The exam will be on 20/3/16… @5:00 pm ….the material included is the first 6 chapters This lec is from the second unit which talked about “statistical inference “

-In the previous lec’s we talked about the constriction of the sample ,main characteristic of the sample , now in the topic “statistical inference “ we have to take Decision ,,,there’s a procedures we have to go through it which we called “hypothesis testing “…and these procedures are the one’s almost always researchers use to reach the decision to accept or reject the null hypothesis in the Randomized clinical trials (RCT’S).

-The evidence we expect in the testing or wherever… as any evidence in the world we read RCT’S or interventional studys ..so we have to know how the researcher takes the decision …

-This unit explains exactly how the researchers take decision and give you some principles important to be understand as probability or the( P) value , alfa (α) .

-----------------------------------------------------------------

Statistical inference defined as “the process involving containing information from samples from population which is Simplified drawn !

\*\*in the Statistical inference all the time we are talking about making hypothesis testing\*\*

We have to know something important about this figure (next page)..

Always when we represent a normally distributed variables or data we have to represent it this way as the figure shows(bell curve )…we also call it perfect world ..



--the zero point is very important point where the mean , median , mode centered there .

-- points where (-3,-2,-1,1,2,3 ) can be called the standered deviation (SD) ,, but to be more accurate we call them z score where we examine the data that located on that area of the curve

-- the areas from -3 to -2 to -1 should equal that from 1 to 2 to 3

--in the perfect world (where the population follows the normal distribution cure ) each z score represent one SD

FOR Eg: from 0 point (mean point) to point 1 …represent one SD …. Which means that one z score translated on one SD (keep in mind that’s only applied on the perfect world).

It’s important how to make an inference according to this figure ..

---------------------------------------------------------------------------------

#if you remember the example about hematoma formation (post reduction bleeding )and heparin use …

the null hypothesis was “there’s no association between anti- coagulant application and hematoma formation “

-the independent variable in this example was the anti-coagulant (the heparin ) while the dependent variable is hematoma formation ..

**NOW if I want to test this hypothesis** : \*note: we will clearly explain these steps in details later on this sheet

-first I have to represent the data …as the previous figure showed .

- put some points (cut points )to examine the date accurately and if these date fit with this hypothesis or not .



-this cut of point (point 1 in the 1’st figure or in the figure above : (µ -α) to (µ +α) ….not TO be confused it’s the mean +/- 1SD ..)…we examine around **68%** from the whole data in this area ! the rest of the date is out of this area ! which’s somehow accepted a little bit but it’s not enough at all because large group of data will not be examined !

-to have more accurate results we go to Distant point from the last one ..where **95%** of the data examined … which’s for sure better as we leave only 5% of the data out of examination …result here valid up to 95%

This cut of point we call it 0.05 ( alfa (α) point ) … (eza mo mfhomeh mmmm …y3ni 100% the full whole data – 95 % which we examined = 5% or 0.05 …n7na bnsmeha hon alfa point )

* So now the result as a probability will be either :
* < the alfa point (0.05 in this eg) ….means it’s out range of the data examined …it’s not the same …it’s different data (y3ni data mohema mo f7snaha )…so this is significant result ( e$i m7rz ☺ ) .
* > the alfa point means the same data which’s not different …so it’s not significant result ( mo m7ra ☹ )

#In this way we examined and do our inference #

-if we want to have more mooore accurate results we take 99% of the data or 99.7% of data so we leave .3% or 0.003 % behind ….here we are more strictly examining the data !

**#remember that :**

\*the cut pointa are the alfa(α) points

\*the result in means of probability are the P Value (we calculate it according to certain Equations)

\*then we compare α with p to take decision whether less or more :

-if **less** than alfa we should **reject** the **null** hypothesis and **accept** the research(the alternative hypothesis ) or the statistical hypothesis

(how to reject the null ..the null says there’s no association to reject it we say there’s association …so there’s a statistically significant relationship between heparin use and hematoma here as an eg. )

-if **more** than alfa we should **accept** the **null** hypothesis and **reject** the research (the alternative hypothesis ) or the statistical hypothesis

For the exam keep the following numbers in your mind :

-68% of the data within +/- 1 SD

-95% of the data within +/- 2SD

99%of the data within +/- 3 SD …we can’t examine 100%the data we examine the majority .

Another e.g :

If we use two dietary diets A &B one experimental and the other is controlled ..and we check the weight before and after and we found for e.g that the experimental group lost 40 kg of there weights from 80 down to 40 …this difference in mean weight visually …we will say oooooh there’s a diffrence …but we want to take a decision in terms of statisticall inference which’s the correct way we should follow because despite the visually result the real statistical result might me different (m$an haik la y3’orkom :p )

-nfs el $i b5sos mndobi el mbe3at 3na lma bns7ona b mntg jded ..oral b , colgate ,or anew device ,,,,etc

Back to z score :

It’s determined where the SD will be exactly

*A standard score of:*

* **Z = 1:** The observation lies one SD above the mean
* **Z = 2:** The observation is two SD above the mean
* **Z = -1:** The observation lies 1 SD below the mean
* **Z = -2:** The observation lies 2 SD below the mean

We have two important z scores ( b9m ..save as => your brain :D )

* +/- 1.96 equal 95% … 95% of the data will be between **-1.96** and **+1.96**.

We talked about Percentile last lec :

* A percentile is a score value above which and below which a certain percentage of value in a distribution fall.
* The most important Percentiles are : 25’th,50’th and 75’th
* Percentiles are symbolized by the letter P, with a subscript indicating the percentage below the score value. Hence, P60 refers to the 60th percentile and stands for the score below which 60% of values fall.
* The statement P40= 55 means that 40% of the values in the distribution fall below the score 55.
* There’s something called Interquartile …we have to read it from the book :/

There’s another concept called slandered error SE ,, as the M capital means the mean of the sample and the µ means the mean of the whole population ,,,the SD means that for the sample while SE that’s for the whole population .

In the probability axioms it’s important to know that the probability located from zero … but in the philosophy of statistic there’s no zero probability because if there’s a zero probability no need to examine the data ! (logic :p )…so it’s located in fractions of zero like 0.000000000001 but not the exact zero …we see it written in research’s/reports as( **less than 0 .0001**) ..and sometime it’s already not in fractions of zero and written as it’s ..the calculated number we got 1,2 ,,, etc

\*\*\*\*\*\*\*\*\*the end of part1 till (23.31)….best of luck ☺.… Maha khaled al-hussban \*\*\*\*\*\*\*\*\*\*