Statistics for Medical Residents

Topics and Concepts Outline

I have listed the major topics below. In three hours, it will be impossible to cover most of these concepts in great detail. These are, however, the types of concepts and methods often encountered in medical articles. Use this outline for ideas on describing the assigned articles.

- Sampling
  - populations vs. samples
  - parameters vs. statistics
  - inference vs description
  - variability
  - statistical independence
- Descriptive/summary measures
  - statistics
    - distribution shape (Where do most of the measurements fall?)
      - symmetric
      - symmetric and bell-shaped (Normal or Gaussian)
      - positively skewed (skewed to the right)
      - negatively skewed (skewed to the left)
    - central tendency (Where is the center?)
      - means (arithmetic, geometric, harmonic)
      - median
      - mode
    - dispersion (How spread out?)
      - standard deviation
      - interquartile range (IQR)
      - range
    - correlation (To what degree do two variables vary linearly together?)
      - Pearson product moment (ratio and interval level)
      - Spearman rank-order (ordinal level and above)
      - correlations from different populations ordinarily CANNOT be compared
    - other
      - quantiles (example: median is the 50^{th} percentile, not affected much by outliers, minimum is 0^{th} percentile)
      - percentile ranks (% of measurements below a specified value)
      - minimum, maximum
  - tables
    - provides details

        - goal is to facilitate comparisons
        - different types of quantities should be in separate columns (for example, "means" and "SDs" should be in separate columns, appropriately labeled)
        - column alignment is important in making comparisons EASY
        - ANYTHING that makes comparisons difficult should be avoided
    - graphs
        - goal is to make seeing important patterns easy
        - use as little "ink" as possible - the data-to-ink ratio should be high
        - avoid annotation - interferes with seeing patterns (if you want detail, make a table)
        - do NOT use pseudo-3D effects - distorts and misrepresents the data
        - examples of graphs:
            - histograms / bar graphs / box plots
            - scatter plots
    - outliers (rogue values)
        - cannot be removed simply because they are extreme
        - often provide essential information
        - published procedures for removal should NOT be used
        - values that have been independently documented as bogus due to a known measurement problem may be removed
    - missing values (nonresponse)
        - missing values should NOT be set to zero!
        - if responses are missing at random, then the study results will not be biased
        - missing values due to systematic factors tend to cause bias
        - in surveys, nonresponders tend to differ systematically from responders
- Measurement levels (higher levels include attributes of lower levels)
    - ratio (true zero point, examples: Kelvin temperature scale, age) - highest level
    - integer (equal intervals, examples: Fahrenheit and Celsius temperature scales, IQ)
    - ordinal (rank order, example: educational levels: grade school, high school, college)
    - nominal (labels, examples: gender and race) - lowest level
- Study design
    - response(s) (we are looking for some change in the response due to other factors)
    - explanatory factors
        - treatment and control groups
        - covariates
    - number of subjects
    - how subjects are allocated to treatment groups
    - experimental
        - researcher intervention

        - randomization
    - observational
        - no intervention
        - randomization not possible
    - cross-sectional vs longitudinal
    - clinical trial
        - dropouts
        - intention to treat analysis
        - efficacy subset analysis
    - power and sample size
- Clustered (hierarchical, multilevel) data
- Statistical models
    - response (dependent variable)
    - explanatory variable (independent variable)
    - response = systematic components + random error
    - parameters in the model describe how the response is affected by the explanatory factors - the goal is to estimate the parameters using data collected in the study
    - cross-sectional vs longitudinal
    - linear/nonlinear models
        - analysis of variance (ANOVA)
            - F-statistics
            - Multiple comparison procedures
                - Fisher's Least Significant Differnce (LSD, all possible t-tests after significant ANOVA F-test)
                - Tukey Honestly Significant Difference (HSD, pairwise comparisons)
                - Newman-Keuls (pairwise comparisons, does not provide confidence intervals, not recommended)
                - many other procedures
        - regression
    - generalized linear models
        - logistic regression (binary response, model parameters are odds ratios, often used in case-control studies)
        - ordinal regression (ordered categorical response, model parameters are odds ratios)
        - Poisson regression (count response, model parameters are odds ratios)
    - survival analysis
        - cross sectional analysis
        - response is time to event occurrence
        - model parameters are hazard ratios

- multilevel (mixed effects, hierarchical) statistical models
  - linear/nonlinear
  - can be longitudinal (correlated measurements over time)
  - often used for meta analysis
- time series models (correlated measurements over time)
- structural equation models (SEMs)
  - manifest (observed) variables
  - latent (unobserved) variables
  - most of the other models listed here can be formulated as SEMs

- Classification
  - sensitivity
  - specificity
  - receiver operating characteristic (ROC) curves
- Contingency table analysis (sometimes referred to as "Chi-square tests")
  - determine whether two factors are dependent
  - uses Chi-square distribution as a large sample approximate test
  - of limited use - Chi-square statistic has no intuitive meaning
  - estimates of effects:
    - odds ratios (exists for case-control studies)
    - relative risk (does not exist for case-control studies)
- Method comparison studies (comparing methods or devices/instruments)
  - measurement error
    - systematic error: bias (lack of accuracy, lack of trueness)
    - random error: imprecision (lack of reliability, lack of repeatability)
    - good agreement between methods = little or no bias and little imprecision
    - Bland-Altman plots
    - SEMs
- Confidence intervals for parameters
  - shows all plausible values for an unknown parameter
  - gauges uncertainty in knowledge of parameter value
  - equivalent to an infinite number of hypothesis tests
- (Statistical) Hypothesis tests
  - questionable usefulness of single hypothesis tests
  - null hypothesis
  - alternative hypothesis
  - significance levels
  - empirical significance levels (P values)
    - "statistical significance" is NOT necessarily of any importance
    - "statistically significant" is often abbreviated as "significant" in research articles

        which causes much confusion
        - a statistically significant effect is NOT necessarily a large or (clinically) important effect
        - saying that an effect is "statistically significant" means only that you are confident in the sign (direction) of the effect
    - parametric tests - distribution shape the errors is assumed known
    - nonparametric tests - no assumption of error distribution shape
- Statistical tests - every statistical test implies a statistical model - focus on the model
    - parametric
        - t-test (single group, two independent groups, paired groups, equal population variances, unequal population variances)
    - nonparametric
        - compare proportions (large sample approximation using Chi-square distribution)
        - Wilcoxon rank sum test (comparison of independent group medians)
        - Wilcoxon signed rank test (paired groups)
        - Kruskal-Wallis test - one-way nonparametric ANOVA
        - Friedman test - two-way nonparametric ANOVA
    - randomization/permutation tests
- Quality control/quality assurance
    - applied to lab work
    - statistical techniques used to ensure accurate and precise measurements

Outline for Describing Each Article

Make copies of this outline and answer the following questions for each article before class. Use the **Topics and Concepts Outline** for ideas on what to look for. Please bring your answers to class and be ready to discuss the articles.

1. Study Goal - Briefly describe the researchers main objective in conducting this study. What was the main research question?

2. Study Design - Briefly describe the statistical design used to answer the research question. (Example: 2 x 3 factorial design)

2.a. Is this an experimental study or an observational study? How can you tell?

2.b. Is this a cross sectional study or a longitudinal study?

2.c. What are the experimental units ("subjects")?

2.d. Name the response(s) (i.e., dependent variable(s)). Is there more than one response for each experimental unit ("subject")? (That is, are the responses repeated?) Is the data clustered (hierarchical, multilevel)?

2.e. What is (are) the measurement level(s) of the response variable(s)?

2.f. Name the explanatory factors (independent variables):

2.g. What is (are) the measurement level(s) of the explanatory variable(s)?

2.h. How many subjects (experimental units)?

2.i. How were subjects allocated to treatment groups?

3. Name the statistical methods used (e.g., ANOVA, regression, t-test, longitudinal mixed effects model, etc.) Example: A t-test was used to compare the treatment to the control. Were appropriate statistical methods used?

4. Name and describe the types of statistical graphics used (e.g., histogram, bar chart, scatter plot, etc.) Example: bar chart for cell counts for each treatment group. Discuss how well the graphics illustrated the results. Do the graphs make it easy to detect differences or trends, or to make comparisons in trends? Are the graphs unnecessarily cluttered? Would the use of color help in the interpretation? Were there any pseudo-3D bar charts or pie charts? (Pseudo-3D is bad because it causes distortion of the data.)

5. If any tables were used, describe them. (Limit this to the two most important tables.) Are the tables easy to read? Are numbers aligned in columns to facilitate comparison? Does the table layout make comparison unnecessarily difficult?

5. What statistical hypotheses, if any, were tested? (State the null and alternative hypotheses.)

6. Single hypothesis tests are of little use in most studies. Confidence intervals are much more informative because they describe the plausible values for the unknown parameters. Did the researchers use confidence intervals to describe their results or simply state P values?

7. If you have any other comments about the article, list them here.