18 points to remember from your stats sessions NERD Block

1. There are different "types" of statistics.
    a. Stochastic: This is what we typically think of "biostatstitics" as, and most of what we will be talking about; distributions, probabilities, independence, etc.
    b. Bayesian: based on prior probabilities or "beliefs" in a hypothesis of study findings.
    c. Qualitative

2. For stochastic statistics, there are assumptions which underlie the use of most tests; They vary, but authors should at least make an effort to evaluate them.

3. Variables in analyses: Independent (the one that is manipulated) and dependent (or outcome) variable

4. Variables types:
    a. Nominal (name based and have no intrinsic numeric value); eg: gender, eye color, etc.
    b. Ordinal: scale based where points on the scale have a relative value to other points eg: pain scale 0-10.
    c. Interval: numbers are exactly relative to each other: eg: in Celsius, 1◦ is $1/100^{th}$ of the difference between freezing and boiling for water (at 1 atmosphere of pressure)
        i. Ratio: an interval scale that has a true zero point. Eg: Kelvin scale for temperature, Blood pressure, lab indices

5. Variables can be coded different ways: In medicine we love dichotomies; many variables are naturally dichotomous – mortality (alive/dead), tumour recurrence (yes/no). Many variables in medicine are continuous (eg: temperature, lab tests, blood pressure, etc) but we dichotomize them (hypotensive SBP < 90 mmHg; febrile – T > 38◦ C; neutropenic – neutrophils < 0.5, etc). When you dichotomize continuous variable, you lose information (eg SBP dichotomized at 90 mmHg gives the same numeric code to a SBP of 91 and 145 mmHg, or 88 and 40, etc.). Because you lose information, a larger sample size will be required to declare a difference "statistically significant".

6. Descriptive statistics:
    a. central tendency
        i. Mean (arithmetic average); total values/# of observations
        ii. Median: 50% of data on either side (not affected by outlier scores).
        iii. Mode: most commonly occurring value.
    b. Dispersion:
        i. standard deviation
        ii. co-efficient of variation (std deviation/mean) - if large (> 50-60%) can tell you that a distribution of data is not normal)
        iii. Intra quartile range (often reported with median) 75^{th}-25^{th} percentile

7. Hypothesis testing: The null hypothesis is always stated that the groups are "equal". i.e. there is no difference between treatments; the odds ratio/relative risk/hazard ratio, etc is not statistically different than the null value (1), or the absolute risk is not statistically different than the null value (0).

8. Hypothesis testing is conventionally based on a P value of 0.05. Which means, essentially, that if you repeated the same study 20 times, from the same population, 19 times out of 20 you would get the same result (in terms of the P value being less than or greater than 0.05). One time out of 20 you would get a different result. There is no reason you couldn't use a P value of 0.1 or 0.001, etc; it is really just based on convention. P value indicates the probability of observing a test result at least as extreme as the one we observed, given that the null hypothesis is true.

9. Confidence intervals (unlike P values) cover a range of possible values for the difference between groups/treatments/etc within which the true value may lie. Usually presented as 95% confidence interval. The lower boundary represents the smallest plausible treatment effect/value/etc that is compatible with the data; the upper boundary represents the largest plausible value. 95% CI can be applied to/calculated for any measure (absolute risk, odds ratio, relative risk, hazard ratio, etc).
    a. Most medical journals require confidence intervals in reporting results; P values are used as well.
    b. Because they arithmetically use the same numbers (set up differently), IF a P value is significant (< 0.05), then the 95% confidence interval MUST NOT include the null value (1 for ratio measures and 0 for additive measures).

10. Measures of association:
    a. Odds ratio – useful because it approximates relative risk/risk ratio. Convenient, because they are easily obtained from logistic regression. They are the main outcome measure available from case-control studies (relative risk and risk ratio are generally not). They are also available from cohort studies and RCTs. The problem there is that the odds ratio will always overestimate the relative risk in studies where the outcome of interest is not rare (> 10% of subjects have it). Say for example, a study of a platelet agent for ACS versus aspirin and the outcome is 60 day re-vascularization, and it occurs in 20 and 30% of subjects in each group, the odds ratio over estimates the relative risk. There is an easy correction (see below – JAMA 1996) and the good news is that if the OR is significant (95% CI does not include 1), then nor will the corrected RR.

What's the Relative Risk?

A Method of Correcting the Odds Ratio in Cohort Studies of Common Outcomes

Jun Zhang, MB, PhD; Kai F. Yu, PhD

Logistic regression is used frequently in cohort studies and clinical trials. When the incidence of an outcome of interest is common in the study population

A and Hospital B by comparing neonatal mortality in very low birthweight neonates between these 2 hospitals.[2] At first

b. Number needed to treat (NNT) and number needed to harm (NNH) are used frequently in trials of therapy. The formula is 1/absolute risk reduction (risk of event in treatment A – minus B) or 1/risk increase (if NNH). For NNT and NNH, you usually WILL NOT see confidence intervals calculated; because of the way it is calculated, the arithmetic for the 95% CI breaks down.

11. T-test, analysis of variance (ANOVA), analysis of covariance (ANCOVA) and linear regression are used for comparisons of means in continuous data; If 2 groups, it is a t-test, 3 or more groups, ANOVA. In situations where a patient can serve as their own control (before-after studies, etc) then a *paired* t-test is used. Linear regression (and multivariable linear regression) and correlation can be used in the (rare in medicine) situation whereby you have a truly continuous outcome. The primary assumptions underlying the use of these tests is that the data is normally distributed (there is s test for this) and that the variances of the groups on the continuous outcome are equivalent.

12. Non parametric statistics (also known as "distribution free statistics") are use used for data which is continuous, but does not meet the assumptions of normality of distribution or equivalent variances. An example from emergency medicine research is "length of stay" which is frequently used as an outcome in ED studies; this outcome is usually skewed, because some people will have very long lengths of stay for various reasons; Means would be affected by these few outliers, but medians/ranks etc would not. Examples you may see are; chi-square, sign test, Wilcoxon signed rank test, Kruskal-Wallis test, Friedman test, etc;

13. Simplest case of 2 treatment groups and 2 outcomes gives you a 2x2 table like this:

	Alive	Dead
Treat A		
B		

And the Chi-square gives you the observed versus expected frequencies in each cell of the table, and a P value. Note that, from this table, you can also calculate the relative and absolute risk as well as the odds ratio (cross products ratio).

If the EXPECTED cell frequencies in any cell is < 5, then Fisher's Exact Test is more appropriate than chi-square

14. More advanced techniques for dichotomous data include:
    a. Log-linear analysis – compares multiple combinations of two variables for data sets
    b. Logistic regression: a favourite in medicine. Regression for a dichotomous outcome and a single (univariable) or, more commonly, multiple predictors. Renders odds ratios form software. The odds ratio from a 2x2 table will be exactly the same as the odds ratio from a univariable logistic regression (one outcome, one predictor). It may be very

slightly different if a continuity correction factor is used. For multiple logistic regression (and same for linear regression) you need about 15 or more X's for each predictor, to ensure that he model is not unstable. For example, a study of out-of-hospital cardiac arrest and survival looking at age, gender, bystander CPR, time to EHS arrival, past history of CAD, and smoking history would require at least 90 subjects (more is better). Logistic regression has several assumptions as well –probably the most important is that the model "fits". This is established by model fit tests (eg: Hosmer-Lemeshow) or maximum likelihood estimation.

c. Survival analysis: probability of survival at a given time point: Life table analysis, Kaplan-Meier curves, etc.
d. Cox regression/Proportional hazards: a type of survival analysis, often called *semi-parametric* because it includes a function for *time-to-event* and multiple predictors. From calculations/software, you obtain *hazard ratios* which are like relative risks (not odds ratios).
e. Regression trees.

15. Propensity score analysis: helps to correct for the predictors that led to a patient getting a specific treatment in <u>non randomized</u> trials. There is still potential for significant latent confounding.

16. Bayesian analysis: based on prior probabilities of "beliefs" in a hypothesis or finding. Sounds subjective, but it is how we function in day-today life. Gaining in popularity; It is often used in the diagnostic literature realm in the form of likelihood ratios.

17. Measures of agreement: Kappa: looks at agreement above chance.

18. Machine learning: Is coming – may be better in some settings.

Using Machine Learning to Define the Association between Cardiorespiratory Fitness and All-Cause Mortality (from the Henry Ford Exercise Testing Project)

Mouaz H. Al-Mallah, MD, MSc[a,b,*], Radwa Elshawi, PhD[c], Amjad M. Ahmed, MBBS, MSc[b], Waqas T. Qureshi, MD, MS[d], Clinton A. Brawner, PhD[a], Michael J. Blaha, MD, MPH[e], Haitham M. Ahmed, MD, MPH[e,f], Jonathan K. Ehrman, PhD[a], Steven J. Keteyian, PhD[a], and Sherif Sakr, PhD[b]

Am J Cardiology 2017. https://doi.org/10.1016/j.amjcard.2017.08.029