Y (outcome)

*Use with caution

O1. Y (continuous)
Assumptions: LINE
A1. Linearity (Not met: transform the outcome (log) or covariates, polynomial terms, categorize a continuous covariate, splines, GAM, etc)
    -Use lowess to assess
A2. Normality (Not met: transform the outcome (log), robust/ sandwich standard errors, etc)
A3. Homoscedasticity (Not met: transform the outcome, robust/ sandwich standard errors, weighted least squares, etc)

O2. Y (binary)
O3. Y (categorical)
O4. Y (rate)
O5. Y (time to event)

Y's independent

Y's dependent

I1a. One binary covariate

I1b. One continuous covariate

I2. One categorical covariate (more than 2 groups)

I3. Multiple covariates

D1. One binary covariate

D2. One categorical covariate (more than 2 groups)

D3. Multiple covariates

1. 2 sample t-test (equal or unequal variance)

2. Linear Regression (same as 2 sample t-test with equal variance)

3. Wilcoxon Rank Sum test (if n is small or no normality)

1. Linear Regression (equivalent to Pearson correlation)

2. Pearson correlation

3. Spearman rank correlation coefficient

1. ANOVA

2. Linear Regression (indicator variables for the categorical variable)

Linear regression

1. Paired t-test

2. Wilcoxon signed- rank test

1. RM ANOVA

2. Linear Mixed Model

1. Linear Mixed Model
2. Generalized Estimating Equations GEE

Concerns and Caution for Linear Regression:
Usually, adjust for confounders & precision variable but not mediator
Consider and check for multicollinearity (VIF): with lots interactions
Check the residuals to see if the assumptions A1-A3 are met
Check for outliers (run the model with and without the outlier)
Model building: AIC, BIC, partial F-test, adjusted R^2, MSE

O2. Y (binary) ← Y (outcome): Assume Independence unless stated otherwise

One binary covariate

Multiple covariates

1. Logistic regression (Wald, Score, LRT)

2. Mantel-Haenszel (categorical confounding factor)

O3. Y (categorical)

1. Multinomial logistic regression (no order Y, compare to base)

2. Ordinal logistic regression (order Y)
2b. proportional odds ordinal logistic regression

O4. Y (rate): # of cases per unit time= incidence rate

1. Poisson regression
-Stationarity: assume incidence rate is constant over time conditional on covariates
-Goodness of fit: Pearson chi square or deviance for moderate # covariate patterns

2. Zero-Inflated Poisson (excess zero counts)

3. Negative Binomial Regression (over dispersion)

1. 2 sample test for binomial proportions

2. Relative Risk

3. Risk Difference

4. Odds Ratio

5. Logistic regression

O5. Y (time to event)
Non informative censoring

1. Kaplan Meier (Survival function)
-Nonparametric
-Confidence interval for $S(t_k)$ with Greenwood or complementary log-log transformation
-Log rank test for differences in survival distribution for binary or categorical covariate
-Other tests: Gehan-Wilcoxon, Peto-Prentice, Fleming-Harrington, Taron-Ware, etc

2. Exponential Survival Regression (hazard)
-Parametric
-Baseline hazard is assumed constant over time

3.Weibull Regression model (hazard)
-Parametric
-Baseline hazard varies as a function of time

4. Cox proportional hazard
-Semiparametric
-Ties: Exact, Efron, Breslow
-Don't need to specify the baseline hazard $h_0(t)$
-Assumes hazard ratio is constant over time
-Plot $\log[-\log(S_j(t))]$ for j=0,1: check parallel over time

Logistic regression:
Model building: AIC, BIC, LRT, *pseudo R^2*

Goodness of fit:
1. Calibration
1a. Pearson chi square
1b. *Hosmer-Lemeshow*
2. Discrimination: c Statistic

Matched design: conditional logistic regression

GLM links: logit, probit, identity, log, complementary log-log