

Anon

Prof. Jack Hester

CEPC 0904

July 21, 2022

Stats / R Studio One Pager

Would use a statistical test when comparing numerical data. To see if there is a difference, how different they are, and what predictions we can make. Most of the ones we will learn will be about if there's a difference, and not how much of a difference there is.

How to apply statistical tests: make a hypothesis (with a null H_0 (no difference) and alternative H_a (significant / there is a difference)), test them with the appropriate test.

All tests are predictors, and independent variables are categorical (label, not number), all outcomes are numerical.

One sample test: comparing a sample to the expected mean.

If you don't know the SD then t test, If $n > 30$ z test, if $n < 30$ t test

Two sample test: comparing two sample means with no relationship

If you don't know the SD then t test, If $n > 30$ z test, if $n < 30$ t test

Two paired sample test: comparing two samples with a relationship

If you don't know the SD then t test, If $n > 30$ z test, if $n < 30$ t test

More than two group comparison: multiple categories; ANOVA test

FOR ALL: Assume that the data is in a normal distribution for the tests to work.

Hypothesis testing using R: test with the appropriate test using null, alternative hypotheses

Significance level (alpha): the probability of rejecting the null hypothesis when it's true

Distribution: representation of probabilities of where a value will lie

P-value: probability of getting a result at least as extreme as the observed result

Critical value: cutoff value in dist. Significance level gives this (from table of calc.)

Confidence interval: range of values which you expect the pop. mean to lie in

Quantile: % of values below a certain level.

Tests in R:

Z test: `z.test(x, y = NULL, alternative = "one/two.sided", mu = 0, sigma.x = NULL, sigma.y = NULL, conf.level = 0.95)`

T test: `t.test(alternative = c("two.sided", "less", "greater"), mu = 0, paired = FALSE, var.equal = FALSE, conf.level = 0.95)`

Applications: understanding data, checking assumptions, look for outliers