10.2 Elements of a Statistical Test

Spse Professor Heiny claims he was an 80% free throw shooter "back in the day". You have serious doubts about this and highly suspect he was less than an 80% free throw shooter. How can we use hypothesis testing to answer this question?

To begin with, form Ho (null hypothesis) and Ha (alternative hypothesis).

Ho: status quo or what is generally accepted to be true

Ha: researcher's claim

In general, we are looking for evidence in favor of Ha (against Ho). Since we think Prof. Heiny is much worse than he claims to be...

$H_0: p = .80$

$H_a: p < .80$

Next, we need to develop a test statistic and a rejection region. It would seem to make sense in this case to have Prof. H shoot some free throws, and let the Y = no. of made free throws be our test statistic.

It would make sense to reject Ho (in favor of Ha) for small values of Y. How small is too small? This will be

our rejection region (RR). Spse we decide to reject if $Y \le 12$.
Is this a good RR? What kind of errors can we make?
How likely are they?

	Reality: H_0 is true	Reality: H_0 is false
don't reject H_0	OK	Type II error $P(\cdot) = \beta$
reject H_0	Type I error $P(\cdot) = \alpha$	OK Power $= 1-\beta$

Judicial system

$\alpha \uparrow, \beta \downarrow$

$\alpha \downarrow, \beta \uparrow$

In our example...

Type I error: Conclude Prof. H makes less than 80% when in fact he does make 80% overall.

Type II error: Decide Prof. H makes 80% overall when he actually makes less than 80%.

Spse we have Prof. Henry shoot 20 free throws.
We will reject H_0 if $Y \le 12$. Find α

$\alpha = P[Y \le 12 \mid p = .80]$ What is the dist. of Y?

$Y \sim Bin(n=20, p=.80)$

$\Rightarrow \boxed{\alpha = .032} = \sum_{y=0}^{12} \binom{20}{y} (.8)^y (.2)^{20-y}$

Table 1

This is a pretty low prob for a type I error (good).
what is the trade-off for this?

Compute β if $p=.7$

$$\beta = P(Y>12 \mid p=.7) = 1-P(Y\le 12 \mid p=.7) = .7732$$

(power $= 1-\beta = .2268$) $\quad 1-\sum_{y=0}^{12}\binom{20}{y}(.7)^y(.3)^{20-y}$

i.e. If Prof. Heiny is only a 70% shooter, we will probably not be able to detect it.

Compute β if $p=.5$

$$\beta = P(Y>12 \mid p=.5) = 1-P(Y\le 12 \mid p=.5) = .132$$

(power $= 1-\beta = .868$)

note: Larger shifts away from H_0 are easier to detect. We have more statistical power for the same level of α.

Spse we decide it is important to detect if Prof. Heiny can only make 70%. We decide to use RR: $Y\le 14$

$$\beta = P(Y>14 \mid p=.7) = \sum_{y=15}^{20}\binom{20}{y}(.7)^y(.3)^{20-y} \approx .42$$

This is a nice reduction from before, and we now have a better than even chance of detecting if Prof. H can only make 70% of his shots.

What is the trade-off for the increase in power?
How much more type I error are we willing to risk?

$$\alpha = P(Y \le 14 \mid p = .8) = \sum_{y=0}^{14} \binom{20}{y} (.8)^y (.2)^{20-y} \approx .196$$

This is too high. If Prof. Heiny can actually make 80%, there is a nearly 1 in 5 chance we would mistakenly conclude he makes less than 80%.

<u>note:</u> If you want to increase power <u>without</u> <u>increasing α</u>, you must <u>increase</u> <u>the sample size</u>. This costs money!

10.6 $H_0: p = .5$
$H_a: p \ne .5$ RR: $|y - 18| \ge 4$ $(n = 36)$

a) $\alpha = P(Y \ge 22 \mid p = .5) + P(Y \le 14 \mid p = .5)$

$$= \sum_{y=22}^{36} \binom{36}{y} (.5)^{36} + \sum_{y=0}^{14} \binom{36}{y} (.5)^{36}$$

$$= .1215 + .1215 = .2430$$

b) Find β if $p = .7$

$$\beta = P[15 \le Y \le 21 \mid p = .7] = .0916$$

10.3 Common Large-Sample Tests

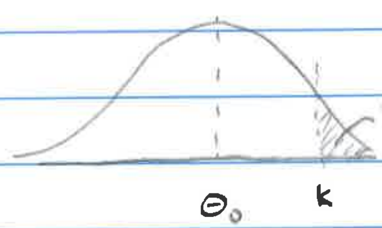
In general, if we want a hypothesis regarding a parameter θ, we will use an estimator $\hat{\theta}$ developed from sample data. In this section, $\hat{\theta}$ will be based on a sample size large enough so that its sampling dist. is approx. normal (CLT).

$\boxed{\text{Table 8.1}}$ contains a list of estimators and their std. errors

Spse we want to test $H_0: \theta = \theta_0$ vs. $H_a: \theta > \theta_0$

test statistic: $\hat{\theta}$

RR: $\{\hat{\theta} > k\}$ for some value of k



$\hat{\theta} \sim N(\mu_{\hat{\theta}} = \theta_0, \sigma_{\hat{\theta}})$

$\alpha = P(\text{type I error})$

For an α-level test, $\boxed{k = \theta_0 + k \cdot z_\alpha}$

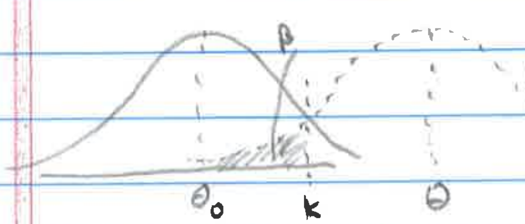
Of course in 2050...

test statistic: $z = \dfrac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} \sim N(0,1)$

RR: $\{z > z_\alpha\}$

note: Both methods are equivalent

If $\theta > \theta_0$ then $\beta = P(\text{type II error})$



our actual sampling dist, but we don't know this e.g. 10.34 Estimate μ = avg. no. of days for inpatient trt. at a hospital. Fed. Agency believes the avg. length is in excess of 5 days.

$n = 500 \quad \bar{y} = 5.4 \quad s = 3.1$

Does the data support the agency's claim? Use $\alpha = .05$

$\begin{pmatrix} \theta = \mu \\ \hat{\theta} = \bar{y} \end{pmatrix}$

$H_0: \mu = 5$

$H_a: \mu > 5$

large sample ($n = 500$)

$\bar{y} \sim N(\mu_{\bar{y}} = \mu, \ \sigma_{\bar{y}} = \sigma/\sqrt{n})$

note: We will use s in place of σ

$$z = \frac{\bar{y} - \mu}{s/\sqrt{n}} = \frac{5.4 - 5}{3.1/\sqrt{500}} = 2.89$$

$z_{.05} = 1.645 \Rightarrow$ Reject H_0 and conclude there is enough evidence (at $\alpha = .05$) that the avg. length of hospital stay is more than 5 days i.e. Yes, the data supports the agency's claim.

note: We could have equivalently set up RR: $\bar{y} > k$

$k = \mu + z_\alpha \cdot \sigma/\sqrt{n}$

$= 5 + 1.645 \, (3.1)/\sqrt{500} = 5.228$

$5.4 > 5.228 \Rightarrow$ Reject H_0

note: We are rejecting H_0. Is it possible we are making a mistake? Yes $\boxed{\text{simulation}}$

10.24 Children's Hospital in Boston: 67% of adults overweight; 15% of children are overweight

r.s. of 100 children, 13 are classified as overweight

Is there sufficient evidence to indicate Children's Hospital percent is too high?

$\begin{pmatrix} \theta = p \\ \hat{\theta} = \hat{p} \end{pmatrix}$

$H_0: p = .15$

$H_a: p < .15$

Is $\hat{p}$ approx. normal?

$p \pm 3\sqrt{\frac{pq}{n}} = .15 \pm 3\sqrt{\frac{(.15)(.85)}{100}}$

$= .15 \pm .107 \rightarrow (.043, .257)$ ✓ Yes

$\hat{p} \sim N\left(\mu_{\hat{p}} = p,\ \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}\right)$

$\hat{p} = \frac{13}{100} = .13 \qquad z = \frac{.13 - .15}{\sqrt{\frac{(.15)(.85)}{100}}} = -.56$

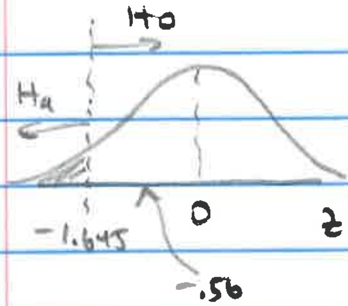
Since this is a left-tail test, RR: $\{z < -z_\alpha\}$

$-z_{.05} = -1.645 \Rightarrow$ don't Reject H_0. There is not enough evidence (at $\alpha = .05$) to conclude that the % reported by Children's Hospital is too high.

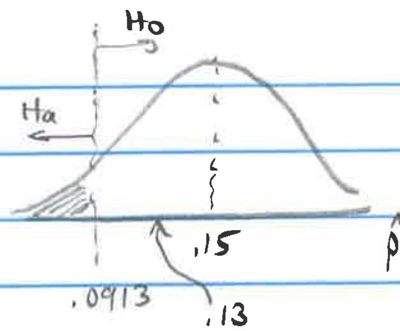
note: Equivalently we could use RR: $\{\hat{p} < k \mid k = p - z_\alpha \cdot \sigma_{\hat{p}}\}$

$k = .15 - 1.645\sqrt{\frac{(.15)(.85)}{100}} = .0913$

$\hat{p} > .0913 \Rightarrow$ don't Reject H_0. If $p = .15$, $\hat{p} = .13$ based on $n = 100$ is pretty reasonable



or



(10.21) Comparing sheer strength for two types of soil.

Soil Type I	Soil Type II
$n_1 = 30$	$n_2 = 35$
$\bar{y}_1 = 1.65$	$\bar{y}_2 = 1.43$
$s_1 = 0.26$	$s_2 = 0.22$

Do the soils appear different with respect to avg. sheer strength? Use $\alpha = .01$

$$\left(\begin{array}{l}\theta = \mu_1 - \mu_2 \\ \hat{\theta} = \bar{y}_1 - \bar{y}_2\end{array}\right)$$

$H_0: \mu_1 = \mu_2 \quad (\mu_1 - \mu_2 = 0)$

$H_a: \mu_1 \neq \mu_2 \quad (\mu_1 - \mu_2 \neq 0)$

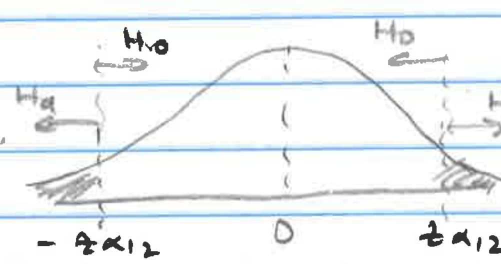
For large sample sizes ($n_1 \geq 30$, $n_2 \geq 30$), $\bar{y}_1 - \bar{y}_2$ is approx. normal.

$$\bar{y}_1 - \bar{y}_2 \sim N\left(\mu_{\bar{y}_1 - \bar{y}_2} = \mu_1 - \mu_2,\ \sigma_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

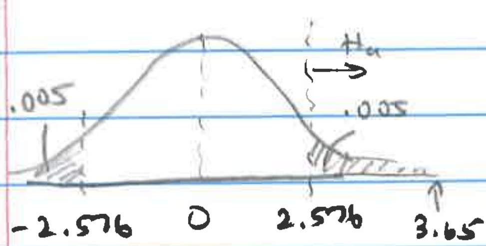
$$\approx \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$z = \frac{\overbrace{(1.65 - 1.43)}^{\hat{\theta}} - \overbrace{0}^{\theta_0}}{\underbrace{\sqrt{\frac{.26^2}{30} + \frac{.22^2}{35}}}_{\sigma_{\hat{\theta}}}} = 3.65$$

For a two-sided alternative ...



⑤

($\alpha = .01$) 

Reject H_0. There is enough evidence at $\alpha = .01$ to conclude that the soil types have different avg. sheer strengths (I > II)

10.4 Calculating Type II Error Probs.

The tests in 10.3 controlled for Type I error, but we did not discuss $\beta = P(\text{type II error})$. In practice, this should always be considered.

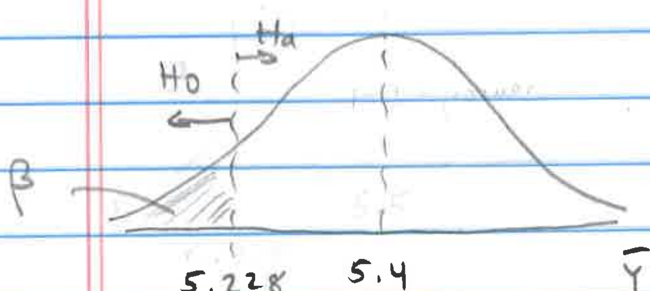
Spse we go back to the avg. length of hospital stay. (10.34)
The federal agency believes $\mu > 5$, but they probably don't care if $\mu = 5.002$. How far above 5 would be considered to be practically significant? This is a judgement call.

Federal agency wants a high prob. of detecting a shift if μ is as high as 5.4

$H_0: \mu = 5$ $\quad n = 500 \quad \sigma \approx s = 3.1$

$H_a: \mu > 5$ $\quad$ with $\alpha = .05$, RR: $\left\{ \bar{Y} > 5 + 1.645 \frac{(3.1)}{\sqrt{500}} = 5.228 \right\}$

If $\mu = 5.4$ then $\bar{Y} \sim N\left(\mu_{\bar{Y}} = 5.4,\ \sigma_{\bar{Y}} \approx \frac{3.1}{\sqrt{500}}\right)$



$\beta = P(\bar{Y} < 5.228)$

$= P\left(Z < \frac{5.228 - 5.4}{3.1/\sqrt{500}}\right)$

$= P(Z < -1.24) = .1075$

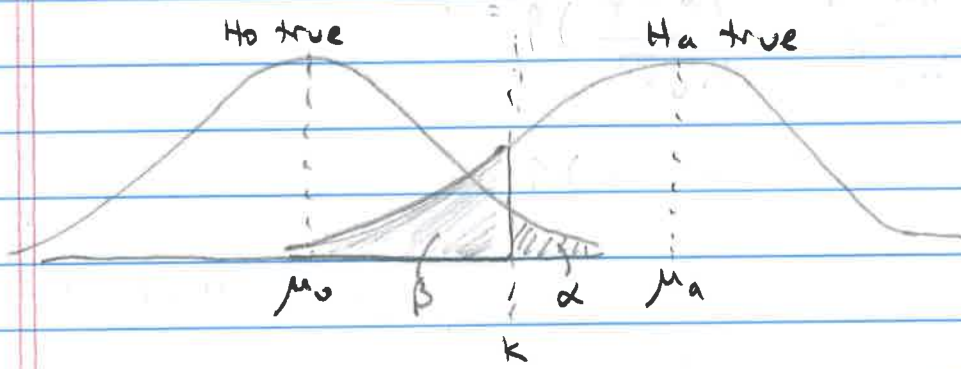
Simulation

What if federal agency decides $\beta \approx 11\%$ is too high?
Tell them n must be higher.

We can arbitrarily increase n and recompute β until we get a satisfactory value, but it would be better if we develop a general result.

$H_0: \mu = \mu_0$

$H_a: \mu > \mu_0$



$\alpha = P(\bar{Y} > k \mid \mu = \mu_0) \Rightarrow k = \mu_0 + z_\alpha \cdot \sigma/\sqrt{n}$

$\beta = P(\bar{Y} \le k \mid \mu = \mu_a) \Rightarrow k = \mu_a - z_\beta \cdot \sigma/\sqrt{n}$

$\Rightarrow \mu_0 + z_\alpha \cdot \sigma/\sqrt{n} \overset{\text{SET}}{=} \mu_a - z_\beta \cdot \sigma/\sqrt{n}$

$\Rightarrow \sigma/\sqrt{n}\,(z_\alpha + z_\beta) = (\mu_a - \mu_0)$

$\Rightarrow \sigma(z_\alpha + z_\beta) = \sqrt{n}\,(\mu_a - \mu_0)$

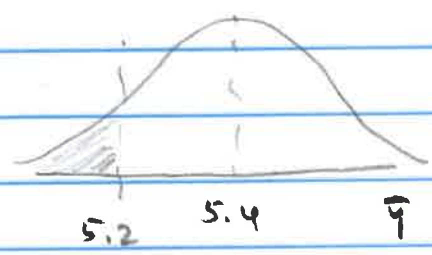
$$\Rightarrow \boxed{n = \frac{\sigma^2 (z_\alpha + z_\beta)^2}{(\mu_a - \mu_0)^2}}$$

Returning to the previous example, if the federal agency wants $\alpha = \beta = .05$ for $\mu = 5.4$ then

$$n = \frac{3.1^2\,(1.645 + 1.645)^2}{(5.4-5)^2} = 650.12 \rightarrow 651$$

Verify $\beta = .05$ for $n = 651$ and $\mu = 5.4$

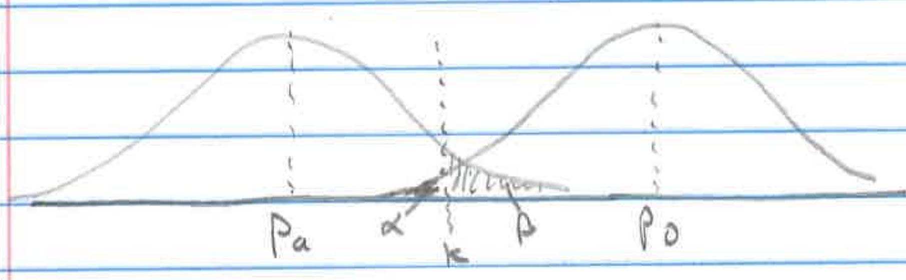
$$RR: \; k = 5 + 1.645\,\frac{(3.1)}{\sqrt{651}} = 5.2$$



$$z = \frac{5.2 - 5.4}{\frac{3.1}{\sqrt{651}}} = -1.646$$

$$P(\bar{y} < 5.2) = P(Z < -1.646) \approx .05 \checkmark$$

Spse for the obesity study ... $H_0: p = .15$ vs. $H_a: p < .15$
we are happy w/ $\alpha = .05$ but for $p = .10$ we don't want β to be higher than $.10$. How large should n be?



$$\alpha = P(\hat{p} < k \mid p = p_0) \Rightarrow k = p_0 - z_\alpha \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

$$\beta = P(\hat{p} \geq k \mid p = p_a) \Rightarrow k = p_a + z_\beta \cdot \sqrt{\frac{p_a(1-p_a)}{n}}$$

$$p_0 - z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}} \overset{\text{SET}}{=} p_a + z_\beta \sqrt{\frac{p_a(1-p_a)}{n}}$$

$$(p_0 - p_a) = \frac{z_\alpha \cdot \sqrt{p_0(1-p_0)} + z_\beta \sqrt{p_a(1-p_a)}}{\sqrt{n}}$$

$$\Rightarrow \boxed{n = \left[\frac{z_\alpha \cdot \sqrt{p_0(1-p_0)} + z_\beta \sqrt{p_a(1-p_a)}}{p_0 - p_a}\right]^2}$$

Back to example...

$$n = \left[\frac{1.645\sqrt{(.15)(.85)} + 1.282\sqrt{(.10)(.90)}}{(.15 - .10)}\right]^2$$

$$n = 377.4 \rightarrow 378$$

10.5 Relationship between Hypothesis Testing and CIs

As discussed in 2050, we can test any hypothesis using an appropriate CI.

$H_0: \theta = \theta_0$ $\quad$ RR: $\left\{ \frac{\hat{\theta}-\theta_0}{\sigma_{\hat{\theta}}} < -z_{\alpha/2} \cup \frac{\hat{\theta}-\theta_0}{\sigma_{\hat{\theta}}} > z_{\alpha/2} \right\}$

$H_a: \theta \neq \theta_0$

$\Rightarrow$ acceptance region is

$$\overline{RR}: \left\{ -z_{\alpha/2} \leq \frac{\hat{\theta}-\theta_0}{\sigma_{\hat{\theta}}} \leq z_{\alpha/2} \right\}$$

with simple algebra we restate accepting H_0 is equivalent to

① $-z_{\alpha/2} \leq \frac{\hat{\theta}-\theta_0}{\sigma_{\hat{\theta}}} \rightarrow \theta_0 - z_{\alpha/2}\cdot\sigma_{\hat{\theta}} \leq \hat{\theta} \rightarrow \theta_0 \leq \hat{\theta} + z_{\alpha/2}\cdot\sigma_{\hat{\theta}} = \hat{\theta}_u$

and

② $\frac{\hat{\theta}-\theta_0}{\sigma_{\hat{\theta}}} \leq z_{\alpha/2} \rightarrow \hat{\theta} \leq \theta_0 + z_{\alpha/2}\cdot\sigma_{\hat{\theta}} \rightarrow \underbrace{\hat{\theta} - z_{\alpha/2}\cdot\sigma_{\hat{\theta}}}_{(\hat{\theta}_L)} \leq \theta_0$

i.e. we accept H_0 (don't reject H_0) if $\theta_0 \in (\hat{\theta}_L, \hat{\theta}_u)$ otherwise we reject H_0.

10.21 99% CI for $\mu_1 - \mu_2$:

$$\bar{y}_1 - \bar{y}_2 \pm z_{.005}\cdot\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(1.65 - 1.43) \pm 2.576\sqrt{\frac{.26^2}{30} + \frac{.22^2}{35}}$$

$$.22 \pm .155$$

$$\boxed{(.065, .375)}$$

for $H_0: \mu_1 = \mu_2$ $(\mu_1 - \mu_2 = 0)$ our CI indicates $\mu_1 - \mu_2 \neq 0$
$H_a: \mu_1 \neq \mu_2$ $(\mu_1 - \mu_2 \neq 0)$ $\Rightarrow$ Reject H_0

note: same decision as before

What about one-sided tests?

- Compute a one-sided upper or lower CI, whichever is appropriate

10.24 Children's hospital

$H_0: p = .15$

$H_a: p < .15$ $\alpha = .05$

$n = 100$ $\hat{p} = .13$ Compute a 95% upper CI for p

$$\hat{p} + z_\alpha \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$.13 + 1.645\sqrt{\frac{(.13)(.87)}{100}} = .13 + .055 = .1855$$

Interpretation: we are 95% confident that $p < .1855$.
Therefore, we cannot say that $p < .15$ with much confidence and we don't reject H_0 (same decision as before)

note: This isn't an exact match w/ the test because we are estimating $\sigma_{\hat{p}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ instead of using $\sqrt{\frac{p(1-p)}{n}}$

10.6 Another way to report the results of a Statistical Test

We know from intro stats, that researchers have some leeway on what α level to use. One way to view α (significance level) is that it quantifies how much evidence is required to reject Ho.

α: .01 ——— .10

.01 ↑ a lot of evidence required

.10 ↑ "some" evidence is required

Naturally 2 different researchers could look at the same data and reach different conclusions.

Def 10.2 If W is a test statistic, the p-value, or attained significance level, is the smallest level of significance α for which the observed data indicate Ho should be rejected.

e.g. Professor Heiny shooting free throws

$H_0: p = .80$

$H_a: p < .80$

Professor Heiny shoots 20 free throws ; Y = no. made

RR: $Y \le 12$

$$\alpha = P(Y \le 12 \mid p = .80) = \sum_{y=0}^{12} \binom{20}{y} (.8)^y (.2)^{20-y} = .032$$

Spse $Y = 14$

$$\text{p-value} = P(Y \le 14 \mid p = .80) = \sum_{y=0}^{14} \binom{20}{y} (.8)^y (.2)^{20-y} = .196$$

i.e. the smallest α could have been while still rejecting Ho for $Y = 14$ is .196.

Spse $Y=10$:

$$\text{p-value} = P(Y \le 10 \mid p = .80) = .0026$$

i.e. the smallest α could have been while still rejecting H_0 for $Y=10$ is $.0026$

In general...

$H_0: \theta = \theta_0$ $\quad$ p-value $= P(W \le \underbrace{w_0}_{\text{observed test stat.}} \mid H_0 \text{ is true})$

$H_a: \theta < \theta_0$

$H_0: \theta = \theta_0$ $\quad$ p-value $= P(W \ge \underbrace{w_0}_{\text{observed test stat.}} \mid H_0 \text{ is true})$

$H_a: \theta > \theta_0$

Conceptually, I like to think of the p-value as quantifying the likelihood of the observed sample data when H_0 is true. Back to free throws...

$H_0: p = .80$

$H_a: p < .80$

Y	15	14	13	12	11	10
p-value	.37	.196	.087	.032	.010	.003

The smaller the p-value, the less likely the data could have come from a pop. where H_0 is true i.e. more inclined to reject H_0.

At what point along here would you conclude (with confidence) that Prof. Henry is not an 80% shooter? This is your level of α.

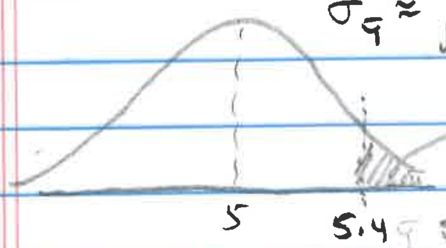
$\boxed{\text{simulation}}$ → can conduct statistical inference conceptually in intro stat

10.34 avg length of hospital stay

$H_0: \mu = 5$

$H_a: \mu > 5$ $\quad n = 500 \quad s = 3.1 \quad \bar{y} = 5.4$

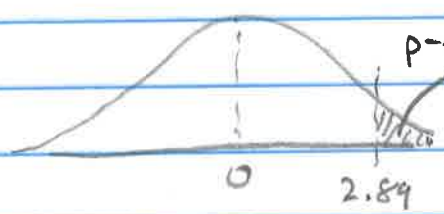
$\sigma_{\bar{y}} \approx \frac{3.1}{\sqrt{500}}$



The actual sampling dist. of $\bar{y}$ under H_0. How likely is it for $\bar{y}$ to get as far above 5 as 5.4?

Equivalent to ..

$$z = \frac{5.4 - 5}{3.1/\sqrt{500}} = 2.84$$



p-value is below our significance level of $\alpha = .05$ ⇒ Reject H_0

interpretation: If the avg. length of stay is really 5 days, there is about a 2-in-a-1000 chance we could have drawn a sample mean as high as 5.4 days.

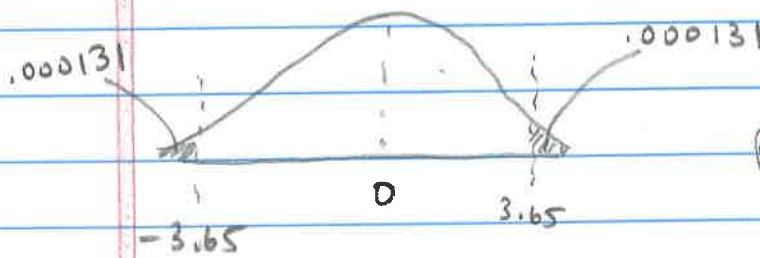
conclusion: $\mu > 5$

10.21 shear strength: 2 soil types

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

$$z = \frac{1.65 - 1.43}{\sqrt{\frac{.26^2}{30} + \frac{.22^2}{35}}} = 3.65$$



p-value = 2(.000131)

= .000262

For a two-sided alternative, we are interested in both directions: ($\mu_1 > \mu_2$ or $\mu_1 < \mu_2$) so multiply tail area by two

Be careful w/ p-values...

e.g. wt. loss study

$y \rightarrow$ wt. loss of each ptcp

μ = avg. wt. loss for pop.

$H_0: \mu = 0$ (diet not effective)

$H_a: \mu > 0$ (diet effective)   use $\alpha = .01$

$n = 950$ $\bar{y} = .9$ $s_y = 7.2$

$$z = \frac{.9 - 0}{7.2/\sqrt{950}} = 3.85$$

p-value = .00006

Just report p-value = .00006 statistically significant !!!

Is this practically significant? Use CI

99% CI for μ (avg. wt. loss)

$$.9 \pm 2.576 \frac{7.2}{\sqrt{950}} = .9 \pm .6 \rightarrow \boxed{(.3, 1.5)}$$

i.e. on avg people will lose between .3 and 1.5 lbs.
not very good but p-value was very small!

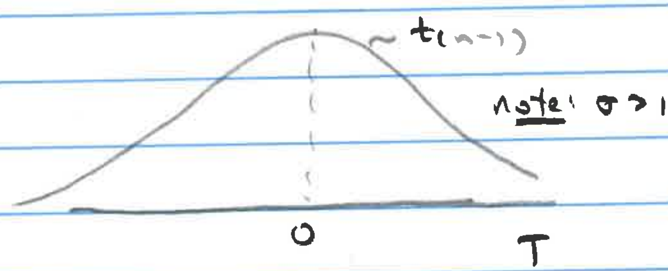
10.8 Small-Sample Hypothesis Tests for μ and $\mu_1 - \mu_2$

as discussed in ch. 7/8, $z = \frac{\bar{Y} - \mu}{S/\sqrt{n}}$ closely follows

a standard normal when $n > 30$. However, for smaller sample sizes this not the case and we need to use a different prob. dist. for our test statistic.

Let $Y_1 \cdots Y_n$ be a r.s. from $Y \sim N(\mu, \sigma^2)$. Then

$$T = \frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$$



It is important to keep in mind that we must be sampling from a normal pop. (or approx. normal). More on this later.

The one-sample T-test for μ is very simaler to the large sample Z-test for μ. area is marked out in the appropriate tails for the rejection region (RR)

$H_a: \mu > \mu_0$ RR: $\{T > t_\alpha\}$

$H_a: \mu < \mu_0$ RR: $\{T < -t_\alpha\}$

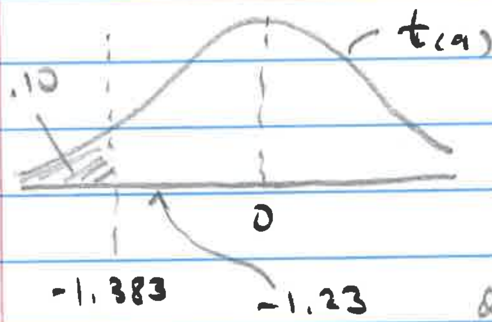
$H_a: \mu \neq \mu_0$ RR: $\{|T| > t_{\alpha/2}\}$

e.g. Co. has developed a new tranquilizer. They claim it slows patient's pulse rate on avg. by 4.50 beats per min. We are a bit skeptical and would like to test this at $\alpha = .10$

r.s. of 10 patients: $\bar{y} = 4.36$ $s = .36$

$H_0: \mu = 4.50$ vs. $H_a: \mu < 4.50$

$$T = \frac{4.36 - 4.50}{.36/\sqrt{10}} = -1.23$$



Don't reject Ho i.e. There is not enough evidence (at $\alpha = .10$) to conclude that the avg. reduction in beats per min. is less than 4.5.

What about a p-value?

- we can bound it using the Table (p-value > .10) and this is good enough to make decision.

or

- we can use technology (T.dist in excel / R)

p-value = .1249 (note: for Z, p-value = .1093) noticeable difference

What about a C.I.?

90% upper CI for μ: $4.36 + 1.383 \dfrac{(.36)}{\sqrt{10}} = 4.517$

i.e. we are 90% confident $\mu < 4.517$, but we cannot (with 90% confidence) conclude $\mu < 4.5$.

Pooled T-test

If we are comparing means between two pops using data collected from independant samples ...

$$z = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ follows approx. a std. normal dist.

However, if n_1 and n_2 are small and σ_1 and σ_2 are unknown we must use a t-dist. for our test statistic

from 8.8 ...

$$T = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n}}} \sim t_{(n_1+n_2-2)}, \quad s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

Assumptions:

1. Pops are at least approximately normal
2. $\sigma_1^2 = \sigma_2^2$

note: $\sigma_1^2 = \sigma_2^2$ seems very restrictive. However, Scheffé showed that for $n_1 = n_2$, T will follow $t_{(n_1+n_2-2)}$ even if $\sigma_1 \neq \sigma_2$. Furthermore, if $\sigma_1 \ll \sigma_2$ or $\sigma_1 \gg \sigma_2$, then comparing μ_1 and μ_2 as your primary research question is probably not appropriate.

e.g. (10.71) Are avg. body temps the same for men and women?

Men	Women
$n_1 = 9$	$n_2 = 9$
$\bar{y}_1 = 97.856$	$\bar{y}_2 = 98.4889$
$s_1^2 = .340$	$s_2^2 = .301$

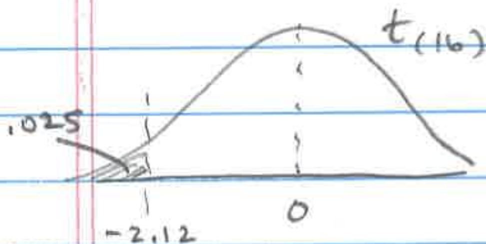
① normality? Probably pretty reasonable for body temp. Similar to ht / IQ, etc.

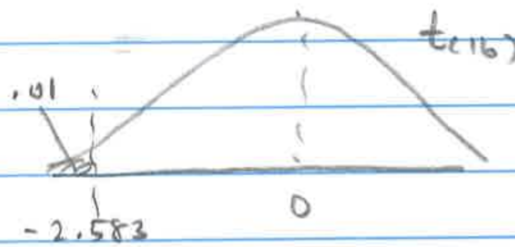
(2) $\sigma_1^2 = \sigma_2^2$? Looks reasonable based on observed s_1^2, s_2^2

$$s_p^2 = \frac{(9-1)(.340) + (9-1)(.301)}{9+9-2} = .3207$$

$H_0: \mu_1 = \mu_2$
$H_a: \mu_1 \neq \mu_2$

$$T = \frac{97.856 - 98.4889}{\sqrt{(.3207)\left(\frac{1}{9}+\frac{1}{9}\right)}} = -2.37$$





$\Rightarrow$ $2(.01) < p\text{-value} < 2(.025)$ $\Rightarrow$ $.02 < p\text{-value} < .05$

i.e. Enough evidence to conclude Men and Women have different avg. body temps at $\alpha = .05$, but not enough evidence at $\alpha = .01$.

95% CI for $\mu_1 - \mu_2$:

$$(97.856 - 98.4889) \pm 2.12\sqrt{(.3207)\left(\frac{1}{9}+\frac{1}{9}\right)}$$

$$-.6333 \pm .5659 \rightarrow \boxed{(-1.2, -.07)}$$

98% CI for $\mu_1 - \mu_2$:

$$-.6333 \pm 2.583\sqrt{(.3207)\left(\frac{1}{9}+\frac{1}{9}\right)}$$

$-.6333 \pm .6845 \rightarrow \boxed{(-1.323, .056)}$

<u>note</u>: The pooled T-test is equivalent to an ANOVA w/ 2 Trts (i.e. $F = t^2$)

<u>Non-Normality</u>

- moderate departures from normality have little effect on the prob. dist. of the test statistic
  i.e. the assumption of normality is <u>robust</u>

- Normality is often found in nature, so the T-test for inference regarding means is very common.

- For large departures from normality, use non-parametric procedures such as Wilcoxon Signed Rank Test and Mann-Whitney test. (T-test has more power if normality robust.)