Tests

Monotonic
- Relationship b/w two vars that need not move at constant rate.
- Spearman corr. coef. uses rank value
  - ↓
  - no assumption on distri

Linear
const-rate

Pearson corr. coef.
numerical feature, dependent var
- ↓
- assumes normal distri
- more sensitive to outliers

degrees of freedom
no of observations − no of parameters being estimated

[OR]

no of independent variables

parametric
assumes underlying Gaussian, data has equal variances

non-parametric
no assumptions

distribution-agnostic

significance tests: check if samples from same distri

t-test	Mann-Whitney U test
paired t-test	Wilcoxon signed rank test
ANOVA	Kruskal Wallis H test
repeated measures ANOVA	Friedman

Parametric independence tests

provide p-value = chance of getting observed difference at random

t-test
H_0: mean of populations same
assume equal variance
dof = n−1

Eg: Mean score of students in two electives same set of students → paired

paired t-test
when each individual from one population appears in the other

t-test
two groups
small samples
independent / dependent (paired) samples

t-statistic: measures difference in means relative to variation within groups

F-test
two or more groups
+
only independent

F-statistic: compares variation within groups to variation between groups

Independence for categorical input var
Categorical → categorical: Pearson's χ^2 test

$E = \frac{(\text{row tot} \times \text{col tot})}{\text{tot}}$

$\chi^2 = \sum \frac{(O-E)^2}{E}$

$dof = (m-1)(n-1)$

Categorical → numerical: ANOVA

" " but ANOVA assumptions ✗ : Kruskal Wallis H test

input independent from output → remove input

sign test
H_0: difference b/w medians is zero
paired observations

Wald-Wolfowitz runs test
H_0: given sequence is random