# **Lecture 3 Biostatistics Notes**

## **Measurement Scales**

### Ordinal Scales

* Allow ordering or measurement
* Do not allow assessment of relative degree difference between items

### Interval Scales

* No *true zero*
* Allow assessment of relative degree of difference between items
* Do not allow assessment of the ratio between items

### Ratio Scales

* Have a *true zero*
* Allow assessment of relative degree of difference between items
* Allow the assessment of the ratio between item

## **Remember what statistics is:**

"The science of learning generalizable knowlegde from data.". It is about inference.

## **Type of data matters**

### The distribution of the measured variable depends on:

* The probability distribution of the variable in the population.
* The sample size
* Sampling procedure (e.g., probability vs non-probability sampling)
* The level of measurement (e.g., count, continuous)

### The Normal Distribution

#### **How do I know when my data follow a Gaussian distribution?**

* Consider the source of scatter.
* Look at data (outliers?).
* Consider type of data (rank or score?).
* Use tests for normality (e.g. chi-square, Kolmogorov-Smironov) (*large saples only*).
* Look at all data available.

## **SD vs SEM:**

* SD refers to the variability (dispersion) between individual observations.
* SEM refers to the precision with which the sample mean estimates the population mean. The precision of estimation which increases as sample size increases.

### Central Limit Theorem

The distribution of sample means (SEM) has an approximate normal distribution, no matter what the distribution of the original data looks like, as long as the sample size is large enough (usually at least 30), and all samples have the same size.

### P-value

* Is a rough numerical guide as to the strenght of evidence against the null hypothesis. It is not the probability as to the trueness of given null hypothesis, nor the falseness of the alternative hypothesis. And furthermore, it does not report on the magnitude or importance of any given observation.

### Type alpha and beta statistical errors

* Type Alpha: claiming significance when there isn't. Alpha=0.05
* Type Beta: claiming no significance when there is. Beta=1-alpha
* Power: refers to the power of detecting a true difference between two populations. *Power*=1-beta.