The lecture was like bullet points!

We said the descriptive measures mainly are measures of variability and measures of central tendency, and that the MEAN is affected mostly by the extreme values, as you know that one of the main principles of the parametric statistics is that MEAN is continuous measures with MEAN's value, the MEDIAN is the 50^{th} percentile or the middle value, the mode is the most repeated value and in case of having two values I have bimodal or multimodal then, and the distribution would be different.

The central tendency has important indications and you have to know how the skewness is affected by the MEDIAN and MODE which usually are close, not exceeding one standard deviation (SD) (<=1.96), if more than two or three SDs then we have skewness which affects the statistical analysis, so if my data is severely skewed I do something called transformation for correction and to have more valid results, we are here comparing mainly with the MEAN to identify any skewness.

Example: in a study of the prevalence of neuropathy of diabetic patients with their relation to smoking, the result showed that nonsmokers are more prevalent to neuropathy which does not make sense, after proper statistical analysis it showed the opposite, here we might have errors in analyzing, coding or even collecting the data like taking a non-representative sample.

Here comes the transformation, we do it to end the problem of skewness.

We said that in ideal situation the values should be close, with no more than one SD, but in real life we usually have more than one so the distribution is skewed, either positively or negatively skewed, mild, moderate or severe, so it's important that whenever you report the MEAN you should report the SD (MEAN+- SD), because if we have more than 1.96 SD that means we have variability between the subjects.

In order to have POWER in you results and strong study, you should have minimal SD.
You have also to pay attention to the RANGE, range of values from 20 to 100 and only one mark is 100 with most of the values are around 50 is will highly affect the results of your analysis, you didn't pay attention to the variability of the samples.

See also where the values are condensed, let's say that an exam results are condensed around 80, that indicates that the exam is either extremely easy or is not discriminative between students.

The Pulse plot is the best way to represent the data and show the outliers.

It is of high importance to report the central tendency measures side by side with the variability measures.

The Range shouldn't be wide or narrow, it's best to be reasonable, for example you should collect your random subjects from a relevant population of relevant ages, and try to minimize each subject's compounding variables (like making a study on the effect of smoking on population and most of your subjects are ages 70 and above, with medical conditions, this kind of study won't give you relevant information), and when say random we mean on a probability basis with rules and principles.

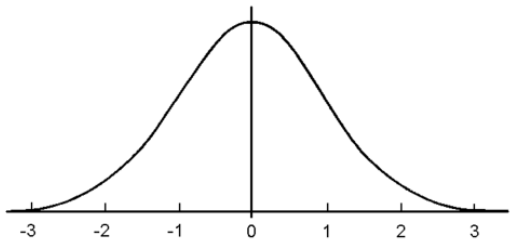
The main Percentiles are the 25^{th} (the 1^{st} quantile), 50^{th} (the 2^{nd} quantile and the MEDIAN), and 75^{th} (the 3^{rd} quantile).

NOW,
The Philosophy behind the statistics

It's important to talk about the normal curve, it's the first assumption in the parametric statistics, having your data within the normal curve, on which the philosophy of statistics is built, mostly it's logic.

In non-parametric statistics it's curve free (not sure), wither positively or negatively skewed not a problem.

In the normal curve we have the standard scores (z-scores) that divide it into equal variances, we usually consider these scores as standard deviation, the zero level gives for the MEAN.



Usually 68% of values are within one standard deviation.
95% within two SDs.
99.7% within three SDs.

If my data is within the first two SDs this indicates the confidence intervals, which means I'm confident about this distribution of results because only 5% is out of the first two SDs, usually this is used in hypotheses testing; the alpha symbol is one of the confidence intervals that we're talking about.

These testing usually are done on the 95% of the distribution; some RCT testing applies on the 99% and more.

We compare the P value with alpha to see the significance of the study, if it's less than 0.05 this is significant difference between the two studied groups, above this value indicates no significant difference.

The Dr said to refer to the book for clearer explanation of this.

When the level of confidence is 0.001 for example, it means the researcher studied 99.999% of the results, more strict then we have more confidence in the results.

The skewness must be transformed as you know.

we have specific equations an indicators to know the positively skewed and the negatively skewed, the first one is pearson's skewness coefficient formula which is the most known and used, $\frac{MEAN-MEDIAN}{Standard\ deviation}$.

The cut of point that I have to have the result within 0.2, if I have more or less indicate severe skewness

The sign negative or positive indicate the positive or negative skewness.

The second one is Fisher's Skewness Coefficient Formula = $\frac{\textbf{Skewness coefficient}}{\textbf{Standard error of skewness}}$ which is a little bit more accurate.

Skewness values within1 .96 SD is normal, more or less indicate severe skewness.

The third one is **ROUGHER =** $\frac{3\ (MEAN-MEDIAN)}{Standard\ Deviation}$**,** 3 is 3 standard deviations.

The normal distribution in ideal world is within the first 3 SDs, even this is not present in real life health and social studies, but I can't take more than 5% (0.05) error.

Alpha in a stricter point of view is 0.001, the confidence then is 99.999%!!
You define Alpha based on the literature review; usually we try to be stricter and included more values.
In very strict cases 100% of values within the first 3 SDs.

We have a non-form and statistical from of hypothesis, in the non-form we say no difference between subjects A and B for example in relation to drug A and drug B, we use the non-form to approve the logical difference in results.

Now the equations of confidence intervals,
how we reject and accept hypothesis.

The types of errors are 1, 2, 3 and 4.

The significance level, we mean here alpha to judge the hypothesis.

Usually the process is, state the hypothesis, choose the way to test the hypothesis, define the degree of risk (alpha) and finally the probability that gives you the P value.

Equations of confidence value, we need to know two,
the first for 95% is the MEAN+-1.96 (SD), the alpha here is 0.05
the second for 95% is the MEAN+- 3.92 (2 SDs).

An example:
66 women were studied if prophylactic anticoagulant they used resulted in hematoma, 6 subjects had hematoma, 31 had no hematoma, looking roughly at the results we can say that there is no relation (the hypothesis).
We sat alpha to 0.05, and when we calculated the probability it was 0.03 whish is less that alpha, so what do we do here? We reject the hypothesis because there is **significant effect** (result) of the anticoagulants on postoperative hematoma.