Permutation Tests
&
False Detection Rate

1

Permutation Tests

- Computer-intensive methods for hypothesis testing
- Used when distribution of the test statistic (under the null hypothesis) is unknown
- Permutation tests maintain the Type I error level without any large sample approximations/assumptions

2

Example - HPV vaccine trial

- 200 uninfected women are randomly assigned 1:1 to HPV vaccine or placebo (i.e., 100 to each group)
- After 1 year subjects are tested for HPV infection (yes/no)

Scientific Question:

Is the risk of infection the same or different in the two groups?

Restate scientific question as statistical hypotheses:

H_0: $p_v = p_p$

H_a: $p_v < p_p$

where p_V = Probability of infection in the vaccine group

p_p = Probability of infection in the placebo group

3

Example - HPV vaccine trial

Results:

	Vaccine	Placebo	Total
HPV+	20	40	60
HPV-	80	60	140
	100	100	200

The overall infection rate is 30%, but we observe 20% and 40% for vaccine and placebo, respectively. What if we repeated the experiment … would we see similar results? We know that sample results are variable. Could the difference go the other way? Could a difference this large be due to chance alone?

4

Example - HPV vaccine trial

We first need a way of summarizing the difference in the infection probabilities between vaccine and placebo groups. A useful summary has these features:

- Summarize the differences between the groups in a single number.

  Example ⇒ $p_v - p_p$
- One particular value (say, 0) of the summary corresponds to the null hypothesis being exactly true.

  Example ⇒ $p_v - p_p = 0$
- We expect values near 0 if the null hypothesis is true; we expect values far from 0 if the null hypothesis is false.
- But how near is near? How far is far?

5

Exercise

1. Name at least two other possible summary statistics that could be used to test the hypothesis H_0: $p_v = p_p$

6

Example - HPV vaccine trial

We need to figure out what sort of distribution of values we would see for our summary statistic if the experiment were repeated many times and the null hypothesis were true.

Imagine the following experiment:

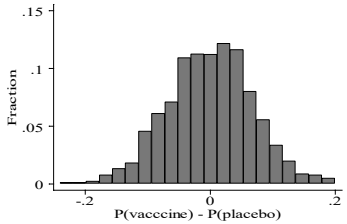
- make up a deck of 200 cards
- mark the word “HPV+" on 60 of them
- shuffle and deal two groups of 100
- form a 2 x 2 table from the results
- calculate your summary statistic
- repeat many times
- plot the results

This experiment should give us an idea of what we expect to see **if the null hypothesis is true**.

7

Example - HPV vaccine trial

Here is the distribution of differences $p_v - p_p$ that we might expect to see if the null hypothesis is true:



Summarize the results by reporting what proportion of the simulated results are as “extreme” or more so than the observed result (p value).

⇒ **only 3/2000 simulated differences were more extreme than the observed difference of -0.2**

⇒ **p = .0015**

8

Example - HPV vaccine trial

Summary:

We have constructed a valid test of the hypothesis, H_0: $p_V = p_P$, using a **randomization test**. There are four steps involved:

1. Restate the scientific question as statistical hypotheses
2. Choose (any) reasonable summary statistic that quantifies deviations from the null hypothesis
3. Resample data assuming the null hypothesis is true and compute the summary statistic for each resampled data set.
4. Compare the observed value of the summary statistic to the null distribution generated in Step 3.

Exercise

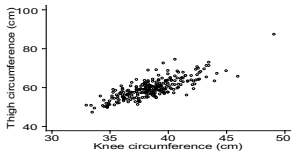
2. The following table provides the data from Pauling's vitamin C study and below that is a stem and leaf plot of the randomization distribution (n = 1000) of the relative risk (RR)from the trial. Find the p-value for testing Ho: RR = 1.

	Cold - Y	Cold - N	Total
Vitamin C	17	122	139
Placebo	31	109	140
Total	48	231	279

Reading a stem and leaf plot:
Each entry gives a RR to 2 digits; the first digit is to the left of the |; the numbers to the right of the | are the second digits. For example, the smallest RR is .34 and the five next smallest are .46.

```
.3 | 4
.4 | 66666
.5 | 0000055555555555555
.6 | 00000000000000000000000006666666666666666666666666666666666666
.7 | 22222222222222222222222222222222222222222222222222222222222222222+63
.8 | 55555555555555555555555555555555555555555555555555555555555555555555+12
.9 | 33333333333333333333333333333333333333333333333333333333333333333333+28
1.0 | 11111111111111111111111111111111111111111111111111111111111111111111+152
1.1 | 99999999999999999999999999999999999999999999999999999999999999999999+15
1.2 | 99999999999999999999999999999999999999999999999999999999999999999999+9
1.3 |  etc
```


Permutation Test for Correlation



We can summarize the "strength of association" between two variables using the correlation coefficient (ρ). The sample estimate is

$$r = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{X_i - \bar{X}}{s_X}\right)\left(\frac{Y_i - \bar{Y}}{s_Y}\right)$$

- No distinction between x and y.
- The correlation is constrained: $-1 \le R \le +1$
- $|R| = 1$ means "perfect linear relationship"
- Correlation doesn't change if there is a linear change in units.

Exercise

3. Describe a permutation test to test the hypothesis Ho: ρ = 0 (no association). What test statistic would you use? How would you find the permutation distribution of that test statistic when Ho is true?

Permutation Tests - Summary

- Useful when we can do resampling under the null hypothesis
- Permutation samples are drawn without replacement
- If the sample size is small, you can enumerate all possible permutations (permutation test)
- If sample size is large, generate a random sample of permutations (randomization test).
- Fewer assumptions than e.g. t-test (i.e., no assumption about skewness or normality of underlying distribution)
- Many standard nonparametric methods (e.g., Wilcoxon Rank Sum Test) are permutation tests based on ranks.
- Good Reference:
  Manly (2007). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman & Hall/CRC.

13

False Discovery Rate

For some studies, answering the scientific question of interest may require testing hundred, thousands, or millions of hypotheses. This is especially true of genetics.

E.g. Hedenfalk et al (2001) screened 3226 genes using microarrays to find differential expression between BRCA-1 and BRCA-2 mutation positive tumors.

Issue: If a traditional hypothesis testing approach is taken and we conduct 3226 tests at the 0.05 level, then we expect (up to) 161 false positive findings. Unfortunately, they are not labeled as such!

Traditional Solution (Bonferroni correction): If we conduct each test at an $\alpha = .05/3226 = .000015$ level then the probability of 1 or more false positive findings will be ~0.05. But, … with such a stringent α level we are likely to miss many true positive results.

New Solution: Don't try to eliminate false positives … control them

14

False Discovery Rate

	Reject null	Fail to reject	
Null true	F	m_0-F	m_0
Alternative true	T	m_1-T	m_1
	S	m-S	m

- false positive rate = F/ m_0
- false discovery rate = F/S

Idea: Control the false discovery rate (q-value) instead of the false positive rate (p-value)

15

False Discovery Rate

E.g. Hedenfalk data

- Order the 3170 p-values (56 genes were excluded from this analysis): p_i , i = 1 … 3170
- Pick a p-value cutoff, say α; reject H_o for all $p_i < \alpha$.

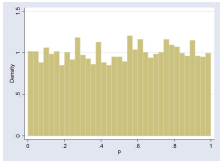
Q: What is the FDR associated with this choice of α?

- FDR = F/S
- S = #{$p_i < \alpha$}
- F = α * m_0
- FDR = q-value = α * m_0 / #{$p_i < \alpha$}
- I know S, I know α, what is m_0?

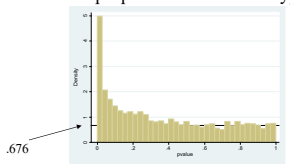
16

False Discovery Rate

Distribution of 3170 p-values when all null hypotheses are true



Distribution of 3170 p-values from Hedenfalk et al. Height of the line gives estimated proportion of true null hyptheses.



$$m_0(\lambda) = \frac{\#\{p_i > \lambda; i = 1...m\}}{(1-\lambda)}$$

17

False Discovery Rate

- $q(\alpha) = \alpha * m_0(\lambda) / \#\{p_i < \alpha\}$

  (technically $q(\alpha) = min_{t \geq \alpha} q(t)$)
- Program QVALUE (http://genomine.org/qvalue/) or p.adjust() in R
- Eg. Hedenfalk et al. ($m_0(.5) = 2143$)

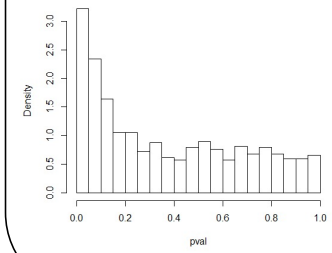
q	α	No. differentially expressed $\#\{p_i < \alpha\}$	expected false pos
.01	.0000126	5	0
.05	.00373	160	8
.10	.0148	317	32

- Using traditional methods Hedenfalk et al. concluded 9-11 genes were differentially expressed.

18

Exercise

4. Here is a plot of 1000 p-values. Use these data to estimate (by eye) the number of true null hypotheses (m_0) and then use that to complete the table below (assume $\alpha = .05$)



	Reject null	Fail to reject	
Null true			
Alternative true			
	161	839	1000

19