15 : distribution of $\bar{x}$

16 : confidence intervals

17 : hypothesis tests

20 : t-scores

21 : distribution of $\bar{x}_1 - \bar{x}_2$

$\bar{x}$

p

Chapter 22 : Inference about Proportions

Ex: 50 % of Americans say that breakfast is the most important meal, but only ~30 % eat breakfast regularly.

Let p be the proportion who eat breakfast regularly. $p = .3$

$$\bar{x} \longrightarrow \mu$$

$$\boxed{\hat{p}} \longrightarrow p$$

Def: In a sample, the proportion of individuals with a certain statistic is written $\hat{p}$ (read p-hat)

Thm: In a sample with n individuals, the distribution of $\hat{p}$ is approximately $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$. As with $\bar{x}$, this is a better approximation as n increases.

<u>Ex</u>: We take an SRS of 1500 Americans and find that 470 regularly eat breakfast. $\hat{p} = \frac{470}{1500} \approx .313$.

$\hat{p}$ is roughly $N\left(.3, \sqrt{\frac{(.3)(.7)}{1500}}\right)$

$$N(.3, .012)$$

Recall: $\bar{x}$

know σ — z^*

didn't know σ — t^*

In the case of $\hat{p}$, $\mu = p$ and $\sigma = \sqrt{\frac{p(1-p)}{n}}$

involves p, the thing we want to approximate

$\Rightarrow$ we will never know σ beforehand.

<u>Prop</u>: the confidence interval for p given $\hat{p}$ and n is

$$\hat{p} \pm \boxed{z^*} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

(approximates $\hat{p} \pm z^* \sqrt{\frac{p(1-p)}{n}}$)

When we did this approximation to $\bar{x}$, we went from $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$ to

$$\bar{x} \pm \boxed{t^*} \frac{s}{\sqrt{n}}$$

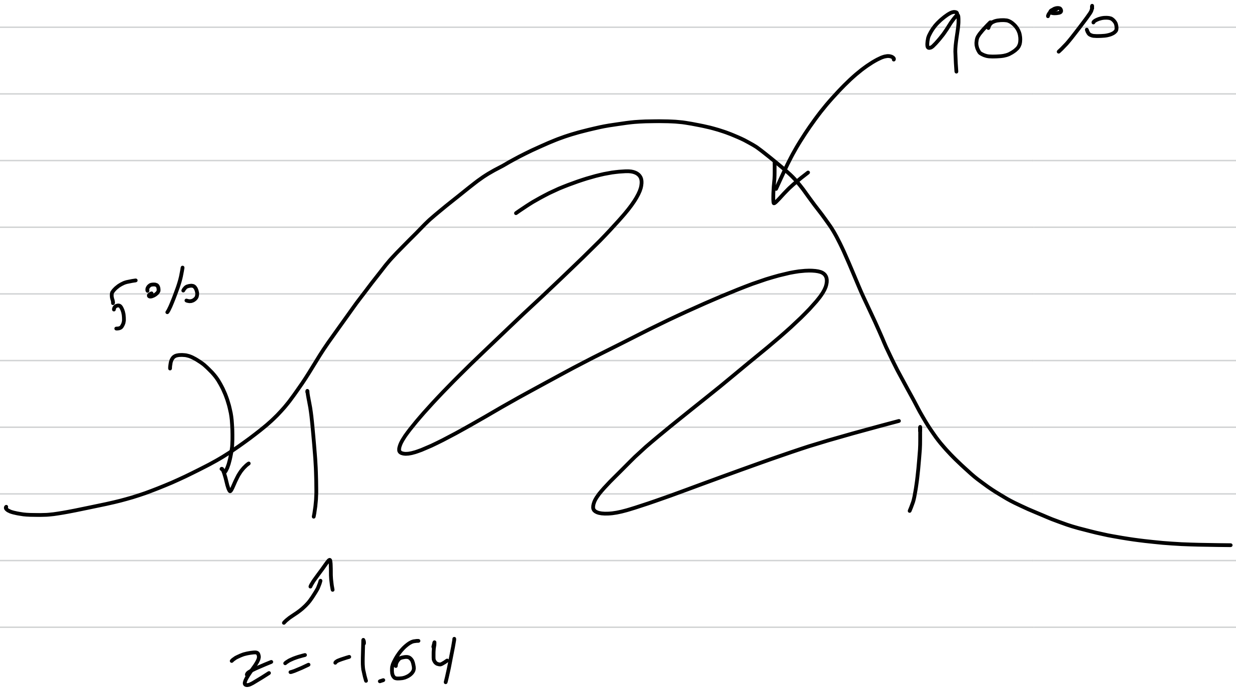
Why are we not using a t-score?

It's because the mean is used in calculating the standard deviation. (Lots of complicated math going on behind the scenes that we don't need to worry about)

Ex: sample 30 Americans on whether they eat breakfast, and 9 do. Find a 90% confidence interval for the proportion of all Americans that eat breakfast.

$\hat{p} = .3$ $\quad n = 30$

CI: $.3 \pm z^* \sqrt{\frac{.3(1-.3)}{30}}$



$z^* = 1.64$

$$.3 \pm 1.64 \sqrt{\frac{.3 \cdot .7}{30}}$$

$$.3 \pm .137$$

Margins of error: $MoE = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$\hat{p}$ depends on n, so we need to approximate $\hat{p}$ before taking a sample.

Two ways to accomplish this:

1. A previous sample was taken and approximated p.

2. Take $\hat{p} \approx .5$ to find n. This is okay because $\hat{p} = .5$ maximizes MoE, so what $\hat{p}$ ends up being

your approximation of p will be at least as good as if $\hat{p}$ were .5.

Ex: Two candidates running for mayor. You take a SRS to find the proportion of the population voting for candidate #1. You want 90% confidence and a margin of error no larger than .03. How many people do you need to survey?

Assume $\hat{p} = .5$

$$MoE = z^* \sqrt{\frac{.5(1-.5)}{n}}$$

$$.03 = 1.64\sqrt{\frac{.25}{n}}$$

$$.03 = 1.64\,\frac{.5}{\sqrt{n}}$$

$$\sqrt{n} = \frac{(1.64)(.5)}{.03} = 27.33$$

$$n = 747.11 \Rightarrow 748$$

Thm: Suppose we have a proportion p_0 of the population that has a certain statistic and a subset of that population whose proportion is p. The null hypothesis that $p = p_0$ has test statistic $z = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$.

Want to use this test when

n is large enough that both np_0 and $n(1-p_0)$ are at least 10

Ex: 20 pairs of dogs and their humans per sheet, 2 sheets — 1 with dogs and humans matched, the other not. Students picked either sheet 1 or sheet 2 based on which they thought had a stronger resemblance. There were 61 students, and 49 chose correctly. If there were no correlation, we'd expect 50% of the students to guess right. Is this sufficient evidence to indicate that the students were doing better than guessing?

$H_0: p = .5$

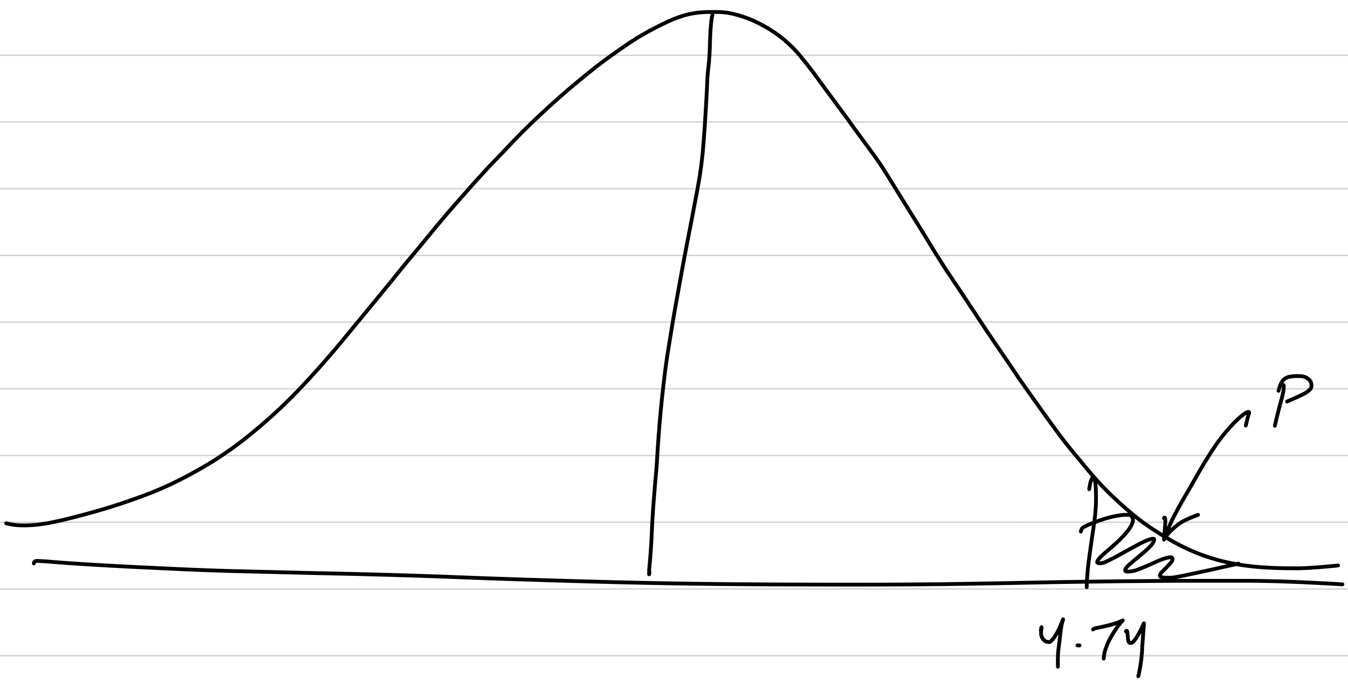
$H_a: p > .5$

$\hat{p} = 49/61 = .8033$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.8033 - .5}{\sqrt{\frac{.5(1-.5)}{61}}}$$

$z = 4.74$

Is this valid? $n p_0 = 61 \cdot .5 = 30.5$ and $61(1-.5) = 30.5$, both of which are ≥ 10, so yes.



p is so small that it is definitely less than .05 (we weren't given a value for α in the problem statement, so we take $\alpha = .05$). Therefore, we reject the null hypothesis.

Ex: Two sample of 29 and 35 people respectively, from two group measure heights

$\bar{x}_1 = 65$ inches

$\bar{x}_2 = 66$ inches

$s_1 = 2$

$s_2 = 3$

distribution $\bar{x}_1 - \bar{x}_2$ is

$$N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

If we want to get a 95 % CI,

we take $\bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

$\min(28, 34) = 28$

$$(65 - 66) \pm 2.048 \sqrt{\frac{2^2}{29} + \frac{3^2}{35}}$$

$$-1 \pm 1.287$$

$\bar{x}_1 - \bar{x}_2$

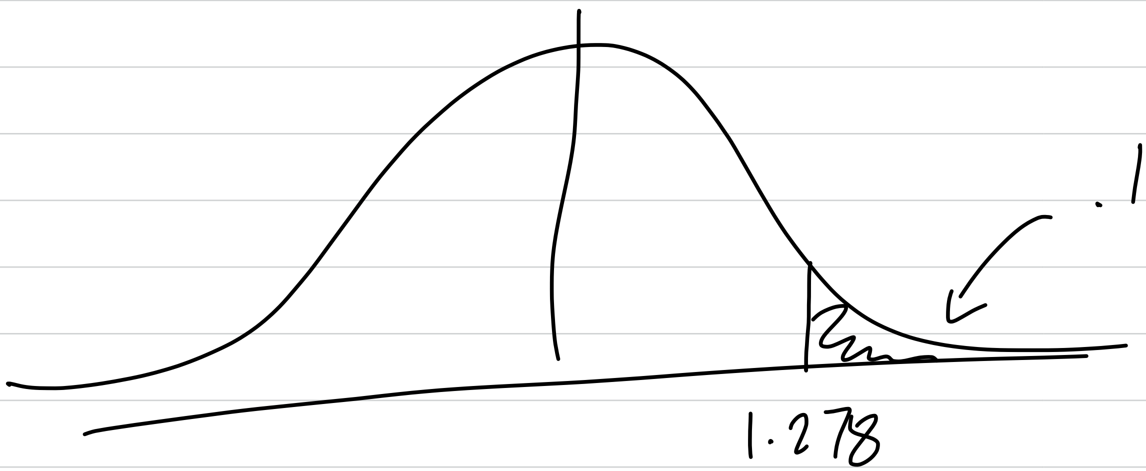
$\text{DOF} = \min(n_1 - 1, n_2 - 1)$

on campus

$H_0: \mu = 2.7$ $\quad$ $H_a: \mu > 2.7$

$\bar{x} = 2.9$ $\quad$ $n = 20$

$$z = \frac{2.9 - 2.7}{.7/\sqrt{20}} = 1.278$$



$$.1 \overset{?}{<} .05 \quad \text{No}$$

Fail to reject

$\bar{x} = 2.485 \qquad s = .819$

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

$$2.485 \pm t^* \frac{.819}{\sqrt{20}}$$

$C = 95\%$ DOF $= 19$

$t^* = 2.093$

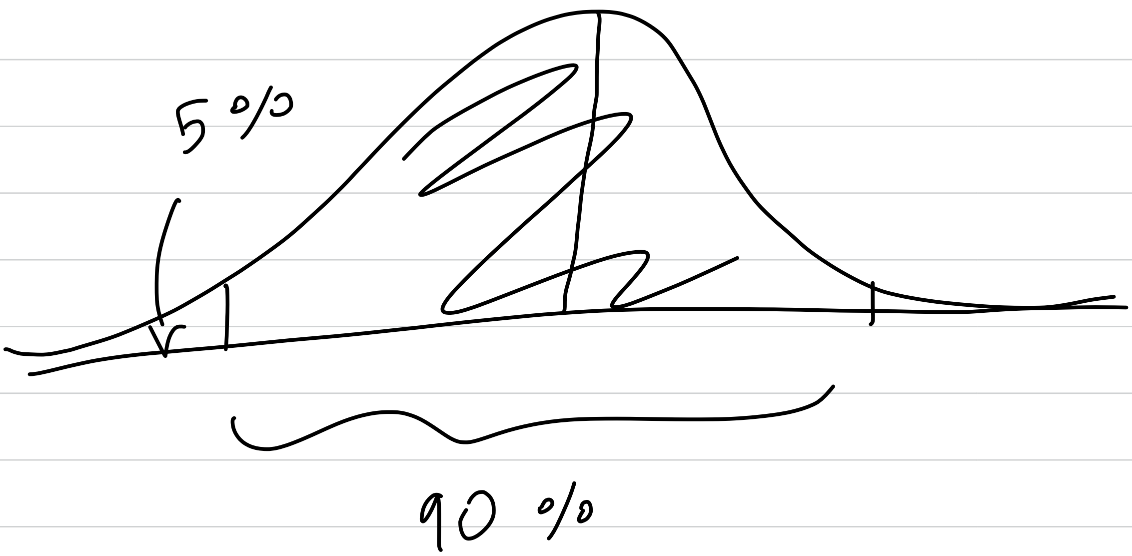
$$2.485 \pm 2.093 \frac{.819}{\sqrt{20}}$$

$2.485 \pm .383$

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

$$26.8 \pm z^* \frac{7.8}{\sqrt{654}}$$

$C = 90\%$



$z = -1.64$

$z = 1.64$

$$26.8 \pm \boxed{1.64 \frac{7.8}{\sqrt{654}}}$$

Know $\mu = 2.7$

Hypothesis: First-year students have a lower GPA than overall average.

$H_0 : \mu = 2.7$ Use $\alpha = .02$

1st-year

$H_a : \mu < 2.7$

one-sided

$\bar{x} = 2.485$

$s = .819$

$n = 20\%$

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

$$t^* = \frac{2.485 - 2.7}{.819/\sqrt{20}} = +1.174$$

DOF = 19

Closest on table on row 19

is $t^* = 1.066$

One-sided p-value: .15

Not less than .02

fail to reject