STA258H5

Al Nosedal
and Alison Weir

Winter 2017

HYPOTHESIS TESTING - CONCEPTS

Example

How rational and consistent is the behavior of the typical American college student? We'll explore whether college student consumers always consider an obvious fact: money not spent now can be spent later.
In particular, we are interested in whether reminding students about this well-known fact about money causes them to be a little thriftier. A skeptic might think that such a reminder would have no impact. We can summarize these two perspectives using the null and alternative hypothesis framework.

Example

H_0: Null hypothesis. Reminding students that they can save money for later purchases will not have any impact on students' spending decisions.
H_A: Alternative hypothesis. Reminding students that they can save money for later purchases will reduce the chance they will continue with a purchase.
We'll explore an experiment conducted by researchers that investigates this very question for students at a university in the southwestern United States.

Example

One-hundred and fifty students were recruited for the study, and each was given the following statement:
Imagine that you have been saving some extra money on the side to make some purchases, and on your most recent visit to the video store you come across a special sale on a new video. This video is one with your favorite actor or actress, and your favorite type of movie (such as a comedy, drama, thriller, etc.). This particular video that you are considering is one you have been thinking about buying for a long time. It is available for a special sale price of $14.99.

What would you do in this situation? Please circle one of the options below.

Half of the 150 students were randomized into a control group and were given the following two options:

(A) Buy this entertaining video.

(B) Not buy this entertaining video.

The remaining 75 students were placed in the treatment group, and they saw a slightly modified option (B):
(A) Buy this entertaining video.
(B) Not buy this entertaining video. Keep the $14.99 for other purchases.
Would the extra statement reminding students of an obvious fact impact the purchasing decision? The Table shown below summarizes the study results.

Table of counts

	decision		
	buy DVD	not buy DVD	Total
control group	56	19	75
treatment group	41	34	75
Total	97	53	150

Table of proportions

	decision		
	buy DVD	not buy DVD	Total
control group	0.747	0.253	1
treatment group	0.547	0.453	1
Total	0.647	0.353	1

We will define a **success** in this study as a student who chooses not to buy the DVD. Then, the value of interest is the change in DVD purchase rates that results by reminding students that not spending money now means they can spend the money later. We can construct a point estimate for this difference as

$$\hat{p}_{trmt} - \hat{p}_{ctrl} = \frac{34}{75} - \frac{19}{75} = \frac{15}{75} = 0.20$$

The proportion of students who chose not to buy the DVD was 20% higher in the treatment group than the control group. However, is this result statistically significant? In other words, is a 20% difference between the two groups so prominent that it is unlikely to have occurred from chance alone?

Simulation

```
# probability of success;
p<-53/150;
#simulated control group;
n1<-75;
p1<-rbinom(1,75,p)/n1;
#simulated treatment group;
n2<-75;
p2<-rbinom(1,75,p)/n2;
diff<-p1-p2;
diff

## [1] -0.02666667
```

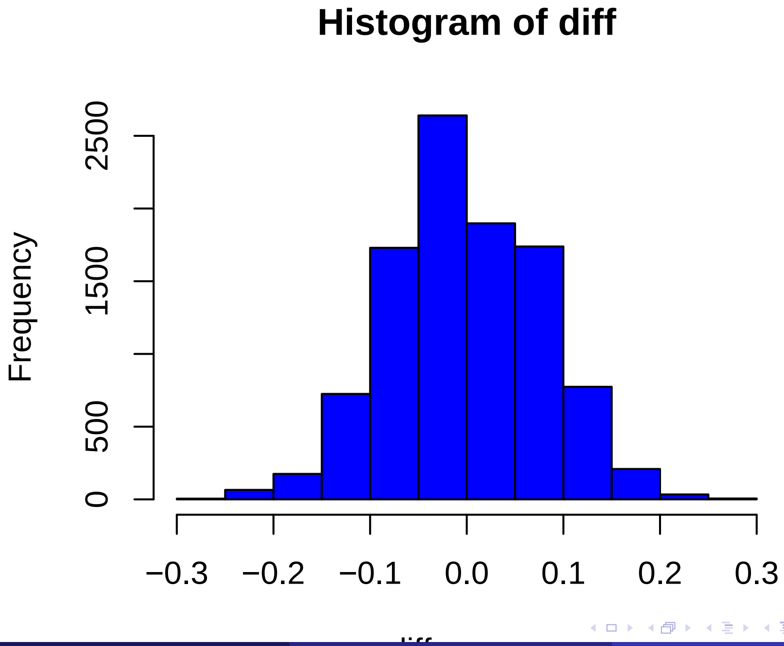
Another Simulation

```
# probability of success;
p<-53/150;
#simulated control group;
n1<-75;
p1<-rbinom(1,75,p)/n1;
#simulated treatment group;
n2<-75;
p2<-rbinom(1,75,p)/n2;
diff<-p1-p2;
diff

## [1] 0.02666667
```

10000 Simulations

```
# number of simulations;
m<-10000;
p<-53/150;
#simulated control group;
n1<-75;
p1<-rbinom(m,75,p)/n1;
#simulated treatment group;
n2<-75;
p2<-rbinom(m,75,p)/n2;
diff<-p1-p2;
hist(diff,xlab="differences",col="blue");
```


Histogram of diff
Frequency
0
500
1500
2500
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3

P-value

```
## P-value

p_value<-length(diff[diff>0.20])/m;

p_value

## [1] 0.0044
```

Conclusion

If there was no treatment effect, then we'd only observe a difference of at least +20% about 0.44% of the time. That is really rare! Instead, we will conclude the data provide strong evidence there is a treatment effect: reminding students before a purchase that they could instead spend the money later on something else lowers the chance that they will continue with the purchase. Notice that we are able to make a causal statement for this study since the study is an experiment.

NULL HYPOTHESIS H_0

The statement being tested in a statistical test is called the **null hypothesis**. The test is designed to assess the strength of the evidence against the null hypothesis. Usually the null hypothesis is a statement of "no effect" or "no difference".

Test statistics

A significance test uses data in the form of a **test statistic**. Here are some principles that apply to most tests:

- The test is based on a statistic that compares the value of the parameter stated by the null hypothesis with an estimate of the parameter from the sample data. The estimate is usually the same one used in a confidence interval for the parameter.
- Large values of the test statistic indicate that the estimate is far from the parameter value specified by H_0. These values give evidence against H_0. The alternative hypothesis determines which directions count against H_0.

P-values

The probability, computed assuming that H_0 is true, that the test statistic would take a value as extreme or more extreme than that actually observed is called the **P-value** of the test. The smaller the P-value, the stronger the evidence against H_0 provided by the data.

Small P-values are evidence against H_0, because they say that the observed result is unlikely to occur when H_0 is true. Large P-values fail to give evidence against H_0.

The P-value Scale

If P-value < 0.001, we have very strong evidence against H_0.
If $0.001 \leq$ P-value < 0.01, we have strong evidence against H_0.
If $0.01 \leq$ P-value < 0.05, we have evidence against H_0.
If $0.05 \leq$ P-value < 0.075, we have some evidence against H_0.
If $0.075 \leq$ P-value < 0.10, we have slight evidence against H_0.

Statistical Significance

If the P-value is as small or smaller than α, we say that the data are **statistically significant at level α.**

Power

The probability that a fixed level α significance test will reject H_0 when a particular alternative value of the parameter is true is called the **power** of the test against that alternative.

Type I and Type II Errors

If we reject H_0 when in fact H_0 is true, this is a **Type I error.**
If we fail to reject H_0 when in fact H_a is true, this is a **Type II error.**
The **significance level** α of any fixed level test is the probability of a Type I error.

The two types of errors in testing hypotheses

		Truth	
		H_0 true	H_a true
Test conclusion	Reject H_0	Type I error	Correct
	NOT rejecting H_0	Correct decision	Type II error