Permutation and Exact Tests

&

False Detection Rate

Exact and Permutation Tests

- Computer-intensive methods for hypothesis testing

- Permutation Test (randomization test):
  Used when distribution of the test statistic (under the null hypothesis) is unknown

- Exact Test:
  Used when sample sizes are small, so standard asymptotic (large sample) procedures do not work well

  Exact test maintains 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)

- Does the probability of infection differ between the two groups?

What is a useful model for these data?

Example - HPV vaccine trial

Vaccine group: Binomial(100, p_V)

Placebo group: Binomial(100, p_P)

Scientific Question:

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

Restate in terms of the model:

$H_0: p_v = p_p$ ("null hypothesis")

vs. $H_a: p_v < p_p$

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?

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 $\Rightarrow$ $p_v - p_p$

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

Example $\Rightarrow$ $p_v - p_p = 0$

- 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?

Example - HPV vaccine trial

We need to figure out what sort of <u>distribution</u> 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**.

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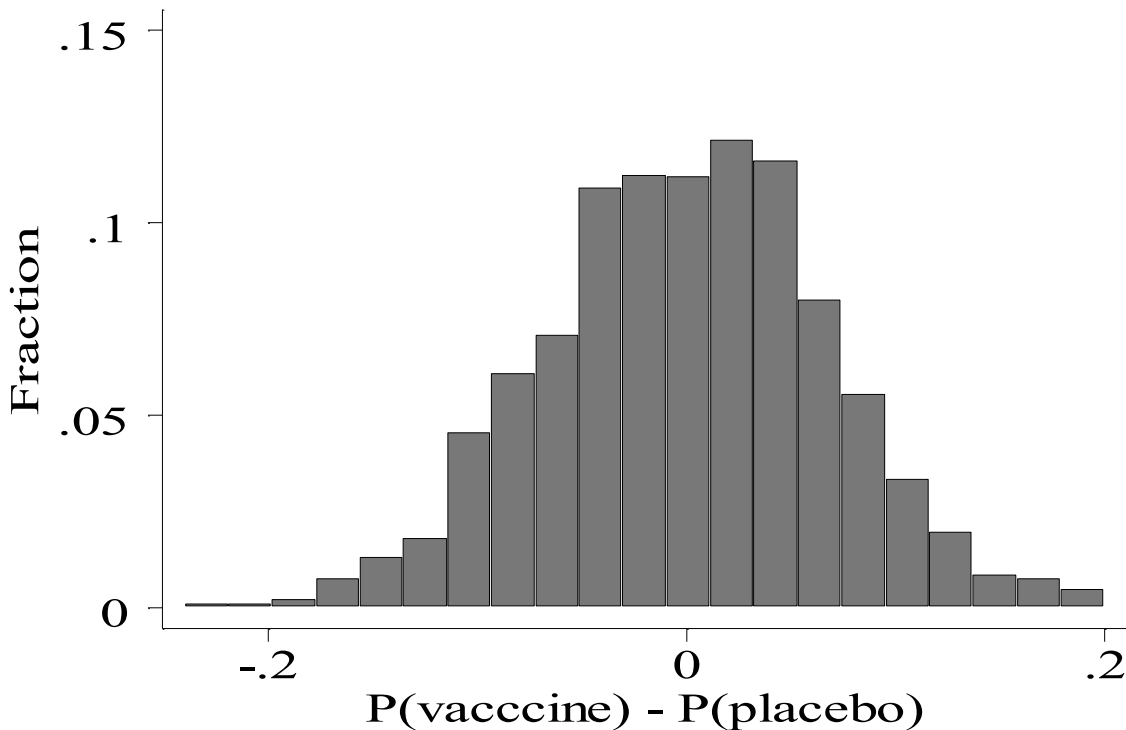
Summarize the results by reporting what proportion of the simulated results are as "extreme" or more so than the observed result (p value).

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

$\Rightarrow$ **p = .0015**

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. Pick a model for the data and restate the scientific question in terms of the model (null hypothesis)

2. Choose (any) reasonable summary statistic that quantifies deviations from the null hypothesis

3. Resample data <u>assuming the null hypothesis is true</u> 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.

Permutation Test for Correlation

Assume data are pairs (X_{1i}, X_{2i}), $i = 1, 2, \ldots, n$

1. Ho: $\rho = 0$

2. Compute $r_{obs} = \text{corr}(X_1, X_2)$

3. Mix up the X_{1i} and X_{2i} ; i.e., for each X_{1i} randomly choose X_{2i}' from all the X_2's. Compute $r_{perm} = \text{corr}(X_1, X_2')$

4. Repeat Step 3 many times and compare r_{obs} to the distribution of r_{perm}

Note: There are n! possible pairings. If n is small, you can enumerate all possible pairings.

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, otherwise generate many permutations.

- 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.

Fisher's Exact Test

Motivation: When a 2×2 table contains cells that have fewer than 5 <u>expected</u> observations, the χ^2 approximation to the distribution of X^2 is known to be poor. This can lead to incorrect inference since the p-values based on this approximation are not valid.

Solution: Use Fisher's Exact Test

	D+	D-	Total
E+	a	b	n_1
E-	c	d	n_2
Total	m_1	m_2	N

Fisher's Exact Test

Example: A retrospective study is done among men aged 50-54 who died over a 1-month period. The investigators tried to include equal numbers of men who died from cardiovascular disease (CVD) and those that did not. Then, asking a close relative, the dietary habits were ascertained.

	High Salt	Low Salt	Total
CVD	5	30	35
Non-CVD	2	23	25
Total	7	53	60

A calculation of the odds ratio yields:

$$OR = \frac{5 \times 23}{2 \times 30} = 1.92$$

Fisher's Exact Test

	D+	D-	Total
E+			n_1
E-			n_2
Total	m_1	m_2	N

If we **fix all of the margins** then any one cell of the table will allow the remaining cells to be filled. Note that a must be greater than 0, less than both n_1 and m_1, and an integer. Thus there are only a relatively few number of possible table configurations if either n_1 or m_1 is small (with n_1, n_2, m_1, m_2 fixed).

Under the null hypothesis,

H_0 : OR = 1

we can use the hypergeometric distribution (a probability distribution for discrete rv's) to compute the probability of any given configuration. Since we have the distribution of a statistic (a) under the null, we can use this to compute p-values. You will *never* do this by hand ….

Fisher's Exact Test

Example: (Rosner, p. 370) Cardiovascular disease.

	High Salt	Low Salt	Total
CVD	5	30	35
Non-CVD	2	23	25
Total	7	53	60

Possible Tables:

0		35
		25
7	53	60

1		35
		25
7	53	60

2		35
		25
7	53	60

3		35
		25
7	53	60

4		35
		25
7	53	60

5		35
		25
7	53	60

6		35
		25
7	53	60

7		35
		25
7	53	60

Chi-square
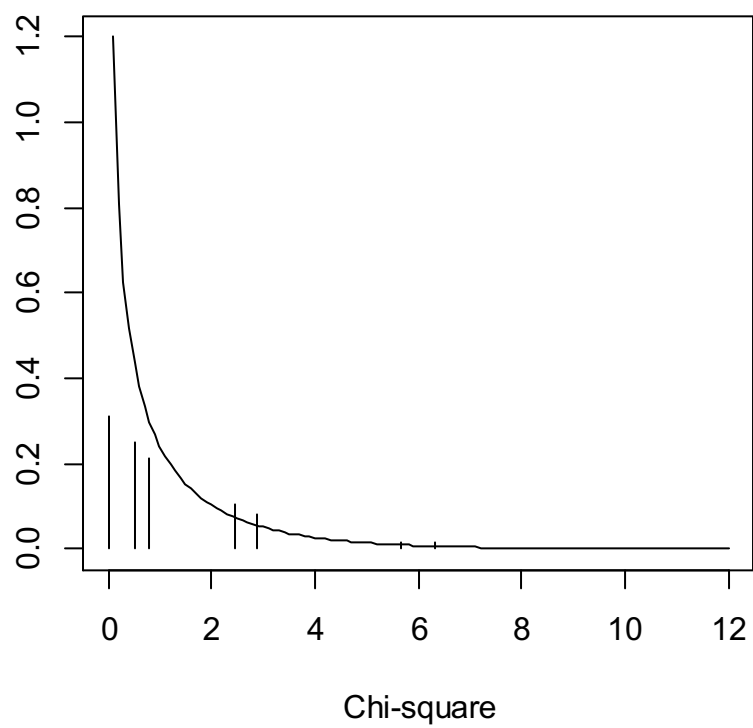
Fisher Exact Test Using Stata

Fisher's exact test.

```
. cci 5 30 2 23, exact

                 |  Exposed   Unexposed  |     Total    Proportion
                 |                        |              Exposed
-----------------+------------------------+----------------------------
           Cases |        5          30  |        35       0.1429
        Controls |        2          23  |        25       0.0800
-----------------+------------------------+----------------------------
           Total |        7          53  |        60       0.1167

                 |  Point estimate  |   [95% Conf. Interval]
                 |------------------+-------------------------
      Odds ratio |     1.916667     |    .2789585     21.62382   (exact)
  Attr. frac. ex.|     .4782609     |   -2.584763     .9537547   (exact)
  Attr. frac. pop|     .068323      |
                    1-sided Fisher's exact P = 0.3747
                    2-sided Fisher's exact P = 0.6882
```

Fisher Exact Test Using Stata

The usual chi-squared test, for comparison.

```
. cci 5 30 2 23
                                                  Proportion
                 |   Exposed   Unexposed  |    Total     Exposed
-----------------+------------------------+------------------------
           Cases |      5          30     |      35      0.1429
        Controls |      2          23     |      25      0.0800
-----------------+------------------------+------------------------
           Total |      7          53     |      60      0.1167

                 |    Point estimate    |    [95% Conf. Interval]
-----------------+----------------------+------------------------
      Odds ratio |      1.916667        |    .2789585    21.62382  (exact)
  Attr. frac. ex.|      .4782609        |   -2.584763    .9537547  (exact)
  Attr. frac. pop|      .068323         |
                         chi2(1) =    0.56   Pr>chi2 = 0.4546
```

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.

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

<u>Issue:</u> 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!

<u>Traditional Solution (Bonferroni correction):</u> 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.

<u>New Solution:</u> Don't try to eliminate false positives … control them

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)
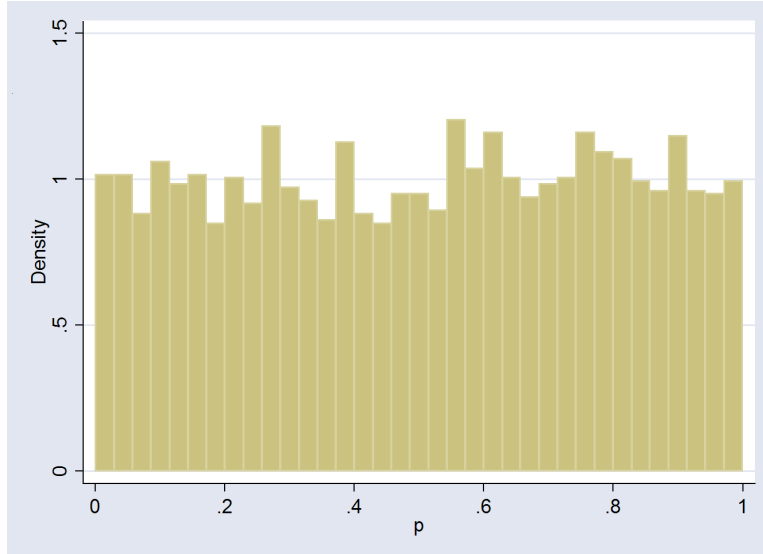
False Discovery Rate

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 $p_i < \alpha$.

  - F will be equal to $\alpha * m_0$

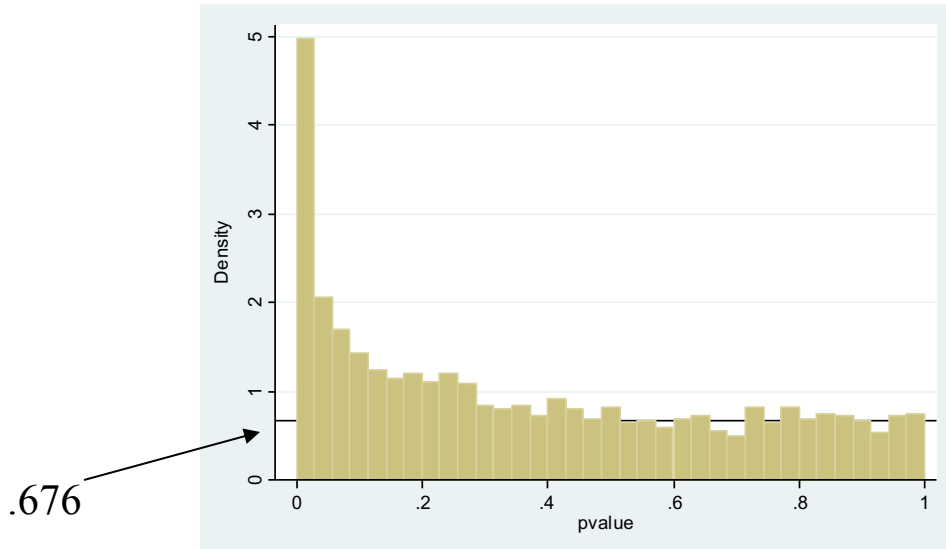Q: What is the FDR associated with this choice of α?

  - 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?

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.

.676

$$m_0 = 3170*.676 = 2142$$

False Discovery Rate

- $m_0(\lambda) = \frac{\#(p_i > \lambda)}{1 - \lambda}$ e.g. $m_0(.5) = \frac{1071}{.5} = 2142$

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

  Program QVALUE (http://genomine.org/qvalue/) or p.adjust() in R

- Eg. Hedenfalk et al. ($m_0(.5) = 2142$)

	No. differentially expressed		expected
q	α	$\#\{p_i < \alpha\}$	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.