

Tier 1 – High likelihood of appearing on the exam:

- Data basics
  - o Identifying cases, variables, and types of variables
- Identifying and describing associations using data visualizations
  - o Know which type of graphs are appropriate for which combination of variables
  - o Know how to apply the definition of association to what you see in a graph
    - Know to discuss form, strength, and direction for associations seen in a scatterplot
    - Know how to compare across groups/categories in boxplots or bar charts
- Interpreting coefficients in simple linear regression and multivariable linear regression models
  - o Understanding how categorical explanatory variables are represented, including one-hot encoding/dummy variables and reference categories.
  - o Adjusted vs. un-adjusted effects and when each is useful
- Confounding, stratification, and conditional vs. marginal effects
  - o Know the definition of confounding, and understand how imbalances in a third variable can influence the marginal relationship between an explanatory and response variable

Tier 2 – Moderate likelihood of appearing:

- Differences between Pearson's and Spearman's correlation.
- Describing the distribution of a single variable
  - o Know that you should discuss center, spread, shape, and outliers for quantitative variables
  - o Know that you should describe categorical variables as either concentrated into one or more categories, or approximately uniform.
- Contingency tables
  - o Conditional proportions, relative risk, and odds ratios as measures of association
- Random assignment and its relationship with confounding

Tier 3 – Low likelihood of appearing (but one or more of these still might appear):

- Coefficient of determination ( $R^2$ )
  - o How it can be used to quantify the strength of association for non-monotonic patterns
  - o How adjusted  $R^2$  can be used to determine if more complex models are warranted
- Ecological fallacy
- Adjusting for confounding via re-weighting the data