**Categorical ↔ Categorical**

Are X groups 1 or 2 more likely to be in Y groups A or B?

- **Significance:** Chi Square $\chi^2$
- **Strength:** Cramer’s V

**Are (1) H.S. or (2) college students more likely to prefer (A) Math or (B) Science?**

![Bar Chart](image)

**Numeric → Categorical**

Are higher/lower values of X more likely to be in Y groups A or B?

- **Significance:** Logistic Regression
- **Strength:** Odds Ratio

**Is (X) age related to the likelihood of studying at the (A) library or (B) home?**

![Area Chart](image)

**Categorical → Numeric**

Do X groups 1 or 2 have higher/ lower values of Y overall?

- **Significance:** t-test / ANOVA
- **Strength:** Cohen’s D

**Do (1) Mason graduates and (2) UVA graduates get different (Y) numbers of job interviews?**

![Box Plot](image)

**Numeric ↔ Numeric**

Are higher/lower values of X associated with higher/lower values of Y?

- **Significance:** Correlation
- **Strength:** Pearson’s $r$

**Is the (X) number of class sessions attended related to (Y) scores on the final exam?**

![Scatter Plot](image)
Categorical ↔ Categorical

Compare or Examine Associations with Group Membership

Chi-Square ($\chi^2$) examines a contingency table (or crosstabulation) of nominal variables (ideally 2-5 groups), comparing the frequencies expected if there were no relationship to actual / observed frequencies.

See also:

- Binomial Test for 1 Proportion (1 variable with 2 levels)
- Fisher’s Exact for 2x2 tables, allows cells having < 5 expected obs
- McNemar’s for repeated measures / paired samples
- Kendall’s tau for Ordinal variables
- Kruskal Gamma for Ordinal variables with many ties
- Log-linear for 3+ Nominal variables

Yates correction is used when cells have < 5 observed values

Categorical → Numeric

Compare Group Means

Analysis of Variance (ANOVA) compares means from independent groups formed by 1+ Nominal IVs. If any IV has 3+ levels, post-hoc tests are used to test pairwise differences. aka Between Subjects ANOVA

See also:

- t-Test ANOVA for just 1 IV having only 2 groups
- Z-Test t-Test for known S and/or large samples
- Kruskal-Wallis H / Mann-Whitney U ANOVA / t-Test for an Ordinal or Non-Parametric DV
- Factorial ANOVA name when testing interactions among 2+ IVs
- MANCOVA Multiple (2+) correlated DVs and a numeric Covariate
- Within Subjects ANOVA/t-Test Compare a change or difference in paired values; aka Repeated Measures, Dependent Samples

Numeric → Categorical

Understand or Predict Group Membership

Logistic Regression models the relationship between a Binary DV and one or more IVs. A linear combination of the values of all IVs produces a predicted likelihood (in log-odds) of being in one group over the other.

See also:

- Multinomial / Ordinal Logistic Regression for nominal / ordinal DVs with 3+ values
- Linear Probability Model name when using Linear Regression to predict probabilities (also a Binary DV)
- Discriminant Functions Analysis (DFA) for a linear combination of Gaussian variables that best discriminates among 3+ groups
- Log-linear Analysis for multiple nominal IVs only; better for testing interactions between the factors

Numeric → Numeric

Understand or Predict Values

[Pearson’s] Correlation evaluates the strength of a linear association, or shared variance, between 2 numeric variables. For Non-Parametric or Ordinal variables, use Spearman’s [Rank-Order] Correlation.

See also:

- [Linear] Regression creates a linear combination/equation of the values of all IVs to produce a predicted value of the DV
- General[ized] Linear Models extension of Regression to allow for non-normal distributions of the response variable or residuals
- Hierarchical Linear Modeling (HLM) when observations have variables for shared group characteristics (e.g., students within schools)

In Regression, nominal variables can be included as numeric IVs when expressed as multiple binary variables (through Dummy Coding).