Review Sheet! NOT COMPREHENSIVE!

(name, what is it for, what does it do, when to use it, how to calculate it, IF applicable)

This value has 50% of the data below it.

This value has 75% of the data below it.

We can balance the histogram at this value if we support it here.

The SD measures the spread around this thing.

The SE measures the spread about this thing.

This measures the spread around the average.

This measures the spread from the median.

This line is a smoothed out version of the point of averages in the y for each given x value.

The regression MUST go through this point.

This line has the lowest r.m.s. error.

This measures how linearly related two variables are.

This says for large number of draws with replacement, the sums or averages will follow a normal distribution.

This says the chance for the average to deviate from the expected average decreases as the number of draws go up.

This is the equation (large number - small number)$\sqrt{\text{(fraction of large number)(fraction of small number)}}$

This is the statement we try to test in z-tests or t-tests.

This is the opposite of the above.

$\sqrt{\frac{\text{\# of ticket in box - \# of draws}}{\text{\# of tickets in box } -1}}$

$\binom{n}{k}p^k(1-p)^k$, (what are the possible values, when can you use it?)

$\sqrt{\frac{\text{\# of draws}}{\text{\# of draws -1}}}$

For 600 coin tosses, we use this to estimate the probability histogram of the total number of Heads.

Used to compare a sample average or sum to a hypothetical average of a box, when the # of draws is small, the average and SD of the box unknown but the content is close to normal.

For the above, what does "# of draws -1" represent?

Used to compared a sample average or sum to a hypothetical average of a large box when the number of draws is large.

This compares the average or sum between two independent samples when the draws are large from 2 large boxes.

r.m.s. of deviation.

sample averages are around _______, give or take _________

sample sums are around ________, give or take _________

the difference between two sample averages are around _______, give or take _________

the difference btween two sample sums are around _________, give or take __________.

We do this because the normal curve is only estimating the probability histogram which sometimes only has discrete values.

This is a special case of average when the data is 0-1.

When we count things (# of reds, # of wins, # of tails, # of spades), we make the box have tickets like these.

When we're classifying things (all people older than 20, all people who are left handed), we make the box have these tickets.

This is calculated from the sample trying to capture the population parameter.

When A happens, B cannot happen.

When A happens, it tells us nothing about B.

$r\frac{SD_y}{SD_x}$

This tests independence between two variables.

This tests how well the data fits a hypothetical model.

The probability of seeing the sample outcome or worst off compared to the null hypothesis assuming the null hypothesis is true.

The calculation of the SE is based on this type of drawing method.

Putting everything on a list then drawing them at random without replacement.

We say this when the p-value is less than 5 %.

We say this when the p-value is less than 1%.

This is when we say we fail to reject the null.

This is when we estimate the box SD by using the sample SD.

This line has slope $\frac{SD_Y}{SD_X}$ and passes through the point of averages.

Why do we take sample of boxes?

(Number of possible outcome for the column variable -1)*(Number of possible outcome for the row variable -1)

(Number of possible outcomes -1) when there are more than one possible outcome.

$\frac{\text{(observe average - expected average)}^2}{\text{SE for average}}$ when the draws is large and the box is even larger.

$\frac{\text{(observe average - expected average)}^2}{\text{SE for average}}$ when the draws is small and the SD of the box is estimated from the data and the box tickets are close to normally distributed.

In a histogram, this represents the proportion of data with each particular value.

In a probability histogram, this represents chance.

$\sqrt{1-r^2}SD_y$, this measures ___________ when the data is football shaped.

$P(A \text{ OR } B) = P(A) + P(B) - P(A \text{ AND } B)$

P(A AND B)=P(A)P(B|A)

When can the equations above be simplified.

Chance error is measured by ______

sample average ±2*SE of average

box average ±2*SE of average

Do this when you need to find out how many possible ways we can roll 2 dice and one is more than twice the number of the other.

We do this to check whether the sample sum of normal enough for examples like 100 draws from a box with one “1” and ninety-nine “0’s”.

A complete hypothesis tests includes these 6 ingredients.

We do this when we’re estimating the sample average when the number of draws is small but we know the box content and the tickets are very close to normal.