Review: Hypothesis Testing

What is a hypothesis test?

- Assess statistical evidence
- Rule out random variation

Hypotheses

Null Hypothesis

$$H_0 : \text{random variation only}$$

Alternative Hypothesis

$$H_a : \text{there is a real effect}$$

$$H_0 : \text{no spending increase}, H_1 : \text{spending increase}$$

Limitations

- Can show a statistical model is not plausible
- Cannot prove a particular statistical model is right
- Statistical significance $\neq$ practical significance

Errors

		Reality	
		H_0 true	H_0 false
Decision based on sample	Reject H_0	Type I error (prob. α)	OK
	Do not reject H_0	OK	Type II error (prob. β)

Steps

1. Parameter(s). Hypotheses. α
2. Conditions.
3. Test statistic and P-value
4. Conclusion.

Conclusion if P is small

If $P \leq \alpha$ reject H_0.

	Reality	
	H_0 true	H_0 false
Reject H_0	Type I error (prob. α)	OK

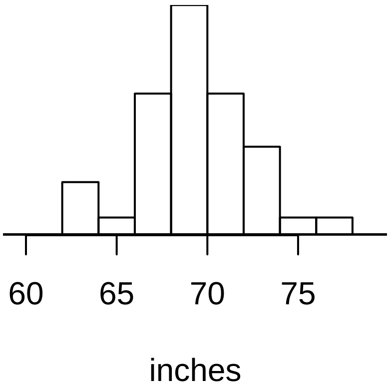
Conclusion if P is not small

If $P > \alpha$ do not reject H_0.

	Reality	
	H_0 true	H_0 false
Do not reject H_0	OK	Type II error (prob. β)

Conservative conclusion: There is insufficient evidence to reject H_0.

Example 1 - Men's Height



$\overline{x} = 69.33, s = 3.02, n = 40$

Is the average height of American men less than 70.5 inches?

Example 1 - Step 1 - Setup

$$\mu = \text{population mean height of American men}$$

$$H_0 : \mu = 70.5, \qquad H_a : \mu < 70.5$$

Test with significance level $\alpha = 0.05$.

A note on hypotheses

This class:

$$H_0 : \mu = 70.5, \qquad H_a : \mu < 70.5$$

The Ott textbook:

$$H_0 : \mu \geq 70.5, \qquad H_a : \mu < 70.5$$

Example 1 - Step 2 - Conditions

Requirements:

1. random sample of data
   - ✓This is a random sample of all American adult men
2. random variable is (approximately) normally distributed
   - ✓The histogram suggests this is reasonable.

Example 1 - Step 3 - Compute

```r
t.test(h, alternative="less", mu = 70.5)
```

```
##
##  One Sample t-test
##
## data:  h
## t = -2.4401, df = 39, p-value = 0.009664
## alternative hypothesis: true mean is less than 70.5
## 95 percent confidence interval:
##      -Inf 70.13942
## sample estimates:
## mean of x
##    69.335
```

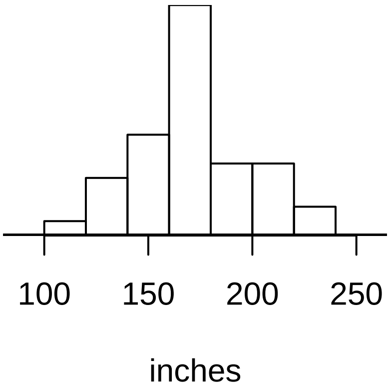
Example 1 - Step 4 - Conclusions

1. Reject H_0 at $\alpha = 0.05$ ($P = 0.00966$). There is statistically significant evidence that the population mean height of American adult men is less than 70.5 inches.

OR

2. The mean height of American adult men is less than 70.5 inches ($P = 0.00966$)

Example 2 - Men's Weight



$\overline{x} = 172.55, s = 26.33, n = 40$

Is the average weight of American men greater than 166 pounds?

Example 2 - Step 1 - Setup

$\mu =$ population mean weight of American men

$$H_0 : \mu = 166, \qquad H_a : \mu > 166$$

Test with significance level $\alpha = 0.05$.

Example 2 - Step 2 - Conditions

Requirements:

1. random sample of data
   - ✓This is a random sample of all American adult men
2. random variable is (approximately) normally distributed
   - ✓OK, histogram shows nice symmetric bell shape.

Example 2 - Step 3 - Compute

```r
t.test(w, alternative="greater", mu = 166)
```

```
##
##  One Sample t-test
##
## data:  w
## t = 1.5735, df = 39, p-value = 0.06184
## alternative hypothesis: true mean is greater than 166
## 95 percent confidence interval:
##  165.5364      Inf
## sample estimates:
## mean of x
##    172.55
```

Example 2 - Step 4 - Conclusions

1. Do not reject H_0 at $\alpha = 0.05$ ($P = 0.0618$). There is not sufficient evidence that the population mean weight of American adult men is greater than 166 pounds.

OR

2. There is not evidence to show the mean weight of American adult men is greater than 166 pounds ($P = 0.0618$).

Is H_0 true?

Does this mean that the null hypothesis is true?

Is $\mu \leq 166$ pounds?

Estimate β first!

What if?

If H_a is true, then what is μ_a?

$$\mu_a = 170 \text{ pounds}$$

Estimating Power in R

$$\text{power} = 1 - \beta$$

```
power.t.test(n = NULL, delta = NULL, sd = 1, sig.level = 0.05,
             power = NULL,
             type = c("two.sample", "one.sample", "paired"),
             alternative = c("two.sided", "one.sided"),
             strict = FALSE, tol = .Machine$double.eps^0.25)
```

Can find one of: n, delta, sd, sig.level, power.

Example 2 - Power Estimate

```
power.t.test( n = 40, delta = 4, sd = 26 , sig.level = 0.05,
              type = "one.sample", alternative = "one.sided")
```

```
##
##      One-sample t test power calculation
##
##               n = 40
##           delta = 4
##              sd = 26
##       sig.level = 0.05
##           power = 0.2455078
##     alternative = one.sided
```

Power Interpretation

- If $\mu_a = 170$, only 25% chance of being correct.
  - Type II error probability: $\beta \approx 1 - .25 = .75$.

Do not accept H_0

Risk of type II error is too high: $\beta \approx .75$

SAFE: do not reject H_0

NOT SAFE: accept H_0

Find sample size for desired power

Choose n so that power$\geq .8$ for smallest worthwhile effect.

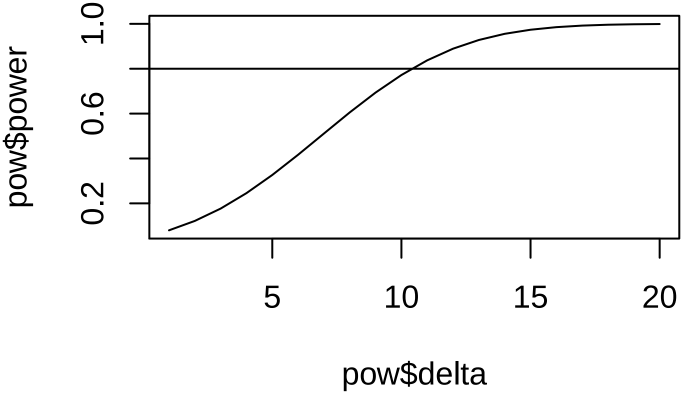
$$\text{power} = .8, \delta = 4, \text{sd} \approx 26, n = ?$$

```
power.t.test( power = .8, delta = 4, sd = 26,
              type = "one.sample",
              alternative = "one.sided")$n
```

```
## [1] 262.5711
```

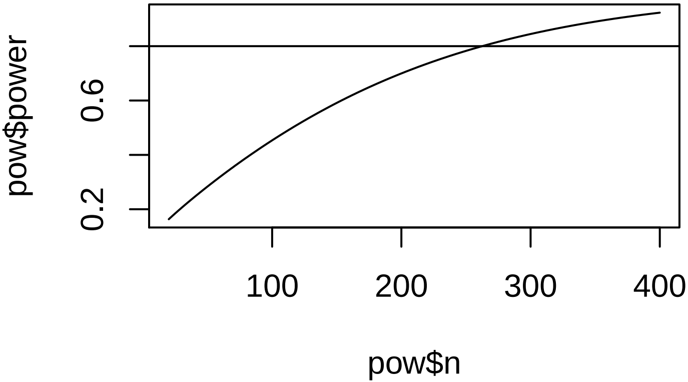
Power Curve - Power vs. Shift

```
pow <- power.t.test( n = 40, delta = 1:20, sd = 26,
              type = "one.sample", alternative = "one.sided")
plot(pow$delta,pow$power,type='l'); abline(h=.8)
```



Power Curve - Power vs. Sample Size

```
pow <- power.t.test( n = 20:400, delta = 4, sd = 26,
               type = "one.sample", alternative = "one.sided")
plot(pow$n,pow$power,type='l'); abline(h=.8)
```



A Common Error

WRONG: P is the probability H_0 is true

RIGHT: P is the probability of observing similar data by chance **if H_0 is true**

Another Common Error

WRONG: A smaller P means a larger effect

RIGHT: Small P means sample is not a plausible outcome of the “null model”

Hypothesis Test vs. Confidence Interval

- Hypothesis Test - The mean is larger than 10.
- Confidence Interval - The mean is between 11 and 13.
- WINNER: Confidence Interval

Why bother with hypothesis tests?

- Very popular
- Useful paradigm when a decision *must* be made
- P is "noise to signal" ratio
- small P may trigger further investigation

Formal Equivalence

$$H_0 : \theta = \theta_0, \alpha = .05$$

- $H_a : \theta \neq \theta_0$ reject H_0 if 95% CI does *not* include θ_0
- $H_a : \theta > \theta_0$ reject H_0 if *90%* CI is *above* θ_0
- $H_a : \theta < \theta_0$ reject H_0 if *90%* CI is *below* θ_0

Two-tailed example

$$H_0 : \mu = 10, H_a : \mu \neq 10, \alpha = 0.05$$

95% confident μ is in (11,13)

$$\Rightarrow \text{ reject } H_0$$

One-tailed example

$$H_0 : \mu = 10, H_a : \mu > 10, \alpha = 0.05$$

- 90% confident μ is in (11.2, 12.8)
- 95% confident *mu* is greater than 11.2
- $\Rightarrow$ reject H_0

One-sided Intervals

```r
t.test(h, alternative="less", mu = 70.5, conf.level = 0.95)$conf.int
```

```
## [1]     -Inf 70.13942
## attr(,"conf.level")
## [1] 0.95
```

```r
t.test(h, alternative="two.sided", mu = 70.5, conf.level = 0.90)$conf
```

```
## [1] 68.53058 70.13942
## attr(,"conf.level")
## [1] 0.9
```

Is $\mu > 100$?

```
x = rnorm(1000,mean=101,sd=10)
t.test(x,mu=100,alternative="greater")$p.value
```

```
## [1] 0.0008727748
```

```
effect_size_d <- (101-100)/10; effect_size_d
```

```
## [1] 0.1
```

$$.2 = \text{small},\ .5 = \text{moderate},\ .8 = \text{large}$$

Practical Significance