BST281: Genomic Data Manipulation, Spring 2019

Wednesday 09: Inference and hypothesis testing

Test statistics summarize data into a single value that captures one or more relevant aspects.

Tests consider a null hypothesis, usually that there is no effect/association/bias/etc.

The test statistic should have some known distribution, the null distribution, when the null hypothesis is true.

P-value: probability of observing a test statistic at least as extreme if null hypothesis true (from null distribution).

Null hypothesis rejected if p-value < critical threshold α, calling the result statistically significant.

For normal distributions with known standard deviation, z-test statistic z = (x - ) / σ is appropriate.

When the standard deviation is unknown, t = (x - ) / follows the Student's t-distribution; the t-test.

One-sided/one-tailed tests consider only values in one direction to be "extreme."

E.g. asking whether a gene's expression is *greater* than another.

Two-sided/two-tailed tests consider both directions; whether a gene's expression is *different from* another.

Non-parametric tests ignore the shape of the distribution, often using a rank transformation.

But the cost is reduced sensitivity - they are discarding potentially useful information!

E.g. Mann-Whitney U test, also known as the Wilcoxon rank-sum test.

Permutation test is used when the null distribution is not nice.

Generate an empirical null distribution directly by calculating test statistic repeatedly in permuted data.

Performance evaluation: perform a test where true values are known (gold standard).

Error rates: false positive rate (Type I errors), and false negative rate (Type II errors).

How well the test calls positives: power, precision, and specificity.

Precision/recall plots and ROC plots capture how well a test does, independent of its parameters.

Area Under the Curve (AUC) commonly used to assess performance (0.5 is random, 1 is perfect).

Multiple hypothesis testing can lead to many false positives.

Bonferroni correction: conservatively only call α false positives among all tests.

False Discovery Rate (FDR) correction: instead control the fraction of false positives among all positives.

# Textbooks

Hypothesis testing: Pagano and Gauvreau, Chapters 10.1-5

T-tests: Pagano and Gauvreau, Chapters 11.1-2

Wilcoxon: Pagano and Gauvreau, Chapters 13.2-4

ANOVA: Pagano and Gauvreau, Chapters 12.1-2

Performance evaluation: Pagano and Gauvreau, Chapters 6.4

# Literature

[Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Subramanian, PNAS 2005](https://www.ncbi.nlm.nih.gov/pubmed/16199517)

[Fewer permutations, more accurate P-values. Knijnenburg, Bioinformatics 2009](https://www.ncbi.nlm.nih.gov/pubmed/19477983)