Concepts from Stats I needed for Stats II:

Essentially all of the "hot words" from stats I are key to the understanding of stats II, for the class builds on assumed knowledge.

To assume normality of a sampling distribution (The central limit theorem):

Means: n≥30

Proportions: n(p-hat)≥15 AND n(1-p-hat)≥15

If these conditions are not met one can use the t distribution (for means only) if the sample comes from a normal distribution. But if the n<30 and the data isn't normal, then no statistical inference can be made (without use of nonparametric methods which you may learn about in stats II).

Probability Rules: Probability is only covered in the beginning of stats II and may be on the first test, probably not more than that.

Event A can be thought of as an outcome that is possible given some set up such as rolling two 1s when rolling two die or getting a full house when playing poker.

P(A)=#of outcomes involving A

number of outcomes possible (called the same space)

It is often useful to draw out the sample space of all possible outcomes. Much of the answers to probability questions can be answered just by looking at the parts of the sample space that is relevant to the problem at hand. Ex: the sample space of two dice rolls is all possible combinations of the two dice (1,1 1,2 1,3…2,1 2,2 …6,6). When creating the sample space you should make a judgment call as to whether or not order matters because it effects the outcome and is situational (most often don't start assuming that order doesn't matter).

P(A^c or A')=the compliment of A=1-P(A)=everything not in A

P(AnB)=intersection of A and B=probability of both A and B happening. If A happening does not affect B happening, the events are said to be independent and thus P(AnB)=P(A)*P(B). However, if A and B share no similar elements (such as both die being even=A and both die being 1s=B) then P(AnB)=0, as A and B are disjoint an thus can not happen at the same time.

P(AUB)=P(A)+P(B)- P(AnB)=the union of two events=P(A or B happening)

P(A|B)= P(AnB)/P(B)=P(A given B). If A and B are independent, then P(A|B)=P(A), as knowing what B is does not affect A's outcome. But if A and B are disjoint, then P(A|B)=0

Population:

Mean (μ): The population mean value such as mean number of dogs owned or slices of pizza eaten in a week. The value is often not known but the objective of statistics is to estimate it.

Variance /Standard deviation (σ^2/σ): The population variance (or the square root of variance, called the standard deviation). It is the expected 'variety' found within the data. Another way of seeing it is the expected squared distance a random observation is from the mean of the population. The standard deviation is just the square root of variance, but is often more useful than the variance as it is the same units that were started with (instead of squared units)

Variance is calculated by:

$$\sigma^2=\sum_{i=1}^{n}(X-\mu)^2$$

σ, the population standard deviation is just $\sqrt{\sigma^2}$

Variance and standard deviations are often quantities of interest. Much of statistics is used to estimate them.

Sample:

X-bar: The arithmetic mean of the data; it is calculated by summing all of the quantitative data points and dividing by the sample size.

P-hat: The categorical analog of X-bar. It is the observed proportion; it is calculated by taking the number of subjects that answered similarly, (such as voted yes for a law or voted for a particular candidate) and dividing that number by the total sample size.

S^2: The sample variance. It is a measure of the variation from the mean of a dataset. The higher the variance, means more variety in the data. It implies that you are more likely to find a data point far from the mean. It is calculated by:

$$S^2=\sum_{i=1}^{n}(X-\bar{X})^2$$

S, the sample standard deviation is simply the $\sqrt{S^2}$

Z Score (with multiple observations from a Normal Population): The Z score is exactly the number of standard deviations away the mean(x-bar or p-hat) is from the mean of the associated Normal distribution. It is calculated by:

$Z=(\bar{X}-\mu)/(s/\sqrt{n})$ for means

$Z=(\hat{P}-P)/\dfrac{\sqrt{P\times(1-P)}}{\sqrt{n}}$ for proportions

if you only want one observation from a normal distribution, simply plug in 1 for n and it simplifies to the regular Z calculation you are used to.

How to read a Z table:

The Z table is set up with the Z value second decimal places on the top row and the whole number and first decimal part of them on left most column. So for the value 1.54 you go look at the first column and find the section that has 1.5 and draw a line going rightward. Then look at the top row and look up the section that has 0.04 and you draw a line going downward. Where your two lines intersect is the associated probability value, or P-value. The way to interpret this value is to view it as the probability of observing a Z value or less than the Z value.

You will often want the value of Z or greater (the tail value) so you will want to take the probability you found in the table (if the probability in the table is the opposite of what you want) and do 1-(Probability you found in the table) to find the tail probability. A little trick that is used often is exploiting the fact that the Z distribution is symmetric about 0. So if your interested in the $P(Z>1.54)$, since the table is set up with $P(Z<1.54)$, you can do $P(Z>1.54)=1-P(Z<1.54)$ or you can observe that $P(Z<-1.54)=P(Z>1.54)$ as the are both the same values at the opposite end of the distribution. Interestingly enough the Z table has $P(Z<-1.54)$ so all you have to look up is $P(Z<-1.54)$ and you'll get the same answer as if you did

$1-P(Z<1.54)$.

<u>How to read a t table:</u>

The t value that you find in the table is entirely dependent upon how confident you wish to be with your observation and the degrees of freedom (df), which will be n-1 most of the stats II course (your teacher will tell you when it is different). To read the t table, first you find the df value that you have located in the far left column, and draw a line rightward. You then look up your a level, (or 1-a CI level, as they are the same value t value) and draw a line downward. Where the two lines meet, this is your t value of associated a level.

Confidence intervals: A more explicit form of testing for a difference, a confidence interval tells you how off your mean (whether it be X-bar or P-hat) is from a true value with a certain amount of confidence. The true power of a confidence interval comes from the fact that with repeated experiments of the same setup (same sample size and confidence level) X% of those intervals will contain the true value (X being the confidence level so 90,95,99 etc) A confidence interval has the general form of :

Proportion:
$P\text{-hat} \pm Za \times SE(p\text{-hat})$
Za=Z table value associated with the confidence level (see how to read a Z/t table)

$$SE(p\text{-hat})=\frac{\sqrt{(phat\times(1-phat))}}{\sqrt{n}}$$

Mean:

$X \pm t\alpha / Z\alpha \times SE(X)$

$t\alpha / Z\alpha$=Z table value associated with the confidence level (see how to read a Z/t table)

SE(X-bar)=S/√n where S is the sample standard deviation(See above)

Hypothesis Tests:

Hypothesis testing is a statistical inference technique for a population parameter that tests to see how likely you are to observe a value that you found in an experiment given that you think (or hypothesized) that the true value of the parameter is a given value.

Setup:

- Read the question and decided what you would like to test against and also decide whether or not you will be working with proportions or means
- Decide what the null and alternative hypotheses are:

  Null: $\mu=\mu o/p=po$. This is your intuitive guess at what the parameter value is. You will be testing against this value

  Alternative: μ/p > or < or ≠ μo/ po. This is what you're seeing if there is enough evidence to decide this. You will be using X-bar/P-hat to compare to your null value to see if there is enough evidence.

- Perform the test:

$Z=(\bar{X}-\mu)/(s/\sqrt{n})$ for means or $t=(X-\mu)/(s/\sqrt{n})$ if n<30 and sample is from Normal distribution.

$Z=(\text{P-hat}-P)/ \dfrac{\sqrt{P\times(1-P)}}{\sqrt{n}}$ for proportions

- Find probability associated with the Z or t value. Refer to how to read Z or t values (above) to find probability. This probability is called the p-value.

P-value: Before we make any conclusions, we much define a p-value, which is quite possibly the mort important concepts to know from stats I is of P-values. The strict definition of p-value is:

P-value: The probability of observing a value as extreme or more extreme than the one we observed, GIVEN Ho IS TRUE.

This is the only interpretation of a p-value, and every single hypothesis test that you will do in stats II (there will be many) has this interpretation. The p-value is the probability of getting what you actually observed with the assumption of Ho being true. If you have a low p-value ($p< a$), this means that there is a low chance of observing the value you found. But you observed your value! So naturally your original assumption of Ho being true should be disregarded. This is equivalent to rejecting Ho.

Conclusion: With established a level, if the p-value(found above) is less than a, then you reject Ho and conclude there is statistically significant evidence that the parameter value is (≥,≤, or ≠) than the Ho value. If $p> a$ then you fail to reject Ho and say there is not enough evidence to say that the parameter value is different from the hypothesized (Ho) value. A common phrase to help students remember whether or not to reject Ho is this: **P-value low, reject that Ho**. It is childish but it's always true. If it helps you, then use it. Regardless, you should always remember what a p-value is, as one every test there will be multiple questions about interpreting p-values (and confidence intervals as well, who's interpretation is above).