EVERYTHING HERE IS IMPORTANT

there will be 10Qs on the study designs, Incidence and prevalence

(I suggest you go over the practicals if you don't have time it will give you an idea about the most imp things)

simple calculations	
OR	**RR**
Odds ratio = ad/bc	**RR = Relative risk = A/ (A+B)/C/(C+D)**
Prevalence	**Incidence**
New and old cases/population	**New cases/population**
Attributable Risk= Incidence in exposed – Incidence in exposed	

What is the difference between sampling and non sampling methods? Sampling method (the probability) in which all the sample units have equal chance to be chosen

What you must have for probability sampling? Sample frame and random number to choose depending on the protocol and your objectives you'll choose one of the techniques (simple, stratified, systemic, cluster and multistage)

The scenario will be given and then you pick the technique

In non probability where are not using any of the random numbers or sample frame or anything

- **For the sample size you must know how to use the tables (which is there in the practical) and what happens to the sample size when variability increases**

What are the measurement scales?
Qualitative (Nominal and ordinal)
Quantitative (Ratio "continuous and discrete" and interval)
In the exam a variable will be given and you have to know which scale we use

- **Diagrams you have to classify based of the type of the data for categorical (Bar chart and pie chart) for continuous quantitative (Histogram stem and leaf Box plot, scatter diagram) where we use scatter diagram? when looking into 2 variables age and blood pressure, high and weight**

- **Summary measure we use: Mean and Median**
- **Variability measure: interquartile and standard deviation**

- **Remember**

when the data skewed we use the median and interquartile Q3-Q1

it's following the normal distribution (symmetrical) mean and standard deviation

negative and positive skew

negative the mean < median

positive the mean > median

Variance = square root of the SD

SD = root of the variance

Standard Error= SD/root of sample size

Normal distrbution main charecterstic:
+/- 1 SD on both sides covers 68
+/- 2 SD on both sides covers 95
+/- 3 SD on both sides covers 99
Zscore = the score - mean / SD
mean and SD (you should know the calculations from the above like the exercises in normal distribution lecture)

null hypothesis (no difference) and alternate hypothesis (there is difference)

suppose you do not know which group is high and which is low > two tailed

one tailed if you knew from the beginning

1 alpha rejecting H0 and 2 beta accepting H0

Alpha is called significance level of the test

1 - beta = power

P-VALUE probability =<0.05 5 out of 100 or 1 out of 20

when pvalue is small we reject the H0 and when it is high we accept H0

Estimation values:

point estimate

interval estimate : p +/- Z alpha x SE (standard error) not standard deviation

how to interpret? we say we are 95% confident that the value will fall in this interval (something-to-something)
or we can say if we repeated the study 100 time 95 times it will be in this interval (something-to-something)

the precision in imprecision of the CI is based on the width which is affected by the sample size, variablity and level of confidence

If we have the alpha level = 0.10 the confidence level will be (1-alpha) so 0.90

and so, alpha level = 0.05 the confidence level = 0.95 (the most used)

alpha level = 0.01 the confidence level = 0.99

what will happen to the interval if we had small sample size? it will be wider, or even if you had large sample size but huge variability it will be wide

p value VS CI

p Value is related to hypothesis testing, where is CI is related to estimation

p Value only tells you if the result is statistically significant or not

p values (hypothesis testing) gives you the probability that the result is merely caused by chance or not by chance, it does not give the magnitude and direction of the difference

Confidence interval (estimation) indicates estimate of value in the population given one result in the sample, it gives the magnitude and direction of the difference

Does the interval contain a value that implies no change or no effect or no association? (no statistical significant)

CI for a difference between two means: Does the interval include 0 (zero)?

CI for a ratio (e.g, OR, RR): Does the interval include 1?

statistical tests

3 for quantitative data (z test, t test and correlation)

4 for qualitative data (z test, chi-square test, Macnemar test and fisher's test)

- **whenever you have quantitive data (mean values)**
- **qualitative/categorical (proportions)**

z test compares sample mean (or proportion) with a population mean (or proportion) or 2 sample means (or 2 proportions) (the sample size must be above 30)

student's t test is the same but only for mean values and in addition you can compare paired samples "before and after" (you *can* have sample size less than 30)

correlation is where you're looking for linear relationship for two different variables (scatter diagram) like hours of studying and exam marks
we have positive and negative correlations

chi-test exclusively for Qualitative (Categorical) data two or more proportions (in z test only two) Sample size: >30 (most exceed 20) Expected frequency: > 5 when we have too sample size (less than 30) and only two proportions we use fisher's exact test
like what we have in the quantitative for the paired samples we used what? student's t test, here for Qualitative we use the macnemer's test (dependent proportions: *CROSS OVER TRIALS*) and the sample size can be any size
So in the exam you have to look if they are related (dependent) or not

Degree of freedom, when we are calculating the degree of freedom ?

Student's t test (one sample) sample size – 1

Student's t test (two samples) n1 + n2 – 2

Chi-square test (number of columns-1)x(number of rows-1)

Macnemar (number of columns-1)x(number of rows-1)

Fisher (number of columns-1)x(number of rows-1)

الله يفتح عليكم، جوهرة المالكي