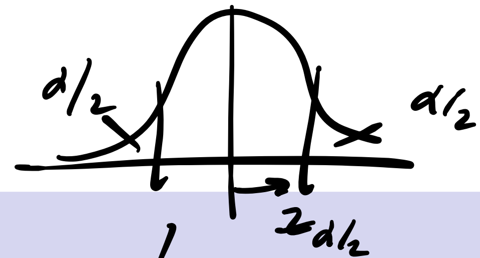
Reject H_0 when $\underbrace{\dfrac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}}_{\text{Test statistic}} < -z_\alpha$

$Z_{test} = \dfrac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}$ ← Asks how many std. devs is $\bar{x}_n$ from μ_0?



μ_0 = "mu nought"

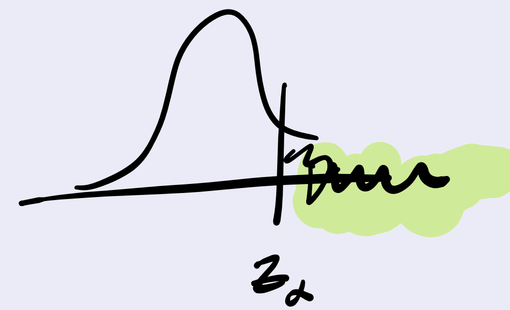
Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, σ^2 known.



Tests about μ when $\boxed{\sigma \text{ is known}}$

For some null value μ_0, define the test statistic

$$Z_{\text{test}} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$



Then the following tests have $P(\text{Type I error}) \leq \alpha$.

	two-sided test	
$H_0\colon \mu \geq \mu_0$ $\boxed{H_1\colon \mu < \mu_0}$	$H_0\colon \mu = \mu_0$ $H_1\colon \boxed{\mu \neq \mu_0}$	$H_0\colon \mu \leq \mu_0$ $H_1\colon \mu > \mu_0$
Reject H_0 if $\boxed{Z_{\text{test}}} < -z_\alpha$	Reject H_0 if $\lvert Z_{\text{test}} \rvert > z_{\alpha/2}$	Reject H_0 if $Z_{\text{test}} > z_\alpha$

The *test statistic* is the quantity based on which we decide to reject/not reject H_0.

The $-z_\alpha$, $z_{\alpha/2}$, z_α (to which we compare the test stat) are called *critical values*.

$H_0: \mu \geq 157$
$H_1: \mu < 157$

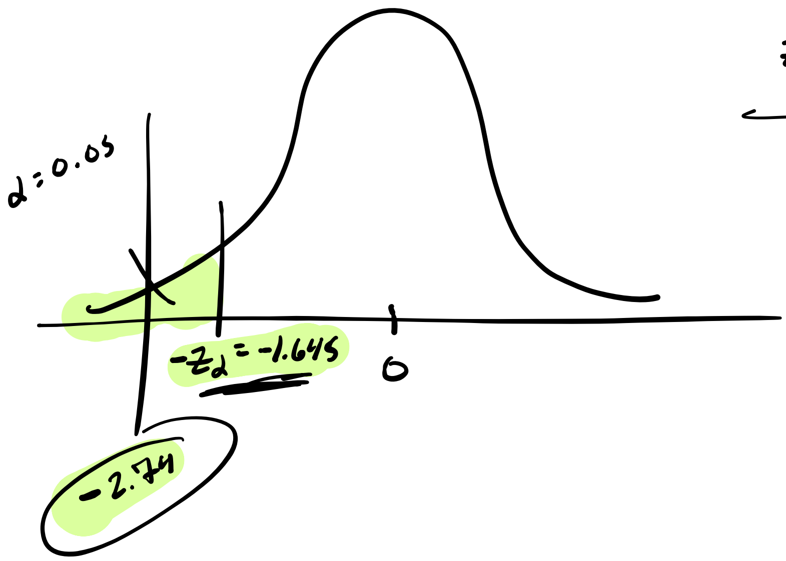
$H_0: \mu = 157$
$H_1: \mu < 157$

some textbooks.

Exercise: Suppose a bottler of soft-drinks claims that its bottling process results in an internal pressure of 157 psi with standard deviation $\sigma = 3$ psi. You want to know whether the mean pressure is less than 157 (Ex 6.92 in [1]).

1. What are the relevant hypotheses?
2. Based on a sample of size $n = 40$ you get $\bar{X} = 155.7$. Suppose $\sigma = 3$. What is your inference at the $\boxed{\alpha = 0.05}$ significance level? $\alpha = 0.01$
3. Identify the following as a correct decision, a Type I error, or a Type II error:
   a. Suppose $\mu = 157.5$ and your data leads you to reject H_0. Type I
   b. Suppose $\mu = 157.5$ and your data leads you to not reject H_0. Correct
   c. Suppose $\mu = 156.5$ and your data leads you to reject H_0. Correct
   d. Suppose $\boxed{\mu = 156.5}$ and your data leads you to not reject H_0. Type II.

$$Z_{test} = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \qquad Z_{test} = \frac{155.7 - 157}{3/\sqrt{40}} = -2.74$$



$z_{.05} = 1.645$

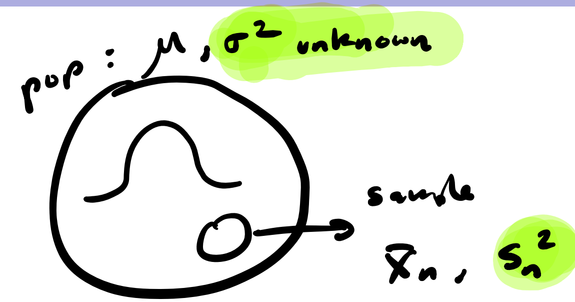
Yes, reject H_0.

Do at home

Exercise: A machine should produce ball bearings with diameters having mean 0.5 in. and std. dev. $\sigma = 0.001$ in. Is the mean truly 0.5 in.? (Ex 6.84 in [1]).

1. What are the relevant hypotheses?
2. Suppose the diameters are Normal and you get $\bar{X}_n = 0.499$ with $n = 5$. Decide whether to reject or not reject H_0 at $\alpha = 0.05$.
3. Compute a 95% CI for μ based on $\bar{X}_n = 0.499$ with $n = 5$.
4. Identify the following as a correct decision, a Type I error, or a Type II error:
   a. Suppose $\mu = 0.52$ and your data leads you to reject H_0.
   b. Suppose $\mu = 0.52$ and your data leads you to not reject H_0.
   c. Suppose $\mu = 0.50$ and your data leads you to reject H_0.
   d. Suppose $\mu = 0.50$ and your data leads you to not reject H_0.

Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, σ^2 unknown.



Tests about μ when σ is unknown

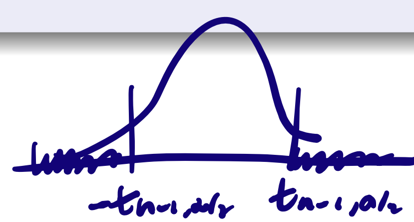
For some null value μ_0, define the test statistic

$$\boxed{T_{\text{test}}} = \frac{\bar{X} - \mu_0}{S_n/\sqrt{n}}.$$

Then the following tests have $P(\text{Type I error}) \leq \alpha$.

$H_0\colon \mu \geq \mu_0$ $H_1\colon \mu < \mu_0$	$\boxed{H_0\colon \mu = \mu_0 \quad H_1\colon \mu \neq \mu_0}$	$H_0\colon \mu \leq \mu_0$ $H_1\colon \mu > \mu_0$
Reject H_0 if $T_{\text{test}} < -t_{n-1,\alpha}$	Reject H_0 if $\lvert T_{\text{test}} \rvert > t_{n-1,\alpha/2}$	Reject H_0 if $T_{\text{test}} > t_{n-1,\alpha}$

t-table.

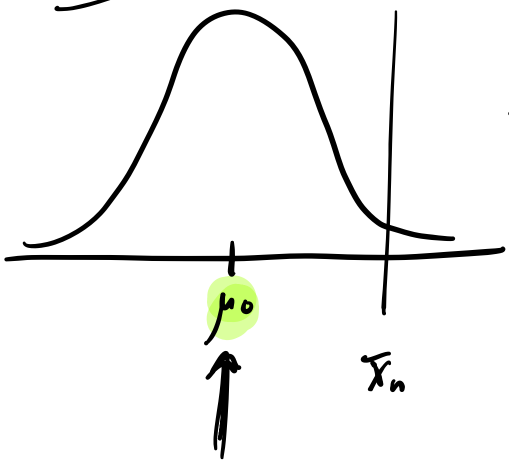


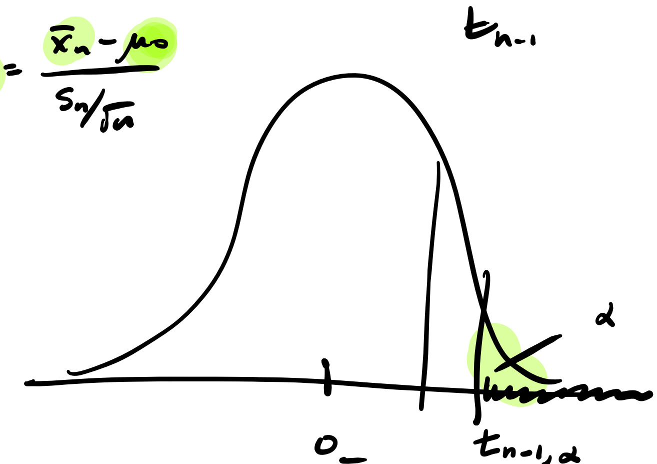
$H_0: \quad \mu \le \mu_0$

$H_1: \quad \mu > \mu_0$

$$Z_{test} = \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}$$

$$T_{test} = \frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}}$$





$$T_{test} = \frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}}$$

Reject H_0 when $T_{test} > t_{n-1,\alpha}$

Exercise: The average height of 14 randomly selected ten-yr-old Loblolly pine trees was $\bar{X}_n = 27.44$ and the sample standard deviation was $S_n = 1.54$. Assume that the heights of ten-yr-old Loblolly pine trees are Normally distributed.

significance level

1. Test the hypotheses $H_0\colon \mu \leq 26$ versus $H_1\colon \mu > 26$ at $\alpha = 0.05$.
2. Test the hypotheses $H_0\colon \mu \geq 26$ versus $H_1\colon \mu < 26$ at $\alpha = 0.05$.
3. Test the hypotheses $\boxed{H_0\colon \mu = 26}$ versus $\boxed{H_1\colon \mu \neq 26}$ at $\alpha = 0.05$.
4. Build a 95% CI for μ.

① $\boxed{H_0: \mu \leq 26}$

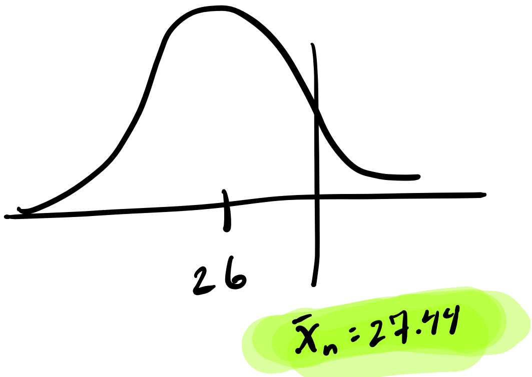
$H_1: \mu > 26$

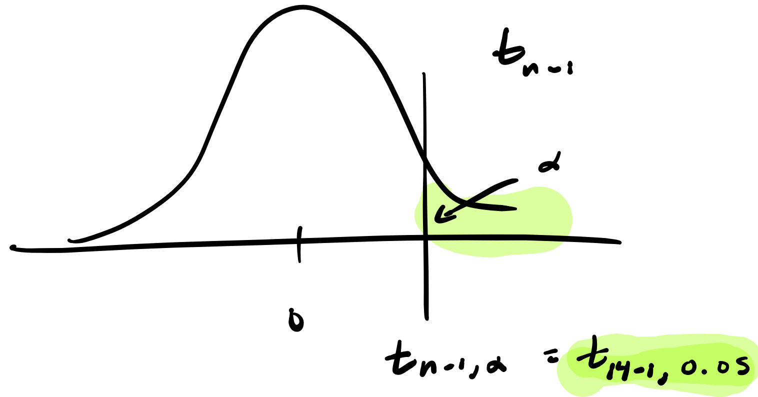
$S_n = 1.54$

$\bar{X}_n = 27.44$

$n = 14$

$$T_{test} = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} = \frac{27.44 - 26}{1.54/\sqrt{14}} = 3.499$$



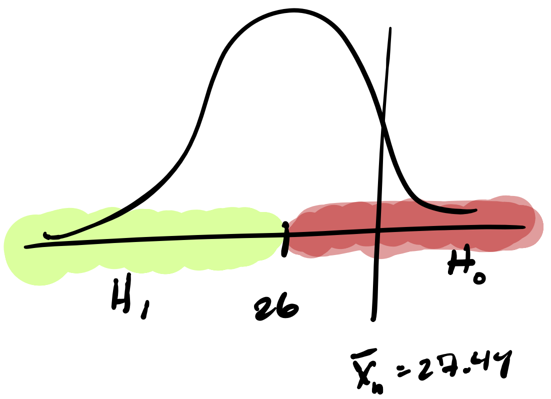


We have $T_{test} = 3.499 > t_{14-1,\,0.05} = 1.77$.

So we reject H_0 at the 0.05 significance level.

② $H_0: \mu \geq 26$

$H_1: \mu < 26$



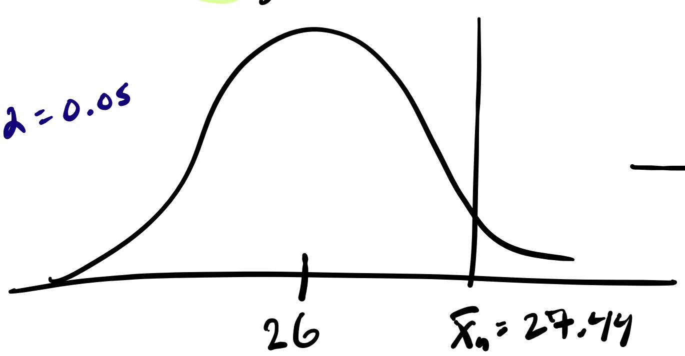
$\leftarrow$ data support H_0, so we fail to reject H_0.

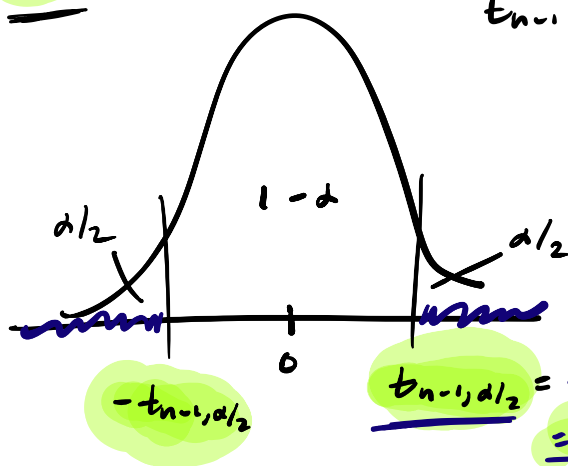
③ $H_0: \mu = 26$ Two-sided

$H_1: \mu \neq 26$

$$T_{test} = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} = \frac{27.44 - 26}{1.54/\sqrt{14}} = 3.499$$

$\alpha = 0.05$





$= t_{14-1,\,0.025} = 2.16.$

For two-sided tests at α, just see if $(1-\alpha)100\%$ CI contains the null value!

For H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$ we have:

- For σ unknown

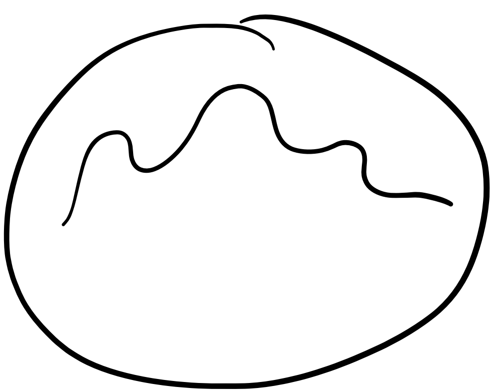
$$\boxed{|T_{\text{test}}| > t_{n-1,\alpha/2}} \iff \mu_0 \notin \left(\bar{X}_n - t_{n-1,\alpha/2}\frac{S_n}{\sqrt{n}}, \bar{X}_n + t_{n-1,\alpha/2}\frac{S_n}{\sqrt{n}}\right).$$

$\frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$

- For σ known

$$|Z_{\text{test}}| > z_{\alpha/2} \iff \mu_0 \notin \left(\bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).$$

1. Introduction to hypothesis testing
2. Testing hypotheses about μ with σ known under Normality
3. Testing hypotheses about μ with σ unknown under Normality
4. **Testing hypotheses about μ when data is not Normal** need large n.
5. Testing hypotheses about p



Since $\sqrt{n}(\bar{X}_n - \mu)/S_n$ behaves like $Z \sim \text{Normal}(0,1)$ for large n...

Tests about μ when data non-Normal and $n \geq 30$

For some null value μ_0, define the test statistic

$$T_{\text{test}} = \frac{\bar{X} - \mu_0}{S_n/\sqrt{n}}.$$

Then the following tests have $P(\text{Type I error}) \leq \alpha$.

$H_0\colon \mu \geq \mu_0$ $H_1\colon \mu < \mu_0$	$H_0\colon \mu = \mu_0$ $H_1\colon \mu \neq \mu_0$	$H_0\colon \mu \leq \mu_0$ $H_1\colon \mu > \mu_0$
Reject H_0 if $T_{\text{test}} < -z_\alpha$	Reject H_0 if $\lvert T_{\text{test}} \rvert > z_{\alpha/2}$	Reject H_0 if $T_{\text{test}} > z_\alpha$

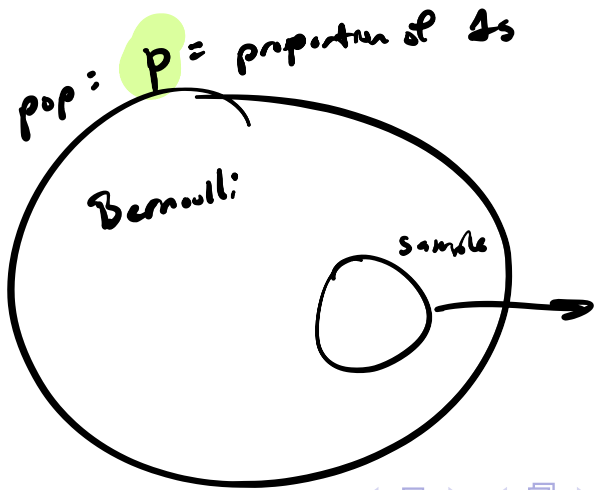
Exercise:

1. Draw a random sample of size $n = 35$ from the 2009 Boston Marathon `women's finishing times` and test the hypotheses

$$H_0\colon \mu \leq 4 \text{ versus } H_1\colon \mu > 4$$

at the $\alpha = 0.05$ significance level.

2. Repeat this 1000 times and record the proportion of times you reject H_0.



$$\hat{p}_n = \frac{X_1 + \dots + X_n}{n} = \frac{\#\{\text{ones}\}}{n}$$

$\mu = p$, $\sigma = \sqrt{p(1-p)}$

Since $\boxed{\sqrt{n}(\hat{p}_n - p)/\sqrt{p(1-p)}}$ behaves like $Z \sim \text{Normal}(0,1)$ for large n...

$\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \to N(0,1)$ as $n \to \infty$.

Tests about p (for $np_0 \geq 15$ and $n(1-p_0) \geq 15$)

For some null value p_0, define the test statistic

$$\boxed{Z_{\text{test}} = \frac{\hat{p}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.}$$

$$Z_{\text{test}} = \frac{\hat{p}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$H_0: p \leq 1/2$
$H_1: p > 1/2$

Then the following tests have $P(\text{Type I error}) \leq \alpha$.

$H_0\colon p \geq p_0$ $H_1\colon p < p_0$	$H_0\colon p = p_0$ $H_1\colon p \neq p_0$	$\boxed{\begin{array}{l}H_0\colon p \leq p_0\\ H_1\colon p > p_0\end{array}}$
Reject H_0 if $Z_{\text{test}} < -z_\alpha$	Reject H_0 if $\lvert Z_{\text{test}}\rvert > z_{\alpha/2}$	Reject H_0 if $Z_{\text{test}} > z_\alpha$

Exercise: Does a female-inhabiting parasite tip the sex ratio of its hosts' offspring in favor of females? A sample of size $n = 500$ offspring from parasite-infected females is collected, among which there are 287 females.

Let p = prop. of females among offspring of parasite-infected hosts.

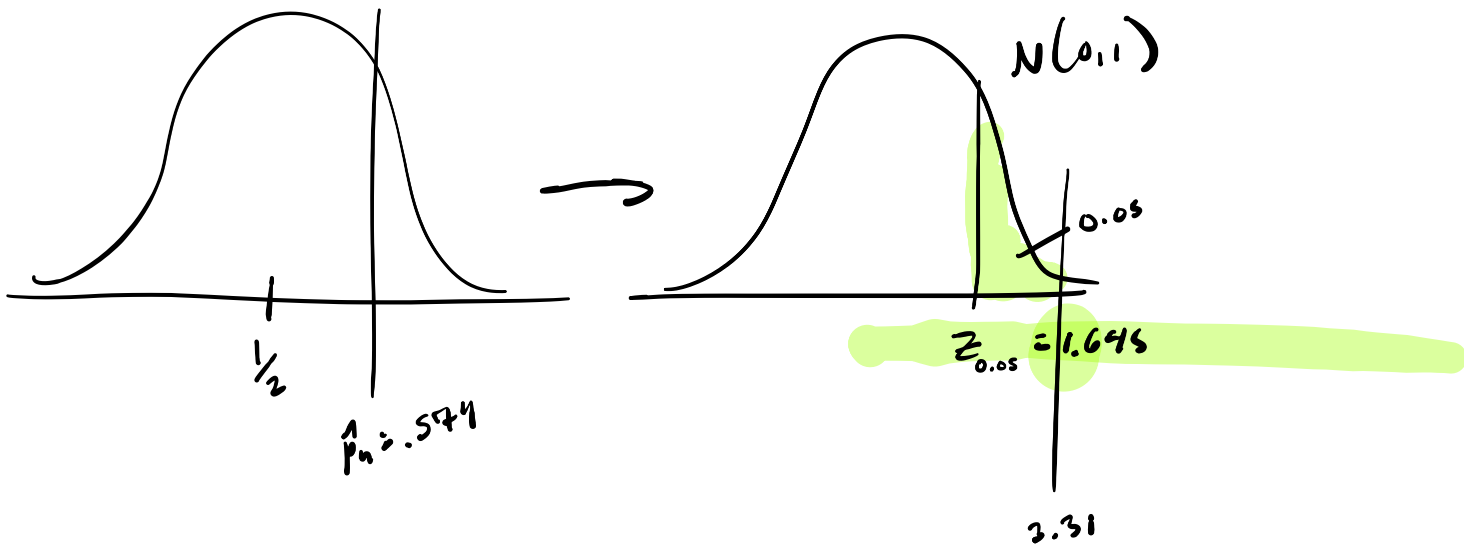
1. What are the relevant hypotheses? $\boxed{H_0: p \le \frac{1}{2} \quad H_1: p > \frac{1}{2}}$
2. Carry out a test of the hypotheses at the $\alpha = 0.05$ significance level.
3. Identify the following as a correct decision, a Type I error, or a Type II error:
   a. Suppose $p = 0.60$ and your data leads you to reject H_0. Correct
   b. Suppose $p = 0.60$ and your data leads you to not reject H_0. Type II
   c. Suppose $p = 0.50$ and your data leads you to reject H_0. Type I
   d. Suppose $p = 0.50$ and your data leads you to not reject H_0.

$n = 500$

$\hat{p}_n = \frac{287}{500} = .574.$

$$Z_{test} = \frac{\hat{p}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.574 - 1/2}{\sqrt{\frac{1/2(1-1/2)}{500}}} = 3.31.$$

$p_0 = 1/2$



1/2
$\hat{p}_n = .574$
N(0,1)
0.05
$Z_{0.05} = 1.645$
2.31

Do at home

Exercise: In a tasting experiment, each o 121 blindfolded students was fed either a red or green gummy bear, (each with probability 1/2) and asked to identify the color from the taste. Of the 121, 97 correctly identified the color(Ex. 8.82 of [1]).

1. If the students guessed "red" or "green" based on flipping a coin, with what probability would they guess the color correctly?
2. Suppose you wish to know if the students are doing better or worse than guessing. What are the relevant hypotheses?
3. Test the hypotheses at the $\alpha = 0.01$ significance level.

J.T. McClave and T.T. Sincich.
Statistics.
Pearson Education, 2016.