Fisher's p-value (Permutation Tests)

BIOS 6611

CU Anschutz

Week 4

Introduction

Introduction

- **Fisher's p-value** is one tool used in data testing procedures
- Fisher used *tests of significance* (versus Neyman and Pearson, which used tests of hypotheses)
  - 1925: Fisher developed tests of significance
  - 1928: Neyman and Pearson developed tests of statistical hypotheses
- Tests of significance combined with tests of hypotheses led to development of **NHST** (null hypothesis significance testing), what is generally used today
  - 1940: NHST developed
  - Fisher never meant the two to be combined
  - Widely used, yet somewhat controversial

Ronald Fisher

- Sir Ronald Fisher (1890-1962): British statistician, geneticist, eugenicist, professor
- Arguably created foundations of modern statistics
- Held strong views on race, insisted on racial differences
- Wrote testimony on behalf of Nazi eugenicist, wrote he believed the Nazi party "wished to benefit the German racial stock, especially by elimination of manifest defectives" and that he would support such a movement (https://doi.org/10.1086/657474)
- Fierce rivalry with Pearson (*The Lady Tasting Tea*, Salsburg)

Fisher's p-value (Permutation Tests)

Fisher's p-value (Permutation Tests)

- Fisher's p-value is based on idea of *permutation tests*, also called *exact tests, randomization tests, or re-randomization tests*
- Idea is to obtain distribution of test statistic under the null hypothesis (*null distribution*) by calculating all possible values of test statistic under all possible rearrangements (permutations) of data
- We will revisit permutation tests in greater detail later

General Steps of Determining Fisher's p-value

General steps of determining Fisher's p-value (permutation tests):

1. Collect random sample: $(X_1, \ldots, X_n)$
2. Calculate observed test statistic: $T_{obs}(X_1, \ldots, X_n)$
3. Calculate "theoretical" test statistics for every possible outcome (every permutation of the data): $T_{theor,i}$ for $i = 1, \ldots, N$, where $N =$ number of possible outcomes
   - The set $\{T_{theor,i}\}$ is the exact distribution of all possible outcomes under the null hypothesis of no statistical association
   - If null is true, then each outcome would be equally likely to occur
4. Fisher's 1-sided p-value is the proportion of permutations with $T_{theor,i} \geq T_{obs}$, i.e.,
$$\text{Fisher's } p\text{-value} = \frac{\sum_{i=1}^{N} 1\{T_{theor,i} \geq T_{obs}\}}{N}$$
5. Fisher's 2-sided p-value is the proportion of permutations (theoretical outcomes) with $|T_{theor,i}| \geq |T_{obs}|$

Example

Example Problem (Chihara and Hesterberg, section 3.3)

Outcome: seconds for a mouse to complete a maze

Group 1: 3 mice given experimental drug

Group 2: 3 mice given placebo

Experimental	Control
30	18
25	21
20	22

Example Problem

```
exp<-c(30,25,20)
control<-c(18,21,22)

mean(exp)
```

[1] 25

```
mean(control)
```

[1] 20.33333

The average time for the experimental group is 25 seconds, for the control groups is 20.33 seconds. Is this difference meaningful?

Example problem

Fisher's idea:

There are $\binom{6}{3} = 20$ possible random permutations of the data (20 ways to choose 3 mice from 6, when order doesn't matter.) These random shufflings represent all possible groupings of the mice.

If there is no association between time to complete and the drug, then each shuffling is equally likely to occur.

Example Problem

Null or *permutation* distribution:

TABLE 3.1 All Possible Distributions of {30, 25, 20, 18, 21, 22} into Two Sets

Drug			Control			$\bar{X}_D$	$\bar{X}_C$	Difference in means	
18	20	21	22	25	30	19.67	25.67	−6.00	
18	20	22	21	25	30	20	25.33	−5.33	
18	20	25	21	22	30	21	24.33	−3.33	
18	20	30	21	22	25	22.67	22.67	0.00	
18	21	22	20	25	30	20.33	25	−4.67	
18	21	25	20	22	30	21.33	24	−2.67	
18	21	30	20	22	25	23	22.33	0.67	
18	22	25	20	21	30	21.67	23.67	−2.00	
18	22	30	20	21	25	23.33	22	1.33	
18	25	30	20	21	22	24.33	21	3.33	
20	21	22	18	25	30	21	24.33	−3.33	
20	21	25	18	22	30	22	23.33	−1.33	
20	21	30	18	22	25	23.67	21.67	2.00	
20	22	25	18	21	30	22.33	23	−0.67	
20	22	30	18	21	25	24	21.33	2.67	
20	25	30	18	21	22	**25**	**20.33**	**4.67** *	Our observed results
21	22	25	18	20	30	22.67	22.67	0.00	
21	22	30	18	20	25	24.33	21	3.33	
21	25	30	18	20	22	**25.33**	**20**	**5.33** *	
22	25	30	18	20	21	**25.67**	**19.67**	**6.00** *	

Rows where the difference in means exceeds the original value are highlighted.

Example Problem

Among 20 permutations of data, 3 have difference in means as large or larger than 4.67 (1-sided p-value).

$$\text{Fisher's } p\text{-value} = \frac{3}{20} = 0.15 \tag{1}$$

If no association between drug and speed, then purely by chance, we would see a difference as or more extreme that what was observed 15% of the time.

Not that unlikely for observed data to have occurred if there is truly no difference between groups.

Summary

Summary

- Fisher's p-value is key building block to modern day null hypothesis significance testing
- Fisher's p-value is based off idea of permutation tests
  - Will see permutation tests again in more detail, and when we learn about non-parametric tests.
  - Big idea: generate a null distribution from the data