Non-parametric statistics

Anthony J. Evans
Professor of Economics, ESCP Europe
www.anthonyjevans.com

Introduction

- So far the data we've looked at has had parameters
  - E.g. mean and variance
- We've used these parameters to utilise a distribution
- Parametric tests assume that the data belongs to some sort of distribution
- Nonparametric statistics allows us to perform tests with an unspecified distribution

- If the underlying assumptions are correct, parametric tests will have more **power**
  - Power = P(Reject H_0 | H_1 is true)
  - Or 1-P(Type II error)
- Nonparametric tests can be more **robust** and allow for new data to be incorporated
  - Robustness = less affected by extreme observations

When to use a non-parametric test

- If we're not sure of the underlying distribution
- When the variables are discrete (as opposed to continuous)

Purpose	Parametric method	Non-parametric equivalent
Compare 2 paired groups	Paired T-test	Wilcoxon signed ranks test
Compare 2 independent samples	Unpaired T-test	Mann-Whitney U
Compare 3+ independent samples	ANOVA/regression	Kruskal-Wallis

See http://www.bristol.ac.uk/medical-school/media/rms/red/rank_based_non_parametric_tests.html

A sign test example

- We are interested in whether the hind leg and forelegs of deer are the same length

Deer	Hind leg length (cm)	Foreleg length (cm)	Difference
1	142	138	+
2	140	136	+
3	144	147	-
4	144	139	+
5	142	143	-
6	146	141	+
7	149	143	+
8	150	145	+
9	142	136	+
10	148	146	+

Example taken from Zar, Jerold H. (1999), "Chapter 24: More on Dichotomous Variables", *Biostatistical Analysis* (Fourth ed.), Prentice-Hall, pp. 516-570

A sign test example

- If they are the same, we should expect as many instances where one is bigger than the other and vice versa.
  - H_0: Hind leg = foreleg
  - H_1: Hind leg ≠ foreleg
- We should expect 5 +'s and 5 -'s
- We observe 8 +'s and 2 -'s
- How likely is this?
- Use a Binomial test to find that:
  -
  -
  -
  -
-
-

Pearson's chi-squared test

$$\chi_c{}^2 = \Sigma \frac{(Oi - Ei)2}{E_i}$$

- Used to test whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories
- Events must be mutually exclusive and collectively exhaustive
- Suitable for
  - Categorical variables *
  - Unpaired data
  - Large samples
- * Potential examples of categorical variables (Mosteller and Tukey 1977):
  - Names
  - Grades (ordered labels like beginner, intermediate, advanced)
  - Ranks (orders with 1 being the smallest or largest, 2 the next smallest or largest, and so on)
  - Counted fractions (bound by 0 and 1)
  - Counts (non-negative integers)
  - Amounts (non-negative real numbers)
  - Balances (any real number)

Student genders

- We expect a PhD programme to have an equal number of male and female students. However, over the last ten years there have been 80 females and 40 males. Is this a significant departure from expectation?

	Female	Male	Total
Observed (O)	80	40	120
Expected (E)	60	60	120
(O – E)	20	-20	0
$(O - E)^2$	400	400	
$(O - E)^2 / E$	6.67	6.67	$x^2 = 13.34$

We can use a chi square table to find a critical value of **3.84**
(Where p=0.5 and n-1 degrees of freedom).
We have a statistically significant finding that the 1:1 ratio is not being met.

Source: http://archive.bio.ed.ac.uk/jdeacon/statistics/tress9.html
Note that n refers to categories so we have 2-1 = 1 degree of freedom.

7

Chi-squared and significance testing

- Chi Square is employed to test the difference between an actual sample and another hypothetical or previously established distribution such as that which may be expected due to chance or probability
- The procedure for a chi-square test is similar to what we've used previously
  - Calculate the test statistic and use a probability table to find the p value
- The key difference is that the "distribution" is based on expected frequency. There is no underlying assumptions about the distributions parameters

- We only use non-parametric techniques for significance tests, we can't use them for estimation.

8

Examples of nonparametric tests

- Analysis of similarities
- Anderson-Darling test: tests whether a sample is drawn from a given distribution
- Statistical bootstrap methods: estimates the accuracy/sampling distribution of a statistic
- Cochran's Q: tests whether *k* treatments in randomized block designs with 0/1 outcomes have identical effects
- Cohen's kappa: measures inter-rater agreement for categorical items
- Friedman two-way analysis of variance by ranks: tests whether *k* treatments in randomized block designs have identical effects
- Kaplan-Meier: estimates the survival function from lifetime data, modeling censoring
- Kendall's tau: measures statistical dependence between two variables
- Kendall's W: a measure between 0 and 1 of inter-rater agreement
- Kolmogorov-Smirnov test: tests whether a sample is drawn from a given distribution, or whether two samples are drawn from the same distribution
- Kruskal-Wallis one-way analysis of variance by ranks: tests whether > 2 independent samples are drawn from the same distribution
- Kuiper's test: tests whether a sample is drawn from a given distribution, sensitive to cyclic variations such as day of the week
- Logrank test: compares survival distributions of two right-skewed, censored samples
- Mann-Whitney U or Wilcoxon rank sum test: tests whether two samples are drawn from the same distribution, as compared to a given alternative hypothesis.
- McNemar's test: tests whether, in 2 × 2 contingency tables with a dichotomous trait and matched pairs of subjects, row and column marginal frequencies are equal
- Median test: tests whether two samples are drawn from distributions with equal medians
- Pitman's permutation test: a statistical significance test that yields exact *p* values by examining all possible rearrangements of labels
- Rank products: detects differentially expressed genes in replicated microarray experiments
- Siegel-Tukey test: tests for differences in scale between two groups
- Sign test: tests whether matched pair samples are drawn from distributions with equal medians
- Spearman's rank correlation coefficient: measures statistical dependence between two variables using a monotonic function
- Squared ranks test: tests equality of variances in two or more samples
- Tukey-Duckworth test: tests equality of two distributions by using ranks
- Wald-Wolfowitz runs test: tests whether the elements of a sequence are mutually independent/random
- Wilcoxon signed-rank test: tests whether matched pair samples are drawn from populations with different mean ranks

Note: list and links taken from Wikipedia

Solutions

A sign test example

- If they are the same, we should expect as many instances where one is bigger than the other and vice versa.
  - H_0: Hind leg = foreleg
  - H_1: Hind leg ≠ foreleg

- We should expect 5 +'s and 5 -'s
- We observe 8 +'s and 2 -'s
- How likely is this?

- Use a Binomial test to find that:
  - P(8) + P(9) + P(10)
  - = 0.04395 + 0.00977 + 0.00098
  - P(0) + P(1) + P(2) *Because it's a two tailed test*
  - = 0.00098 + 0.00977 + 0.04395

- Thus ***p* = 0.109375**
- Since this is above 0.05 we fail to reject H_0

- This presentation forms part of a free, online course on analytics

- http://econ.anthonyjevans.com/courses/analytics/