Permutation Tests & False Detection Rate

Session 10

Module 1 Probability & Statistical Inference

The Summer Institutes
DEPARTMENT OF BIOSTATISTICS
SCHOOL OF PUBLIC HEALTH
UNIVERSITY of WASHINGTON

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

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

EXAMPLE

HPV Vaccine Trial

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

	Vaccine	Placebo	
HPV+	20	40	60
HPV−	80	60	140
	100	100	**200**

The overall infection rate is 30%, but we observe

- 20% for vaccine
- 40% for placebo

What if we repeated the experiment ... would we see similar results? Could a difference this large be due to chance alone?

EXAMPLE

HPV Vaccine Trial

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

We need a way of summarizing the difference in infection probabilities between vaccine and placebo groups.

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 close is close? How far is far?

EXAMPLE

HPV Vaccine Trial

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

We need to figure out what 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:

1. make up a deck of 200 cards
2. mark the word “HPV+" on 60 of them
3. shuffle and deal two groups of 100
4. form a 2 x 2 table from the results
5. calculate your summary statistic
6. repeat #s 3-5 many times
7. plot the results

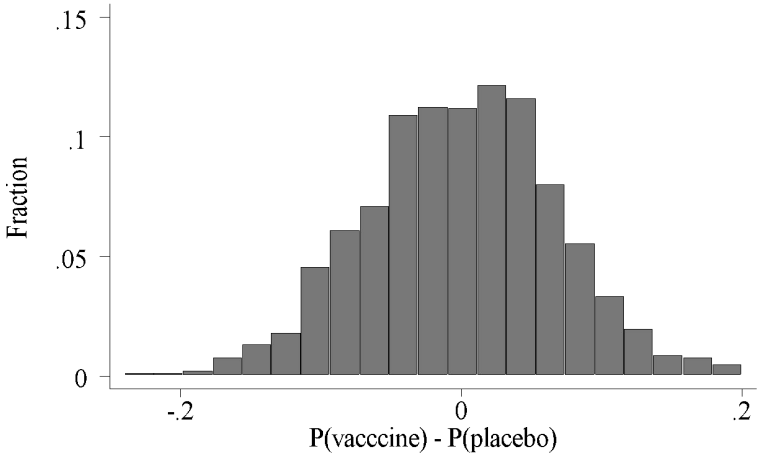
This experiment should give us an idea of what we expect to see **if the null hypothesis is true.**

EXAMPLE

HPV Vaccine Trial

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

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



Here is the distribution of differences $p_v - p_p$ that we might expect to see, assuming the null hypothesis is true.

- only 3/2000 simulated differences were more extreme than the observed difference of -0.2
- p = 0.0015

EXAMPLE

HPV Vaccine Trial

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

<u>Summary</u> We have answered our scientific question by using a permutation test.

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.

Summary: Permutation Tests

- Useful when we can do resampling under the null hypothesis
- Fewer assumptions than e.g. t-test (i.e., no assumption about skewness or Normality of underlying distribution)
- 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)
- Permutation samples are drawn without replacement
- 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.

Break #1

**Pause the video,
take a break, stretch,
then continue on!**

Image Credit: indg0.com

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.

> 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 $\alpha = 0.05/3226 = 0.000015$ 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 eliminate false positives … control them.**

False Discovery Rate

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

	Reject null	Fail to reject null	
Null true	F	m_0-F	m_0
Null not true (Alternative true)	T	m_1-T	m_1
	S	m-S	**m**

- **false positive rate** = F/ m_0 = type I error rate = α
- **false discovery rate** = F/S = q

Idea

Control the false discovery rate (q-value)
instead of the false positive rate (related to the p-value)

EXAMPLE

False Discovery Rate

Hedenfalk *et al* (2001)

Hedenfalk et al (2001)
Screened ~~3226~~ 3170 genes using microarrays to find differential expression between BRCA-1 and BRCA-2 mutation positive tumors.

- Order the 3170 p-values: **p_i , i = 1 ... 3170**
  (56 genes were excluded from this analysis)

- Pick a p-value cutoff, say α : reject H_o for all $p_i < \alpha$.

What is the false discovery rate (FDR) associated with this choice of α?

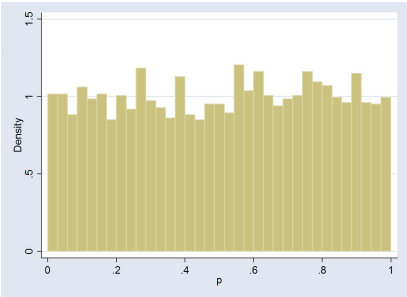
- FDR = F / S
- S = #{$p_i < \alpha$}
- F = $\alpha * m_0$
- FDR = q-value = $\alpha * m_0$ / #{$p_i < \alpha$}
- I know S, I know α, what is m_0?

EXAMPLE

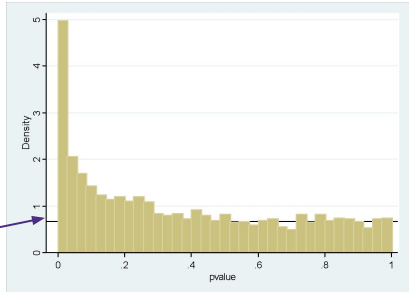
False Discovery Rate

Hedenfalk *et al* (2001)

Hedenfalk et al (2001)
Screened ~~3226~~ 3170 genes using microarrays to find differential expression between BRCA-1 and BRCA-2 mutation positive tumors.



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



0.676

Distribution of 3170 p-values from Hedenfalk *et al.*

Height of the line gives estimated proportion of true null hypotheses.

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

False Discovery Rate

Hedenfalk et al (2001)
Screened ~~3226~~ 3170 genes using microarrays to find differential expression between BRCA-1 and BRCA-2 mutation positive tumors.

	Reject null	Fail to reject null	
Null true	F	m_0-F	m_0 =3170*0.676 =2143
Null not true (Alternative true)	T	m_1-T	m_1
	S = #{p_i<α}	m-S	**m=3170**

- **false positive rate** = F/ m_0 = type I error rate = α (we set alpha)
- **false discovery rate** = F/S = q → **F = q * S**

Idea

Control the false discovery rate (q-value)
instead of the false positive rate (related to the p-value)

False Discovery Rate

Hedenfalk et al (2001)
Screened ~~3226~~ 3170 genes using microarrays to find differential expression between BRCA-1 and BRCA-2 mutation positive tumors.

> **$q(\alpha) = \alpha * m_0(\lambda) / \#\{p_i < \alpha\}$**
> [technically $q(\alpha) = \min_{t \geq \alpha} q(t)$]

> Package **`qvalue`** in R

Example : Analysis of data from Hedenfalk *et al* (using $m_0(0.5) = 2143$)

q false discovery rate	α false positive rate	#{ $p_i < \alpha$ } expected # of positives	expected # of false positives
0.01	0.0000126	5	0
0.05	0.00373	160	8
0.10	0.0148	317	32

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