Case studies, Method and Assumption Checks, and Executive Summaries

Russell Millar

One focus of STATS 20x is learning how to clearly and concisely write up your analysis and the results, using plain English. Specifically, in your assessments you will be required to write **Method and Assumption Checks** and/or **Executive Summaries**. These can be worth 50% or more of the total marks.

The Case Studies in this coursebook are a valuable resource for learning how to score well on **Methods and Assumption Checks** and **Executive Summaries**. Also, see the Model Answers from previous tests and exams (under Modules in CANVAS) for a wide variety of additional examples.

The requirements for scoring full marks on **Method and Assumption Checks** and **Executive Summaries** are given below.

Method and Assumptions Checks

The exact content will depend on the particular situation and modeling steps taken, but would typically include mention of:

- Justification of the steps taken to arrive at the fitted model that is used for inference. For example,
  - Inspection of plots to assess features of the data
  - Decisions made in choosing the best fitted model (e.g., transformations, variable selection, removal of outliers, use of a quasi model, etc)
  - Clear statement of the form of the chosen model (e.g., simple linear, t-test, power law, two-way ANOVA, interaction, quasi-Poisson, etc)
  - Validity, or not, of assumptions
- Equation of the fitted model (unless told otherwise)
- Statement about R^2 (the proportion of variation explained. This is only relevant to linear models - doesn't apply to generalized linear models)

Executive Summary

Typically:

- Brief introduction.
- Clear statement of the conclusions expressed in a complete sentence (e.g., ``There was a significant linear relationship between test score and exam score'').

- A clear quantification of the effect of the significant terms in the model, expressed as confidence intervals (in STATS 20x the actual estimated values are not required)
- Prediction intervals if requested. If there is reason to doubt the validity of the prediction intervals then this should be mentioned. Reasons include:
  - Residuals from a linear model being clearly non-normal (this could arise in a situtation where there is a large number of observations, so non-normality is OK for estimation of effects and confidence intervals, but not for prediction of new observations).
  - The intervals don't make sense. E.g., they have lower or upper bounds that are impossible, such as predicting a percentage and getting bounds below 0% or above 100%.
- Be very careful with your interpretation of the intervals. E.g.,
  - Is the effective additive or multiplicative?
  - Is the interval for a predicted value, or an expected value, or a median value, or a log-odds?
- ALSO, be sure to address any additional specific questions of interest that may have been asked.