

STAT 582 - General Guidelines for Analyzing Data

- How were the data collected/generated?
 - Are they from a designed experiment or observational study?
 - If a designed experiment, how was the experiment run?
 - If observational, what is the study population?
 - Design:
 - How was the randomization performed?
 - What is/are the experimental unit(s)?
 - Material to which a treatment is applied in a single trial of the experiment – Cochran and Cox (1957)
 - Observational:
 - How was sampling performed?
 - Are there any potential confounders?
 - Are data grouped or clustered?
- Why were the data collected?
 - What sorts of results are expected? What are the statistical hypotheses?
 - Are there primary and secondary goals?
 - Is there interest in looking for patterns or unanticipated results?
- Plot/summarize the data (EDA!!)
 - Look at data versus treatment factors, blocking factors, time, location, technician, predictors, etc.
 - Use interaction plots, superimposed plots, scatterplot matrix, etc.
 - Do these plots seem to support the hypotheses of interest?
 - If not, do the patterns suggest revised or new hypotheses?
 - Are there any suspicious patterns that suggest a violation or concerns about model conditions? Potential outliers?
- Fit the data as appropriate
 - Start simple and build towards more complex models if necessary
 - Maybe look at subsets of the data first and then combine
 - Remember that the analysis should be guided by the needs (and possibly the statistical ability) of the domain expert, the way the data were collected/generated, and the data itself
 - A simpler model that serves the researcher's needs is almost always better than a more complicated model the researcher can't explain.
- Check conditions necessary for trustworthy inference
 - Are the conditions reasonable given how the experiment was conducted?
 - Check the residuals using various plots (see above)
 - Look for correlation and outliers
 - Should a transformation or weighted analysis be performed?