Statistical Inference means: learning from data about the population or the process generating the data

Confidence intervals

Hypothesis testing

STAT 515 Lec 14 slides

Hypothesis testing

Karl Gregory

University of South Carolina

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.

does her data tell her?

(f) If the true proportion of drones in the hive were equal to 0.15, with what probability would the beekeeper obtain 40 or more drones in her scoop of 294 bees?

i. Compute this probability exactly, assuming that there are 30,000 bees in the hive.

ii. Compute this probability ignoring the fact that she is sampling without replacement.

iii. Compute an approximation to this probability using the Normal distribution.

i) X = # drones in sample $\sim$ Hypergeometric$(N = 30{,}000,\ M = 4500,\ k = 294)$

$$P(X \geq 40) = 1 - P(X < 40)$$
$$= 1 - P(X \leq 39)$$
$$= 1 - \text{phyper}(39, \boxed{})$$

see notes

$$P(X = x) = \frac{\binom{4500}{x}\binom{30{,}000 - 4500}{294 - x}}{\binom{30{,}000}{294}}$$

ii) $X \sim \text{Binomial}(n = 294,\ p = 0.15)$

$$P(X \geq 40) = 1 - P(X \leq 39) = 1 - \text{pbinom}(39, 294, 0.15)$$

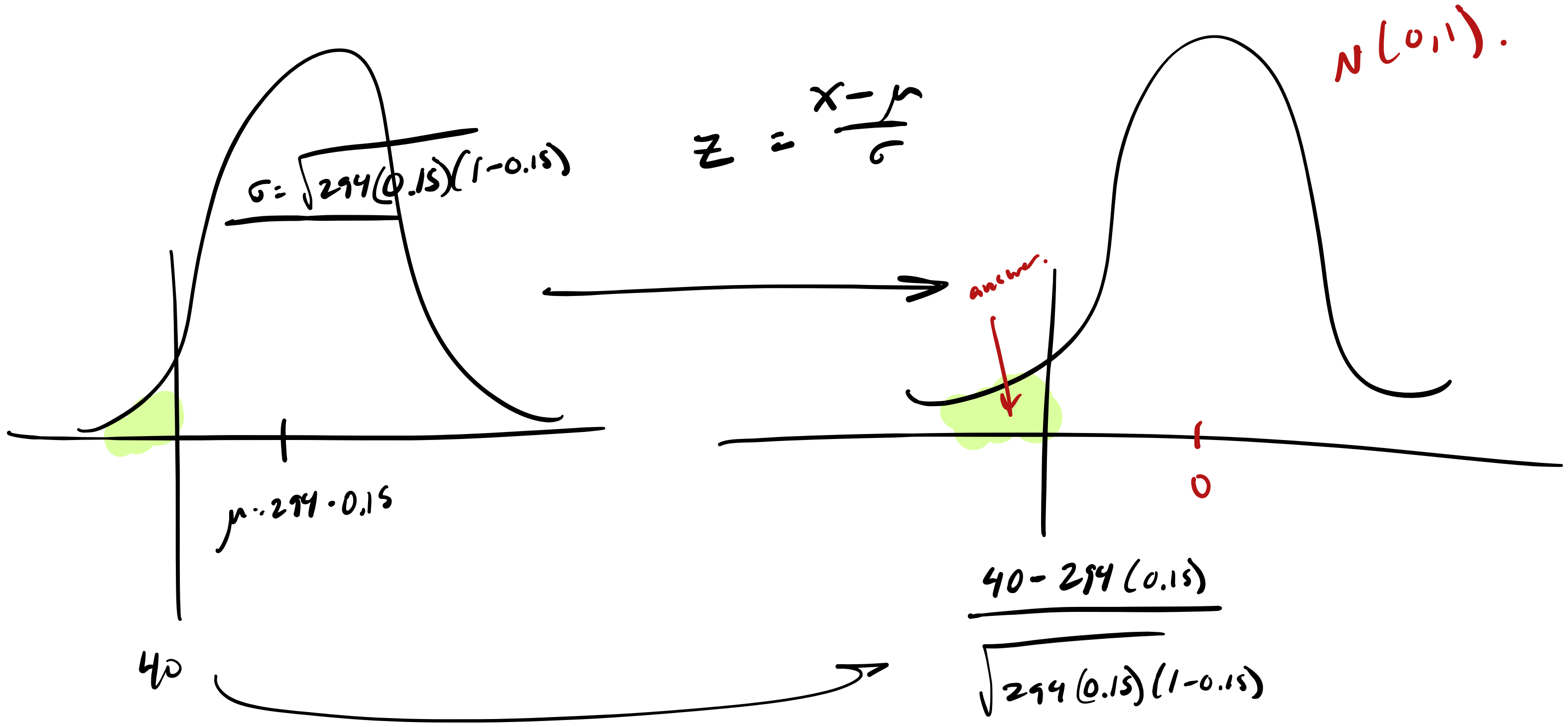
iii) $X: \underbrace{n}_{\text{\# draws in sample}} \hat{p}_n \overset{\text{approx}}{\sim} N\left(np,\ np(1-p)\right)$

$n = 294$

$p = 0.15$

$$X \overset{\text{approx}}{\sim} N\Big(\underbrace{294 \cdot 0.15}_{\mu},\ \underbrace{294 \cdot 0.15(1-0.15)}_{\sigma^2}\Big)$$

$P(X \geq 40)$



Statistical inference about μ and p

H_0 — null hypothesis

H_1 — alternate hypothesis

1. Reject H_0
2. Fail to reject H_0.

null value of μ

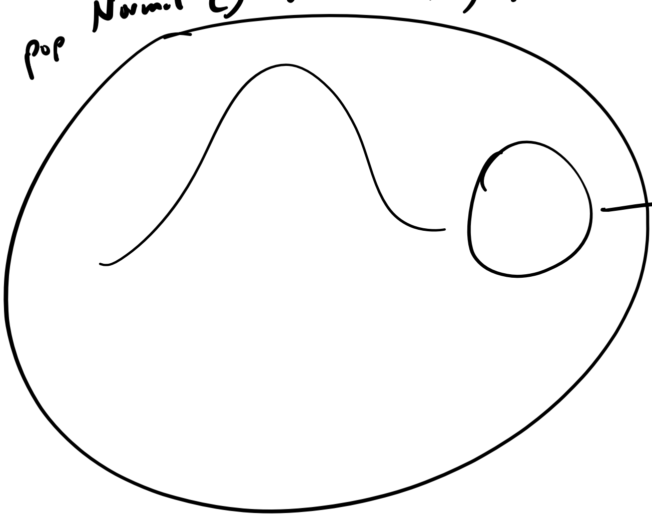
$\downarrow$

μ_0

$H_0: \mu \geq \mu_0$ $H_0: \mu = \mu_0$ $H_0: \mu \leq \mu_0$

$H_1: \mu < \mu_0$ $H_1: \mu \neq \mu_0$ $H_1: \mu > \mu_0$

pop Normal(μ, σ^2), μ, σ^2 unknown



$\rightarrow \bar{x}_n, S_n^2$

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

A *statistical inference* is a conclusion about a $\overbrace{\text{pop. parameter}}^{\mu \text{ or } p}$ based on a rs. random sample

Specifically, a decision concerning contradictory statements about the parameter:

- The *null hypothesis* H_0.
- The *alternate hypothesis* H_1.

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

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

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

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.

$\boxed{\text{Either (1) Reject } H_0 \quad \text{(2) Fail to Reject } H_0.}$

(3) Reject H_1 and Conclude H_0?

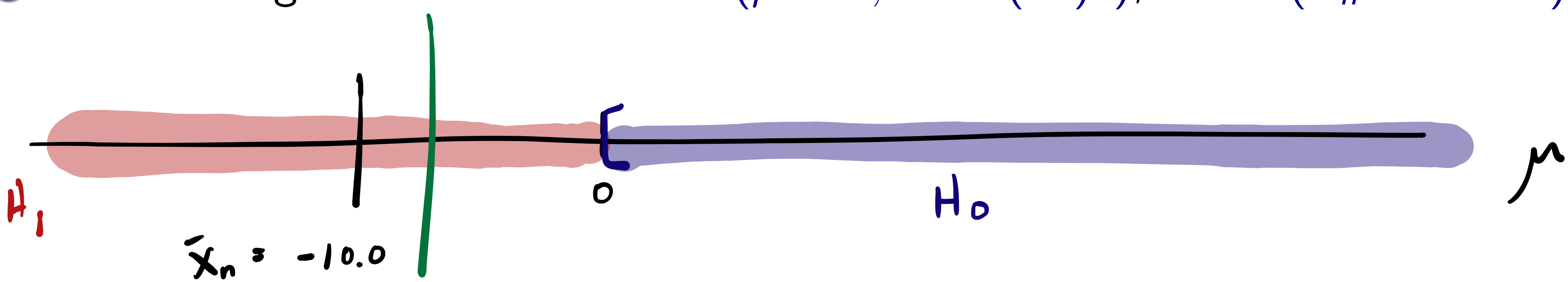
Exercise: Is a treatment effective in lowering cholesterol levels? Let μ represent the average difference (after-minus-before treatment) in cholesterol levels.

We want to test

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

Suppose we obtain $\bar{X}_n = 10.0$ from $n = 100$ subjects. Discuss the following:

1. Reject or fail to reject H_0?
2. What if we had observed $\bar{X}_n$ equal to -10.0?
3. If the changes in chol. level are $\mathrm{N}(\mu = 0, \sigma^2 = (25)^2)$, find $P(\bar{X}_n < -10.0)$.



Either
1 Reject H_0

2 Fail to Reject H_0.

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.

TRUTH

INFERENCE

	H_0 true	H_0 false
Reject H_0	Type I error	Correct decision
Fail to reject H_0	Correct decision	Type II error

Now:

Calibrate tests of hypotheses so that

$$P(\text{Type I}) \le \alpha.$$

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

unknown

We will consider null and alternate hypotheses of the form

$$H_0\colon \mu \geq \boxed{\mu_0} \quad \text{or} \quad H_0\colon \mu = \boxed{\mu_0} \quad \text{or} \quad H_0\colon \mu \leq \boxed{\mu_0}$$
$$H_1\colon \mu < \mu_0 \qquad\qquad H_1\colon \mu \neq \mu_0 \qquad\qquad H_1\colon \mu > \mu_0.$$

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*

value of μ_0 in H_0

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

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

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

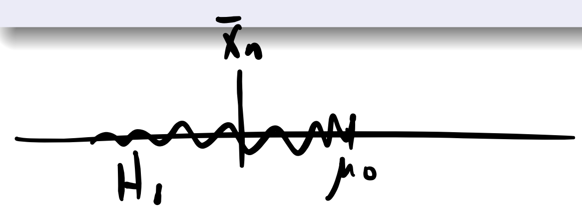
For some null value μ_0, define the $\boxed{\text{test statistic}}$

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

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

Two sided hypotheses

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



X = psi $\sim N(\mu, \sigma^2)$.

? ?

$\mu_0 = 157$

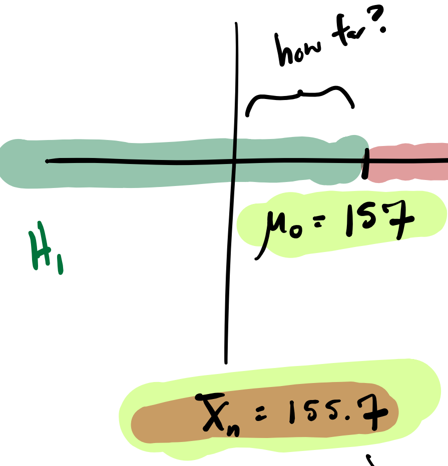
Exercise: Suppose a bottler of soft-drinks claims that its bottling process results in an internal pressure of 157 psi. You want to know whether the mean pressure is less than 157 (Ex 6.92 in [1]).

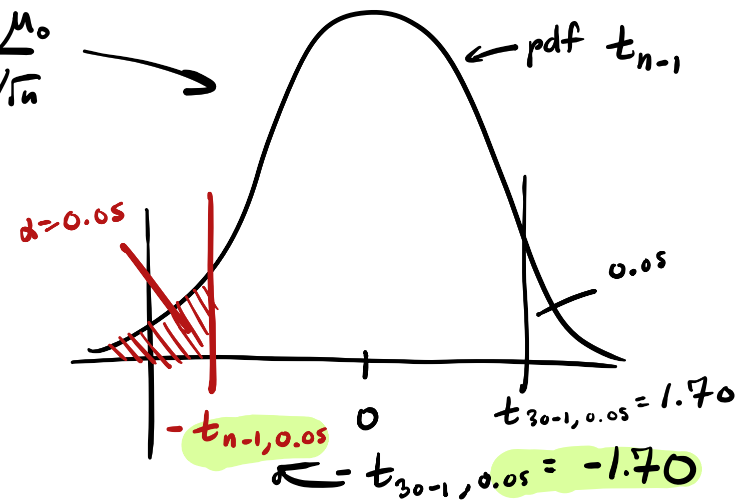
1. What are the relevant hypotheses?
2. Based on a sample of size $n =$ 30 you get $\bar{X}_n = 155.7$ and $S_n = 3.0$. What is your inference at the $\alpha = 0.05$ significance level?
3. Identify the following as a correct decision, a Type I error, or a Type II error:
   a. Suppose $\mu = 157.5$ and 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 $\mu = 156.5$ and your data leads you to not reject H_0. Type II error.

$\underbrace{H_0: \mu \geq 157}_{\text{always has an equality } (\leq, \geq, =)}$

$\underbrace{H_1: \mu < 157}_{\text{always has a strict inequality } (<, >, \neq)}$

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





$$\frac{155.7 - 157}{3.0/\sqrt{30}} = -2.37$$

$S_n = 3.0$

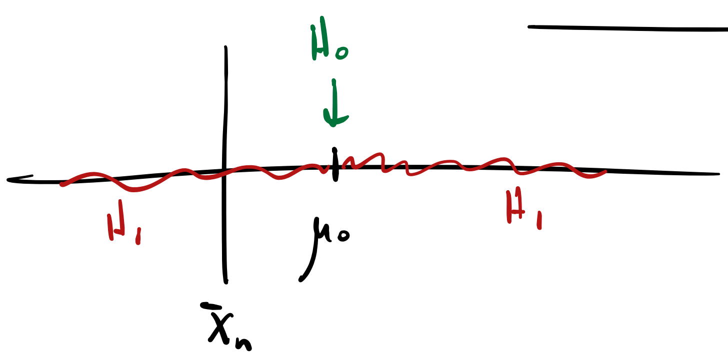
$n = 30$

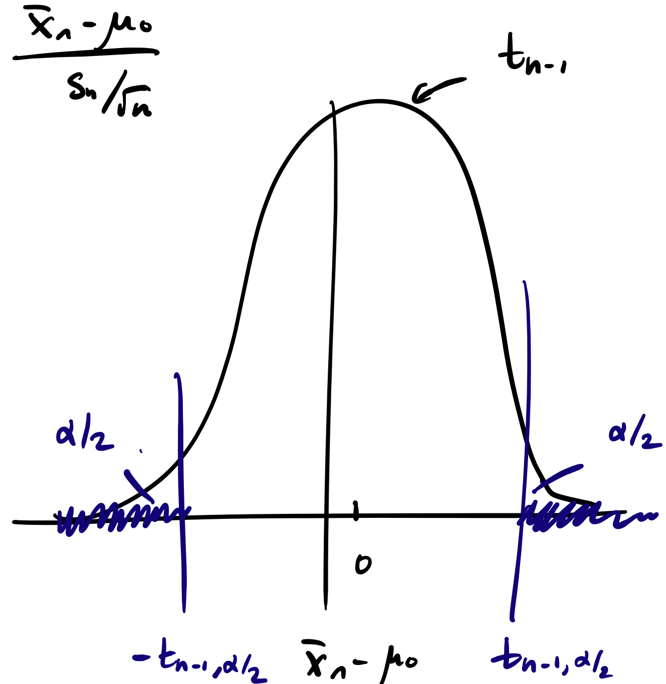
Since $-2.37 < -1.70$, I should reject H_0.

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

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





Reject H_0 if $\frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} < -t_{n-1,\alpha/2}$ or $\frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} > t_{n-1,\alpha/2}$

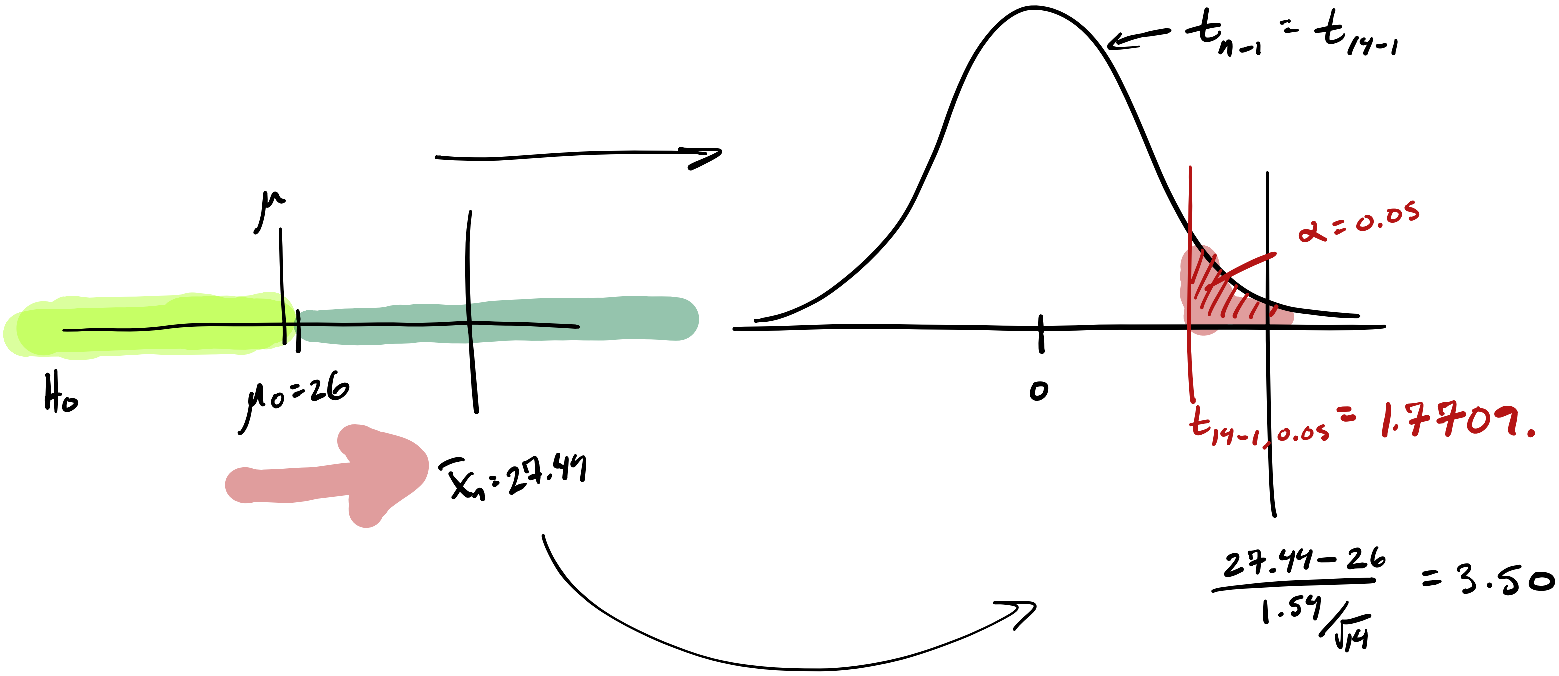
④ $\bar{X}_n \pm t_{n-1,\alpha/2} \frac{S_n}{\sqrt{n}} = 27.44 \pm \underbrace{t_{14-1,\frac{0.05}{2}}}_{2.1604} \frac{1.54}{\sqrt{14}} = \underline{(26.61, 28.27)}$

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.

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

① $\bar{X}_n = 27.44$, $S_n = 1.54$, $n = 14$ $\qquad \alpha = 0.05.$

Test $H_0: \mu \leq \underset{\mu_0}{26}$ vs $H_1: \mu > 26$.

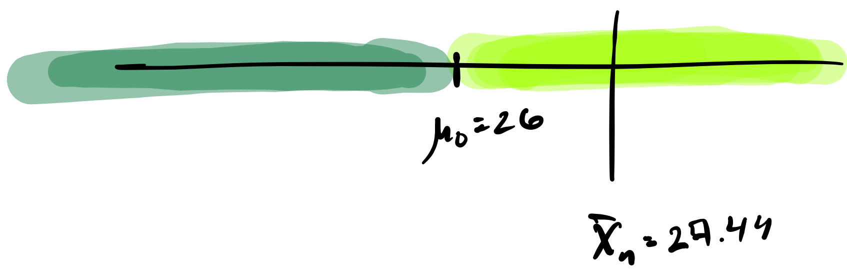


Reject H_0 if $T_{stat} = \dfrac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} > t_{n-1, \alpha}$

Yes, reject.

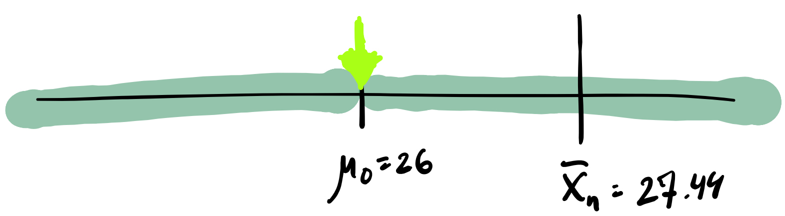
② $H_0: \mu \geq 26$

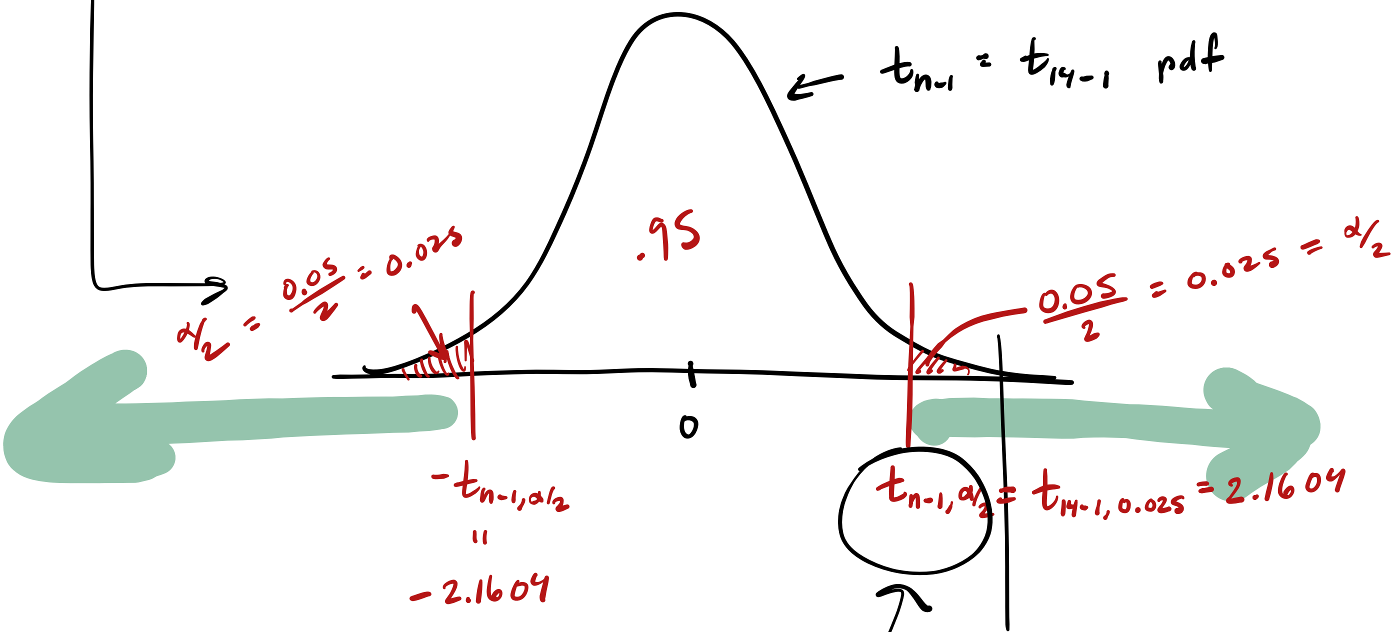
$H_1: \mu < 26$



Fail to reject H_0. Data support H_0, so we don't need to do anything.

③ $H_0: \mu = 26$ vs $H_1: \mu \neq 26$ (Two-sided hypotheses)





$$\frac{27.44 - 26}{1.54/\sqrt{14}} = 3.50$$

Reject H_0 if $T_{test} = \dfrac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} > t_{n-1,\alpha/2}$

or $T_{test} < -t_{n-1,\alpha/2}$.

Reject H_0 if $|T_{test}| > t_{n-1,\alpha/2}$

	H_0 true	H_0 false
Reject H_0	Type I error	Correct
Fail to reject H_0	Correct	Type II error

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

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:

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

1. Reject H_0 if $|T_{\text{test}}| > t_{n-1,\alpha/2}$

$\iff$

2. Reject H_0 if μ_0 is not in the C.I. $\bar{X}_n \pm t_{n-1,\alpha/2}\frac{S_n}{\sqrt{n}}$.

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

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

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:

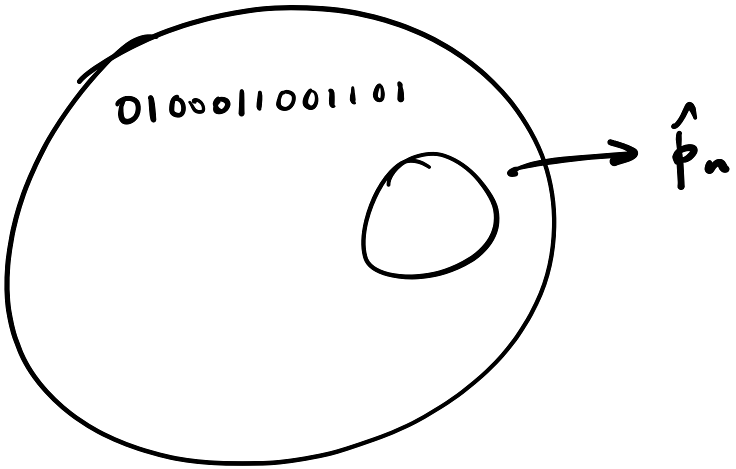
1. Draw a random sample of size $n = 35$ from the 2009 Boston Marathon `women's finishing times` and test the hypotheses

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

at the $\alpha = 0.05$ significance level.

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

p = proportion of 1's



$H_0: \; p \geq p_0$

$H_1: \; p < p_0$

$H_0: \; p = p_0$

$H_1: \; p \neq p_0$

null value

$H_0: \; p \leq p_0$

$H_1: \; p > p_0$

Ask: $\underbrace{\text{how far is } \hat{p}_n \text{ from } p_0 \text{ ?}}_{\text{send this into the Z world.}}$

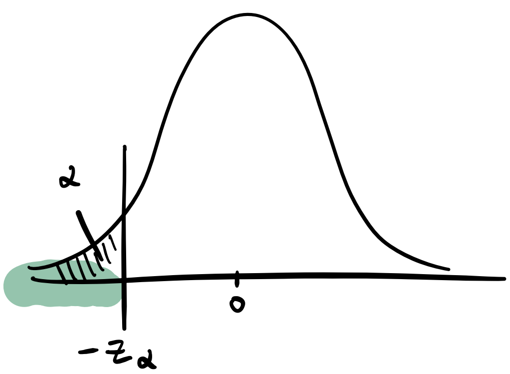
"Test statistic"

$$Z_{test} = \frac{\hat{p}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{\text{approx}}{\sim} N(0,1)$$

if H_0 is true.

$H_0: \; p \geq p_0$

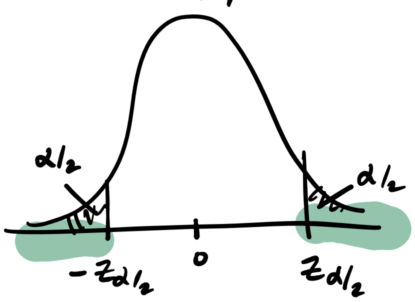
$H_1: \; p < p_0$



Reject H_0 if $Z_{test} < Z_\alpha$

$H_0: \; p = p_0$

$H_1: \; p \neq p_0$

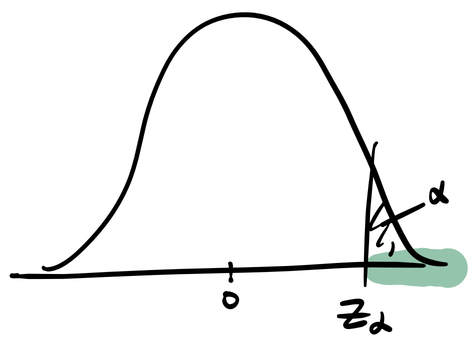


Reject H_0 if

$|Z_{test}| > Z_{\alpha/2}$

$H_0: \; p \leq p_0$

$H_1: \; p > p_0$



Reject H_0 if

$Z_{test} > Z_\alpha$

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

ensures that n is large enough for the central limit to take effect.

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

For some null value μ_0, define the test statistic

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

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

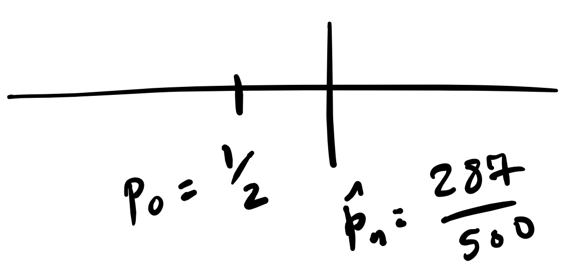
$H_0\colon p \geq p_0$ $H_1\colon p < p_0$	$H_0\colon p = p_0$ $H_1\colon p \neq p_0$	$H_0\colon p \leq p_0$ $H_1\colon p > p_0$
Reject H_0 if $Z_{\text{test}} < -z_\alpha$	Reject H_0 if $\lvert Z_{\text{test}}\rvert > z_{\alpha/2}$	Reject H_0 if $Z_{\text{test}} > z_\alpha$

Let p = prop. of females in offspring

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.

1. What are the relevant hypotheses? $H_0: p \leq \frac{1}{2}$ $\quad H_1: p > \frac{1}{2}$ (p_0)
2. Carry out a test of the hypotheses at the $\alpha = 0.05$ significance level. Reject H_0.
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. Correct.

(2)

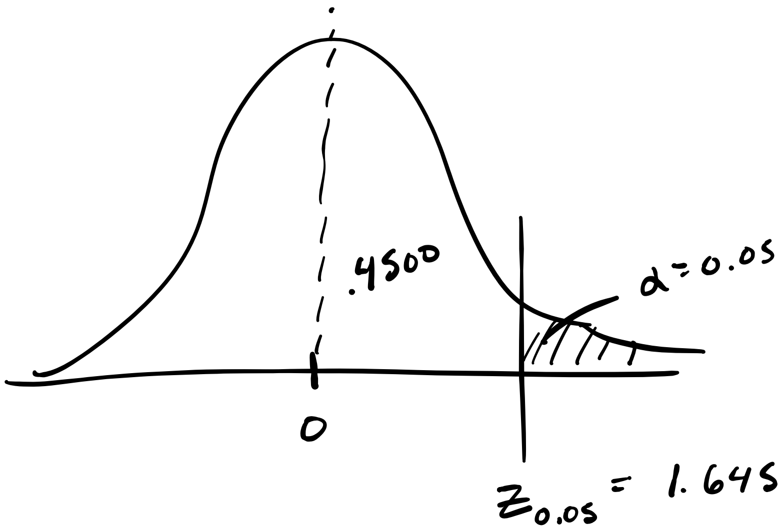


$n = 500$ $\quad \hat{p}_n = \frac{287}{500}$

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

$$Z_{\text{test}} = \frac{\frac{287}{500} - \frac{1}{2}}{\sqrt{\dfrac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{500}}}$$

$$= 3.31$$



Exercise: In a tasting experiment, each of 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? $1/2$
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.

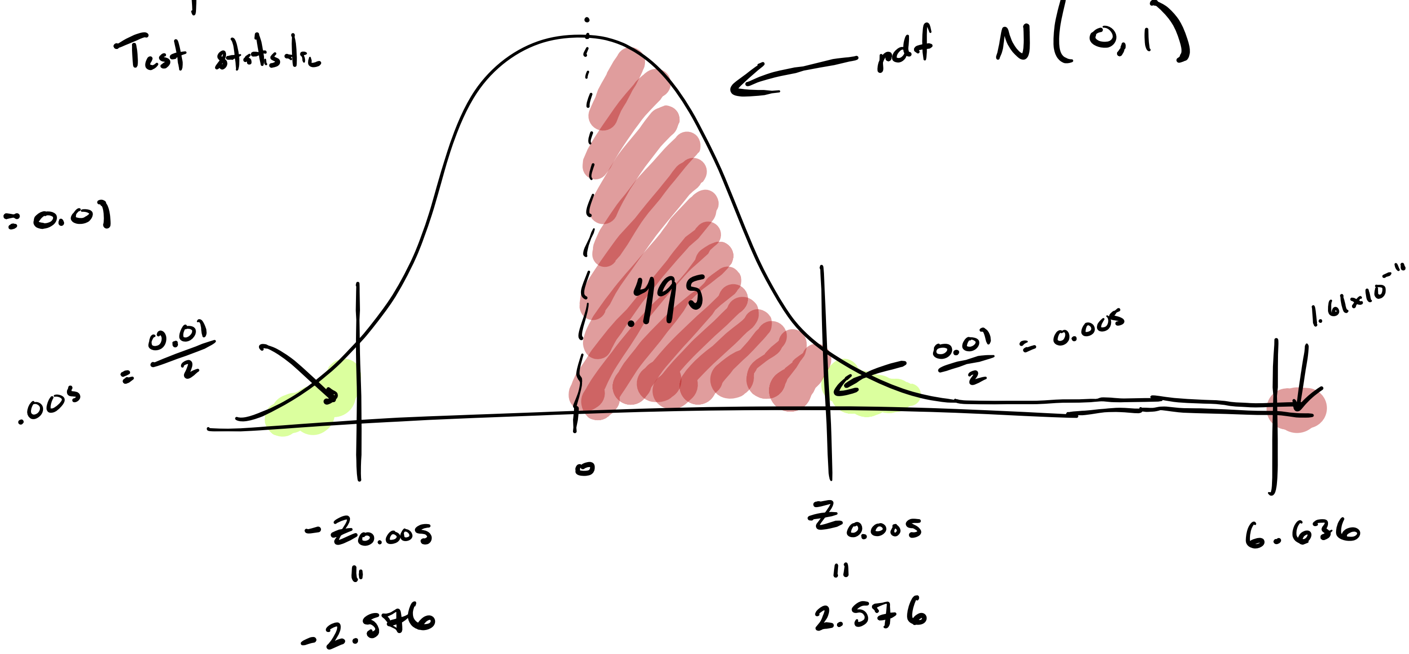
② $H_0: p = \frac{1}{2}$ vs $H_1: p \neq \frac{1}{2}$

③ $n = 121$, $\hat{p}_n = \frac{97}{121}$, $p_0 = \frac{1}{2}$

$$Z_{test} = \frac{\hat{p}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{97}{121} - \frac{1}{2}}{\sqrt{\frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{121}}} = 6.636$$

Test statistic

$\alpha = 0.01$



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