Study Guide for Exam 2 Key

Part V: Sampling and Inference Chapters 10-11
Sample Surveys—
Random Samples are best for the same 2 reasons that randomized experiments are best:
1. **They eliminate selection bias**
2. **They can be translated into box models so you can attach SE's to your estimates.**

Box Model for Sample Surveys:

- The box has 1 ticket for every person in the population.
- A random sample of n tickets is drawn from the box without replacement (because you don't want to sample the same person twice).
- You know the average or percent of your sample and you use it to estimate the average or percent in the whole population.
- Of course, the average or percent in your sample won't be *exactly* the same as that of the population, because of chance error (samples will vary because of the luck of the draw). As long as the sample size is big enough, the probability histogram for the sample average and percent will follow the normal curve so we can attach SE's to our estimates and build confidence intervals.
- For small samples from approximately normal populations with unknown SD, the probability histogram of the sample average (not percent) will follow the t distribution, so we can improve our estimates by using the t curves to attach SE's to our estimates to build confidence intervals. (See chart on page 68 for when to use t)

Note: The size of the population doesn't affect the accuracy of our estimates, only the size of the sample matters. The bigger our sample size, the smaller the SE for averages and percents (smaller by a factor or the square root n). This is apparent in the SE formulas for sample averages and percents because we divide the SD by $\sqrt{n}$, where n is the sample size (not the population size)

Sample Questions:

1) City A has **1 million** people and City B has **9 million** people. A simple random sample of **1000** people is taken from City A and a simple random sample of **9000** is taken from City B. This is for a pre-election poll looking at the percent of people who will vote for a certain candidate. All other things being equal the sample from City A is __________ the sample from city B.

a) 9 times *more* accurate **b)** 3 times *more* accurate **c)** the same accuracy as **d)** 9 times *less* accurate **e)** 3 times *less* accurate

more people $\Rightarrow$ more accurate $\sqrt{9}=3$

2) A recent Pew Research Center Poll asked a random sample of 1,211 adults nationwide the following question: "Do you think a woman should be able to get an abortion if she decides she wants one no matter what the reason."
We posted the same question on last semester's Bonus Survey. Here's the results of both surveys:

	Yes	No	Sample Size
Pew Research Center	18%	82%	1211
Bonus Survey	46%	54%	631

a)As you can see, the results of the 2 polls are quite different. Which survey gives a better estimate of the percentage of all US adults who would answer "yes" to this question? ***Choose one:***
i) The Pew Research survey because the sample size was larger.
ii) The Bonus Survey because we can be sure it was an anonymous survey.
iii) The Pew Research survey because the people were ***randomly*** drawn from all adults nation-wide.

b) What is SE of the sample percent for the Pew Poll? ***Choose one:***
i) It's not possible to calculate a SE for this sample because we don't know the SD of the sample.
ii) It's not possible to calculate a SE for this sample because we don't know the size of the population.
iii) The SE of the sample percent is approximately 13.4%
iv) The SE of the sample percent is approximately 1.1%

$$SE_{\%} = \frac{\sqrt{.18 \times .82}}{\sqrt{1211}} \times 100\% = 1.1\%$$

n population

3) A recent Gallup poll asked a simple random sample of 900 adults nationwide how much they spent on Black Friday. The sample average was \$400 with a SD of \$300.

a) What most closely resembles the relevant box model?

- i) It has 900 tickets marked with "0"s and "1"s.
- ii) It has about millions of tickets marked with "0"s and "1"s..
- iii) It millions of tickets. On each ticket is written a \$ amount. The exact average and SD are unknown but are estimated from the sample.
- iv) It has 900 tickets. The average of the tickets is \$400 and the SD is \$300.

b) 900 draws are made ___________replacement.

Choose one: i) With ii) Without

c) What is the SE of the sample average?

$$SE_{avg} = \frac{\text{SD of box}}{\sqrt{n}} = \frac{300}{\sqrt{900}} = 10$$

i) \$100 ii) \$10 iii) \$3 iv) \$0.33 v) \$30,000.

d) A 92% CI for the true population average = \$400 ± 1.75 * SE.
Fill in the 2 blanks with the correct numbers. (Hint: Use the normal table for the second blank.)

$z = 1.75$ for 92%

e) To which of the following populations would the above 92% confidence interval apply?

- a) All US females
- b) All US adults
- c) All Illinois adults
- d) All middle class US adults
- e) All of the above

* sample is only representative of the pop. it was drawn from

f) How would a 99% CI compare to the 92% CI we calculated in part d?
a) It would be wider b) It would be narrower c) It would be the same.

more confident ⇒ larger z ⇒ wider

g) Suppose we wanted to use SE^+, instead of SE to calculate our CI's, you'd multiply your answer in part c above by

i) $\sqrt{\frac{900}{899}}$ ii) $\sqrt{\frac{899}{900}}$ iii) $\sqrt{\frac{300}{299}}$ iv) $\sqrt{\frac{400}{399}}$

$SD^+ = \sqrt{\frac{n}{n-1}} \times$ SD of sample

v) None of the above because you cannot use SE+ to calculate a CI if you're using the normal curve.

h) Suppose we had a small sample size (n< 25) with the same sample average and SD as above.
Should we use the t curves to compute Confidence Intervals?
i) Yes, because we know the SD of the sample.
ii) Yes, because we don't know the SD of the population.
iii) No, because it doesn't say that it comes from a normal population.

Check for 3 conditions

small sample ✓

normal pop. X

unknown pop. SD ✓

n pop.

4) A CBS News Poll asked a random sample of 1,600 adults nationwide the following question: **"Do you think the distribution of money and wealth in this country is fair or you do you think wealth should be more evenly distributed among more people?"** 26% answered "Fair" 74% unfair

a) What most closely resembles the relevant box model? Draw box for pop.

~~a)~~ It has 1600 tickets, 26% are marked "1" and 74% are marked "0"
~~b)~~ It has 1600 tickets with an average of 0.
c) It has millions of tickets marked "0" and "1", but the exact percentage of each is unknown and estimated from the sample.

b) The draws are made ___________ replacement. a) With b) Without

$SE_{\%} = \frac{|1-0|\sqrt{.26 \times .74}}{\sqrt{1600}} \times 100$

$= 1.1\%$

c) Which one of the statements below is true?

a) The expected value for the percent of registered Democrats who would answer "Fair" to the question is 26%.
b) The expected value for the percent of corporation executives who would answer "Fair" to the question is 26%.
c) The expected value for the percent of Chicago residents who would answer "Fair" to the question is 26%.
~~d)~~ All of the above are true.
e) None of the above are true. only adults nationwide

d) Is it possible to compute a 95% confidence interval for the percent of all US adults who would answer "Fair" to the question?

z=2

$95\%\ CI = \text{sample}\ \% \pm 2SE_{\%}$

$= 26 \pm 2(1.1)$

i) Yes, a 95% confidence interval is approximately 26% +/- 1.1%
ii) Yes, a 95% confidence interval is approximately 26% +/- 2.2%
~~iii)~~ No, because we're not given the SD of the sample.
~~iv)~~ No, because we cannot infer with 95% confidence the answers of 200 million Americans from data based on a sample of only 1,600 randomly selected Americans.

e) If 1000 people all took random samples of 1600 and computed 95% CI's, about how many of their intervals would capture the true population percent? $1000 \times .95 = 950$

i) All of them ii) 9999 iii) 995 iii) 950 iv) 10 v) 50 vi) 100 vii) Impossible to predict.

$\sqrt{4} = 2$

f) If the researcher decreased his sample size by a factor of 4 (to n=400) then the width of the 95% confidence interval would ... $SE_{\%} = \frac{SD}{\sqrt{n}} \times 100$

i) increase by a factor of 2 ii) increase by a factor of 4 ~~iii)~~ decrease by a factor of 2 ~~iv)~~ decrease by a factor of 4 SE↑ more error

g) If our sample size was small (n < 25) would it be appropriate to use the t curves instead of the normal curve to compute CI's?

i) Yes ii) No, it's never appropriate to use the t curves with 0-1 data

n

5) To estimate the average IQ of students at a large public high school of 2000 students, a random sample of 17 students is taken. The sample average = 102 with a SD =16. Compute a 95% CI using the t distribution.

a) SE+ = 4

$95\%\ CI = \text{sample avg} \pm t^* SE^+_{avg}$

b) How many degrees of freedom? 16

$n-1 = 17-1 = 16$

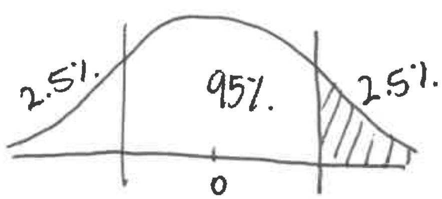
$SE^+ = \frac{SD^+}{\sqrt{n}} = \frac{16.49}{\sqrt{17}} = 4$

c) What is the t* (the critical value of t)? 2.12 (use the t-table in your notes)

d) 95% CI = (93.52 to 110.48) Put the lower number first.

$SD^+ = \sqrt{\frac{n}{n-1}} \times SD$

$= \sqrt{\frac{17}{16}} \times 16 = 16.49$



$95\%\ CI = 102 \pm 2.12(4)$

M of error = $z \times SE$

for percents: $n = \left(\frac{100 \times z \times SD}{M of E}\right)^2$

Choosing how many people to poll

6) In a pre-election poll in a close race, how many people would you have to poll to get… (Assume SD= 0.5)

a) a 95% CI with a 3% margin of error? $n = \left(\frac{0.5 \times 100 \times 2}{3}\right)^2 = 1111.11$ or 1112

b) an 80% CI with a 3% margin of error? $n = \left(\frac{0.5 \times 100 \times 1.3}{3}\right)^2 = 469.44$ or 470

c) Let's say the SD = 0.4, would we need more or less people than we did assuming the SD=0.5 ?
i) More ii) Less iii) Same

Part VI Significance Tests – are statistical checks to decide whether some difference we observe is "real" (due to some particular cause) is just due to chance variation.

Chapters 12-The one sample Z Test

$$\text{Z test-statistic} = \frac{\text{Observed} - \text{Expected}}{\text{SE}}$$

Look at the sampling distribution of Z under the null and see how likely it would be to get our data or something even more extreme if the null were true. That's called the p-value.
The convention is to reject the null when $p < 5\%$ and call the result "statistically significant" and when $p < 1\%$ call the result "highly significant". There's no particular justification for those values. In other words, a p-value of 4.9% isn't really much different than a p-value of 5.1%, people just like to draw the line somewhere.

7) Ellen thinks she has no musical ability but Karle thinks she does. To find out Ellen took a musical memory test online that had 36 questions. For each question she had to choose whether a sequence of notes were the same or different. She answered 24 of the 36 questions correctly. The null hypothesis is that she was just guessing.

↳ Draw box for a single play.

a) Which of the following most accurately describes the null box?
i) It has 36 tickets, 24 marked "1" and 12 marked "0"
ii) It has 36 tickets marked either "1" or "0" but the exact percentage of each is unknown.
iii) It has 2 tickets, 1 marked "1" and 1 marked "0"

b) The draws are made _________ replacement. i) with ii) without

Assuming the null hypothesis to be true, you would expect Ellen to answer _____questions correct, give or take______ questions.
c) Fill in the first blank in the above sentence with the correct expected value.

i) 12 ii) 18 iii) 21 iv) 24 v) 18

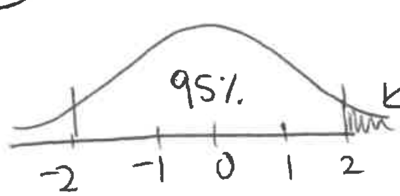
d) Fill in the second blank in the above sentence with the correct $SE_{sum} = \sqrt{n} \times SD$ of box $= \sqrt{36} \times 0.5 = 3$

i) 1 ii) 2 iii) 3 iv) 4 v) 5

e) The Z -statistic for testing the null hypothesis is $Z = \frac{obs - exp}{SE_{sum}} = \frac{24 - 18}{3} = 2$

i) 6/SE for the average ii) 6/ SE for the sum iii) 7/SE for sum iv) 6/SD of the box

f) The p-value for the one-sided alternative is …
i) 2.5% ii) 5% iii) 16% iv) 21% v) 11.5%



$\frac{100 - 95\%}{2} = 2.5\%$

8) An internet access company that serves millions of customers claims that it takes an average of only 1.8 attempts to connect with their service, but customers think it takes more. To test the company's claim, a consumer advocate looked at a random sample of 400 connections and recorded the number of attempts required to establish each connection. The average of the 400 observations is 2.1 and the SD is 5.0. We want to decide whether the observed difference between 2.1 (the sample average) and the claimed box average of 1.8 is due to chance or not.

a) The null hypothesis box is best described as:
   i) containing millions of tickets, each marked 1 or 0, where 1 denotes that a connection was made.
   ii) containing 400 tickets, each marked 1 or 0, where 1 denotes that a connection was made.
   iii) containing millions of tickets with whole number values such as 1, 3, 5, 2, ...
   iv) containing 400 tickets with whole number values such as 1, 3, 5, 2 ...

b) The average of the null hypothesis box is: a) 1.8 b) 2.1

c) The SE of the sample ~~average is~~ closest to:
   a) 0.05 b) 0.25 c) 0.50 d) 5.0 e) 20.0

$$SE_{avg} = \frac{SD}{\sqrt{n}} = \frac{5}{\sqrt{400}} = 0.25$$

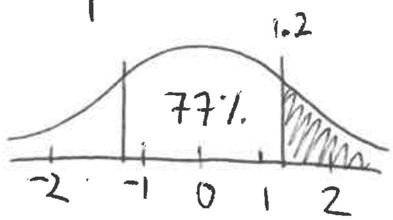
d) The Z-statistic is closest to:
   a) 0.15 b) 0.12 c) 0.6 d) 1.2 e) 6.0

e) The p-value is closest to: a) 77% b) 23% c) 11.5%

f) Conclusion:
   a) Reject the null, there is very strong evidence that the company's claim is false and the average number of attempts is greater than 1.8
   b) Cannot reject the null, it's reasonable to think that observed difference could is simply due to chance. $p > 5\%$

$$Z = \frac{\text{Obs} - \text{exp}}{SE_{avg}} = \frac{2.1 - 1.8}{0.25} = 1.2$$

p-value



$$\text{p-value} = \frac{100 - 77}{2} = 11.5\%$$

Chapter 13: The t test (see chart on page 76)
W use the SE+ and the t distribution when we have

1. **You have a small sample n<25**
2. **The population (contents of the box) roughly follows the normal curve**
3. **The SD of the population (null box) is unknown, all you know is the SD of the observed sample.**

When the sample size is small, using the sample SD to estimate the SD of the box is not very accurate. It's likely to be too low so we use $SD^+ = \frac{\sqrt{n}}{\sqrt{n-1}} \times SD$ **instead.** $SD^+ > SD$ **but the difference becomes negligible as n gets large.**

$$t-\text{statistic} = \frac{\text{Observed avg- Expected avg}}{SE^+_{avg}} \quad \text{where } SE^+_{avg} = SE^+_{avg} = \frac{SD^+}{\sqrt{n}} = \frac{SD}{\sqrt{n-1}}$$

When the null is true the sample distribution of the t statistic follows the t curve with **n-1 degrees of freedom.**

9)A factory that packages corn flakes is supposed to put the flakes in the boxes so that the boxes weigh an average of 16 ounces and a standard deviation of 1 ounce. An inspector randomly chose 12 boxes from one day's output of 2500 boxes. These 12 had an average weight of 15 ounces. The inspector wishes to test the null hypothesis that the factory is doing what it is supposed to on this day.

a. Which of the following best describes the null box?
- **i)** The box has 12 tickets, with an average of 180/12 = 15 ounces.
- **ii)** The box has 12 tickets, with an average of 16 ounces.
- **iii)** The box has 2500 tickets, but we do not know exactly the average.
- **iv)** The box has 2500 tickets, with 16% 1's and 84% 0's.
- **v)** The box has 2500 tickets, with an average of 16 ounces.

b. The SE for the average of the draws is closest to
i) 0.367 ii) 0.288 iii) 3.46 iv) 4 v) .02

$SEavg = \frac{SD}{\sqrt{n}} = \frac{1}{\sqrt{12}} = 0.288$

c. What test statistic would you use?
i) z-statistic ii) t-statistic

small sample ✓
normal data
unknown pop SD

* 3 conditions are not met

d. The test statistic is -3.47. What conclusion do you draw?
- i) Accept the null hypothesis.
- ii) There is not enough evidence to suspect there is anything wrong.
- iii) Reject the null hypothesis, there is strong evidence that the factory is not doing what it is supposed to. $p < 5\%$
- iv) The p-value is larger than 5%.

10) Now suppose the factory makes the same claim as above, that the boxes weigh 16 ounces on the average, but the factory doesn't make any claim about the SD and the factory says the histogram for the population is close to normal. The inspector computes the SD of the 12 boxes and finds the SD =1 ounce.

a) What is the best estimate of the SD of the 2500 boxes?
i)1 ounce ii) 1.044 ounces iii) 1.4 ounces

$SD^+ = \sqrt{\frac{12}{11}} \times 1 = 1.044$

b) What test statistic should the inspector now use?
i) z-statistic ii) t-statistic

* 3 conditions are met

c) If he decides to use the t-statistic, how many degrees of freedom are there?
i) 2499 ii) 12 iii) 11 iv) 6

$n-1 = 12-1 = 11$

c) What is the value of the t-statistic?
i) 3.3 ii) -3.47 iii) -3. 9

$t = \frac{obs - exp}{SE^+} = \frac{15-16}{1.044/\sqrt{12}} = -3.3$

d) Which test yields a larger p-value for the same data, the t-test or the z-test?
i) t test ii) Z test iii) they always yield exactly the same p-value

always b/c t-dist. has fatter tails + z is further from 0 than t

Chapter 14- The 2 sample Z test Used to compare averages and percents of 2 populations
H_0 is that the 2 populations have the SAME average or percent
2-sided H_a is that they're not the same, 1-sided H_a specifies which is larger.

$$\text{Z stat} = \frac{\text{Observed difference} - \text{Expected difference}}{SE_{\text{difference}}}$$

Where SE difference is the square root of the sum of the squares of each sample's SE

11) A study on the amount of time teenagers spend watching TV took a nation-wide random sample of 25 girls and 20 boys and found the following:

	Girls	Boys
Average hours per day spent watching TV	2.6 hours	2.1 hours
SD	1 hour	1 hour

The null hypothesis is that the average time girls' and boys' spend watching TV is the same in the population.
The alternative hypothesis is that girls watch more TV on the average than boys in the population.

a. Which of the following most accurately describes the null box(es)?

i) There is one null box with 164 tickets, 100 marked "1" and 64 marked "0"
ii) There is one null box with millions of tickets each marked with the amount of hours spent watching TV.
iii) There are 2 null boxes, each with millions of tickets. One box has an average of 2.6 and the other has an average of 2.1
iv) There are 2 null boxes, each with millions of tickets. The 2 boxes have the same average.
v) There are 2 null boxes, each with millions of tickets marked "0" and "1".

$$SE_{diff} = \sqrt{\left(\frac{1}{\sqrt{25}}\right)^2 + \left(\frac{1}{\sqrt{20}}\right)^2} = 0.3$$

b. The SE of the difference of the 2 sample averages is
i) 0.09 ii) 0.16 iii) 0.2 iv) 0.3

c. The Z statistic for testing the null hypothesis is closest to
a) 0 b) 1 c) 1.63 d) 1.67

$$Z = \frac{\text{obs diff} - \text{exp diff}}{SE_{diff}} = \frac{(2.6-2.1)-0}{0.3} = 1.67$$

d. The p-value is 4.75%. If the significance level is set at 5%, we would

i) Reject the null and conclude girls watch TV more than boys 95.25% of the time.
ii) Reject the null and conclude that if the average time girls' and boys' spend watching TV were the same in the population, the probability that we'd see a 0.5 hour difference or more in our sample is less than 5%.
~~iii)~~ There's good evidence that there is no real difference between the amount of time boys and girls spend watching TV.

e. Suppose we had chosen a 2-sided alternative hypothesis at the start of the problem. What would be our p-value?

2×4.75% = 9.5%

12) Suppose you wanted to use SE+ and the t-test instead of the z test in Question 11. (See Chap 15)

Girls $SD^+ = \sqrt{\frac{25}{24}} \times 1 = 1.02$

Boys $\sqrt{\frac{20}{19}} \times 1 = 1.026$

a)

$$SE^+_{diff} = \sqrt{\left(\frac{1.02}{\sqrt{25}}\right)^2 + \left(\frac{1.026}{\sqrt{20}}\right)^2} = 0.307$$

b) The t statistic for testing the null hypothesis is closest to
a) 0 b) 1 c) 1.63 d) 1.67

$$t = \frac{\text{obs diff} - \text{exp diff}}{SE_{diff}} = \frac{(2.6-2.1)-0}{0.307} = 1.63$$

c) To find the p-vale you'd look at the t curve with 19 df $n_1 - 1$ where n_1 is the smaller sample size

d) Is the p-value ______ < 4.75% (the p-value using the z test). i) Yes ii) No iii) Not enough info

p-value for t is always > p-value for z

13) Gallup asked a random sample of 400 men and 400 women nationwide the following question: "If you were taking a new job and had your choice of a boss, would you prefer to work for a man or a woman?"
H_0: % of all US women who would prefer a male boss = % of all US men who would prefer a male boss.
H_a: % of all US women who would prefer a male boss ≠ % of all US men who would prefer a male boss. two sided

In our sample we found 50% of the women and 45% of the men said they would prefer a male boss.

a) Which of the following most accurately describes the null box(es)?
- **i)** There is one null box with 800 tickets, marked with "0"s and "1"s
- **ii)** There is one null box with millions of tickets, marked with "0"s and "1"s
- **iii)** There are 2 null boxes, each with millions of tickets. One box has 45% "1"s and 55% "0"s and the other has 50% "1"s and 50% "0"s
- **iv)** There are 2 null boxes, each with millions of tickets. The 2 boxes have the same percentage of "1"s and "0"s.

b) The SE for the 2 sample percentages are both about 2.5%.
The SE for the difference of the 2 sample percentages is closest to

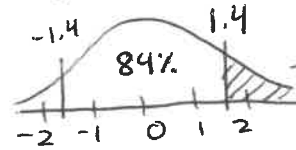
a) 2.5% b) 0% c) 5% d) 3.5%

$SE_{diff} = \sqrt{2.5^2 + 2.5^2} = 3.5\%$

c) The p-value for testing the null hypothesis is closest to

$z = \frac{(50-45)-0}{3.5} = 1.4$

a) 0% b) 2% ~~c) 8%~~ d) 16% e) 84%



$\frac{100-84}{2} = 8\%$

2 sided p-value = 2×8% = 16%

Chapter 16
Part I-The Chi-Square Goodness-of-Fit Test
Used to decide whether the observed data fits a specified model when the model has more than 2 catergories.
With 2 categories (0-1 box) we use the one sample z test.
Null Hypothesis: The observed data fits the model "good". (The difference between the observed and expected is just due to chance.)
Alternative Hypothesis: The observed data does NOT fit the model "good". (The difference between the observed and expected are too big to be due to chance.)

Chi-Square Statistic = sum of (observed frequency – expected frequency) ^{2}/ expected frequency
Degrees of freedom = # of categories -1

Part II- The Chi-Square Independence Test
Use to compare the percent composition of 2 or more variable when each variable has 2 or more categories. With 2 variables and 2 categories you can use either a 2 sample z-test or a chi-sq ind test.

(You can think of the Chi-Square Goodness-of-fit Test as a 1 sample test, comparing the sample percents to a null box that has multiple categories and you can think of the Chi-Square Independence Test as a 2 sample test, comparing the percent composition of 2 populations when each population has multiple categories.)

Null Hypothesis: The 2 variables are independent. (The 2 populations have the SAME percent composition; the difference between observed and expected frequencies are just due to chance.)
Alternative Hypothesis: The 2 variables are dependent. (The 2 populations have different percent compositions; the difference between observed and expected are too big to be due to chance.)

Chi-Square Statistic = sum of (observed frequency – expected frequency*) ^{2}/ expected frequency
Degrees of freedom = (# of rows -1) x (# of columns -1)

*To figure the expected frequency for each cell: multiply the row total x column total/overall total

14) A certain University has 30% freshman, 25% sophomores, 25% juniors and 20% seniors. A group of 200 students are chosen for a survey. The group has 30 freshman, 40 sophomores, 60 juniors and 70 seniors. The null hypothesis is the students were chosen at random.

	Expected Percentages	*Observed #*	*Expected #*
Freshman	10%	30	.10(200) = 20
Sophomores	20%	40	.20(200) = 40
Juniors	30%	60	60
Seniors	40%	70	80
Total	100%	200	200

$\chi^2 = \Sigma$ (obs-exp)²/exp

a) To test the null hypothesis that the students were chosen at random we'd do
- i) the chi-square test for "goodness -of-fit"
- ii) the chi-square test for independence
- iii) the one-sample z test
- iv) the two-sample z test

The table above is missing 3 values. Fill in the missing values by answering the following 3 questions:

b) What is the ***expected*** number of *freshman*?
i) 10 ii) 20 iii) 30 iv) 40 v) 50

c) What is the ***expected*** number of *sophomores*?
i) 10 ii) 20 iii) 30 iv) 40 v) 50

d) To compute the proper test statistic you'd sum 4 terms: 5 + 0 + 0 + ___. The term for **seniors** is missing. What should it be?
i) 0 ii) 1 iii) 1.25 iv) 1.43 v) 2.5

$$\frac{(obs-exp)^2}{exp} = \frac{(70-80)^2}{80} = 1.25$$

e) The number of degrees of freedom is
i) 2 ii) 3 iii) 4 iv) 5 v) 6

df = # categories − 1

f) What do you conclude?
i) Reject the null because p < 5% ii) Reject the null because p > 5% iii) Cannot reject the null because p > 5%

$\chi^2 = 6.25$ df = 3 ⇒ use table to find p

15) A simple random sample of 148 Stat 100 students were asked whether or not they thought they would ever use statistics again in their lives. Assume the students were chosen from a population of 2000. The following table gives the results:

	Would use	Would not use	
Men	47	21	68
Women	64	16	80
	111	37	148

The chi-square statistic to test the null hypothesis that sex and anticipated use are independent is 2.32. = χ^2

a. To compute this statistic, expected frequencies were calculated. What is the expected frequency for the men who answer, "would use"?
a) 51 b) 47 c) 44

$$\frac{\text{row total} \times \text{column total}}{\text{overall total}} = \frac{68 \times 111}{148} = 51$$

b. How many degrees of freedom does the chi-square statistic have?
a) 1 b) 2 c) 3 d) 4

df = (2-1) × (2-1)

c. Can you reject the null hypothesis?
a) Yes b) No

p > 5% from χ^2 table

d. If we had done a 2-sample z test with a 2-sided Ha, would we have gotten the same exact p-value?
i) Yes ii) No, we would have gotten half the p-value iii) No, we would have gotten twice the p-value

16) The table below shows the results of a recent nationwide poll of Hispanic adults who were asked;
"All in all, do you think the situation for the younger generation of Hispanic or Latino Americans is better, worse, or about the same as their parents' situation was when they were the same age?"
You may assume that the data are from a simple random sample of 200 people, of whom 100 were over 35 years old and 100 were 18-34 years old.

	18-34	Over 35
Better	49%	39%
Worse	37%	45%
About the same	14%	16%

To answer the question of whether the answers are really different for young and old adults, you use

i) the one-sample z test
ii) the two-sample z test
iii) the chi-square test for "goodness-of-fit" which specifies the contents of the box
iv) the chi-square test for independence

Chapter 17--Significance tests can only tell you whether or not a difference is likely to be due to chance, not whether a difference was important or what caused the difference, or whether the experiment was properly designed

By definition, significant results will appear by chance with enough tests. A p-value of 5% means that even when the null is true, you'll reject it 5% of the time.

17) Which of the following does a test of significance deal with?

a. Is the difference due to chance? only!
b. Is the difference important?
c. Was the experiment properly designed?
d. What are the probable causes of the difference?

18) 100 investigators each set out to test a different null hypothesis. Unknown to them, all the null hypotheses happen to be true.

a. About how many of them would you expect to get statistically significant results?

i. None, if they did the test correctly they would all confirm that the null hypothesis is true.
ii. 1
iii. 5 $0.05 \times 100 = 5$
iv. 95
v. Impossible to predict.

b. About how many of them would you expect to get highly statistically significant (p<1%) results?

i. None, if they did the test correctly they would all confirm that the null hypothesis is true.
ii. 1
iii. 5 0.01×100
iv. 95
v. Impossible to predict.

Part VII: Type I and Type II errors and Power (Chapters 18-20)

19) Significance tests are always subject to 2 types of unavoidable mistakes: Type I and Type II errors. Where do the errors below in the Table below?

		The Truth	
		H_A is True	H_0 is True
Our Decision	Reject H_0	A	B
	Fail to Reject H_0	C	D

a) Type I errors (α) belong in cell … Choose one: A B C D

b) Type II errors (β) belong in cell… Choose one: A B C D

c) Which cell does Power belong in? Choose one: A B C D

d) Is it possible to calculate the Type II error from the Type I error? Choose one:

i) Yes, since they're opposites of each other.
ii) Yes, if the sample size is known
iii) Yes, if the SD is known
iv) No

e) A Type I error is a false positive and Type II error is a false negative. Fill in each blank with one of the options below.

i) false negative ii) false positive iii) true negative iv) true positive

f) Is it possible to calculate Power from a Type II error?
i) Yes, they must sum to 100%
ii) Yes, they're always the same in absolute value but opposite signs.
iii) No

g) If you fill in the cells with probabilities which must sum to 100%? Circle all that do.
i) A + B ii) A+ C iii) A+ D iv) B + C v) B + D vi) C+ D

20) Since there's no standard way to write the above chart, it's good to be able to read it no matter how the rows and columns are labeled. So do the same problem as above with this chart.

		Our Decision	
		Reject H_0	Fail to Reject H_0
The Truth	H_0 is True	A	B
	H_A is True	C	D

a) Type I errors (α) belong in cell … Choose one: A B C D

b) Type II errors (β) belong in cell… Choose one: A B C D

c) Which cell does Power belong in? Choose one: A B C D

d)) If you fill in the cells with probabilities which must sum to 100%? Circle all that do.
i) A + B ii) A+ C iii) A+ D iv) B + C v) B + D vi) C+ D

21) A significance test is performed to analyze the results of a randomized experiment to determine whether students learn more or less from watching a lecture online compared to attending the same lecture in person. Subjects are randomly assigned to treatment (online lecture) and control (in person lecture) and then given the same exam afterwards.

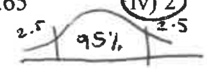
a) Fill in the blanks to complete the null and alternative hypotheses below:
H_0: The difference in mean exam scores between the treatment and control groups in the population ___ 0
Choose one: i) > ii) < iii) = iv) ≠

H_A: The difference in mean exam scores ~~between the~~ treatment and control groups in the population ___ 0
Choose one: i) > ii) < iii) = iv) ≠

b) A significance level of alpha = 0.02 means ~~when~~ the null is true the probability of making a Type I error= α
Circle one: **i)** 0% **ii)** 1% **iii)** 2% **iv)** 4% **v)** 96% **vi)** 98% **vii)** not enough info

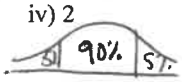
and when the null is false the probability of making a Type II error = β.
Circle one: **i)** 0% **ii)** 1% **iii)** 2% **iv)** 4% **v)** 96% **vi)** 98% **vii)** not enough info

c) If we set a =0.05 (null cut-off at 5%) for a 2-sided H_A then the critical value of our test-statistic, Z^*= ______
Choose closest answer. i) 0.85 ii) 1.3 iii) 1.65 iv) 2 v) 2.35 vi) 2.6



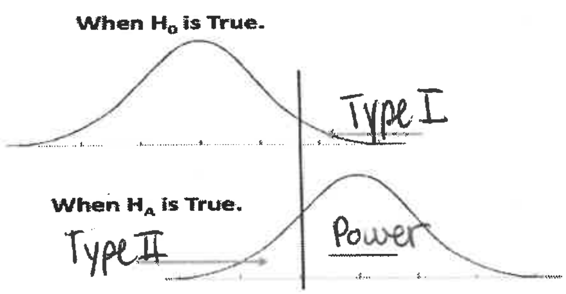
both tails together = 5%.

d) Repeat (c) above with a 1-sided H_A keeping all else the same. Choose closest answer.
i) 0.85 ii) 1.3 iii) 1.65 iv) 2 v) 2.35 vi) 2.6



each tail = 5%.

22) Look at the histograms below. Label the 3 areas (indicated by arrows) that represent Type I and Type II errors and Power by writing "Type I", "Type II", or "Power" above each arrow.



(Which of the following will increase the Power of the test? Circle either "yes" or "no".

a) Increasing the probability of a Type I error i) Yes ii) No

b) Increasing D, the effect size (H_A-H_0) i) Yes ii) No

c) Increasing the sample size i) Yes ii) No

d) Increasing the SD i) Yes ii) No

Options a and b in the last problem increased our Power but they're not worth doing (because we usually don't want to increase the probability of a Type 1 error and we can't increase our effect size beyond what we believe to be true).
So our only real option is increasing our sample size, which reduces our SE and so the 2 curves don't overlap as much.

To do power calculations we use this relationship: $\frac{|H_a - H_0|}{SE} = |Z_\alpha| + |Z_\beta|$ where $SE_{avg} = \frac{SD}{\sqrt{n}}$ or $SE_{\%} = \frac{SD}{\sqrt{n}} * 100$

There are 2 types of Power calculations. In both types you're always given a specified effect size $|H_a - H_0|$, significance level α, so you can calculate $|Z_\alpha|$, and the SD's of both distributions then you have to either:

1. Calculate Power given n : 1. Use n to calculate SE 2. Solve for $|Z_\beta|$ 3. Use Normal curve to find Power.

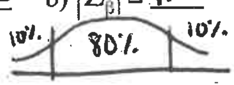
2. Calculate n given Power 1. Use Power to find $|Z_\beta|$ 2. Then solve for SE 3. Then solve for n in SE formula

23) Let's say I think male students may be inflating their heights on our surveys, They claim to have an average height = 71" and SD= 3". But I believe their average height is at least 1.5" *shorter* than what they claim with the same SD=3". I plan to take a random sample of men in the class and measure how tall they are. How big a sample would I need to detect an effect size of at least 1.5" with 90% Power at significance level α= 0.05. $\beta=10\%$.

$\frac{|H_a-H_0|}{SE} = |Z_\alpha|+|Z_\beta| \Rightarrow \frac{1.5}{SE} = 1.65+1.3$

SE=0.51

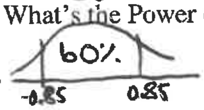
a) $|Z_\alpha|$ = 1.65 b) $|Z_\beta|$ = 1.3 c) SE = 0.51 d) n = 35



$\frac{|H_a-H_0|}{SE} = |Z_\alpha|+|Z_\beta| \to \frac{1.5}{3/\sqrt{n}} = 1.65+1.3$

n=34.81 or 35

24) Suppose I could only manage to take a sample of n=25. What's the Power of the test?



a) SE = 0.6 b) $|Z_\beta|$ = 0.85 c) Power = 80%

$SE_{avg} = 3/\sqrt{25} = 0.6$ $\frac{|H_a-H_0|}{SE} = |Z_\alpha|+|Z_\beta| \Rightarrow \frac{1.5}{0.6} = 1.65+|Z_\beta| \Rightarrow |Z_\beta| = 0.85$

25) Karle and Ellen are planning to do a randomized experiment to see whether students benefit more from in person exam review or online exam reviews. We'll ask for volunteers to participate in the study and then randomly assign half the volunteers to attend a final exam review lecture in person and the other half to watch the same final exam review lecture online. Then we'll compare the final exam averages of the 2 groups to see which group did better. How many students do we need to detect a difference of at least 2 points with 80% power? Assume SD =10 on both Final exams Assume = 0.05 as usual.

Both groups have the same $SE_{avg} = \frac{SD}{\sqrt{n}}$ so taking the square root of the sum of their squares gives $SE_{diff} = \frac{\sqrt{2} * SD}{\sqrt{n}}$

$|Z_\alpha|$ = 2 (Be careful here, we didn't specify a direction so 5% significance level means 2.5% in each tail.)

$|Z_\beta|$ = 0.85 (This is always one-sided because we're computing the smaller tail.)

SE = 0.7 $\frac{|H_a-H_0|}{SE} = |Z_\alpha|+|Z_\beta| \to \frac{2}{SE} = 2+0.85 \Rightarrow SE = 0.7$

n = 407 in each group.

$\frac{|H_a-H_0|}{SE} = |Z_\alpha|+|Z_\beta|$

$\frac{2}{\frac{\sqrt{2}\times 10}{\sqrt{n}}} = 2+0.85$

n=406.125 or 407

Question 26 is a problem from an old exam *(27 points total)*

Part I (15 pts.) Suppose the average number of times students reported texting during lecture is **5** with a SD = **3** and the data roughly follows the normal curve. I think students are underestimating how much they text and the true average is 6.6 texts or higher, so I decide to choose **6.6** texts for my alternative hypothesis and assume the SD is still **3** texts. I decide to do a significance test by randomly choosing **25** students to be carefully observed in lecture and record how many times they text.

a) (3 pts.) Assuming H_0 to be true, I'd expect the sample average to be 5 texts with a SE_{avg} = 0.6 texts. (Show work for SE.)

$$SE_{avg} = \frac{SD}{\sqrt{n}} = \frac{3}{\sqrt{25}} = 0.6$$

b) (3 pts.) Assuming H_A to be true, I'd expect the sample average to be ≥ 6.6 texts with a SE_{avg} = 0.6 texts

c) (2 pts) What is the effect size? D = 1.6 texts

$$\text{effect size} = |Ha - Ho| = 6.6 - 5 = 1.6$$

d) (2 pts.) What is the effect size in Standard Units? In other words, what is Zdistance?

D_z = 2.67. *Show work.*

$$Z\text{distance} = \frac{|Ha - Ho|}{SE} = \frac{1.6}{0.6} = 2.67$$

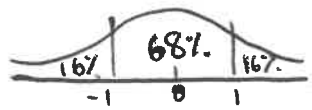
e) (5 pts.) If I set the significance level $\alpha = 5\%$, what is the Power of the test?

i) (2 pts.) First find $|Z_\alpha|$. $|Z_\alpha|$ = 1.65.

$$Z\text{distance} = |Z_\alpha| + |Z_\beta|$$
$$2.67 = 1.65 + |Z_\beta|$$

ii) (3 pts.) So $|Z_\beta|$ = 1.02 which means β= 16 %, so Power = 84 %.

$1-\beta$

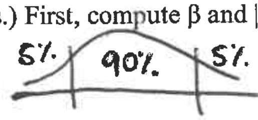


26) Part II (12 pts.)
Now suppose we keep the same $\alpha = 5\%$, same SD = 3 texts and the same effect size (D) as before, but we want Power = 95%.

a) (2 pts.) Will we need a larger sample size than in Part I? (i) Yes ii) No

Calculate what size sample you'll need to achieve Power = 95% by following the steps below. $\beta = 5\%$

b) (4 pts.) First, compute β and $|Z_\beta|$. β = 5 %, and $|Z_\beta|$ = 1.65



c) (2 pts.) What is the effect size in Standard Units (Zdistance)? D_z = 3.3. *Show work. Round to 2 decimal places*

$$Z\text{distance} = |Z_\alpha| + |Z_\beta|$$
$$= 1.65 + 1.65 = 3.3$$

d) (2pts.) SE_{avg} = 0.485 *Round to 3 decimal places. Show work.*

$$\frac{|Ha - Ho|}{SE} = |Z_\alpha| + |Z_\beta|$$
$$\frac{1.6}{SE} = 1.65 + 1.65$$

e) (2 pts.) How large an n will give us that small a SE_{avg}? *Show work.*

$$\frac{|Ha - Ho|}{SE} = |Z_\alpha| + |Z_\beta|$$
$$\frac{1.6}{3/\sqrt{n}} = 1.65 + 1.65$$

n = 38.29

Round n to 2 decimal places, not up to the nearest whole number as usual.

Chapter21 ANOVA to compare group means

Question 27

9 numbers are divided into 3 groups as shown below.

$n=9$
$g=3$

Group 1	Group 2	Group 3	
0	4	5	
2	6	7	
4	8	9	
Mean = 2	Mean = 6	Mean = 7	Overall Mean = 5

SST= 66

given

a) Compute SSB $= \sum_{i=1}^{n} (\bar{y} - \bar{\bar{y}})^2$

$$= (2-5)^2 + (2-5)^2 + (2-5)^2 + (6-5)^2 + (6-5)^2 + (6-5)^2 + (7-5)^2 + (7-5)^2 + (7-5)^2 = 42$$

b) Compute SSW (same as SSE)

$$SSW = \sum_{i=1}^{n} (y_i - \bar{y})^2 = (0-2)^2 + (2-2)^2 + (4-2)^2 + (4-6)^2 + (6-6)^2 + (8-6)^2 + (5-7)^2 + (7-7)^2 + (9-7)^2 = 24$$

given

c) The SST = 66. Use the SST to compute the SD. (Hint: The SST is the sum of the squared deviations. In other words, SD=sqrt(SST/n)

$$SD_y = \sqrt{\frac{SST}{n}} = \sqrt{\frac{66}{9}} = 2.71$$