COURSE GLOSSARY

Hypothesis Testing in Python

Alternative hypothesis: The competing hypothesis (denoted HA) that expresses a new claim or effect we seek evidence for and is true when the null hypothesis is not

ANOVA (Analysis of Variance): A statistical test that assesses whether there are any statistically significant differences among the means of three or more groups by comparing between-group and within-group variability

Bonferroni correction: A simple multiple-comparison adjustment that reduces the chance of false positives by multiplying each p-value (or dividing alpha) by the number of tests, making it more conservative

Bootstrap distribution: An empirical sampling distribution generated by repeatedly resampling the observed data with replacement and recomputing a statistic to estimate its variability and standard error

Chi-square goodness-of-fit test: A one-sample chi-square test that assesses whether the observed frequency distribution of a categorical variable matches a specified hypothesized distribution

Chi-square test of independence: A test that evaluates whether two categorical variables are statistically independent by comparing observed cell counts in a contingency table to the counts expected under independence

Confidence interval: A range of values, computed from sample data, that quantifies plausible values for a population parameter at a specified confidence level (e.g., 95%), typically equal to 1 minus the chosen alpha

Hallucination: When a model produces confident but incorrect or fabricated information, often due to gaps or biases in its training data or reasoning process

Hallucination: When a model produces confident but incorrect or fabricated information, often due to gaps or biases in its training data or reasoning process

Kruskal–Wallis test: A non-parametric analogue of ANOVA that tests whether three or more independent groups come from the same distribution by comparing the rank sums across groups

Non-parametric test: A hypothesis test that does not assume a specific parametric form (such as normality) for the population distribution and often operates on ranks rather than raw values

Null hypothesis: The default or status-quo hypothesis (denoted H0) that specifies a particular value or relationship for a population parameter and is assumed true until evidence suggests otherwise

p-value: The probability, assuming the null hypothesis is true, of observing a test statistic at least as extreme as the one obtained from the sample

Paired vs unpaired t-test: Paired t-tests compare two related measurements taken on the same units (e.g., before/after) by testing the mean of their differences, whereas unpaired (independent) t-tests compare means from two separate independent groups

Pooled proportion: A weighted average of two sample proportions used under the null hypothesis in two-sample proportion tests to estimate a common population proportion for calculating the standard error

Proportion (p and p-hat): p denotes a population proportion (unknown), and $\hat{p}$ (p-hat) denotes the observed sample proportion used to estimate or test hypotheses about p

Significance level (alpha): A pre-chosen threshold (commonly 0.05, 0.01, or 0.10) that specifies the maximum acceptable probability of a Type I error and is used to decide whether to reject the null hypothesis

Standard error: The standard deviation of a sampling distribution of a statistic, estimating how much that statistic would vary from sample to sample

t-distribution: A family of probability distributions like the normal but with heavier tails, used for test statistics when the population standard deviation is unknown and must be estimated from the sample

Test statistic: A numeric summary computed from sample data (e.g., a z-score or t-statistic) that quantifies how far the observed sample result is from the value expected under the null hypothesis

Type I error (false positive): The error of rejecting the null hypothesis when it is actually true

Type II error (false negative): The error of failing to reject the null hypothesis when the alternative hypothesis is actually true

Wilcoxon signed-rank test: A non-parametric paired test that ranks absolute differences between paired observations, uses the signs of those differences, and evaluates whether the median difference is zero

Wilcoxon–Mann–Whitney (rank-sum) test: A non-parametric test for comparing two independent samples that assesses whether values from one group tend to be larger than values from the other by comparing ranks

z-score / z-distribution: A standardized test statistic that measures the number of standard errors a sample statistic is from the hypothesized population value, referenced to the standard normal (z) distribution with mean 0 and standard deviation 1