

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.  $p = .3$

$$\bar{x} \rightarrow \mu$$

$$\boxed{\hat{P}} \rightarrow P$$

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

$N(p, \sqrt{\frac{p(1-p)}{n}})$ . As with  $\bar{x}$ ,

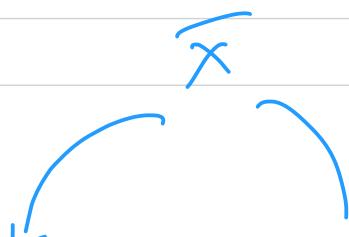
this is a better approximation as  $n$  increases.

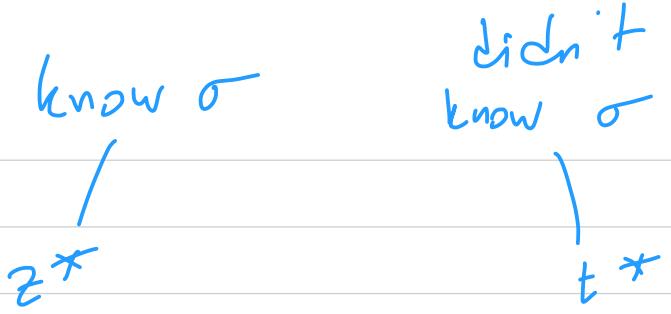
Ex: we take an SRS of 1500 Americans and find that 470 regularly eat breakfast.  $\hat{p} = \frac{470}{1500} = .313$ .

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

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

Recall:





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  $\sigma$  beforehand.

Prop: the confidence interval for  $p$

given  $\hat{p}$  and  $n$  is

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

(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

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

Why are we not using a t-score?

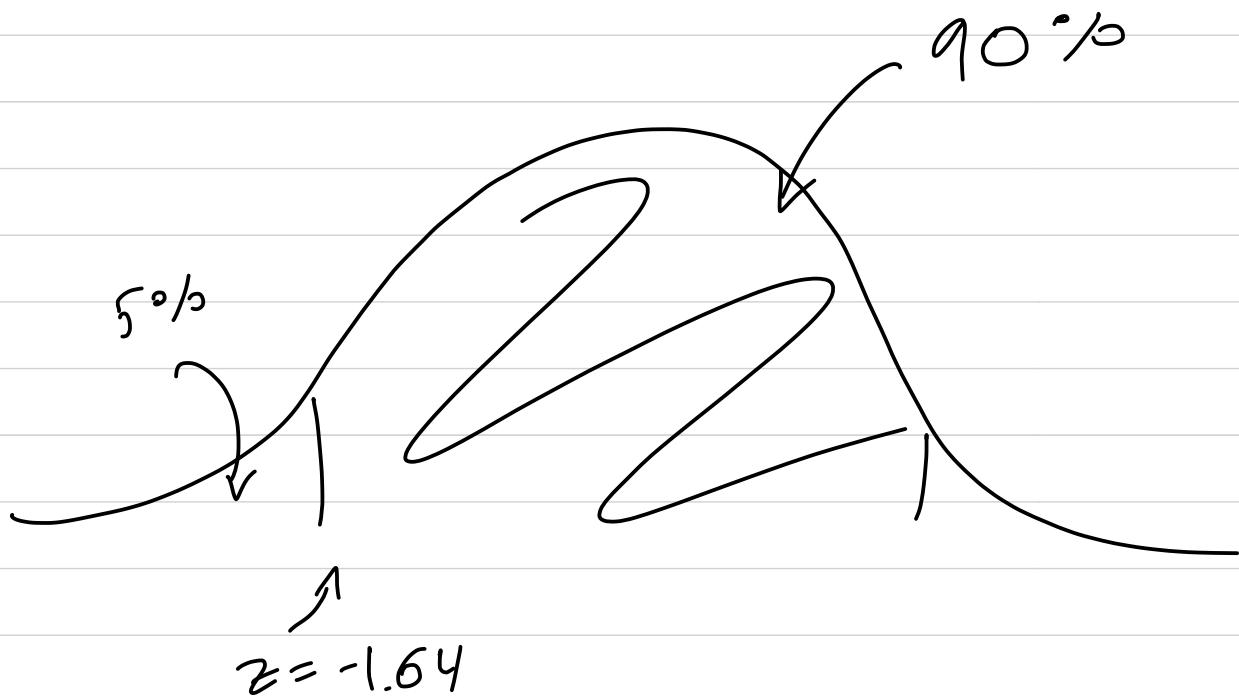
It's because the mean is used in calculating the standard deviation. (Lots of complicated math going on behind the scenes that we don't need to worry about)

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:  $MoE = z^* \sqrt{\hat{p}(1-\hat{p}) / n}$

$\hat{p}$  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  $\hat{p} = .5$

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

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

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

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

$$n = 747 \quad \Rightarrow \quad 748$$

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

Want to use this test when

$n$  is large enough that both  
 $n p_0$  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?

$$H_0 : p = .5$$

$$H_a : p > .5$$

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

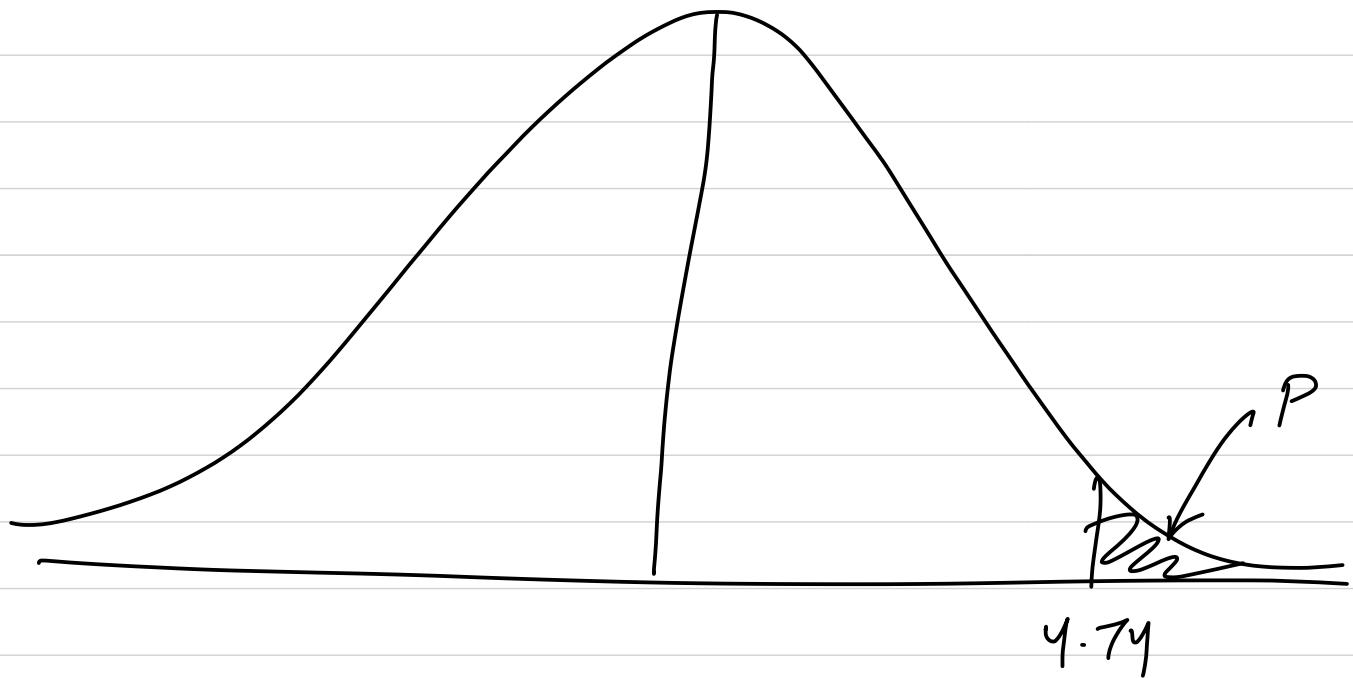
$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.8033 - .5}{\sqrt{\frac{.5(1-.5)}{61}}}$$

$$z = 4.74$$

Is this valid?  $n p_0 = 61 \cdot .5 = 30.5$

and  $61(1-.5) = 30.5$ , both of

which are  $\geq 10$ , so yes.



$P$  is so small that it is definitely less than .05 (we weren't given a value for  $\alpha$  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

$$\bar{x}_1 = 65 \text{ inches}$$

$$\bar{x}_2 = 66 \text{ inches}$$

$$s_1 = 2$$

$$s_2 = 3$$

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

$$N\left(\mu_1 - \mu_2, \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)$$

If we want to get a 95% CI,

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

$$\min(28, 34) = 28$$

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

$$-1 \pm 1.287$$

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

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

on campus

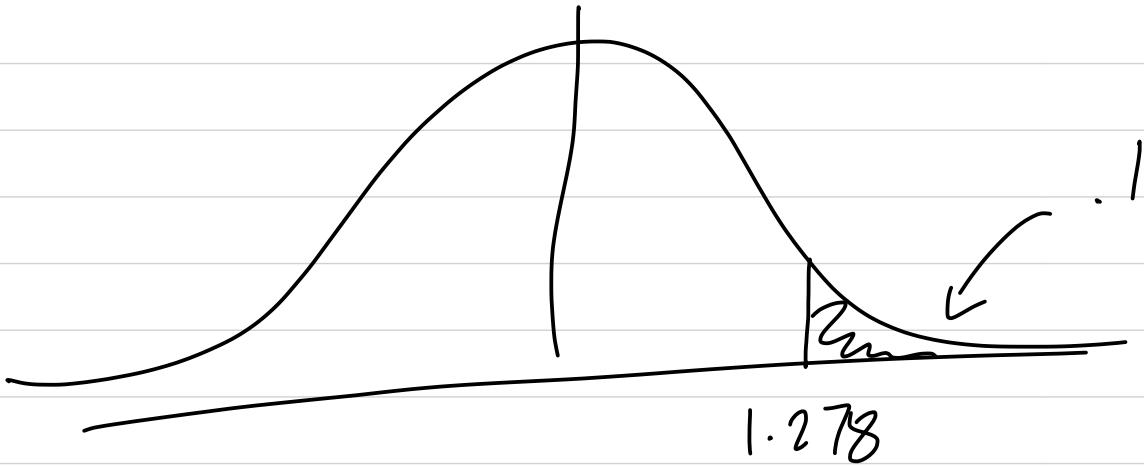


$$H_0: \mu = 2.7$$

$$H_a: \mu > 2.7$$

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

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



?  
 $.1 < .05$       No

Fail to reject

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

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

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

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

$$t^* = 2.093$$

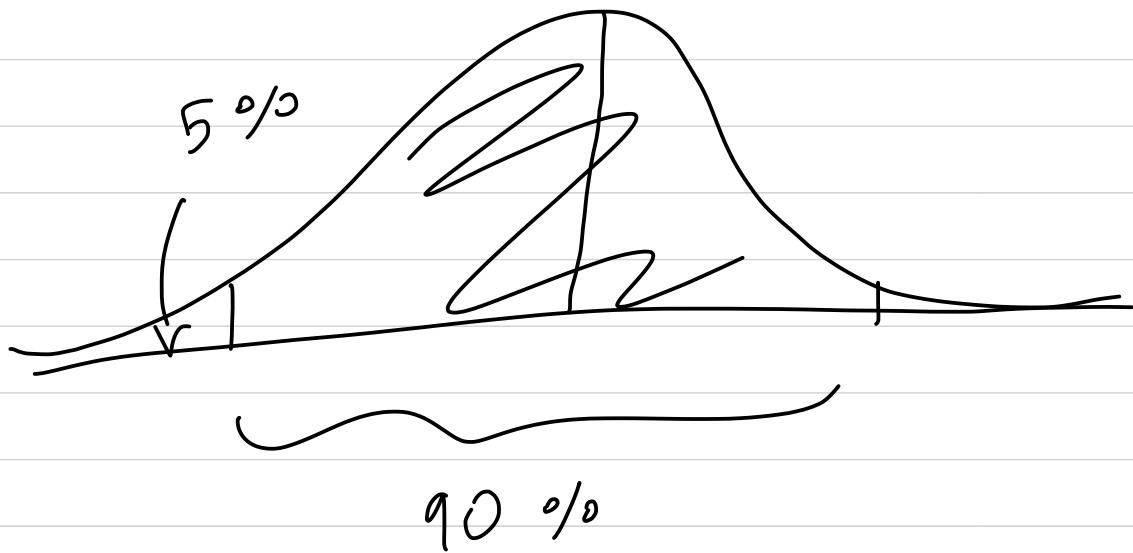
$$2.485 \pm 2.093 \frac{-819}{\sqrt{20}}$$

$$2.485 \pm .383$$

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

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

$$C = 90\%$$



$$z = -1.64$$

$$z = 1.64$$

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

Know  $\mu = 2.7$

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

$$H_0 : \mu = 2.7 \quad \text{Use } \alpha = .02$$

1st-year

$$H_a : \mu < 2.7$$

One-sided

$$\bar{x} = 2.485$$

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

$$s = .819$$

$$n = 20 \%$$

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

$$DOF = 19$$

(Closest on table on row 19)

is  $t^* = 1.066$

One-sided p-value: .15

Not less than .02

fail to reject

## Quiz 4

①  $\bar{x}_1 - \bar{x}_2$  is roughly  $N(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$

$$\min(n_1 - 1, n_2 - 1)$$

②  $\mu_0 = 2.7$        $\sigma_0 = .7$

one-sided  
P-value

$$H_0: \mu = 2.7$$

students on campus

$$H_a: \mu > 2.7$$

$$\bar{x} = 2.9, \quad n = 20, \quad \alpha = .05$$

$\bar{x}$  is roughly  $N(2.7, \frac{.7}{\sqrt{20}})$

$$z = \frac{2.9 - 2.7}{.7/\sqrt{20}} = \frac{.2}{.7/\sqrt{20}} = 1.278$$



from z-score table  
↓

$$z = -1.278 \Rightarrow .1$$

$$P\text{-value} : p = .1$$

Is  $.1 < .05$ ? No, so we fail to reject the null hypothesis.

③  $n=20, \bar{x}=2.485, s=.819$

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

$$t^*: C = 95\%, \text{DOF} = 19$$

$$t^* = 2.093 \quad \leftarrow \begin{array}{l} \text{t-score table, row 19,} \\ \text{column 95\%} \end{array}$$

$$\text{CI: } 2.495 \pm 2.093 \cdot \frac{.819}{\sqrt{20}} = 2.495 \pm 3.83$$

## Chapter 23: Comparing Two Proportions

Recall:  $p$ : population proportion

$\hat{p}$ : sample proportion

$\hat{p}$  is roughly  $N(p, \sqrt{\frac{p(1-p)}{n}})$ , so approximately  $N(p, \sqrt{\frac{\hat{p}(1-\hat{p})}{n}})$ .

With  $p$ , we use  $Z^*$  rather than  $t^*$ .

Ex: We want to study the difference in proportions of UO and OSU students who live on campus. We take a sample of 30 students from UO and 35 from OSU and find that

9 of the UO students and 12 of the OSU students live on campus. How do we estimate the difference in proportions among all students?

$p_1$  : proportion of UO students living on campus

$p_2$  : proportion of OSU students living on campus

Want to estimate  $p_1 - p_2$  using  $\hat{p}_1 - \hat{p}_2$ .

Thm: the distribution of  $\hat{p}_1 - \hat{p}_2$  is approximately  $N(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}})$

$n_1, n_2$  = sample sizes of  $\hat{P}_1$  and  $\hat{P}_2$

Ex

$$n_1 = 30$$

$$n_2 = 35$$

$$\hat{P}_1 = 9/30 = .3$$

$$\hat{P}_2 = 12/35 = .343$$

So the distribution of  $\hat{P}_1 - \hat{P}_2$  is approximately

$$N(-.043, \sqrt{\frac{.3(1-.3)}{30} + \frac{.343(1-.343)}{35}}), \text{ so}$$

$$N(-.043, .116)$$

To take a confidence interval for

$P_1 - P_2$ , take the standard error

$$SE = \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}. \text{ Then the}$$

confidence interval is  $\hat{P}_1 - \hat{P}_2 \pm z^* SE$ .

Ex: If we want a 99% CI for the difference in proportions of students living on campus between UO and OSU,

we take  $\hat{P}_1 - \hat{P}_2 = -.043$ ,  $SE = .116$ ,

and  $z^* = 2.57$ . So we have a  
b/c  $\uparrow$   
 $-2.57$  gives .005 on the table

$$CI \text{ of } -.043 \pm .116 \cdot 2.57 = -.043 \pm .298.$$

Taking a hypothesis test of the difference in proportions: first, find the pooled sample proportion  $\hat{P} = \frac{\# \text{ total successes}}{\# \text{ total individuals}}$  between both groups. Then the z-score (assuming

$$H_0 \text{ is true) is } z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

What exactly is  $H_0$  in this case? It's always that  $P_1 = P_2$  (so  $\hat{P}_1 - \hat{P}_2 = 0$ )

Ex: Use the data about students living on campus to assess the claim that a different proportion of students live on campus at OSU compared to UO. Use a significance level of .1.

$$H_0: P_1 = P_2$$

$$H_a: P_1 \neq P_2$$

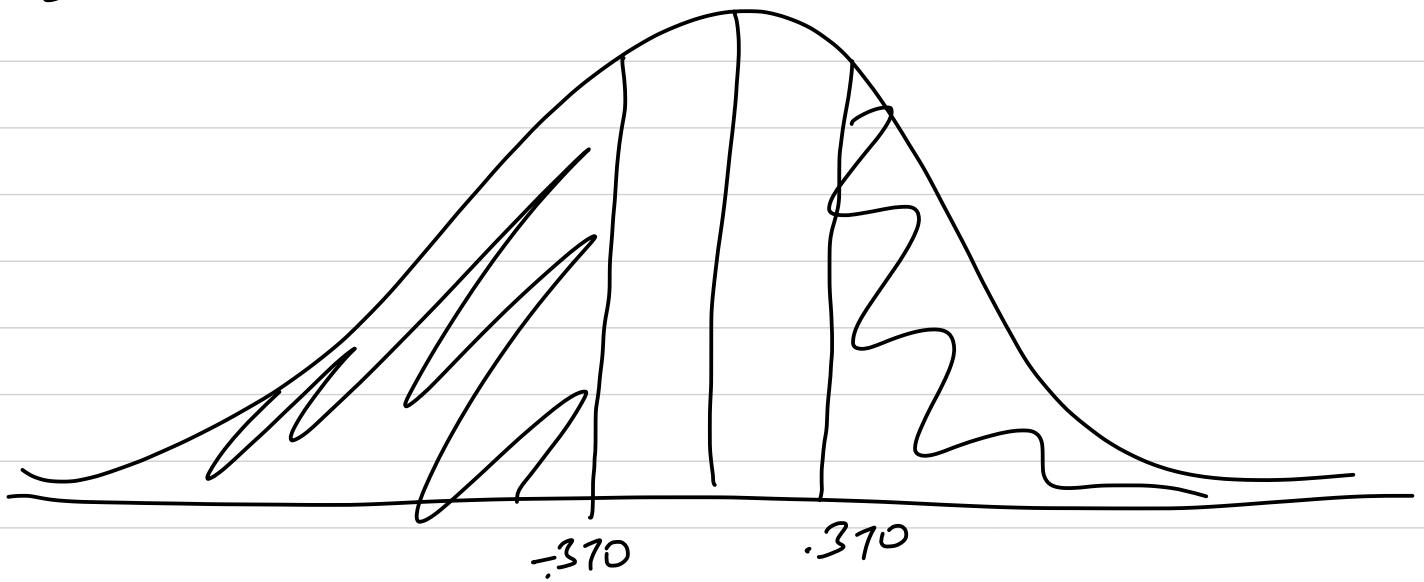
2-sided  
P-value

$$z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{-0.043}{\sqrt{.323(1-.323)\left(\frac{1}{30} + \frac{1}{35}\right)}}$$

$$\hat{P} = \frac{\# \text{ total students living on either campus}}{\# \text{ total students in sample}}$$

$$\hat{P} = \frac{9 + 12}{30 + 35} = .323$$

$$z = - .370$$



$$z = - .37 \Rightarrow .3557 \quad \text{by table}$$

$$P = 2 (.3557) = .7114 \neq \alpha = .1$$

So we fail to reject  $H_0$ .

①

Quiz 4  $\bar{X}_1 - \bar{X}_2$  is approximately  $N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$ .

$$\bar{x} : \text{DOF} = n - 1$$

$$\bar{X}_1 - \bar{X}_2 : \text{DOF} = \min(n_1 - 1, n_2 - 1)$$

②

$$\mu = 2.7$$

$$\sigma = .7 \Rightarrow$$

z-score

$$\bar{x} = 2.9$$

$$s = ? \Rightarrow t\text{-score}$$

$$n = 20$$

$$\alpha = .05$$

In general, if you know  $\sigma$ , use a z-score b/c it gives a better

approximation. Otherwise, try to use a t-score.

Hypothesis test: find z/t-score, find p-value, compare against  $\alpha$ .

$$H_0: \mu \xrightarrow{\text{on-campus}} = 2.7$$

$$H_a: \mu \xrightarrow{\text{on-campus}} > 2.7$$

$$z = \frac{\bar{x} - 2.7}{\cdot 7/\sqrt{20}} = \frac{2.9 - 2.7}{\cdot 7/\sqrt{20}} = 1.28$$



$$z = -1.28 \Rightarrow .1003$$

$$P = .1003$$

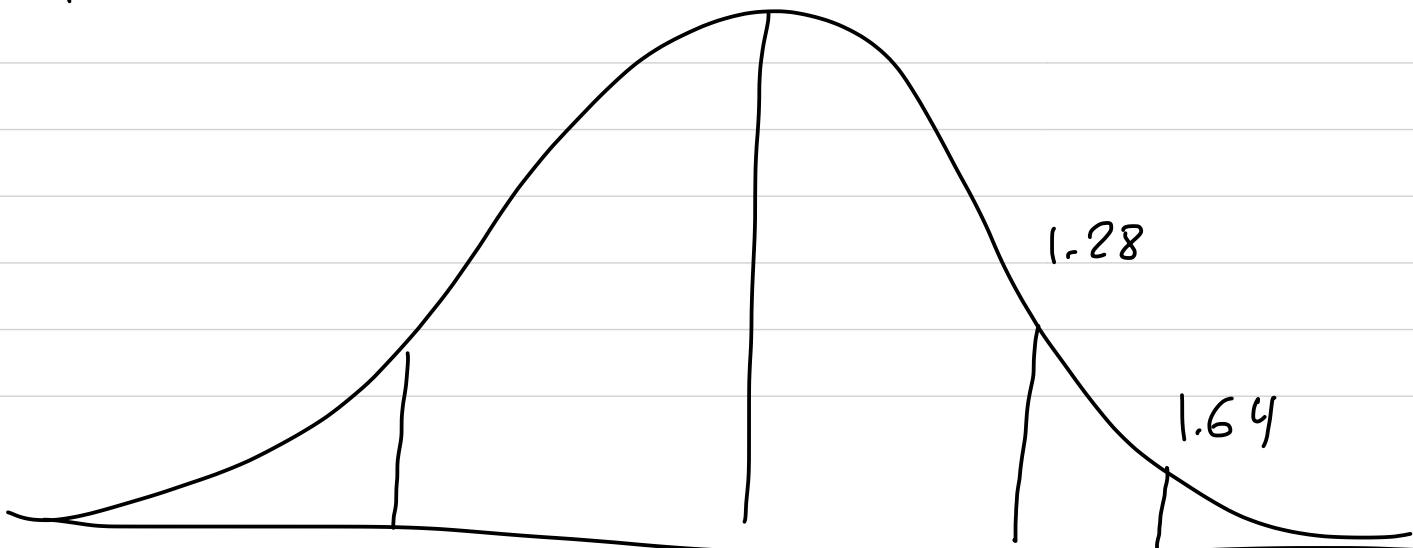
Is  $P < \alpha$ ? No, so we fail to reject  $H_0$ .

Alternatively:  $\alpha = .05 \Rightarrow z = -1.64$

We want  $z = 1.64$  So, we can

compare our  $z$ -value of 1.28 to

1.64



$$\mu = ?$$

$$\sigma = ?$$

$$\bar{x} = 2.485$$

$$s = .819 \implies t\text{-score}$$

$$n = 20$$

$$C = 95\%$$

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

don't know this

$$\text{Table: } t = 2.093$$

$$2.495 \pm 2.093 \quad \frac{.819}{\sqrt{20}}$$

$$2.495 \pm .383$$

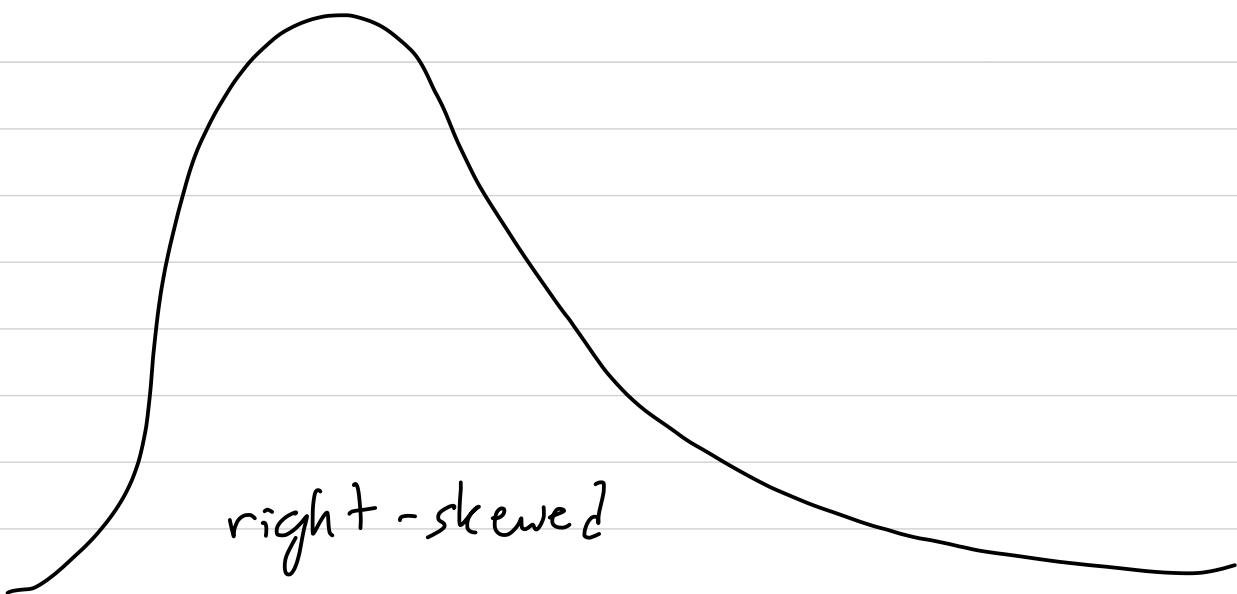
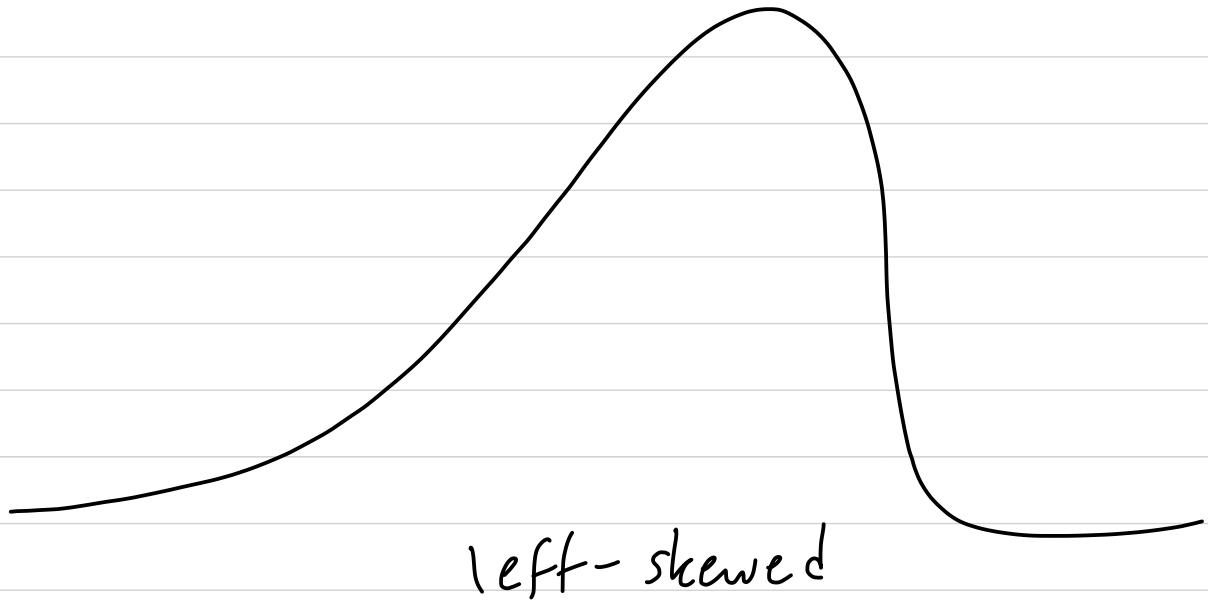
## Course summary

Chapter 1: Defs of individuals, statistics, variables, etc.

Distributions: give the information of what values a variable can take and how often it takes those values. Can be categorical or quantitative (i.e. categories vs. measurements) and discrete or continuous (i.e. finite # of possible values vs. possible values along a line or curve).

# Pie charts / bar graphs / histograms

Skews: the direction (if any) where the tail of a distribution extends



Center: mean or median

Variability: standard deviation or 5-number summary

Measures are resistant when they aren't significantly affected by outliers/skews

Stem-and-leaf plots

## Chapter 2

Def of mean, median, standard deviation,  
5-number summary

Method to find outliers given 5-num  
summary

Box-plots and a modified version

with outliers marked as circles:

once the outliers are removed and marked as dots, then the "whiskers" of the plot extend to the max and min non-outliers

Ex: 1, 2, 2, 5, -1, 20, 0, -2, -4, 0

-4, -2, -1, 0, 0, 1, 2, 2, 5, 20  
 $\underbrace{-1.5}_{-1.5}$        $\underbrace{.5}_{.5}$        $\underbrace{3.5}_{3.5}$

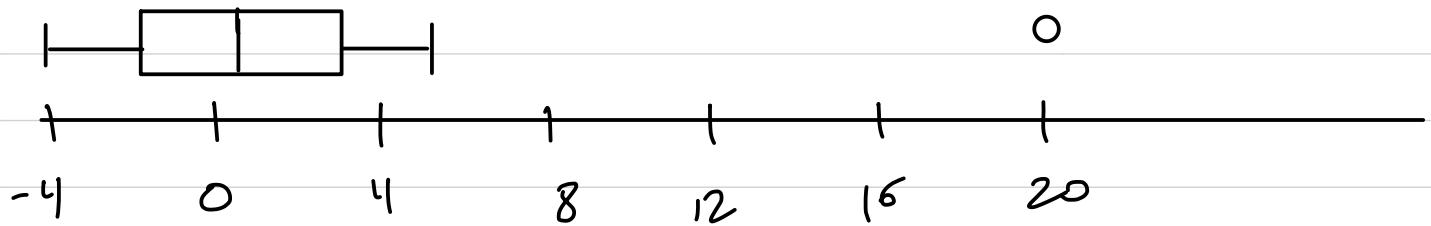
$$IQR = 3.5 + 1.5 = 5$$

$$-1.5 - 1.5 IQR \rightarrow 3.5 + 1.5 IQR$$

$$-9 \rightarrow 11$$

20 is an outlier

5-num: -4, -1.5, -5, 35, 20



## Chapter 3

z-scores: 
$$z = \frac{x - \mu}{\sigma}$$

	mean	std dev
pop	$\mu$	$\sigma$
sample	$\bar{x}$	$s$

68-95-99.7

# Z-score table

## Chapter 8

Population, sample, bias (i.e. one group being over/under represented)

SRS: equal chance for anyone to be chosen

Stratified Random Sample: split into groups of similar individuals, then take an SRS of each group

Clustered Random Sample : group by clusters like location and sample clusters

## Chapter 9

Observational studies vs. experiments (i.e. whether you attempt to change variables)

Correlation  $\neq$  causation

Lurking variables can make it look like an implication exists where it doesn't.



Explanatory and response variables:  
variables that we change and measure,  
respectively, in an experiment.

Control groups: groups of subjects  
that do not receive treatment

Randomization: using an SRS to divide  
subjects into groups

Replication: having the entire experiment  
replicated in other labs to verify the  
results.

Chapter 12

Sample space: the set  $S$  of all possible outcomes

Ex: 3 dice rolled, record the 3 numbers

$$S = \{(1,1,1), (1,1,2), \dots\}$$

Event: any outcome or group of outcomes in the sample space

$$S = \{(a, b, c) \mid a, b, c \in \{1, 2, 3, 4, 5, 6\}\}$$

Disjoint: two events that cannot occur simultaneously

Probability  $P(A) = \frac{\# \text{ of outcomes with } A}{\# \text{ of outcomes total}}$

Disjoint probabilities add: if  $A$  and  $B$  are disjoint, then  $P(A \text{ or } B) = P(A) + P(B)$ .

Inverting probabilities:  $P(\text{not } A) = 1 - P(A)$

Random variable: placeholder for the possible outcomes.

Ex: if  $X$  is the random variable corresponding to the result of rolling a 6-sided die, then we could write  $P(X=2) = \frac{1}{6}$ .

$P(X)$  ← meaningless

$P(X = x)$  ← a value that depends  
on  $x$

## Chapter 13

A and B are independent if knowing one occurs doesn't change the likelihood that other does.

Independent probabilities multiply:  
if A and B are independent,  
then  $P(A \text{ and } B) = P(A)P(B)$

Ex: You flip a fair coin twice.

Let A be the event of flipping heads on the first flip, B the event of tails on the first flip, and C the event of tails on the second flip. Then A and B are disjoint, A and C are independent, and B and C are also independent.

$$P(A) = .5$$

$$P(A \text{ or } B) = .5 + .5 = 1$$

$$P(B) = .5$$

$$P(A \text{ and } C) = (.5)(.5) = .25$$

$$P(C) = .5$$

$$P(B \text{ and } C) = .25$$

A and C being independent means:

$$P(A|C) = P(A) \quad \text{and} \quad P(C|A) = P(C)$$
$$\frac{.5}{.5}$$

Ex: if winning a lottery has probability .1, what is the chance of winning at least once across 5 plays? Not disjoint, but are independent so reframe the question to use independence.

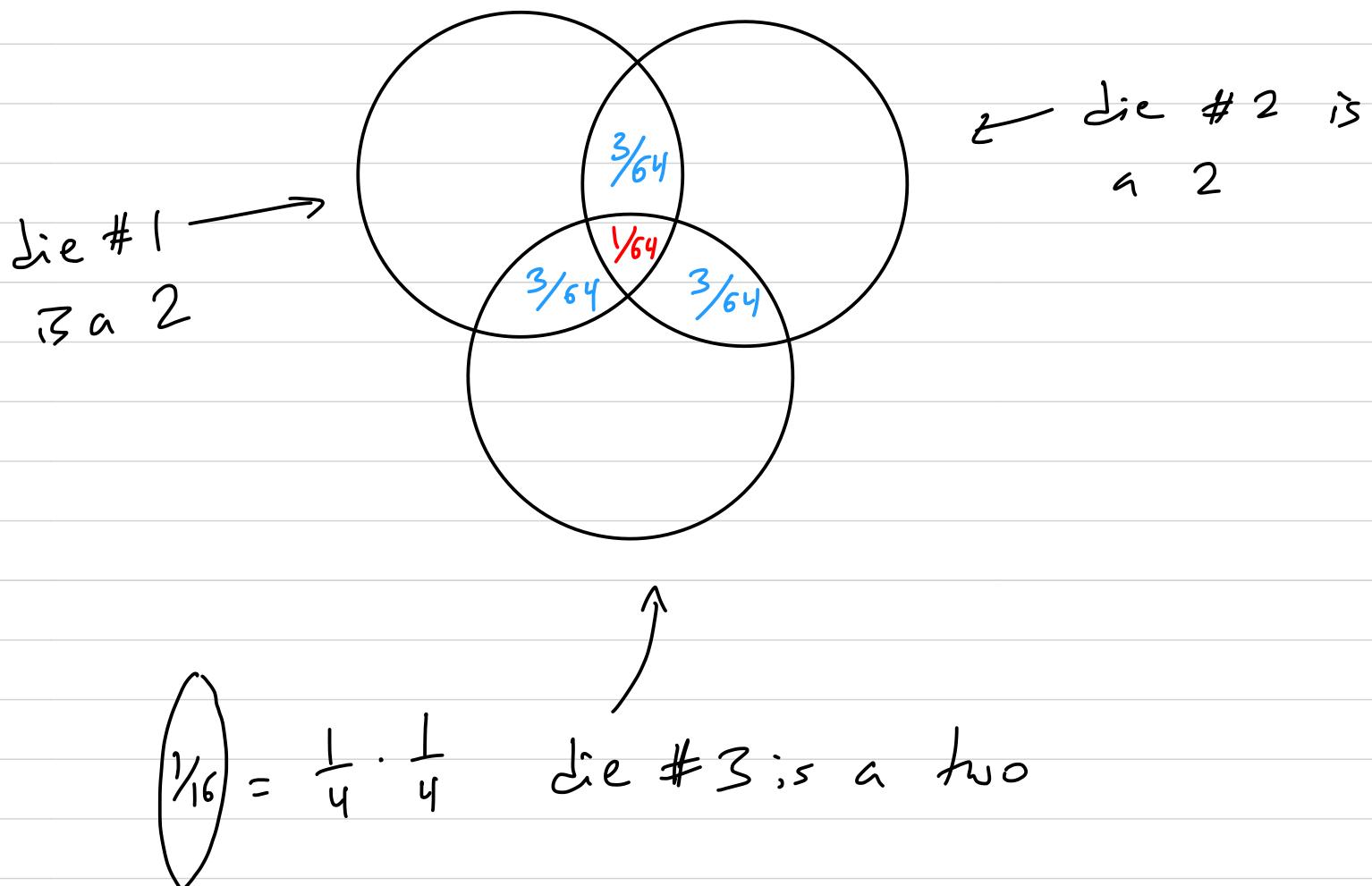
$$P(\text{not winning at least once}) = P(\text{losing 5 in a row})$$
$$= (.9)^5$$

$$\text{So } P(\text{winning at least 1}) = 1 - (.9)^5$$

Venn diagrams: use if a probability question isn't immediately obvious (and you can't Bayes' rule)

Venn diagrams make all probabilities disjoint

Ex: Roll 3 4-sided dice. What's the chance of getting exactly 2 2s?



$$\left(\frac{4}{64}\right) - \left(\frac{1}{64}\right) = \left(\frac{3}{64}\right)$$

$$\text{Total} = \frac{3}{64} + \frac{3}{64} + \frac{3}{64} = \frac{9}{64}$$

$P(A)$  = the probability that A occurs

$P(A|B)$  = the probability that A occurs  
after B has already occurred

Different from  $P(A \text{ and } B)$  = the probability  
that both A and B occur at  
once

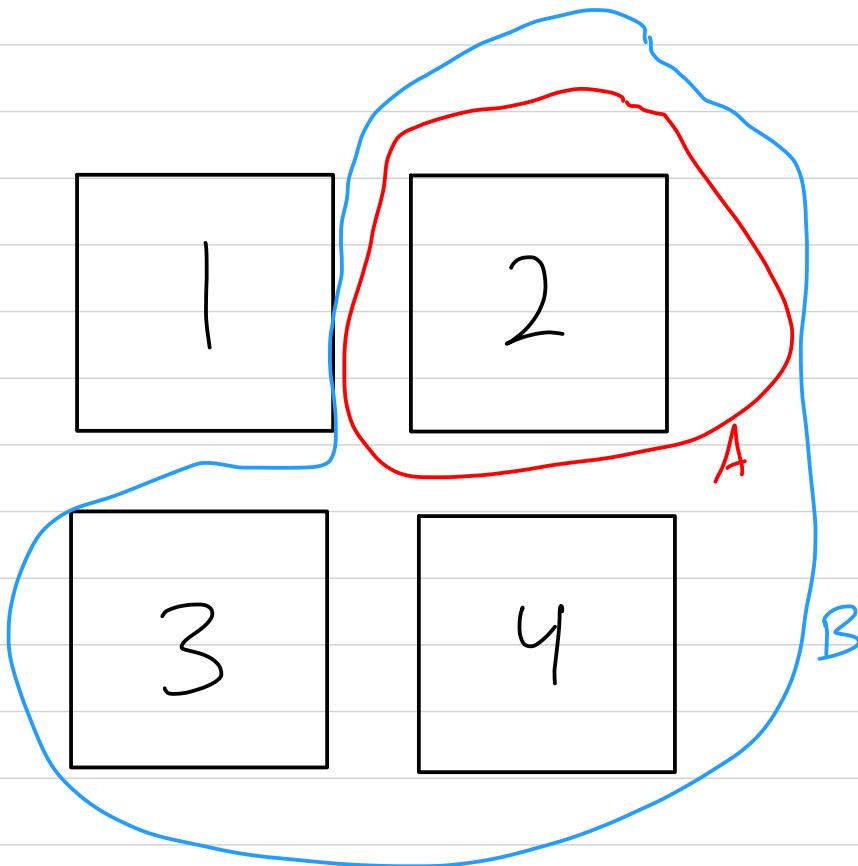
Let A be the event of rolling  
a single 2 on a single roll of a  
4-sided die. Let B be the event  
of not rolling a 1.

Then:  $P(A \text{ and } B) = \frac{1}{4}$

$P(A|B) = \frac{1}{3}$

$P(B|A) = 1$

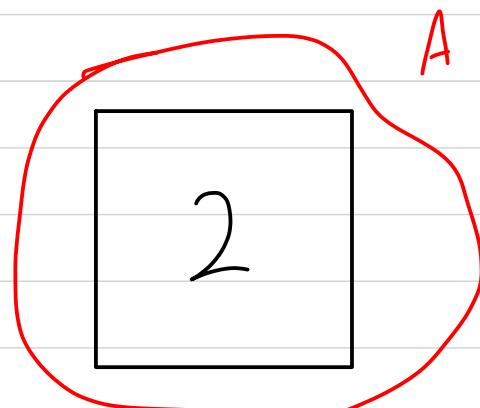
Given that you rolled a 2, what's the chance that the roll is not a 1?



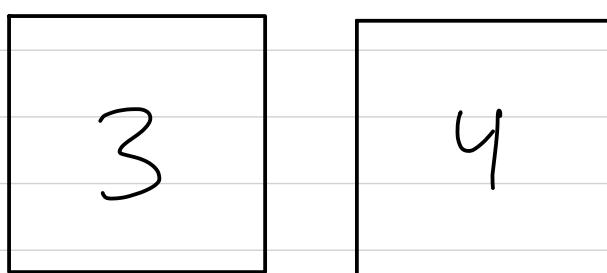
$$P(A) = \frac{1}{4}$$

$$P(B) = \frac{3}{4}$$

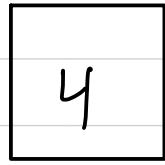
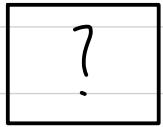
$$P(A \text{ and } B) = \frac{1}{4}$$



$$P(A|B) = \frac{1}{3}$$

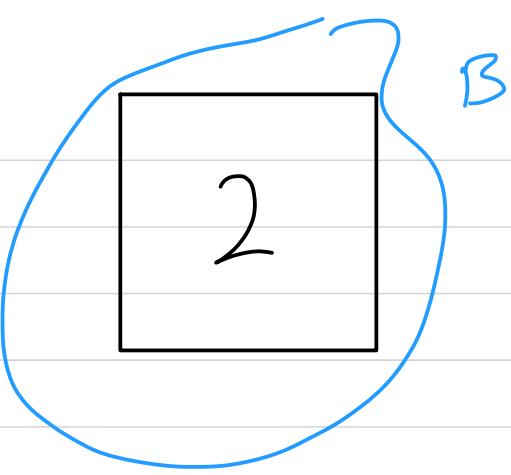


Ex: If you rolled 3 dice and wanted to find the probability of getting at least two 2s given that die #3 is a 4



$$P(2) = \frac{1}{4} \quad P(?) = \frac{1}{4}$$

$$\frac{1}{16}$$



$$P(B|A) = 1/1 = 1$$

You can think of  $P(A|B)$  as  $P(A)$   
where the sample space is  $B$ .

Ex: Bag of dice

Some blue, some 8-sided

$$P(B) = .2$$

$$P(E) = .1$$

$$P(E|B) = .15$$

If we remove all dice except the 8-sided ones, what is the chance that a random one will be blue?

$$P(B|E) = \frac{P(E|B) \cdot P(B)}{P(E)}$$
$$= \frac{.15 \cdot .2}{.1} = .3$$

## Chapter 15

the distribution of  $\bar{x}$  is approximately  $N(\mu, \sigma/\sqrt{n})$

single individual  $x : N(\mu, \sigma)$

sample mean  $\bar{x} : N(\mu, \sigma/\sqrt{n})$

sample standard deviation  $s$ : (didn't discuss)

"sample standard deviation":  $s$

"standard deviation of the sample mean":  $\sigma/\sqrt{n}$

$n=3$  sample from  $N(2, 4)$

Ex: 2, 3, -4

$$\bar{x} = \frac{2+3-4}{3} = \frac{1}{3}$$

$$s^2 = \frac{1}{3-1} \left( \left(2 - \frac{1}{3}\right)^2 + \left(3 - \frac{1}{3}\right)^2 + \left(-4 - \frac{1}{3}\right)^2 \right) \\ = 14.33$$

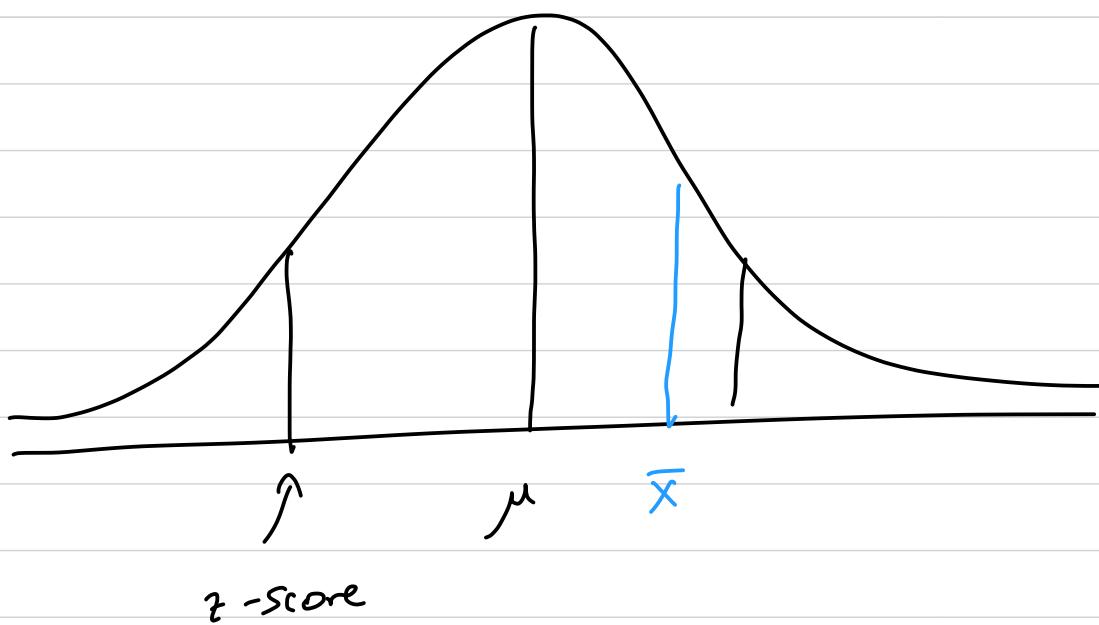
$$s = 3.786$$

$\bar{x}$  comes from  $N\left(2, \frac{4}{\sqrt{3}}\right)$   
 $= N(2, 2.31)$

## Chapter 16

level  $C$  confidence interval: interval centered at some  $\bar{x}$  that extends far enough in both directions to have probability  $C$  of containing  $\mu$ .

If we know  $\sigma$ , then we find the  $z$ -score corresponding to  $C$ :



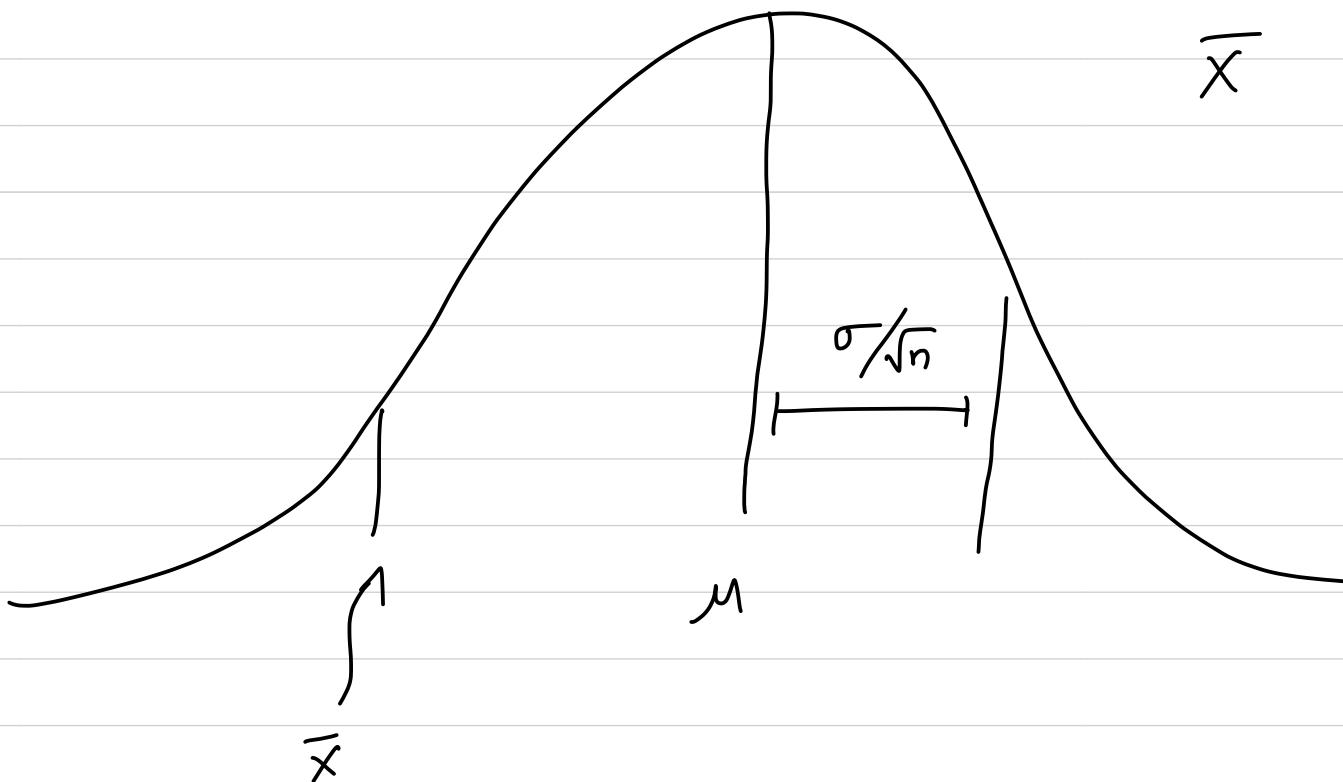
$$\text{proportion} = \frac{1-C}{2}$$

Ex:  $C = .95$ , want z-score corresponding to .025

$$CI: \bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

If we don't know  $\sigma$ , we have to use a t-score: take C and  $DOF = n-1$  and use the table to find t.

$$CI: \bar{x} \pm t \frac{s}{\sqrt{n}}$$



mean of  $n$  numbers

$\sigma/\sqrt{n}$  tells you how far away from  $\mu$  we can expect any given  $\bar{x}$  to be

given a sample,  $s$  tells you how far away from  $\bar{x}$  you can expect the individuals in the sample to be

## Chapter 17

Hypothesis test: state  $H_0$  (status quo) and  $H_a$  (the result you'd like to prove). Assuming  $H_0$ , find the chance of the given event occurring, and compare it to the  $\alpha$ -value. If the p-value is smaller, we reject  $H_0$ .

Know a population has mean 3 and std dev 1, want to evaluate the claim that a subset has mean  $> 3$  given that a sample of 20 people had  $\bar{x} = 4$ , using  $\alpha = .05$

$$H_0: \mu = 3$$

$$z = \frac{4 - 3}{\sqrt{20}} = 4.47$$

$$H_a: \mu > 3$$

$p = \text{very small} (\sim \frac{1}{100000})$

p-value : definitely less than .05, so we reject  $H_0$ : if  $H_0$  were true, then the chance of a 20-person sample having mean 4 would be less than .05

If  $p > .05$ , then there would be a substantial chance that given  $H_0$ , this  $\bar{x}$  would have occurred, so we say there isn't sufficient evidence to reject  $H_0$ .

## Chapter 21

$\bar{X}_1 - \bar{X}_2$  is approximately

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

when using t-score, use  $DOF = \min(n_1 - 1, n_2 - 1)$

Hypothesis test:  $H_0: \mu_1 = \mu_2$

## Chapter 22

$\hat{p}$  is approximately  $N(p, \sqrt{\frac{p(1-p)}{n}})$

- Always use standard error

( $\therefore \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ ), but use z-scores  
always

## Chapter 23

$\hat{P}_1 - \hat{P}_2$  is approximately  $N(\hat{P}_1 - \hat{P}_2, \sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}})$

Again, use z-scores and never t-scores.

Ex: to find a CI for  $p$ :

$$\hat{P} \pm z \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

$$\bar{x} \pm z \frac{s}{\sqrt{n}}$$

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

$$\bar{y}_1 - \bar{y}_2 \pm z \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$\bar{x}$ : problems involving means

$\hat{p}$ : problems involving proportions

Ex: Among the population of UO students

problem about the mean GPA of the  
second-years :  $\bar{x}$

problem about the proportion of  
second-years with  $GPA \geq 2.5$ :  $\hat{p}$

$$\hat{P} = 5/20 = .25$$

$$C=99.5$$

$$CI: \hat{P} \pm z \sqrt{\frac{\hat{P}(1-\hat{P})}{20}}$$

$$\therefore .25 \pm z \sqrt{\frac{.25 \cdot .75}{20}}$$

$$.25 \pm .0968 z$$

$$\rightarrow .25 \pm .0968 \cdot 2.81 = .25 \pm .272$$



$$z = -2.81$$

$$CI: \hat{p} \pm z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$



2%

.02

$z = 2.05$

$$CI = .4375 \pm 2.05 (.124017)$$

$$= .4375 \pm .254$$

P

$$s_1 = 139$$

$$\bar{x}_1 = 2.91$$

$$s_2 = .239$$

$$\bar{x}_2 = 3.285$$

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

$$.375 \pm t (.063)$$

$$3.12, \quad 3.305, \quad 3.45$$
$$Q_1, \quad \text{med} \quad Q_3$$

$$\text{IQR} = .33$$

$$Q_3 + 1.5 \text{ IQR} = 3.45 + .495 = 3.945$$

$$C = 99.5$$

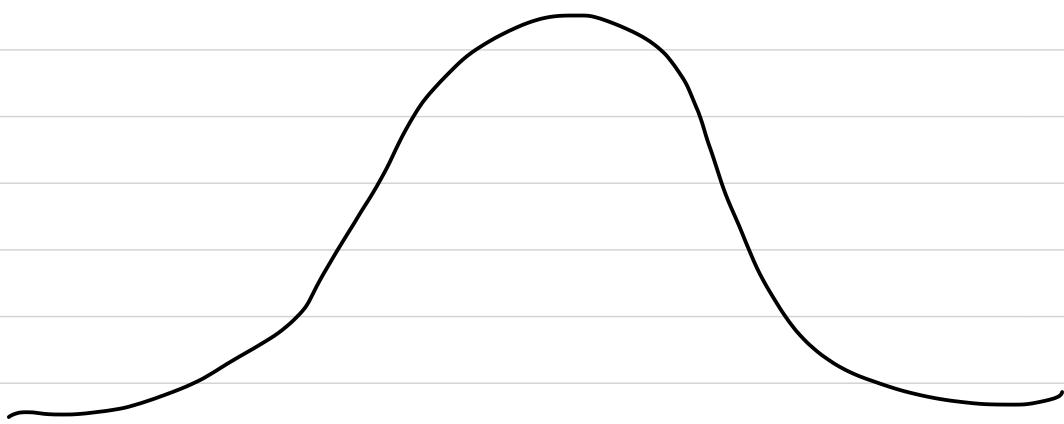
$$DDF = 18$$

$$t = 3.197$$

$$-.375 \pm 3.197 \cdot .063$$

$$\hat{P} \pm 2 \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

$$.184 \pm 2 \sqrt{\frac{.184(1-.184)}{125}}$$



$$.184 \pm 1.96 \sqrt{\frac{.184(1-.184)}{125}}$$

$$.184 \pm .0679$$

$$.184 - .0679 \rightarrow .184 + .0679$$