

VOCAB & CONCEPTS !!

1	1	What are categorical variables?	They are a quality, or a category, like eye color and music preference.
1	2	What are quantitative variables?	These are numeric measures, like height and IQ, you can find a mean and sd.
1	3	what is a case?	an individual that has been measured or data has been gathered about
1	4	what is frequency?	the number of times the category comes up, like "32 blue candies"
1	5	What is relative frequency?	the percentage of time the category comes up like "14.5% of the candies are blue"
1	6	contingency table	shows distributions across 2 variables (male/female across hip hop/country/ classical)
1	7	marginal distribution	overall distributions of a single variable in contingency table (out in margins)
1	8	How can you tell if variables in a contingency table are independent?	If the distributions are the same across the variables..
1	9	How do you describe distributions (histograms)?	Shape-Center-Spread-Outlier-Gaps ---- GSOCS... where's yo GSOCS?
1	10	Shape description?	unimodal, bimodal, multimodal, uniform, symmetric, skewed,
1	11	Center description?	mean (balance), median (splits area in half), mode (peaks... if bimodal, talk about both modes)
1	12	Spread description?	range, IQR, stand dev, variance,
1	13	mean/med/sd/iqr... what to use?	when unimodal and symmetric, mean and sd... skewed or outliers? Median and IQR
1	14	Which is more sensitive to outliers and skewed?	Mean is more influenced by outliers... median is RESISTANT, RESILIENT, ROBUST!!
1	15	When can you round?	AT THE VERY END!!! (keep 3 digits until end!)
1	16	How can you match boxplots to histograms?	USE THE FISH TANK METHOD!
1	17	where are the "outlier fences?"	1.5 IQR above and below Q1 and Q3
1	18	If asked to compare distributions, what should you write about?	GSOCS.. Shape center spread gaps outliers.. SCSGO
1	19	What is a standard deviation?	average distance to the mean
1	20	what is a z score?	the number of standard deviations away from the mean
1	21	what is the mean/median/mode helper diagram?	a skewed left distribution with mean/median/mode labeled in order from L to R
1	22	what happens if you multiply all of a data set by a constant?	it is scaled.. Everything is effected. Mean/ median/ stand dev/ iqr/ all multiplied by that constant. Center, spread and all individual values are changed.
1	23	what happens if you ADD a constant to each value in a data set?	it is SHIFTED only. This effects all of the data values and measures of center. IT DOES NOT CHANGE THE SPREAD!
1	24	what is the empirical rule?	mean... 68-95-99.7 yeahh....
1	25	what are the percentiles from left to R	2.5-16-50-84-97.5
2	26	which is response?	y.. Vertical axis.. It "responds" to the x
2	27	which is explanatory variable?	x... horizontal axis... it "explains" what happens to y
2	28	does correlation mean causation?	HELLS NO
2	29	How is r calculated?	$r = \frac{\sum(ZxZy)}{(n-1)}$ --- the sum of rectangle areas on standardized axes
2	30	outliers in regression?	doesn't follow the "flow"
2	31	is r sensitive to outliers?	yes. A single outlier can make it seem like there is a relationship (out in x direction..)
2	32	does high r value mean anything?	NOT IF IT ISN'T LINEAR... and there aren't outliers.. LOOK AT THE DATA
2	33	how can you check for "straight enough?"	residuals plot fool!
2	34	Look for lurking variables...	think hot chocolate sales in caf at wachusett and ski accidents at wachusett
2	35	How to describe association? scatterplot	DIRECTION... FORM... STRENGTH
2	36	direction?	positive or negative
2	37	form?	straight, curved?
2	38	strength?	give the r value (if straight), or say... "tightly packed... loosely packed"
2	39	association or correlation?	
2	40	what is a residual?	ACTUAL-PREDICTED... A-P... like this class.. AP (get it?)
2	41	what is the LSRL	the "least squares regression line"
2	42	what is the line that you plot?	IT IS A MODEL!
2	43	what is a linear model?	it is an equation you can use... or a line of a graph, but it is just a model that says what kind of happens, and can be used to ESTIMATE WHAT MIGHT HAPPEN
2	44	how do you interpret y intercept?	if there were no [x stuff] this is how much [y stuff] you'd have
2	45	how do you interpret slope?	for an increas of 1 [unit of x] there is an (increase/decrease) of [SLOPE] [units of y]
2	46	how to interpret slope EQUATION?	for each increase of 1 st dev in x direction, you go r st dev in y direction.
2	47	interpret r squared	"r squared" % of variability in y can be explained by the model. The rest is in residuals..
2	48	if you switch x and y does r change?	NO. The strength stays the same.
2	49	if you mult or divide the x's or y's (shift/scale) does r change?	no. the strength remains the same. (If you log or square it, it will change, but just adding or multiplying won't change it)
2	50	if you switch x and y will slope change?	YES- slope is r_{sy}/s_x ... to get new slope you do: $(r \text{ squared})/\text{old slope}$

2	51	what does "regression to the mean" mean?	predictions for y are closer to the mean y (\bar{y}) than the actual x is to the mean x (in s.d)
2	52	does high r squared mean a good model?	no, you should check your plot and residuals to make sure model is appropriate and no outliers present.
2	53	what's up with extrapolation?	not a good idea..... sometimes it's all you can do, but still, NOT GOOD
2	54	how do you straighten scatters?	do things to L2 and replot and check condish... follow ladder of powers. Make sure to put the transformation into the equation!
2	55	last resort?	log log... log both! (DON'T ROUND AT ALL! GO OFF TO LIKE 10 DIGITS)
2	56	what about your calculator for using curves to fit curved data?	sure.. Quadreg, cubicreg, lnreg, etc... just be careful when substituting while writing the equation given.
2	57	what is b1 and bo ?	b1 is the SLOPE, and bo is the intercept. Remember that bo can be thought of as "b old" it is the old b... the intercept in $y=mx+b...$ so it is still the intercept.
2	58	what does influential mean?	It means that the point, when added or removed to data, will influence the SLOPE.. Generally these are outliers in the x direction.... Far left or right.
2	59	what is leverage?	leverage just means it is far away from \bar{x} ... far right or left. Some leverage points are not influential if they go along with the flow of the scatter.
2	60	will residual plots always show outliers? (will outliers always have large residuals?)	No... sometimes the outlier has so much leverage and is so influential that it pulls the LSRL right up to it, so it looks like a small residual.
2	61	what is the "ladder of powers?"	the order of transformations to try... it goes in order, and is based on exponents on y... 2, 1, 1/2, 0, -1/2, -1
2	62	what do the exponents all mean in ladder of powers?	2 means square the y, 1 means do nothing, 1/2 means take square root of y, 0 means log y, -1/2 means take $1/\sqrt{y}$, and -1 means just do $1/y$. Try them each and see who has most random resid plot.
2	63	what should we look for in resid plot?	curve or pattern.. Also, it should have equalish scatter from left to right
2	64	are any populations actually normal?	no, nothing is normal, just normalish. The only normal thing is the model we use.
2	65	are there any normal samples?	no, nothing is normal, just normalish. The only normal thing is the model we use.
3	66	What is the problem with convenient sampling?	The sample may not be representative as it is not randomized to include every type of person. E.G Friends and family are convenient but they likely share similar opinions and thus the sample is not representative of a population.
3	67	What do observational studies and experiments have in common?	In both, you are making OBSERVATIONS.. recording data... doing statistical analysis...
3	68	What is bias? What are some common errors?	It's any systematic failure of a sampling method. COMMON ERRORS: Voluntary response, undercoverage of the population, nonresponse bias and response bias. We use randomness and methods like stratifying to reduce these.
3	69	How many trials should you run to have an accurate simulation?	At least 20-30.
3	70	What is wrong with using volunteers in an experiment?	Not much. In an experiment, we are not looking for a sample that is like the population... We just want to see the effectiveness of a treatment. It is fine if the subjects are all similar. In fact it is best sometimes when they are!
3	71	What is wrong with using volunteers in a survey?	Those who volunteer may not be like the rest of the population. An example may be, if you're trying to find out how often people volunteer for things. So you ask for volunteers to take the survey.... A question may be "when was the last time you volunteered for something?" Well. they all just volunteered for the survey!
3	72	What is the difference between response bias and nonresponse bias?	Response is when the person's response is influenced by the question or questioning method, non response is when the people who don't respond might have different opinions/views than the people who did.
3	73	Samplin Method Types?	SRS, stratified, clustered, systematic, multistage, convenience, voluntary
3	74	What is the difference between a cluster sample and random sample?	A cluster sample is when the population is first divided into sections of clusters that have traits similar to the population (the clusters are heterogeneous and have all types within them). Then we randomly select an entire cluster or clusters, and include all of the members of the clusters in the sample. As for random sample is when each member of the population, and each possible group is equally likely to be included.
3	75	What's the difference between lurking and confounding?	Lurking variables, on one hand, infer the assoiation between the two variables; confounding variables, on the other hand, make it unclear which variable has had an impact on which in an experiment.
3	76	What are humans bad at ?	Humans are bad at generating random numbers.
3	77	What four things do you need in an experimental design?	NEED: control , randomization, replication. Use blocking when appropriate
3	78	What is the placebo effect?	When those who get the placebo show improvements, or show the effects of the treatment. This often happens to up 20% of participants!
3	79	What is the standard sampling method?	A Simple Random Sample (SRS) is our standard. Every possible group of n individuals has an equal chance of being our sample. That's what makes it simple.
3	80	what is the best way to reduce bias?	randomness. sophisticated answer: make as many things as random as possible

3	81	What is Placebo used for?	Placebo is used for control in an experiment. the purpose of placebo is to determine the change between the controlled treatment and the other treatments
3	82	Why randomize in an experiment?	To avoid bias. An experimenter might want their treatment to work, so may chose the subjects that might respond best.
3	83	What's a useful alternative when you can't run an experiment? What are they useful forms of this, and how do you preform them respectively?	An alternative of an experiments could be an observational study. There's two forms: prospective and retrospective. A prospective observational study is when you identify subjects in advance and record data as you go along. A retrospective observational study is when you analyze observations from the past.
3	84	What are the two types of observatinal studies?	Retrospective, and Prospective
3	85	What is a level in an experiment?	A level is a specific value(s) that the experimenter chose for a factor that is manipulated. ex. Factor is sleep, level(s) would be how many hours the subjects were aloud to sleep.. 4 hours, 6 hours, 8 hours.. 3 levels
3	86	What's the difference between cluster and stratified?	Stratified- you divide the population up into groups according to traits, called strata (groups with similar traits- homogeneous groups) and randomly choose from each strata. Cluster- grab clusters of the population.. each cluster should be like the population.
3	87	What is systematic sampling?	Systematic Sampling is one of four different ways to make a survery sample random. Systematic sampling includes picking every Nth number of what you are sampling (for example people.). You must still start on a random person and then from then on take every Nth person. So you can take every 10th person in a line in order to take a survey as long as you also start on a random individual.
3	88	How is Blocking in an Experiment Similar to Stratifying in a Sample?	The two are similar because they divide the subjects into homogenous groups where the subjects are all similar
3	89	Does a census make sense?	It's not practical to attempt a census because the population rarely remain the same - takes a long time - and is often EXPENSIVE or IMPOSSIBLE
3	90	You want to simulate the likelihood of more than 4 psychology majors being on a full bus that seats 30. 1 in 9 students are psych majors.	use single digits on a random number table. Each digit represents a student on the bus. Ignore the zeros. Let 1 be a psych major, and 2 through 9 be other students. Trials end when you have reached 30 students. Count the number of psych majors (ones) in the trial. Record this. Do this 20 times. Find the percent of times there were 4 or more psych majors on the "bus." If this occured in 5 trials.. then the likelihood is 5 in 20, or 25%
3	91	How is clustering and stratifying different when doing a sample?	Clustering is when chosen at random a group from the population that looks like the population, clusters should be heterogenous. While Stratifying is slicing a population into homogeneous groups(strata). Then randomly sample within each stratum before the results are combined.
3	92	What is the difference between a study and an experiment?	In a study you are basically just watching and in an experiment you are manipulating factors and (hopefully randomly) assigning treatments
3	93	What is response bias? How do you avoid it?	Response bias is any influence that may sway the respondent to give a more favorable answer e.g wording of the question, interviewer's behavior/background. Therefore, in a survey, ask questions that allow respondents to answer comfortably and honestly. Keep the wording "indifferent" or neutral in some way in order to unduly favor one response over another.
3	94	Why is this sentence incorrect: "The true percentage of all Statistic Students who enjoy the homework is called a 'Population statistic.'?"	It would be the called the 'population parameter'. 'statistics' describes samples. Ppppp population ppppp parameter ssssss sample ssssssss statistics.. The "TRUE" mean, proportion, average, is the parameter... you're talking about what's out there in the population.
3	95	What is the magic number for samples?	100 subjects. But, we will learn later that smaller samples give us pretty good estimates of population parameters, but not as precise.
3	96	What is a mutlistage sample?	A sample that combines several sampling methods
3	97	Explain two types of experimental design.	1.)Randomized Block Design: randomization occurs within the blocks only. 2.) Completely Randomized Design: all of the experimental units have the same chance at recieving a treatment.
3	98	What type of study would find relationship between Verbal and Math SAT?	You could take all of the SAT Math and Verbal scores and run a regression and find the r-quared value and linear model. This would be a Retrospective Study.
3	99	What's the difference between a prospective and a retrospective study?	A retrospective study takes a group and looks back at its history while a prospective study watches a group for a period of time and records the data. RETRO-REVERSE, PROspective- PResent and On..
3	100	What is a simple random sample?	A sample where every possible group has the same chance of becoming a part of a sample.
3	101	To make a survey to tell of a restaurant is good, would you ask the people coming out of the restaurant?	People at the restaurant are probably there because they already like it. If you asked the question "Is this your first time dining here?" and if they say "yes" you survey them, that would be a better method. But then again.. the people wouldn't go into an Italian restaurant if they didn't like that type of food.
3	102	what is a simulation component?	is the simple, individual parts of a simulation

3	103	explain one of the four principles of experimental design	one of the principles would be the control, which are the factors that the experimentors keep constant in each trial because they believe it would effect the outcome of the experiment. Also having a group that is not getting treatment helps to control because it measures the effects of the natural environment.
3	104	What is a control group?	A group in an experiment without the treatment that is compared to groups with treatments to make results or conclusions. The control group helps us see what would happen anyway... without any treatment so that we can see the true effect of the treatment.
3	105	What is the purpose of matching?	Matching, like blocking, reduces unwanted variation. In a retrospective or prospective study, subjects who are similar in ways not under study may be matched and then compared with each other on the variables of interest.
3	106	Why do you have to block?	You don't have to.. But you might want to if you feel that the experimental units (subjects) may respond differently to the treatment.
3	107	What is the difference between experiments and samples?	Samples try to represent the entire population and experiments try to assess the effectiveness of treatments.
3	108	Who can be blinded?	Subjects. Those delivering treatments. Those assessing effectiveness of treatments. and three mice.
3	109	What is difference between subject and experimental unit?	Humans who are experimented on are commonly called subjects in an experiment. Subjects like dogs, days, plants and anything not human are called Experimental Units
3	110	What are the 3 ways we used random numbers?	1. To simulate the likelihood of an event occurring. (ch 11) 2. To choose a sample that is representative of the population and avoid bias.(Ch 12) 3. To assign subjects (experimental units) to treatments to evenly distribute variability and help reduce possible confounding variables.(Ch 13)
3	111	what's the difference between response bias and nonresponse bias?	response bias is anything in a survey design that influences responses falls under the heading of response bias (wording of questions). Nonresponse bias is bias introduced to a sample when a large fraction of those sampled fails to respond.those who respond are likely to not represent the entire sample.
3	112	Why do you have to Stratify?	You don't have to.. But you might want to if you feel that a simple random sample might not be representative of the population . You want your sample to be like the population.. a representative sample (it represents the population well).
3	113	What is the difference between single-blind and double blind?	Single blinding is when all individuals in either one of the classes are blinded; double-blinded is when everyone in BOTH classes are blinded. Classes are: subjects, treatment givers, evaluators...
3	114	Why does it make sense to double-blind an experiment?	It reduces bias in an experiment. If subjects don't know what treatment they're receiving, they won't change their habits based on that knowledge. If evaluators don't know which treatment each subject is receiving, they won't bias the true results based on the results they expect to see
3	115	why shouldnt you spend your entire budget on the first run	you should always pretest a survey before you make an official one for your experiment. this way you can see the factors that would be right, the effects you need to control and about the unanticipated confoundings. Work out the kinks..
3	116	What is statistically significant?	When an observed difference is too large for us to believe that it is likely to have occurred naturally (or just randomly). Basically it is Statistically Significant when we don't think it happened randomly
3	117	What is the difference between confounding and lurking?	Confounding is to experiment, it may seem as both factors affect the response.Lurking is to sample, y and x makes it appear that x may be causing y
3	118	What is prospective study?	Prosepective study is when you study the experimental unit's present and futrue response variable.
3	119	What is retrospective study?	A retrospective study is a study that looks backwards in time. They focus on estimating differences between groups or variable association because they are not based on random samples.
3	120	when does a trial of a simulation end?	Generally there are two cases: 1. You want to know the probability of having x successes in n attempts (getting 3 smokers in a group of 5 students). Trials end when you get to n (get to 5 students). You record the number of smokers for each trial. 2. You want to know how many attempts it takes to get f successes. Trials end when you get f successes. Record the number of attempts.
3	121	What is sample size and how does it compare with the fraction of a population?	Sample size is the number of individuals in a sample. The sample size determines how well the sample represents the population, not a fraction of the population sampled. The fraction of the population that you've sampled doesnt matter. Its the sample size its self thats most important.
3	122	What is the main purpose of a placebo ?	To blind the subject that is being experimented on to avoid influence to the given variable therefore altering the response variable . When people think they're getting help, they often improve anyway..
3	123	what is a simulation?	Basically a test based on reality with a sequence of random outcomes that model it. Like an imitation.
3	124	What is a factor?	A variable in an experiment that the experimenter manipulates
3	125	How can you estimate the probability of an event occurring?	run a simulation. Find the percent of trials that you observed the event occur.

3	126	Can you stratify in an experiment?	NO. stratification is a sampling method, blocking is method used in experiments. They are similar ideas.
3	127	What is the sure way to assign treatments correctly?	Figure out about how many will be in each treatment group. Number the subjects 1-n... then use a random number table, or randint and then fill the treatment groups one at a time, when you get to the last group, throw in all the subjects whose numbers hadn't come up yet.
3	128	What is undercoverage?	Undercoverage is when either one part of the population is not included in a survey or is underrepresented in the survey
3	129	What is a quality of SRS that is not a quality of Systematic, Stratified or Clustering?	In an SRS, all groups are possible, and ALL POSSIBLE GROUPS have the same chance of being picked. The other methods have lots of "impossible groups" SRS has no impossible groups. -Stratified- an impossible group would be all girls (you're taking some boys and girls) -Clustered- an impossible group would be all girls (each cluster has boys and girls) -systematic- an impossible group would be 4 people that are right next to each other (you are taking every nth person)
3	130	Is it always better to do a census or a sample?	it is better to do a sample since a census is expensive to execute, and because populations are always changing it is hardly more accurate than a sample
4	131	what is probability?	THE LONG RUN RELATIVE FREQUENCY!!
4	132	what is the law of large numbers?	guarantees that in the long run.. The relative frequency settles down to true probability.. (you'll have 50% heads after an infinite number of coin flips with a fair coin)
4	133	What is the law of averages?	a misinterpretation of the law of large numbers. Using this law, if you flipped 4 heads in a row, you'd expect the next one to be a tails because it should even out in the long run. Not true, 5 flips is not the long run. Infinity is. The next flip still has a 50% chance of being another head.
4	134	Does sample size matter, or percent of population?	Sample size. A sample of 150 will say as much about a population of 2,000 as it will about a population of 2,000,000. The sample size determines level of confidence and interval widths..
4	135	what is representative?	It means that the sample statistics will be kind of like the population parameters.. The sample "looks like" the population.
4	136	what is a complement?	the probability that it doesn't happen. $1-P(\text{it happens})$. (they add to 1)
4	137	what is disjoint?	not joined.... They can't both happen at the same time! (being over 5 feet and under 4 feet)
4	138	what is independent?	when $P(A)=P(A B)$... When the probability of A is the same even when B is also true... Knowing B does not affect the probability of A.
4	139	How to find $P(\text{at least } 1)$?	$1-P(\text{none})$
4	140	What is probability first success is on 7th try?	$q^6 p$
4	141	what is a trial?	a single attempt of a random phenomenon
4	142	what is an event?	a collection of "overall" outcomes of trials.. Often called A or B or C
4	143	What is diff between trial and event?	Well, being a female is an event overall, like the probability of being a female in Stats is 0.6. In this case, while observing if one person is female is a trial, if she is, the outcome is F.(female)
4	144	can disjoint events be independent?	NO.. If they are disjoint then knowing one tells you that the other couldn't happen so they are always NOT INDEPENDENT
4	145	can independent events be disjoint?	NO, if they are independent, then knowing one doesn't change the probability of the other, but if they are disjoint, knowing one makes the other impossible, so it does change the probability to 0
4	146	what is a probability model?	a list of all possible values of random variable with respective probabilities. The probabilities should add to 1!
4	147	How do you find mean and sd of probability model?	put values in L1, probabilities in L2, and run "1-var stats L1,L2" and you get it!
4	148	how do you combine probability models?	add or subtract the means, and then ADD THE VARIANCES ALWAYS...
4	149	what is pythagorean theorem of stats?	st dev of combined model is: $\sqrt{\text{st dev squared} + \text{st dev squared}}$
4	150	Do we add or subtract st dev when combining models?	neither... you always just add variances. Square the st devs, add them, then take sqrt.
4	151	what does geometric model tell us about	it is about FIRST SUCCESS
4	152	what does binomial model tell us about	exactly x successes in K trials.
4	153	when can you expect the first success? (this is mean of geo model)	$1/p$ this tells you, on average, when the first success will occur.
4	154	geopdf	(p,x) ... probability of FIRST SUCCESS being ON the Xth trial
4	155	geocdf	(p,x) ... Probability of the FIRST SUCCESS being ON OR BEFORE the Xth trial.
4	156	binopdf	(n,p,x) Probability of exactly X successes in N trials.
4	157	binocdf	(n,p,x) Probability of X OR LESS successes in N trials.
4	158	what is that $\binom{n}{k}$ thing in the binomial equation?	"n choose k" it tells you how many ways you can choose k objects from a set of n things. The formula is $\frac{n!}{k!(n-k)!}$ the two numbers on bottom add to the number up top
4	159	what is n! ?	it is "n factorial" example: $5! = 5*4*3*2*1 = 120$. tells you how many ways you can arrange n objects.

5	160	what is a statistic	some numerical summary of a sample.. Could be the mean of a sample, the standard deviation of a sample, the proportion of successes in a sample, the slope calculated from a sample, a difference of 2 means from 2 samples, a difference of 2 proportions from 2 samples, a difference of 2 slopes from 2 samples.. you can make sampling distributions for any of these, and they will all be centered around the parameter...
5	161	what is a parameter?	some numerical summary of a population. Often called "the parameter of interest." It is what we are often trying to find.. It doesn't vary. It is out there and STUCK at some value, it is the truth, and you'll probably not ever know it! We try to catch them in our confidence intervals, but sometimes we don't (and we don't know it!). It Could be the mean of a population, the standard deviation of a population, the proportion of successes in a population, the slope calculated from a population, a difference of 2 means from 2 population, a difference of 2 proportions from population
5	162	When can we use a normal model for sampling distributions?	only when conditions are met!
5	163	What are the conditions that have to be met in order to use a normal model for the distribution of sample proportions? (sampling distribution of proportions).. (the distribution of p-hats)..	1. Randomization (this helps with assumption of independence) 2. 10% condition (this is the upper limit of our sample size. above this, the sampling distribution starts looking leptokurtic (thinner and taller), not normal) 3. success/failure: np and nq > 10. this is the lower limit of our sample size. It is when the sampling distribution starts looking normal.
5	164	what is difference between assumptions and conditions?	Assumptions must be made in order to perform inference. We need to assume independent sample values and a large enough sample (but not too large). We check conditions to help support our assumptions.
5	165	What is statistical inference?	Using a statistic to infer something about a parameter.. Basically, using a sample to say something about a population.
5	166	N (?1 , ?2) what does this mean?	it means NORMAL models centered at ?1 With a standard deviation of ?2
5	167	Are models what really happens?	No. A model train is not a real train. We use models to say what kind of happens.
5	168	What is the Fundamental Theorem of Statistics?	The CLT!! The Central Limit Theorem!
5	169	What does Central Limit Theorem Say?	It basically says.. NO MATTER WHAT SHAPE THE POPULATION IS (normal, bimodal, uniform, skewed, crazy..)... If you make a histogram of a bunch of means taken from a bunch of samples, that histogram will be unimodal and symmetric.. Close to normal. So.. A nerdy way to say it is: The sampling distribution of means is approximately normal no matter what the population is shaped like. The larger the sample size, the closer to normal. (the normal curve is just a model.. the sampling distribution is close to it, but not it! we use the model anyway!)
5	170	What does the CLT say about the actual sample data?	Nothing... The sample will be distributed similar to the population. The CLT only talks about distributions (histograms) of samples, which are groups of data, NOT OF INDIVIDUALS!!!!
5	171	Describe the distribution of a sample	It will look like the population. The distribution of a sample is a histogram made from the sample, which will look kind of like the population. If the population is bimodal, then the distribution of the sample is bimodal. The SAMPLING distribution of a bunch of means, however, will look normalish.
5	172	What is the difference between the distribution of a sample and a sampling distribution?	A distribution of a sample is just a histogram of a sample. A sampling distribution is made from an infinite amount of samples. It is the distribution of the statistic that was calculated from those infinite samples.
5	173	What are the mean and standard deviation of a sampling distribution for a mean?	mu and sigma/root n (look at formula sheet) N(mu, sigma/rootn)
5	174	What are the mean and standard deviation of a sampling distribution for a proportion?	p and root pq/n... (look at formula sheet) N(p, root (pq/n))
5	175	What is a standard error?	It is calculated like the standard deviation, but we are using sample statistics.. We don't know the true parameters, so we estimate with statistics adding error to our calculation
5	176	What is sampling variability?	The natural variation of sample statistics.. Samples vary... so do their statistics.. Parameters do not vary!
5	177	What is sampling error?	same as sampling variability. We call it error EVEN THOUGH YOU MADE NO MISTAKES!!!
5	178	What is a confidence interval?	it is a parameter catcher.. Like a fishing net.... We stand at our statistic, and reach up and down a margin of error, and hope to CATCH the parameter... sometimes we do, sometimes we don't... but we never know.. Moooo hooo hooo haaaa haaa haaa (evil laugh)
5	179	What are we confident in?	our confidence lies in our interval... if we took another sample.. We'd have a different interval..
5	180	what does 95% confidence interval mean?	It means if we took a ton of samples, and made confidence intervals from each of them, 95% of the intervals would contain the parameter, 5% would not.
5	181	What are confidence intervals for?	They are an attempt to say what the true population parameter is.. It is our best guess... "We think that there will be between 6 and 12 inches of snow..."
5	182	How wide is a confidence interval?	It is 2 margins of error wide

5	183	How is a confidence interval made?	statistic +/- margin of error Statistic +/- (crit * s.d)..... Stand at the statistic, reach out a margin of error, and hope that you catch the parameter.
5	184	What is a critical value?	It is the amount of standard deviations (errors) you'll reach out, depending on your confidence. Example.. 68% crit z = 1 For 95% crit z = 2 (well, 1.96).. For means.. Use t crits
5	185	What is a margin of error?	critical * s.d. It is how far you reach out in a confidence interval.. You reach up and down one of these, so the interval is actually 2 margins of error wide.
5	186	What if you want more confidence?	get a bigger net.. (wider confidence interval)
5	187	What if you want more confidence with same size interval?	increase your sample size
5	188	Do parameters vary?	NO!!! Statistics do... they vary from sample to sample... PARAMETERS DO NOT VARY!
5	189	Will 95% of other statistics be within my interval?	NO!!! You have no idea where your interval is in regards to true parameter
5	190	Can you make a 100% confidence interval?	Sure, I'm 100% confident that it will snow between 0 and 500 feet tomorrow...
5	191	What are the 3 steps in hypothesis testing AFTER YOU CHECK CONDITIONS?	1. Make your Ho and Ha 2. Make a Null Model (centered at null, use your Ho as center, use your sample size).. This is a sampling distribution for the statistics if the null were true. 3. CHECK... Calculate your statistic (p-hat, x-bar, phat1-phat2, xbar1-xbar2, b)
5	192	What is a null model?	It is a sampling distribution. It tells us how sample statistics would vary if the null were true. It is centered at the null.
5	193	What is a p-value	It is the probability of getting your sample randomly if the null were true. Basically, how likely is it that your sample statistic came from the Null Model.
5	194	One tail or 2 tailed? How do you tell?	if it just says "changed" or "different".. Then it is 2 sided.. DOUBLE THE P VALUE! If it says "more" "less than" "greater" etc.. Then it is just one sided..
5	195	Do you use p-hat or p-null when you calculate your standard deviation?	use p-null..
5	196	Do you use p-hat or p-null when you check the success/failure condition?	use p null
5	197	Can you prove a null hypothesis true?	NO
5	198	How do you write conclusion if you reject?	With such a low p-value, I reject the null hypothesis. There is strong evidence that the proportion of students who eat rice has changed.
5	199	How do you write conclusion if you fail to reject?	With a p-value this high, I fail to reject the null. (I retain the null). There is not enough evidence to say that more students like eggs now.
5	200	Can you accept a null hypothesis?	Never accept a Ho
5	201	Which hypothesis says what you're trying to prove?	Ha.. The alternative is what you are trying to prove. You can only prove an alternative, not a null.
5	202	What is alpha?	It is the rejection threshold. You reject p-values below it.. It is how willing you are to make a Type 1 error... $\alpha = P(\text{Type I error})$
5	203	What is beta?	It is probability that you'll make a Type II error.. $P(\text{Type II error})$
5	204	Can you draw the type 1 error type 2 error square?	see page 483... know how to draw and label this
5	205	What is Type 1 error?	We reject when we shouldn't have (the null was actually true)
5	206	What is Type 2 error?	We fail to reject when we should have rejected (the null was actually false)
5	207	If the null is true, what is the only error you could make?	Type 1
5	208	If the null is false, what is the only error you could make?	Type 2
5	209	If you reject, what is the only type of error you could make?	Type 1
5	210	If you fail to reject, what is the only type of error you could make?	Type 2
5	211	Can you draw the alpha/beta/power diagram?	See page 486. Be able to draw and label this.
5	212	What is power?	The probability that you correctly rejected a false null
5	213	How else can you explain power?	It tells you, If the null were actually false.. How likely is it that you'd reject?
5	214	How are power and alpha related?	they go up and down together
5	215	how are alpha and beta related?	as one increases, the other decreases, and vice versa
5	216	how are beta and power related	as one increases, the other decreases, and vice versa... They have to because they BOTH ADD TO ONE!!! $\text{Power} + \text{Beta} = 1$
5	217	what is a test statistic?	a t or z score (or chi squared) that you use to find a p value
5	218	How can you increase power?	Increase alpha... increase sample size..
5	219	How can you decrease alpha and beta at the same time?	increase sample size... this will also increase power...

5	220	Can you decrease alpha while increasing power (even though they move together?)..	Yes.. They move together with constant sample size. If you increase the sample size, alpha will decrease and Power will increase..
5	221	What is "statistically significant?"	When p-value is below the alpha, we say "statistically significant".. Low p-values are statistically significant. When our sample most likely didn't happen randomly, that is statistically significant.
5	222	What is effect size?	Difference between null and true parameter... something we don't know
5	223	When we are looking at differences of proportions, what is the sampling distribution a distribution of?	You have to imagine taking a a pair of samples, say.. Of girls and boys, subtracting phat girl-phat boy, and then writing that difference down. Do this over and over again, and you will have a list of differences. Now make a histogram of that list of differences, and that is your sampling distribution. It is an imagined distribution of an infinite amount of differences (of sample pairs)..
5	224	Where did the s.d. of differences of proportions that is on the formula sheet come from?	From the added variances of the the sampling distributions of the 2 proportions...
5	225	How do statistics from big samples compare to small?	Larger samples have less variability, so their statistics are probably closer to parameter. Statistics from smaller samples are more likely to be far away from true parameter.
5	226	What is a 2 sample t interval?	Well.. Suppose you were trying to find the difference between the IQ of math teachers and IQ of English teachers. You sample 50 of each and find math \bar{x} = 125 and English \bar{x} =115. So, the difference of the samples is 10 points. That is your statistic. 10 points. You now have to add on a margin of error, let's say.. 4. so, you'll say something like "I'm 90 % confident that math teachers score between 6 and 14 points higher on IQ tests."
5	227	Do you pool with means (t test)?	No you don't have to do it. Only pool with hyp tests for props.
5	228	If you were going to pool with means (t), when would you?	When you have reason to believe the variances of both populations are equal.
5	229	What is a t-crit?	It is the same as z crit. It is the number of sd you reach out in your CI. To find it, do INVT(area in one tail, degrees of freedom)
5	230	How do you find df in 2 samples?	you have to run an interval or a test on your TI and read the output (unless you want to use the equation....)
5	231	When do you use 1 prop z test instead of one prop t test?	There is no 1 prop t test. You use Z for props. T for means.
5	232	Why pool?	Pooling allows you to increase your sample size... sort of, increasing DF, allowing more precise predictions
5	233	How do you pool with means?	Don't
5	234	Do alpha and beta work with means?	Yep, alpha, beta, power, type 1, type 2 all go along with means and proportions
5	235	Why does the book use \bar{y} instead of \bar{x} ?	I honestly don't know. I'll figure it out and get back to you
5	236	\bar{x} and μ in t-test?	\bar{x} is your sample mean, μ is your hypothesized population mean
5	237	what is a sampling distribution	a theoretical distribution of statistics taken from an infinite amount of random samples
5	238	how do you find z and t crit?	for z crit.. INVNORM(area in 1 tail)... for t crit.... INVT(area in 1 tail, deg freedom)
5	239	when do you need crits?	in confidence intervals..... (and old fashioned hyp tests....)
5	240	how do you find deg freedom?	n-1 for one sample... 2 samples you must use calculator
5	241	when is data "paired"	when you have 2 measurements on the same subject (or matched subjects)
5	242	when do you pool with means?	never. Don't do it. But if you become a pro, and can prove equal variances, then you can get more precision in your results by increasing your deg freedom.
5	243	how are t models like Normal models?	bothe unimodal and symmetric. T models have more area in tails, that's why you have to reach out a little further than z for same confidence.
5	244	what happens to t models as n gets larger?	The models look more like the normal model. An infinite sample size would give a t model identical to the normal model.
5	245	who invented the t model?	Bill Gosset, guiness brewing company.
5	246	how are 2 samp t and paired t different?	2 samp t you are looking at a difference between 2 averages, with a paired test, you are looking at an average of a bunch of differences.
5	247	What is the normal enough condition?	for smaller sample sizes, it must be plausible that the sample came from a normalish population.
5	248	what are the conditions that have to be met for t procedures?	independent groups, random, nearly normal.
5	249	how do you check nearly normal.	normal prob plot on calculator, should be straightish and diagonal.... Boxplot should be symmetrical... histogram should be unimodalish and symmetricish...
5	250	how can you decide the right test? What are the 3 questions?	1 or 2 samples? Proportions (z) or Means (t)? Test or Interval?