Selecting Statistical Procedures

Inference about "centers" (means or medians) of distributions of quantitative variables*

design	assuming normality	nonparametric rank-based	nonparametric sign-based	randomization test (CIs by bootstrap method)
one sample	*t*	signed-rank	sign	bootstrap CI [hypothesis test as for paired]
paired sample	paired *t*	analyze differences as one sample		randomize signs of differences
two independent samples	*t*	rank sum	Mood's median	randomize observations between samples
three or more samples	ANOVA	Kruskal-Wallis	Mood's median	randomize observations among samples
pre-planned hypotheses	contrast	[none other than comparing two means/medians by two-sample test]		
unplanned comparisons				
all pairwise	Tukey's or stepdown	use set of two-sample tests with Bonferroni adjustment of significance level		
to control	Dunnett's	use set of two-sample tests with Bonferroni adjustment of significance level		

* confidence intervals can be obtained for all these procedures except the tests among three or more samples.

Inference about relationship between two variables

design	assuming normality	nonparametric	randomization test
two quantitative variables			
response-explanatory model desired	LS regression	robust regression (several types)	randomize one variable relative to the other
only measure of association desired	correlation	Spearman's rank correlation, Kendall's	
quantitative response, categorical explanatory	analyze as two or more samples, as in table above		
categorical response, quantitative explanatory	reverse the response and explanatory roles and analyze as two or more samples, as in table above [or logistic regression]		
two categorical variables	analyze proportions as in table below		

Inference about probabilities (proportions)

design	
one sample, binary outcome	binomial or normal-approximation test, normal-approximation CI
two samples, binary outcome	normal-approximation test and CI, or contingency-table test
more than two samples and/or more than two outcomes	contingency table: X^2, Fisher's exact test [, *G* test]

Selecting Statistical Procedures

The table on the reverse covers only the situations covered in this course, involving no more than two variables. It includes only the most commonly used procedures, those covered in this course [plus, in brackets, a few others].

What general sorts of procedure(s) to use should be primarily determined ***by the scientific question(s) you want to answer***. In particular:

- Nearly all of the procedures in the table can be used for **estimation** (**confidence intervals**) as well as for **significance testing**. Which is more appropriate depends on your purpose.
- Most of the procedures in the table provide inference about measures of "**central tendency**" (means, medians, population probabilities, etc.). Under some circumstances other features of a population or distribution, such as the degree of **variation**, may be more interesting or important. There are special procedures which can address some of these.

Once the general sort of procedure is determined, based on the purpose of the analysis, the choice of a specific method will depend on how the data were obtained — the design of the study — and on the features of the data, such as the shapes of distributions. In other words, you need to check that the data meet the assumptions of the procedure you choose. The general assumptions, from most to least important, are:

- Independence of observations.
  - This generally depends on the design of the study, but can be empirically assessed in some cases (e.g. plotting regression or ANOVA residuals against temporal or spatial order).
  - Some forms of dependence can be handled by more complicated analyses (e.g. blocked or nested ANOVAS).
- Linearity of relationship – regression and Pearson's correlation only.
  - Assessed by scatterplots and residual plots; a smoother can help.
  - Nonlinearity can be handled by transformations, polynomial or nonlinear regression, or tests for monotonic relationships (e.g. Spearman's or Kendall's)
- Constant/equal variances – regression and ANOVA, Kruskal-Wallis; the null hypothesis in rank-sum and randomization tests includes equal variances.
  - Assessed by comparing standard deviations of samples, or by inspecting scatterplot or residual plot.
  - Uneven variation can be handled by transformations, modified procedures not assuming homogeneous variation (e.g. Welch's ANOVA), or weighted least squares methods.
- Normality – "parametric" procedures only.
  - Assessed by normal-quantile-quantile plots, histograms, boxplots, etc. The most important problems are outliers and severe asymmetry.
  - Most parametric procedures are fairly robust to non-normality, especially if sample sizes are not small. The nonparametric and resampling procedures do not assume normality: this usually is the <u>only</u> assumption they avoid. Other remedies for non-normality include transformation, or methods design for a specific non-normal distribution.

The final factor to be considered in selecting a procedure is **power**. For most data parametric and randomization procedures will be the most powerful, rank-based nonparametric procedures will be as powerful or nearly so, and sign-based nonparametric procedures will be least powerful. With heavy-tailed distributions or outliers, parametric procedures are less powerful than randomization and rank-based nonparametric procedures; with very heavy tails, sign-based procedures may be most powerful.