What is your objective?
- Publication → Choosing a journal? Finding reviewers? The review process?
- Solve a Pragmatic Problem
- Increase Knowledge
- Satisfy PhD Requirement → Finding out what those requirements are?

Problem identification and justification
- Literature Search
- Discussion with Experts and colleagues

Enumerating your resources
- Money → Grant writing
- Time → How many hours per week?
- Time → How many weeks?
- People → Research assistants?

Problem development and hypothesis generation
- Choose model(s)
- Choose definition(s)
- Formulate hypothesis

Choose your methodology
- Type
  - Empirical research
    - Case study
    - Simulation
    - Field Study
    - Controlled Experiment
      - Interview
      - Observation
      - Audio/Video
      - Questionnaires
  - Theoretical Conceptual analysis
  - Literature review
  - Meta-analysis
- **3A**
- Pragmatics
  - Appointments
    - One-on-one
    - Group
    - Location
    - Length
    - Number
  - Response recording
    - Digital
    - Physical
  - Sample Characteristics → **1A**

1A

Availability
- Disabilities
- Occupational concerns
- Health concerns
- Likelihood of participation → Incentives
- Recruitment → Posters
- Recruitment → Advertisements
- Recruitment → Pre-existing pool

Required Number

What type of study are you running?

→ Correlational

Do you know the likely effect size?
- Yes → Calculate minimum sample size needed to see the effect, based on Cohen's d
- No → Guess: Recruit as many as possible

Experimental

Do you know the likely effect size?
- Yes
- No → Guess: Recruit as many as possible

Can you use a control group or a seperate control treatment?
- Yes
- No → Time-series trial , n = guess

Is your measure reliable over the intervention period?
- Yes
- No → Posts-only trial $n = 2x\ 2(1.65^2)(SD/d)^2$

Will the intervention loose its effect in an acceptable time for crossover?
- Yes
- No → Fully controlled trial $n = 2x\ 4(1.65^2)(sd/d)^2$

Is the measure reliable over the intervention / effect loss period?
- Yes
- No → Fully controlled crossover $n = 4(1.65^2)(sd/d)^2$

Are you limited by subjects or resources?
- Yes → Simple crossover $n = 2(1.65^2)(e/d)^2$
- No → Fully controlled crossover $n = 4(1.65^2)(sd/d)^2$

3A

Collect your data

Collate your data

Do you need to score your measures (i.e. 20 item scale to one number outcome)?

Yes → Folder containing raw measure data and single file containing single number outcomes, sharing the same name

No →

Do you have multiple measures from each participant?

Yes → Combine all relevant outcome numbers into a single file

No → Single file containing raw data

Prepare your data

Check to ensure your final datasheet is laid out as required for analysis (i.e. multiple rows per participant for hierarchical linear modlling)

Perform your analysis → **4A**

Think about results in theoretical context

Have you addressed your hypothesis /hypotheses?

Yes

No

Did an unforseen problem with your sample or methods stop you from doing do?

Yes

No

Have a little cry, then return to methodology development

Has something unexpected, yet interesting popped up?

Yes

No

Carry out additional analyses

Study Writeup

Have your study reviewed

Make changes where

Celebrate!

These steps certainly need filling in, but I lack the experience or citations to do so

4A → Is your data nested (i.e. repeated measures, or children sorted by school)?

- Yes → **13C**
- No → How many variables does the problem have?
  - One variable
  - Two variables → **5A**
  - More than two variables → **10A**

One variable → This will be a summary of univariate data properties.

How do you want to treat the variable with respect to scale of measurement?

Note: I have no idea how Bayesian statistics fit into all of these frameworks; they do not make assumptions like normality etc. I'd recommend consulting either Smithson, or his honours student looking at Bayesian stats, Cloudy, to get a better idea of how they could be woven in.

Nominal

What do you want to know about the distribution?
- Central Tendency → Mode
- Dispersion → Relative Frequencies
- Frequencies → Percentages, as well as the absolute frequencies.

Ordinal

What do you want to know about the distribution?
- Central Tendency → Percentages, median, absolute frequencies.
- Dispersion → Inter-Quartile Deviation
- Frequencies → Median, percentages, n-tiles

Interval

What do you want to know about the distribution?
- Symmetry → Skewness
- Dispersion → Standard deviation, coefficient of variation, range
- Central Tendency → Do you want to treat outlying cases differently from others?
  - Yes → Windzorised mean, trimmed mean, Hampel estimate of location, Biweight mean
  - No → What is the form of the distribution?
    - Symmetric → Mean
    - Skewed → Mean and Median
- Peakedness → Kurtosis
- Frequencies → Percentages, absolute frequencies, N-tiles
- Normality → D'Agnostino-Pearson K-squared, Jarque-Bera, Kolmogorov-Smirnov one sample test, Chi-square goodness of fit test

5A

How do you want to treat the variables with respect to scale of measurement?

- Both interval
- Both ordinal → **6A**
- Both nominal → **6A**
- One interval, One ordinal → **7A**
- One interval, One nominal → **8A**
- One ordinal One nominal → **9B**

Both interval → Is a distinction made between a dependent and an independent variable?
- Yes → **13A**
- No → You will look at the structure of your data

You will look at the structure of your data → Do you want to test for the presence of latent variables?
- Yes
- No → Do you want to test for structural equivalence?
  - Yes → Orthoganal Procrustes Rotation
  - No → Do you want to test whether the means on the two variables are equal?
    - No → Do you want to treat the relationship as linear?
    - Yes → Is the relationship between the two variables dependent or independent?

Is the relationship between the two variables dependent or independent?
- Independent → Independent samples t-test,
- Dependent → Paired samples t test, Wilcoxon

Do you want to treat the relationship as linear?
- No → There's nothing you can do
- Yes → What do you want to measure?

What do you want to measure?
- Agreement → Should there be a penalty if the variables don't have the same distribution?
  - Yes → Robinson's A and Intraclass Correlation Coefficient
  - No → Krippendorff's coefficient of agreement
- Covariation → How many of the variables are dichotomous?

How many of the variables are dichotomous?
- None → Pearson Product-Moment r; Do Fisher's r to Z transformation and refer the critical ratio of Z to a table of the unit normal curve.
- One → Is the dichotomous variable a collapsing of a continuous variable, and do you want to estimate what the correlation would be if it were continuous?
  - Yes → Biserial R
  - No → Pearson Product-Moment r (correlation coefficient)
- Both → Are the variables collapsings of continuous variables and do you want to estimate what the correlation would be if they were continuous?
  - Yes → Tetrachoric r
  - No → Pearson Product Moment r Phi Refer critical ratio for phi to a table of the unit normal curve

Do you have a theoretical reason to test for specific latent factors?
- Yes → Confirmatory Factor Analysis
  - Canonical rotation
  - N iterations
  - Orthogonal rotation
- No → Principal components analysis, exploratory factor analysis.

6A → Both nominal → Are the variables both two-point scales?
- Yes → What do you want to measure?
  - Symmetry → McNemar's Test of Symmetry
  - Covariation → Yule's Q, Phi, Fisher's exact test
- No → Do you want a statistic based on the number of cases in each category or on the number of cases in the modal categories?
  - Statistic based on number of cases in each category → Goodman and Kruskal's tau b
  - Statistic based on number of cases in modal categories → Asymmetric lambda

13B ← Yes — No → Are you trying to predict group membership?

6A → Both ordinal → Is a distinction made between a dependent and an independent variable?
- No → What do you want to measure?
  - Agreement → Do you want to look at observed vs expected outcome?
    - Yes → Chi square goodness of fit
    - No → Should there be a penalty if the variables do not have the same distribution?
      - Yes → Scott's Coefficient of Agreement
      - No → Cohen's Agreement Coefficients (Kappas)
  - Symmetry → McNemar's Test of Symmetry
  - Independence → Chi square test of independence
  - Covariation → Do you want a statistic based on the number of cases in each category or on the number of cases?
    - Number of cases in modal categories → Symmetric Lambda (AB)
    - Number of cases in each category → Do you want a statistic whose upper limit varies with the number of categories and whose upper limit may be less than one?
      - Yes → Contingency Coefficient, Pearson chi-square
      - No → Cramer's V, Pearson chi-square

- Yes → Is a distinction made between a dependent and an independent variable?
  - No → Are you comparing groups?
    - No → **7B**
    - Yes → Is it a between subjects design?
      - No → Wilcoxon or sign test
      - Yes → Do you have more than 20 people per group?
        - Yes → Mann-Whitney U
        - No → Rank Sum
  - Yes → Are you trying to predict group membership?
    - Yes → **13B**
    - No → Is there one dependent variable, or two?
      - One → Percentages, goodness of fit chi square tests, somer's d
      - Two → Percentages, contingency table, test of association chi square, somer's d

7A

One interval, One ordinal

Is the ordinal variable a two-point variable?

→ Yes → Is a distinction made between a dependent and an independent variable?
- Yes → Are you trying to predict group membership?
  - Yes → **13B**
  - No → Do you want to treat the relationship as linear?
    - Yes → Regression Coefficient
    - No → Curvilinear Regression Coefficients
- No → Do you want to test whether the means on the two variables are equal?
  - Yes → T Test
  - No → Do you want to treat the relationship as linear?
    - Yes → What do you want to measure?
    - No → There's nothing You can do

→ No → Do you want to treat the ordinal variable as if it were based on an underlying normally distributed interval variable?
- No → Do you want to treat the ordinal variable as if it were a monotonic transformation of an underlying interval variable?
  - Yes → Mayer and Robinson's Mu
  - No → There's nothing You can do
- Yes → Jaspen's Coefficient of Multiserial Correlation

What do you want to measure?
- Covariation → How many of the variables are dichotomous?
  - None → Pearson Product-Moment r
  - One → Is the dichotomous variable a collapsing of a continuous variable, and do you want to estimate what the correlation would be if it were continuous?
    - Yes → Biserial R
    - No → Pearson Product-Moment r
  - Both → Are the variables collapsings of continuous variables and do you want to estimate what the correlation would be if they were continuous?
    - Yes → Tetrachoric r
    - No → Pearson Product Moment r, Phi
- Agreement → Should there be a penalty if the variables don't have the same distribution?
  - Yes → Robinson's A and Intraclass Correlation Coefficient
  - No → Krippendorff

7B

What do you want to measure?
- Agreement → the data may be transformed to ranks and r or Krippendorff's r used.
- Structural equivalence → Multidimensional scaling, followed by generalised procrustes analysis
- Covariation → Do you want to treat the ranks of the ordered categories as interval scales?
  - Yes → Spearman's rho
  - No → Kendall's tau-a/b/c, Goodman and Kruskal's gamma, Kim's d

8A

One interval, One nominal

Are you trying to predict group membership? → No

13B ← Yes ← No ← Is the interval variable dependent?

No

Yes

Do you want a measure of the strength of relationship between the variables or a test of the statistical significance of differences between groups?

Test of Significance

Measure of Strength

Are you willing to assume that the interval scaled variable is normally distributed in the population?

Do you want to describe the relationship in your data or estimate it in the population which you've sampled?
- Describe → Eta Squared
- Estimate → Omega Squared / Intraclass Correlation Coefficient / Kelley's Epsilon Squared

Yes

No

Do you want to test the equality of means or of variances of the dependent variable for different categories of the independent variable?
- Variance → ANOVA, Bartlett's test
- Means → Do you want to assume homoscedasticity across levels of the independent variable?
  - Yes → F-test for significance
  - No → T test

Do you want to test the equality of variances of the dependent variable for different categories of the independent variable?
- Yes → ANOVA
- No → Is the nominal variable a two-point scale? → There's nothing you can do

Is the nominal variable a two-point variable?
- Yes → Is a distinction made between a dependent and an independent variable?
  - Yes → Do you want to treat the relationship as linear?
    - Yes → Linear regression
    - No → Curvilinear Regression Coefficients
  - No → Do you want to test whether the means on the two variables are equal?
    - Yes → T Test
    - No
- No → There's nothing you can do

Do you want to treat the relationship as linear?
- Yes → What do you want to measure?
  - Covariation → **9A**
  - Agreement → Should there be a penalty if the variables don't have the same distribution?
    - Yes → Robinson's A and Intraclass Correlation Coefficient
    - No → Krippendorff
- No → There's nothing you can do

9A

How many of the variables are dichotomous?

- **None** → Pearson Product-Moment r (correlation coefficient)

- **One** → Is the dichotomous variable a collapsing of a continuous variable, and do you want to estimate what the correlation would be if it were continuous?
  - **Yes** → Biserial R
  - **No** → Pearson Product-Moment r (correlation coefficient)

- **Both** → Are the variables collapsings of continuous variables and do you want to estimate what the correlation would be if they were continuous?
  - **Yes** → Tetrachoric r
  - **No** → Pearson Product Moment r, Phi

9B

Is a distinction made between a dependent and an independent variable?

- **Yes** → Are you trying to predict group membership?
  - **Yes** → **13B** → Is the ordinal variable dependent?
    - **Yes** → Are the cases (e.g. people) in one category of the nominal variable matched to the cases of each of the other categories of that variable?
      - **Yes** → Friedman Test
      - **No** → Freeman's Coefficient of Differentiation
    - **No** → Is the nominal variable two-point?
      - **Yes** → Are the cases (e.g., people) in one category of the nominal variable matched to the cases in the other category of that variable?
        - **Yes** → Sign and Wilcoxon Tests
        - **No** → Somer's d
      - **No** → Freeman's Coefficient of Differentiation
  - **No** → (to Is the ordinal variable dependent?)

- **No** → Is the nominal variable two-point?
  - **No** → There's nothing you can do
  - **Yes** → the nominal variable may be treated as ordinal or as interval
    - Treat the nominal variable as interval → **7A**
    - Treat the nominal variable as ordinal → **6A**

10A

Is a distinction made between dependent and independent variables?

Yes → Is there more than one dependent variable?

No → (down to "Do you want to measure agreement?")

Is there more than one dependent variable?
- Yes → Is there more than one independent variable?
- No → **11A**

Is there more than one independent variable?
- Yes → Do you want to treat the relationships among the variables as additive?
- No → **12B**

Do you want to treat the relationships among the variables as additive?
- Yes → **10B**
- No → **10C**

Do you want to measure agreement?
- Yes → How do you want to treat the variables with respect to scale of measurement?
- No → Do you want to predict group membership?

How do you want to treat the variables with respect to scale of measurement?
- All nominal → Light's Agreement Coefficient
- All ordinal → Kendall's Coefficient of Concordance
- All interval → Robinson's A and Intraclass Correlation Coefficient
- All other → There's nothing we can do

Do you want to predict group membership?
- No → Do you want to test whether the means (or proportions) on all variables are equal?
- Yes → Multiway frequency analysis

Do you want to test whether the means (or proportions) on all variables are equal?
- Yes → **12C**
- No → **12D**

10C

Do you want to treat all the dependent variables as interval or ratio scaled?
- Yes → Do you want to treat all the independent variables as nominal and test a set of prespecified relationships?
- No → There's nothing we can do

Do you want to treat all the independent variables as nominal and test a set of prespecified relationships?
- Yes → Multivariate Analysis of Variance
- No → Do you want to treat all the independent variables as nominal or ordinal and do you want to do an empirical search for strong relationships?

Do you want to treat all the independent variables as nominal or ordinal and do you want to do an empirical search for strong relationships?
- Yes → Multivariate Binary Segmentation Techniques
- No → There's nothing we can do

10B

Do you want to treat all the dependent and independent variables as interval?
- Yes → Do you want to treat all the relationships as linear?
- No → There's nothing we can do

Do you want to treat all the relationships as linear?
- No → There's nothing we can do
- Yes → (down)

Does the analysis include at least one intervening variable?
- Yes → Does your analysis include at least one latent (i.e., unmeasured) variable?
- No → Canonical Correlation (Wilks' lambda, Roy's greatest root criterion, and Pillai-Bartlett V and Chi_square)

Does your analysis include at least one latent (i.e., unmeasured) variable?
- Yes → There's nothing we can do
- No → Path Analysis

11A

Do you want to statistically remove the linear effects of one or more covariates from the dependent variable?

→ Yes → Do you want to treat the relationships involving the covariate(s) as additive?

→ Yes → Do you want to treat the dependent variable and the covariate(s) as interval, and the independent variable(s) as nominal?
- Yes → Covariance Analysis
- No → Heirarchical regression

→ No → There's nothing we can do

→ No → Do you want to treat the relationships among the variables as additive?

→ Yes → How do you want to treat the dependent variable with respect to scale of measurement?
- Nominal → Do you want to treat all independent variables as interval?
  - No
  - Yes
- Ordinal → There's nothing we can do
- Interval

Do you want to treat the relationships among the independent variables as linear?
- Yes → Multiple Discriminant Function
- No → There's nothing we can do

Is the dependent variable two-point?
→ Yes
→ No → Do you want to treat all of the independent variables as nominal?
- Yes → Multidimensional Contingency Table Analysis
- No → There's nothing we can do

Is there a very high proportion in one category of the dependent variable (e.g., 90%)?
- Yes → Dummy Variable Regression
- No → Do you want to assume homoscedasticity?
  - Yes → Multiple Classification Analysis or Dummy Variable Regression
  - No → Dummy Variable Regression

→ No → Do you want to do an empirical search for strong relationships or to test a set of prespecified relationships?

→ Search → How do you want to treat the variables with respect to scale of measurement?
- Dependent: Nominal or Interval Independent: Nominal or Ordinal → Binary Segmentation Techniques
- Other → There's nothing we can do

→ Test → Do you want to treat the dependent variable as ordinal?
- Yes
- No → Do you want to treat all the independent variables as nominal?
  - Yes → Multidimensional Contingency Table Analysis based on the Cumulative Logistic Distribution
  - No → There's nothing we can do

Do you want to treat any of the independent variables as ordinal? → **12A**

12A

Do you want to treat any of the independent variables as ordinal?
- Yes → There is nothing we can do
- No → Do you want to treat the dependent variable as interval and all the independent variables as nominal and assume homoscedasticity?
  - Yes → Analysis of Variance
  - No →

Do you want to treat all the variables as nominal?
- Yes → Multidimensional Contingency Table Analysis
- No → Do you want to treat the dependent variable as interval and all independent variables as nominal?
  - Yes → Analysis of variance using weighted least squares
  - No → Multidimensional contingency table analysis using weighted least squares

12B

Do you want to treat the independent variable as nominally scaled and all the dependent variables as interval scaled?
- Yes → Do you want to test only whether the vectors of means are equal for all categories of the independent variables?
  - Yes → Multivariate Analysis of Variance
  - No → Profile Analysis
- No → There is nothing we can do

12C

Are all the variables two-point?
- Yes → Cochran's Q
- No → Analysis of Variance with Repeated Measures

12D

Do you want to treat the relationships among the variables as additive?

Do you want to analyze patterns existing among variables or among individual cases (e.g., persons)?
- Yes
- No

Do you want to treat all of the variables as nominal?
- Yes → Multidimensional Contingency Table Analysis
- No → There is nothing we can do

Variables → Do you have two or more sets of variables and do you want to measure the strength of the association between those sets?
- No → Q-type Factor Analysis
- Yes → Do you want to treat the variables as measured on interval scales and relationships among them as linear?
  - Yes → Canonical Correlation
  - No → There is nothing we can do

Cases → Do you want to treat the variables as measured on interval scales and relationships among them as linear?
- Yes → Q-type Factor Analysis
- No → Clustering Techniques

13A

Do you want to treat the relationship as linear?
- Yes → Regression, Coefficient, F test
- No → Curvilinear, Regression, Coefficient, F test

13B

Is there one or multiple dependent variables (groups)?
- One → Are there continuous, discrete, or continuous and discreet predictors (of group membership)?
  - Multiple continuous → Are there covariates?
    - Yes → Sequential one-way discriminant function
    - No → One-way discriminant function
  - Multiple discrete → Multiway frequency analysis (logit)
  - Multiple discrete and continuous → Are there covariates?
    - Yes → Sequential logistic regression
    - No → Logistic regression
- Multiple → Are there covariates?
  - Yes → Sequential factorial discriminant function
  - No → Factorial discriminant function

13C

Is your data nested by time, or by another grouping variable?

- Time → Is time your dependent variable?
  - Yes → Do you have one or more independent variables?
    - Yes → Survival analysis with predictors
    - No → Survival analysis with life tables
  - No → Is time your only independent variable?
    - Yes → Forecasting time series analysis
    - No → Intervention time series analysis

- Another grouping variable → Do you have covariates?
  - Yes → Do you have nominal or ordinal level variables?
    - Yes → Is the relationship between your variables linear?
      - Yes → Hierarchical multilevel linear modelling with dummy coding
      - No → Non-linear hierarchical multilevel modelling with dummy coding
    - No → Is the relationship between your variables linear?
      - Yes → Hierarchical multilevel linear modelling
      - No → Non-linear hierarchical multilevel modelling
  - No → Do you have nominal or ordinal level variables?
    - Yes → Is the relationship between your variables linear?
      - Yes → Multilevel linear modelling with dummy coding
      - No → Non-linear multilevel modelling with dummy coding
    - No → Is the relationship between your variables linear?
      - Yes → Multilevel linear modelling
      - No → Non-linear multilevel modelling