Final Exam Study Guide
(Note that this is not a comprehensive list)

- **Epidemiology:**
  - Definition:

- **Causality:**
  - Bradford Hill Criteria:
    - What are the seven criteria? What do they mean?

- **Types of Epidemiological Studies**
  - Clinical case series
    - Definition:
  - Cross-sectional study
    - Definition:
    - Measure of association:
  - Case-control study
    - Strategy:
    - Suitable for what kinds of diseases?
    - How do we find cases and controls?
    - Measure of association:
    - Advantages:
    - Disadvantages:
  - Cohort study
    - Participants are selected based on what?
    - Measure of Association:
    - Advantages:
    - Disadvantages:
  - Randomized controlled trial
    - Definition of clinical trial:
    - Intention to treat rule:
    - Randomization:
      - Definition:
      - Masking/Blinding:
      - Sample Size: pros/cons of large/small sample size:
      - What is gained by randomizing?
    - Advantages:
    - Disadvantages:

- **Measures of Dz occurrence: Incidence, Prevalence,...**
  - Definitions:
  - Formulae:

- **Prevalence**
  - Definition:
  - Formula:

- **Odds Ratio**
  - 2x2 contingency table (attention how to set up)

	A	B
	C	D

    - Formula for OR:
    - Formula for 95% CI:
    - Null Hypothesis:
    - Interpretation of OR
      - <1:
      - =1:
      - >1:
    - Interpretation of 95% CI:
    - Factors that influence the size of the OR and CI
      - Sample size:
      - Measurement error:
      - Case misclassification:

- **Relative Risk**
  - 2x2 contingency table (attention how to set up)

	A	B
	C	D

  - Formula for Relative Risk:
  - Formula for 95% CI:
  - Null Hypothesis:
  - Interpretation of RR:
  - Interpretation of 95% CI:

- **Chi-Square Test**
  - Evaluates the difference between …?
  - Null Hypothesis:
  - Assumptions:
    - 1)
    - 2)

When to employ **Fisher Exact Test**?

- **Power**
  - What is the interpretation of alpha and beta in hypothesis testing?
    - Type I error:
    - Type II error:
    - Power:
  - Why don't we make alpha and beta as small as possible?
    - How are alpha and beta related?
    - How is sample size affected by alpha and beta?

- **Sample Size**
  - What is needed to find the sample size? List:

- **Non-parametric tests**
  - Definition:
  - Spearman Correlation: non-parametric version of _____ , based on the _____ of the data instead of the data themselves
  - Wilcoxon Signed Rank Test? (hypotheses, test statistic, interpret results..)
  - Wilcoxon Rank Sum: non-parametric version of the _____ test
    - Null Hypothesis:
    - W_s is the rank sum of the smaller or larger group?
  - Advantages:
  - Disadvantages:

- **Survival Data Analysis**
  - Basic concepts
  - Survival function
    - Definition:
  - Censored data
    - Definition and examples
  - Kaplan-Meier Approach
    - Probability of surviving beyond month t = ?

- **Review of statistical tests from before midterm**
  - What kind of test to use?

    - T test (assumptions: normal distribution, independent observations, variances are equal {there is a special test for unequal variances})
      - Paired
        - Special case where the T distribution is used for the situation that every subject observed in the first group is observed again in the second group of data points
        - Null: mean before=mean after

        - Alternative: mean before ≠ mean after
    - One sample
        - When we want to compare the mean of a variable to a specific value (say, if we calculated the change in a variable, and then tested if the mean change=0)
        - Null: mean=0 (or other value)
        - Alternative: mean ≠ 0
    - Two sample
        - When we want to compare means from two different study samples
        - The test determines whether the difference between the observed means is too large to have occurred by chance
        - The null hypothesis is that mu1=mu2. The alternative is that mu1 and mu2 are not equal.
- ANOVA
    - Null: all means are equal
    - Alternative: at least one mean is not equal
    - Assumptions
        - Independent observations
        - Normal distribution in the dependent variable
        - Homogeneity of variance across the groups (rule of thumb: s1/s2<2)
    - Analysis of variance
        - Gets its name from its ability to determine whether the variability within groups is greater or lesser than between groups (between groups is more desirable)
        - The outcome variable is the dependent variable, which must be continuous and the explanatory variables are the independent variables.
        - If there are only two groups, the ANOVA p value is the same as a 2-sample t test p value (if there are no other variables in the model)
    - Within group variation: Within each group, the total deviation of each individual's score from the group mean
    - Between group variation: the deviation of each group mean from the overall mean
    - If the variability due to the independent variables (between group) is greater than the variability that is unaccounted for in the model (within group) then the groups are different with respect to their means.
    - Bonferroni
        - This test addresses the problem of multiple comparisons among the groups
        - If we want to determine which groups are different from others we can use Bonferroni

        - We don't want to inflate the alpha level, so we need a way to adjust the pair-wise alpha so that it keeps the experiment-wise alpha at 0.05
        - Bonferroni achieves this by dividing alpha by the number of pairs of means that will be compared
- Simple Linear Regression
    - Assumptions
        - The relationship between the outcome and the predictors can be described by a linear relationship
        - The errors are normally distributed (which implies that Y is normally distributed)
        - The errors have the same variance
        - All of the data points are independent
    - Goal: to find a linear equation relating a dependent variable to an independent variable
    - Residual or error: sum of the differences of each observation to the corresponding predicted point on the line
    - Both variables must be continuous
    - Null: rho=0
    - Alternative: rho ≠ 0 (which is the same thing as slope=0 and slope≠0)
    - Interpretation of slope and intercept
    - Outliers
    - Extrapolating
    - Correlation (Rho estimated by 'r' the Pearson corr coef)
        - The degree to which two continuous variables are linearly related
        - Rho: ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation)
        - Rho=0 means that there is no linear relationship between the two variables
        - Perfect correlation means that knowing one variable allows perfect knowledge of the other variable
        - Used with two continuous variables
        - Not necessarily cause and effect
        - When the causal relationship is known or strongly suspected (and linear), we use regression to predict the levels of one variable, given the values of the other
    - Degree of correlation

        Calculate the Pearson correlation coefficient r
        - 0 – absence
        - 1/-1 – perfect correlation
        - 0.75 to 1/-0.75 to -1 – high degree
        - 0.25 to 0.75/-0.25 to -0.75 – moderate degree
        - 0 to 0.25/0 to -0.25 – low degree

- R^2 (Note: R^2 same as r^2 for simple linear regression only
    - Describes how much of the variability in the dependent variable was accounted for by the model
    - The more variables in the model, the larger R^2 becomes (once a variable is in the model and another is added R^2 cannot get any smaller)

- How to find normality?
    - Visually
    - Mathematically
- P-value and test statistic
    - If p<=alpha then reject null hypothesis and conclude the alternative
    - If p>alpha, then fail to reject the null
    - Likewise, if the calculated test statistic is less than the critical value (from F table, or t table…), then we fail to reject the null hypothesis. If the calculated statistic is greater than the critical value, we reject the null hypothesis and conclude the alternative hypothesis.