Multiple Testing

Grinnell College

April 12, 2024

Exam

Some comments:

- A brief, random persual suggests that students did better than the looks on their faces at the end would suggest
- I will handle incomplete questions and note the impact of time constraint on quality of other solutions
- Grades will be curved
- Plan to have them back by Wednesday

Final Projects

- Group have been confirmed
- Proposal and research question due Wed 4/17
- Have data collected with exploratory analaysis by Mon 4/29
  - Two weekends to collect data
  - 3-4 variables (including your outcome)
  - $\sim$30-40 observations

Rest of Semester

The rest of the tools we will learn about in class invovles testing for *differences* or *associations* between groups which may help inform your project goals

Type	Continuous	Categorical
Simple Test	t-test	Single Proportion
2 Groups	Two-sample t-test, paired test	Difference in Proportion
Multiple Groups	ANOVA	χ^2 Test
Mixed variables	Regression	Regression

Today

1. What are some of the basic properties of probabilities?
2. What is multiple testing, and how is this related to the problem of Type I errors?
3. What is the Family-Wise Error Rate (FWER)?
4. What adjustments can we make for multiple testing

Probability Basics

A **random event** has outcomes that we cannot predict but have a regular distribution of outcomes over many repititions

The **probability** of an event is the proportion of times that event occurs in many repeated experiments of the same random event

Random Events:

- Flipping a fair coin, with probability of $1/2$ for each outcome
- Rolling a dice, with probability $1/6$ of each outcome
- Flipping a coin 2 times, the probability of getting heads twice is $1/4$

Probability

The sum of probabilities for all possible events must sum to 1

Flipping a coin has two events: heads and tails

$$P(\text{Heads}) + P(\text{Tails}) = \frac{1}{2} + \frac{1}{2} = 1$$

How does this look when flipping a coin twice?

Probability – Independence

A sequence of random events is said to be **independent** if the result of one outcome does not influence

If I flip a coin twice, the probability of flipping heads on my second toss is the same, regardless of what the first flip was

Gambler's Fallacy

Probability – Successive Independent Events

If a sequence of random events is independent, the probability of seeing a sequence is the product of each event's probability

Because coin flips are independent, the probability of flipping heads 3 times in a row is

$$
\begin{aligned}
P(\text{Flip heads three times}) &= P(H) \times P(H) \times P(H) \\
&= \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) \\
&= \frac{1}{8}
\end{aligned}
$$

Probability – Compliments

The **compliment** of an event A is the probability that A *does not* occur

$$P(A^C) = 1 - P(A)$$

For example, the compliment of *not* flipping heads three times is

$$\begin{aligned} P(\text{Not flip heads three times}) &= 1 - P(\text{Flip heads three times}) \\ &= 1 - \frac{1}{8} \\ &= \frac{7}{8} \end{aligned}$$

Probability – Compliments

Compliments are immensely useful in situations in which the probability of (typically a series of) events is complicated, but determining its compliment is trivial

If I roll a die 8 times, what is the probability that *at least* one of the rolls lands on a 1?

Birthday Paradox

Basics of probability

- The **probability** of a random event is the proportion of times we would expect an event to occur if repeated multiple times
- The sum of probabilities for all possible events must equal 1
- A series of events are said to be **independent** if the result of one does not influence any of the others
- The complement of an event describes a situation in which it *does not* occur

Type I Error

Consider conducting 2 hypothesis tests, each with a Type I error rate of 5%

For any given test, the probability of *not* making an error is

$$P(\text{No type I error}) = 0.95$$

1. What is the probability that neither test has a Type I error?
2. What is the probability that *at least* one test has a Type I error?

Example

Suppose that I am interested in testing if there is a non-zero correlation between cost and average faculty salary in each of the 8 regions of our college dataset

Suppose further we are testing for significance at the level $\alpha = 0.05$

	Region	*p*-value
1	Far West	0.7667
2	Great Lakes	0.0085
3	Mid East	0.0001
4	New England	0.0061
5	Plains	0.9487
6	Rocky Mountains	0.7394
7	South East	0.0143
8	South West	0.0344

Example

Suppose that I am interested in testing if there is a non-zero correlation between cost and average faculty salary in each of the 8 regions of our college dataset

If my Type I error rate for each test is 5%, what is the probability that I make at least one Type I error?

$$
\begin{aligned}
P(\text{At least one Type I error}) &= 1 - P(\text{Probability of no Type I errors}) \\
&= 1 - (1 - 0.05)^8 \\
&= 33.6\%
\end{aligned}
$$

That is, instead of making a Type I error 1 in 20 times, we are now making it 1 in 3 times

Family-wise error rates (FWER)

For a collection of independent hypothesis tests, the **family-wise error rate (FWER)** describes the probability of making one or more Type I errors

For m independent tests with a Type I error rate of α, the FWER is defined as

$$\text{FWER} = 1 - (1 - \alpha)^m$$

FWER Correction

Just as we control the Type I error rate of a single hypothesis test with α, we also have an interest in controlling the FWER

For m hypothesis tests controlled at level α, the correction $\alpha^* = \alpha/m$ is known as the **Bonferonni Adjustment**

If instead for a series of m tests we reject the null hypothesis when $p < \alpha^*$, we will control the FWER at level α

Assuming the 8 regions of our hypothesis test are independent, our Bonferonni adjustment for $\alpha = 0.05$ should be

$$\alpha^* = 0.05/8 = 0.00625$$

Testing $p < \alpha$

	Region	p-value
1	Far West	0.7667
2	Great Lakes	0.0085
3	Mid East	0.0001
4	New England	0.0061
5	Plains	0.9487
6	Rocky Mountains	0.7394
7	South East	0.0143
8	South West	0.0344

Testing $p < \alpha^*$

	Region	p-value
1	Far West	0.7667
2	Great Lakes	0.0085
3	Mid East	0.0001
4	New England	0.0061
5	Plains	0.9487
6	Rocky Mountains	0.7394
7	South East	0.0143
8	South West	0.0344

Review

1. What are some of the basic properties of probabilities?
2. What is multiple testing, and how is this related to the problem of Type I errors?
3. What is the Family-Wise Error Rate (FWER)?
4. What adjustments can we make for multiple testing