BIOS 662 Fall 2018

Tests of Hypotheses

David Couper, Ph.D.

david_couper@unc.edu

or

couper@bios.unc.edu

https://sakai.unc.edu/portal

Basic Approach

1. Set up a hypothesis
2. Collect data
3. Infer from the data whether the hypothesis is plausible

- Examples:

  Is mean BP the same in diabetics and non-diabetics?

  Will folic acid supplementation reduce the risk of stroke?

Null Hypothesis: H_0

- Null hypothesis H_0: the hypothesis to be tested
- Example: Null hypothesis for folic acid study

  The incidence of stroke is the same in those taking folic acid supplements as in those not taking folic acid supplements
- See Note 4.16 in the text. Typically H_0 is:
  - the prevailing view or straw man, or
  - the most parsimonious hypothesis

Null and Alternative

- In a test of a hypothesis, we are testing whether some population parameter has a particular value
- For example,

$$H_0 : \theta = \theta_0$$

where θ_0 is a specified constant

- The alternative hypothesis is the complement of the null hypothesis

$$H_A : \theta \neq \theta_0$$

Test Statistic

- Once the data are collected, we compute the value of a *test statistic* related to θ, say $S(\hat{\theta})$
- $S(\hat{\theta})$ is a random variable, because it is computed from a sample
- $S(\hat{\theta})$ will have a particular probability distribution under the assumption H_0, say $F_0(S(\hat{\theta}))$

Test Statistic

- Under F_0, we compute the probability that we would observe $S(\hat{\theta})$ or a value more extreme than $S(\hat{\theta})$ if the null H_0 is true
- If this probability is large, the data are consistent with H_0
- If this probability is small, there are two possibilities:
  1. An unlikely event has occurred
  2. H_0 is not true

Interpretation

- Usually if the probability is small, we conclude H_0 is not true; i.e., we "reject" H_0
- If the probability is large, we have not proved H_0. We say that "we failed to reject H_0"
- We can never prove H_0 is true!
- Also: we don't "accept the alternative"

Significance Level

- How do we decide if the probability is too small?
- Prior to seeing the data, we select a value α such that:

  If the computed probability is less than or equal to α, we reject H_0
- α is known as the *significance level*

Critical Region and Value

- We have a statistic $S(\hat{\theta})$ with distribution F_0 under the null hypothesis
- We specify α and under F_0 determine a *critical region* or *rejection region* C_α such that

$$\Pr[S(\hat{\theta}) \in C_\alpha | H_0] = \alpha$$

- Values at the boundaries of C_α are called *critical values*

Critical Region and Value

- From the data we compute the value of $S(\hat{\theta})$
- If $S(\hat{\theta}) \in C_\alpha$, we reject H_0
- If $S(\hat{\theta}) \notin C_\alpha$, the data are consistent with H_0 (or, at least, not *inconsistent* with H_0) and we do not reject H_0

Tests of Hypotheses: Seven Steps

1. Design study (sample size depends on steps 2-4)
2. Establish null hypothesis
3. Determine test statistic to be employed
4. Choose significance level α and establish C_α
5. Carry out study and collect data
6. Compute statistic from data
7. If statistic is in C_α, reject H_0

Example

- Does calcium supplementation affect blood pressure in African Americans with high blood pressure?
- Study: Enroll 10 AA men with hypertension; measure their BP; ask them to take calcium tablets for 3 weeks and re-measure their BP
- Aside: Later in the semester we will look at how to determine whether $n = 10$ is a large enough sample size to provide a reasonable test of the hypotheses.

Example cont.

- Let θ denote the mean BP change after 3 weeks
- Hypotheses

$$H_0 : \theta = 0 \text{ vs } H_A : \theta \neq 0$$

- Let Y_i = BP at 3 weeks $-$ BP at baseline for the i^{th} individual in the study, $i = 1, \ldots, 10$
- $\hat{\theta} = \bar{Y}$

Example cont.

- Intuition: We want to reject H_0 if $\bar{Y}$ is far from $\theta_0 = 0$, i.e., if
$$\left|\bar{Y}\right| > c$$
for some constant c

- In particular, we want c such that
$$\Pr\left[\left|\bar{Y}\right| > c \mid H_0\right] = \alpha$$

Example cont.

- Equivalently, want c such that

$$\Pr\left[\left|\frac{\bar{Y}}{s/\sqrt{n}}\right| > \frac{c}{s/\sqrt{n}} \mid H_0\right] = \alpha$$

- Assuming Y_i are iid $N(\theta, \sigma^2)$, under H_0,

$$\frac{\bar{Y}}{s/\sqrt{n}} \sim t_{n-1}$$

- Thus choose c such that

$$\frac{c}{s/\sqrt{n}} = t_{n-1,1-\alpha/2}$$

Example cont.

- So we reject H_0 if

$$\left|\bar{Y}\right| > c = t_{n-1,1-\alpha/2} s/\sqrt{n}$$

that is, if

$$\frac{\left|\bar{Y}\right|}{s/\sqrt{n}} > t_{n-1,1-\alpha/2}$$

- Equivalently

$$C_\alpha = \{T : |T| > t_{n-1,1-\alpha/2}\}$$

where

$$T = \frac{\bar{Y}}{s/\sqrt{n}}$$

Example cont.

- In the calcium supplementation example,

$$S(\hat{\theta}) = S(\bar{Y}) = \frac{\bar{Y} - \theta_0}{s/\sqrt{n}} \sim t_9$$

- If $\alpha = 0.05$, the critical region is

$$C_{0.05} = \{T : |T| > t_{9,0.975} = 2.26\}$$

where

$$T = \frac{\bar{Y} - 0}{s/\sqrt{10}}$$

Example cont.

Calcium supplementation in African-American men

ID	treatment	before	after	change
1.	calcium	107	100	-7
2.	calcium	110	114	4
3.	calcium	123	105	-18
4.	calcium	129	112	-17
5.	calcium	112	115	3
6.	calcium	111	116	5
7.	calcium	107	106	-1
8.	calcium	112	102	-10
9.	calcium	136	125	-11
10.	calcium	102	104	2

Example using R

```
> x <- c(-7,4,-18,-17,3,5,-1,-10,-11,2)
> se <-  sd(x)/sqrt(length(x))
> mean(x)/se
[1] -1.808411

> t.test(x)

        One Sample t-test

data:  x
t = -1.8084, df = 9, p-value = 0.104
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -11.254545   1.254545
```

Example using SAS PROC UNIVARIATE

```
proc univariate;
   var x;
```

```
The UNIVARIATE Procedure
Variable:  x

                            Moments

N                            10    Sum Weights                 10
Mean                         -5    Sum Observations           -50
Std Deviation        8.74325137    Variance            76.4444444

           Tests for Location: Mu0=0

Test              -Statistic-    -----p Value------

Student's t    t  -1.80841    Pr > |t|    0.1040
```

Example using SAS PROC TTEST

```
proc ttest;
   var x;

The TTEST Procedure

Variable:  x

  N        Mean     Std Dev      Std Err         95% CL Mean

 10     -5.0000      8.7433       2.7649    -11.2545   1.2545


 DF    t Value    Pr > |t|

  9      -1.81      0.1040
```

Example cont.

- Because the observed $t = -1.8084$ is not in the critical region, we do not reject H_0
- The data are consistent with no effect of calcium on BP

Types of Errors

		Nature / Truth	
		H_0 true	H_A true
Decision	Do not reject H_0	$\surd$	Type II
	Reject H_0	Type I	$\surd$

- Type I error: Reject H_0 when H_0 true
  that is, false positive
- Type II error: Do not reject H_0 when H_A true
  that is, false negative

Types of Errors

- Type I error

$$\alpha = \Pr\left[S(\hat{\theta}) \in C_\alpha \mid H_0\right]$$

- Type II error

$$\beta = \Pr\left[S(\hat{\theta}) \notin C_\alpha \mid H_A\right]$$

- Power

$$1 - \beta = \Pr\left[S(\hat{\theta}) \in C_\alpha \mid H_A\right]$$

that is, the probability of rejecting H_0 when H_A is true

Power

- Recall $H_A : \theta \neq \theta_0$
- Power: $\Pr\left[S(\hat{\theta}) \in C_\alpha \mid H_A\right]$
- Power depends on the value of θ

$$\Pr\left[S(\hat{\theta}) \in C_\alpha \mid \theta\right] \equiv P(\theta)$$

- Note

$$P(\theta_0) = \alpha$$

Alternative Hypotheses

- Different possible alternatives

$$H_A : \theta \neq \theta_0 \ \ \text{(two-sided)}$$

$$H_A : \theta > \theta_0 \ \ \text{(one-sided)}$$

$$H_A : \theta < \theta_0 \ \ \text{(one-sided)}$$

Alternative Hypotheses

- In the calcium supplementation example,

$$H_A : \theta \neq 0; \quad C_{0.05} = \{T : |T| > 2.26\}$$

- If we took a 1-sided alternative,

$$H_A : \theta < 0; \quad C_{0.05} = \{T : T < -1.83\}$$

- Because $T = -1.8084$, we do not reject H_0 in either case

Alternative Hypotheses

- 2-sided test addresses: Does calcium **change** BP?
- 1-sided test addresses: Does calcium **lower** BP?
- If we applied this 1-sided test and obtained $T = 2.5$, we would not reject H_0 because our test did not ask if calcium raised or changed BP
- We must choose the alternative hypothesis **before seeing the data**
- Friedman et al. (p. 98) "In general, two-sided tests should be used unless there is strong justification for expecting a difference in only one direction"

P-value

- Definition 4.24. The *p-value* is the smallest significance level α for which the observed data indicate the null hypothesis should be rejected
- Probability of obtaining test statistic as unlikely or more unlikely than the observed test statistic if the null hypothesis is true

Calcium Example Revisited

- Recall $T \sim t_9$; $t_{9,0.975} = 2.26$; $t_{9,0.95} = 1.83$
- p-value for 2-sided test $= 2\Pr[T < -1.8084] = 0.1040$
- p-value for 1-sided test $= \Pr[T < -1.8084] = 0.0520$
- R

```
> 2*pt(-1.8084,9)
[1] 0.1039981

> pt(-1.8084,9)
[1] 0.05199907
```