Lecture 13

Error and One-Sample T-Tests

Today’s Updates / Reminders

- PRWA #1 (Graph) – part 2 – closes at the end of the day… TODAY!
  - Recall that you should review each paper. If you just give it a 100 thinking you’ll be a “generous grader” and help out your classmates, you will likely lose points since the accuracy of your grading depends on how you compare to your peer graders. Also being too harsh negatively impacts some students.
- Homework 6 is due at the end of the day TODAY.
- Lab 5 using Excel is due on Friday.
- Next week is chill! 😎 Only assignment is Lab 6, opens today, due next Friday.
- Last day to drop is coming up… Wednesday, March 25. Come see me if you want to talk over this decision.

Z vs. T

Summary Confidence Intervals and Hypothesis Tests

* If df = n-1 is not on table→ round down df

σ is known: Z-test	σ is not known: T-test
From sample (size n) we find $\bar{x}$	From sample (size n) we find $\bar{x}$ and s
$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \to N(0,1)$ (standard normal distr.)	$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \to t(n-1)$ (t-Distribution, df = n-1)
CI with C% confidence level: $\left(\bar{x} - z^* \frac{\sigma}{\sqrt{n}}, \bar{x} + z^* \frac{\sigma}{\sqrt{n}}\right)$ z^* : from bottom t-table, column C%	CI with C% confidence level: $\left(\bar{x} - t^* \frac{s}{\sqrt{n}}, \bar{x} + t^* \frac{s}{\sqrt{n}}\right)$ t^* : from t-table, row df = n-1, column C% if df = n-1 is not in t-table → round down df
Remember the 4 steps for HT, use z as test-statistic and the standard normal table (table A) to calculate the P-value if P-value < α → reject H_0; if P-value > α → don't reject H_0	Remember the 4 steps for HT, use t as test-statistic and the t-table (row df = n-1) to calculate the P-value if df = $n-1$ is not on t-table → round down df if P-value < α → reject H_0; if P-value > α → don't reject H_0
$H_a: \mu < \mu_0$ → P-value = P(Z < z) → z-table	$H_a: \mu < \mu_0$ → P-value = P(T < t) if t < 0 P-value = P(T < t) = P(T > | t |) → t-Table, row df = n-1 if t > 0 P-value = P(T < t) = 1- P(T > t) → t-Table, row df = n-1 *
$H_a: \mu > \mu_0$ → P-value = P(Z > z) = 1 - P(Z < z) → z-table	$H_a: \mu > \mu_0$ → P-value = P(T > t) if t < 0 P-value = P(T > t) = 1 – P(T > | t |) → t-table, row df = n-1 if t > 0 P-value = P(T > t) → t-table, row df = n-1 *
$H_a: \mu \neq \mu_0$ → if z < 0: P-value = 2 P(Z < z) → z-table if z > 0: P-value = 2 P(Z < -z) → z-table	$H_a: \mu \neq \mu_0$ → if t < 0 P-value = 2 P(T > | t |) → t-table, row df = n-1 if t > 0 P-value = 2 P(T > t) → t-table, row df = n-1 *

Using SALT for z and t

Finding z/t or area using distribution calculators

Homework 6, #13, 14

13. [–/5 Points] DETAILS DEVORESTAT9 7.3.028.MI.S.

Determine the values of the following quantities. (Round your answers to three decimal places.)

USE SALT

(a) $t_{0.10,\ 12}$

(b) $t_{0.05,\ 12}$

14. [–/6 Points] DETAILS DEVORESTAT9 7.3.029.S.

Determine the t critical value(s) that will capture the desired t-curve area in each of the following cases. (Assume that central areas are centered at $t = 0$. Round your answers to three decimal places.)

USE SALT

(a) Central area = 0.95, df = 15

±

(b) Central area = 0.95, df = 30

±

The notation explained:

$t_{\alpha,df}$ = t for a given alpha and degrees of freedom

Section 8.1 (continued)

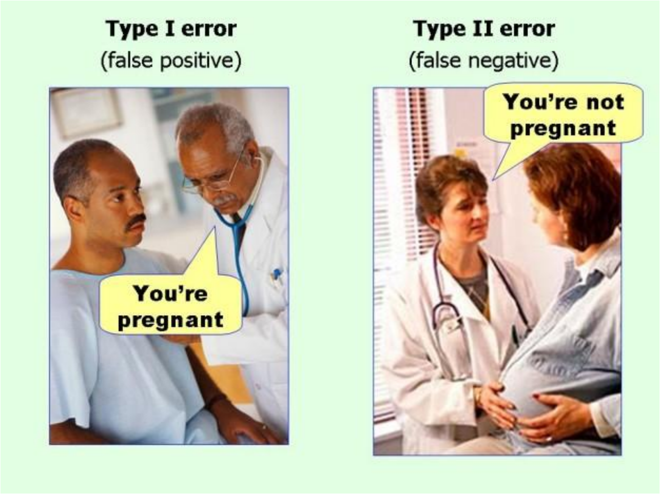
Hypothesis Tests and Error

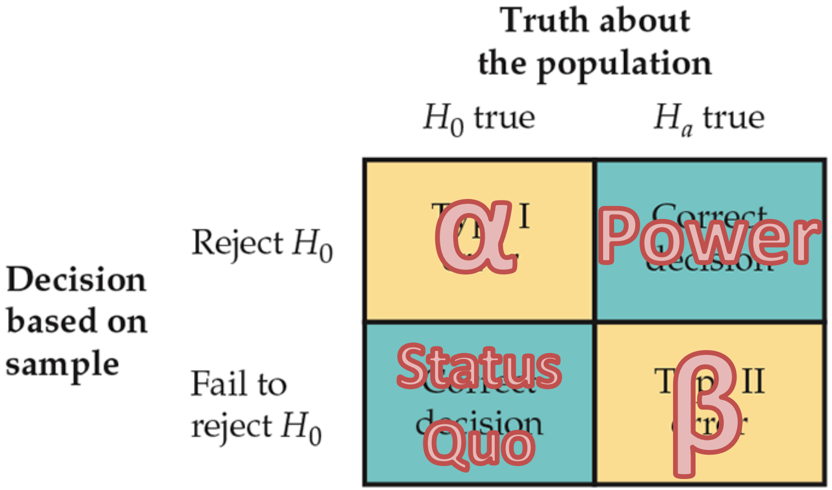
Error

- When we do hypothesis testing, there is a chance we make a mistake in our conclusions.
- If we think of our hypothesis conclusions in the frame of a court of law, we have two possible undesirable outcomes:
  - An innocent person goes to prison
  - A guilty person walks free
- In fact, you can spend some time meditating over a cup of tea about the ethics of this. For example, El Salvador has a new president that prefers the former error and Honduras has followed suit.

Error Types

- If we reject H_0 when H_0 is true, we have committed a Type I error.
- If we fail to reject H_0 when H_0 is false, we have committed a Type II error.





Hypothesis Test Example

- Your brand new car is supposed to get 35 miles per gallon. You have measured your fuel mileage on 10 tanks of gas and you are getting 31.9 mpg with a standard deviation of 5.3. This seems low… but is it really that bad? But, it is the average of 10 tanks tho. Use significance level of 0.05.
  - Write the null and alternative hypotheses.

Hypothesis Test Example

$\mu = 35$, $\bar{x} = 31.9$, $s = 5.3$ (in fact, let's assume this is σ)

- Calculate the test statistic.

- Find the p-value.

Hypothesis Test Example

- While you're looking at the table, we said $\alpha = 0.05$. What would z have to be in order for the p-value to be ≤ 0.05?

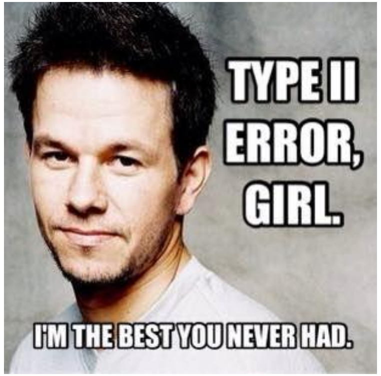
- What is your conclusion in technical terms?

- Now, you're ready to go get snooty with the car salesman. Given that we rejected H_0, describe the Type I error. Describe the Type II error.

Alpha

- The probability of a Type I error is the probability of rejecting H_0 when it really is true. **This is exactly the significance level of the test.**
- The significance level α of any fixed-level test is the probability of a Type I error. That is, α is the probability that the test will reject the null hypothesis H_0 when H_0 is, in fact, true.
- This value is chosen from the outset (Step 0!) and cannot be modified by changing sample size, decreasing standard deviation, or changing your hypotheses.
- Consider the consequences of a Type I error before choosing a significance level. By this, we mean:
  - "What if an error would cause a loss of human life?" Very low α
  - "What if an error caused the snack company to not stock enough Doritos in the snack machine?" 'Less low' α.
  - Common values for α are 0.01, 0.05, and 0.10… 0.05 is most common.

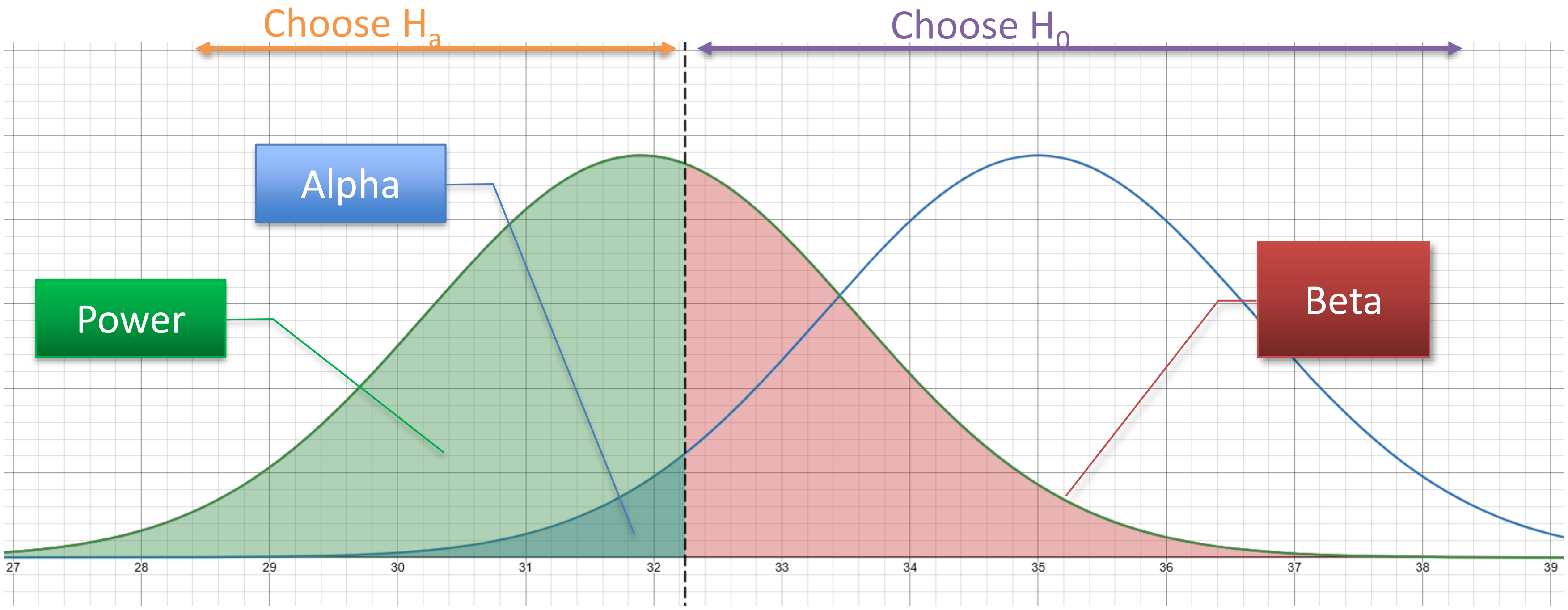
Beta



- A significance test makes a Type II error when it fails to reject a null hypothesis that really is false (β). There are many values of the parameter that satisfy the alternative hypothesis.
- The easiest way to reduce β is to increase sample size.
- We can calculate the probability that a test does reject H_0 when any specific alternative is true. This probability is called the **power** of the test against that specific alternative.
- In the Wienermobile example, we can show the probability that I choose $\mathbf{H_0}$ when I should have chosen H_a was 0.4189. So what is the probability that I choose $\mathbf{H_a}$ when I <u>should have</u> chosen H_a? P(A) = P(A|B) + P(A|B')

Alpha, Beta, and Power

Desmos



Beta (β)

- $\mu = 35$, $\bar{x} = 31.9$, $\sigma = 5.3$, $n = 10$
- Steps to find β
  1. Using alpha, find critical z for H_0. We call this z_α
  2. Using $x = z_\alpha \frac{\sigma}{\sqrt{n}} + \mu$, find critical x.
  3. Use critical x to find the z-score with respect to the "new" μ (μ_A or H_a). We call this z_β. Use: $z_\beta = \frac{x_{crit} - \mu_A}{\sigma/\sqrt{n}}$
  4. Find the area on the H_0 side of z_β

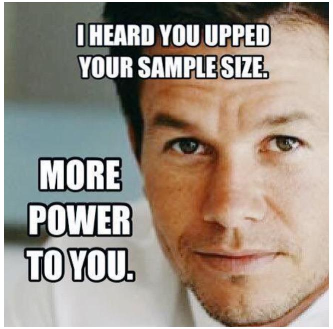
Increasing the Power / Decreasing β

- Suppose we have performed a power calculation and found that the power is too small. There are four ways to increase power:

1. **Increase the significance level α.** It is easier to reject a null hypothesis with a larger α level.
2. **Consider another value for μ that is farther from the null value.** Values of μ that are farther from the hypothesized value are easier to detect.
3. **Increase the sample size.** More data will provide better information about the sample average, so we have a better chance of distinguishing values of μ.
4. **Decrease σ.** Improving the measuring process and restricting attention to a subpopulation are possible ways to decrease σ.

Sample Size Determination

- We see sample size is the easiest way to get power up.
- How large a sample would be required so that the power is at least 0.9?
- Sample size required for an α-level test with power = 1 – β



For 1-tailed test, $n = \frac{(Z_\alpha + Z_\beta)^2 \cdot \sigma^2}{(\mu_A - \mu_0)^2}$

For 2-tailed test, $n = \frac{(Z_{\alpha/2} + Z_\beta)^2 \cdot \sigma^2}{(\mu_A - \mu_0)^2}$

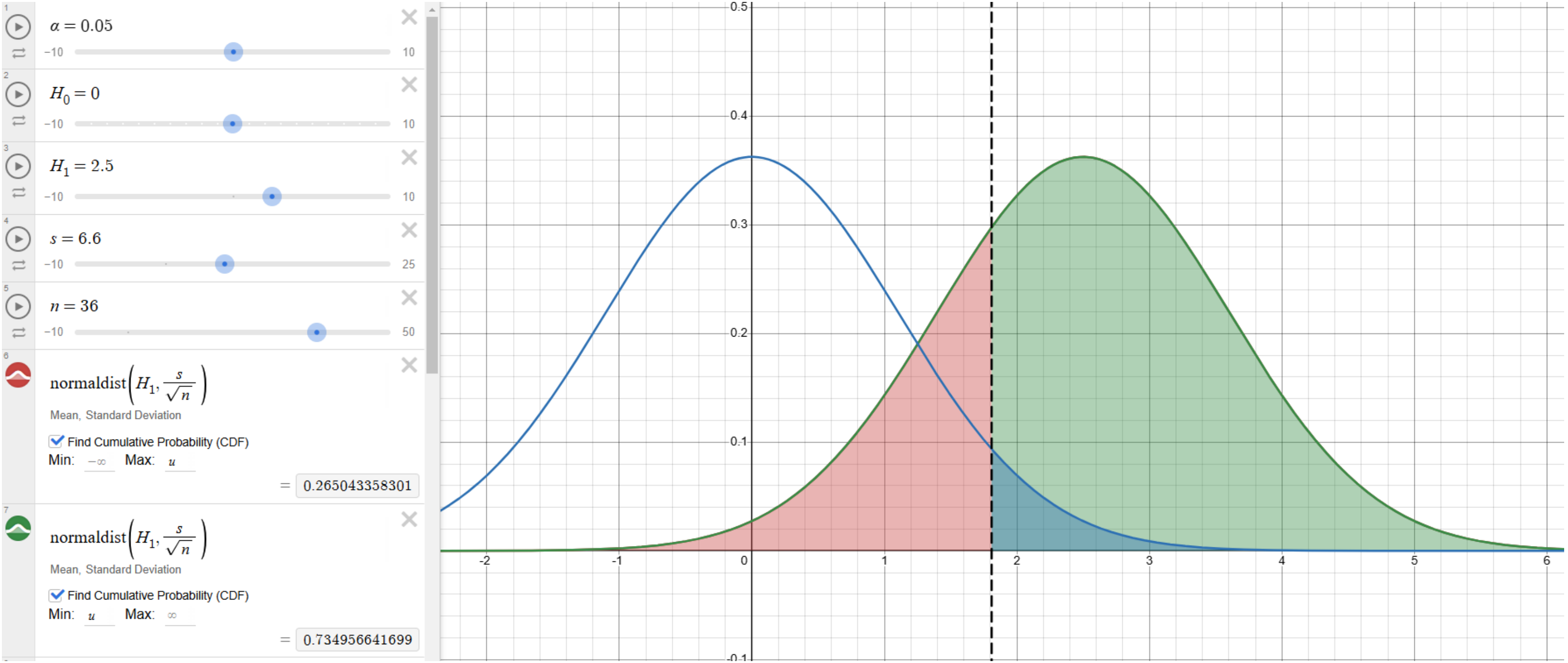
- where μ_0 and μ_A are the hypothetical values for the population mean under H_0 and H_A, respectively.

	Critical z (Two Tails)	Critical z (One Tail)
α = 0.10	z = ±1.645	z = -1.28 or z = 1.28
α = 0.05	z = ±1.96	z = -1.645 or z = 1.645
α = 0.01	z = ±2.576	z = -2.33 or z = 2.33

Power Calculation Example

- DJ Burns wishes to conduct the following test: H_0: $\mu \leq 0$ vs. H_a: $\mu > 0$ using $\alpha = 0.05$.
- Assume that the population SD $\sigma = 6.6$ and $n = 36$.
- What is the power of the test to detect a true $\mu = 2.5$? Also stated as $\beta(2.5)$.

Desmos Results



Alpha / Beta / Power Using R

NC STATE UNIVERSITY ST 370

Power Calculation Example

- DJ Burns wishes to conduct the following test: H_0: $\mu \leq 0$ vs. H_a: $\mu > 0$ using $\alpha = 0.05$.
- Assume that the population SD $\sigma = 6.6$ and $n = 36$.
- What is the power of the test to detect a true $\mu = 2.5$? Also stated as $\beta(2.5)$.

```
     One-sample t test power calculation

              n = 36
          delta = 2.5
             sd = 6.6
      sig.level = 0.05
          power = 0.7202509
    alternative = one.sided
```

We talked about how we must use σ to avoid using numerical methods to find beta or power. Using R, you can quickly run this calculation using t-procedures.

Go to R and try these commands:

power.t.test(n=36, delta=2.5, sd=6.6, sig.level=0.05, type='one.sample’, alternative='one.sided')

Chapter 8.3

The One-Sample T-Test

Cucumbers Again…

- We built a confidence interval for this problem in lecture 11.

One-sample mean HT, *t*-table approach

En.wikipedia.org

A study is carried out to measure the production of cucumbers in a farm. You are told that the mean production of cucumbers produced in the whole region is 42. Studying **10** plots of land, it is found that the **mean** number of cucumbers produced is **34** with a **standard deviation** of **12.75.** Conduct a hypothesis test with **α = 0.05** to determine if the mean production of cucumbers in the whole region is **less than 42.**

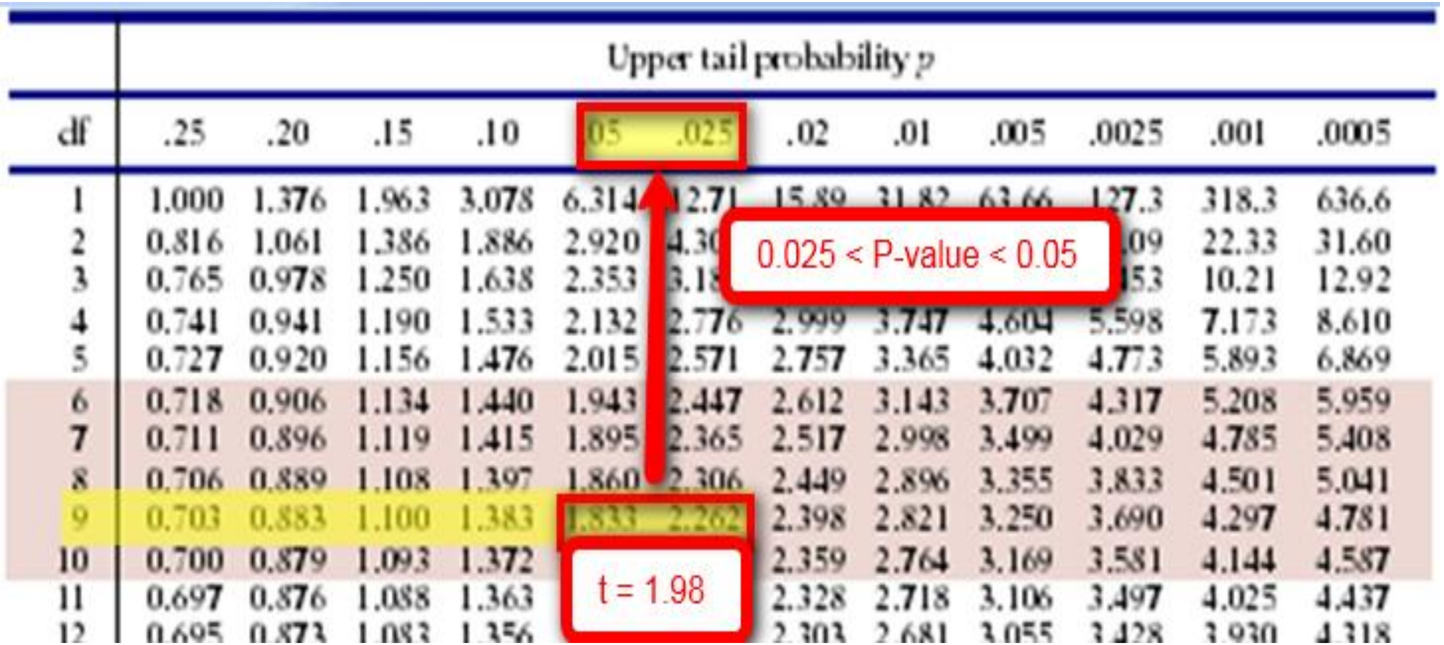
1

2

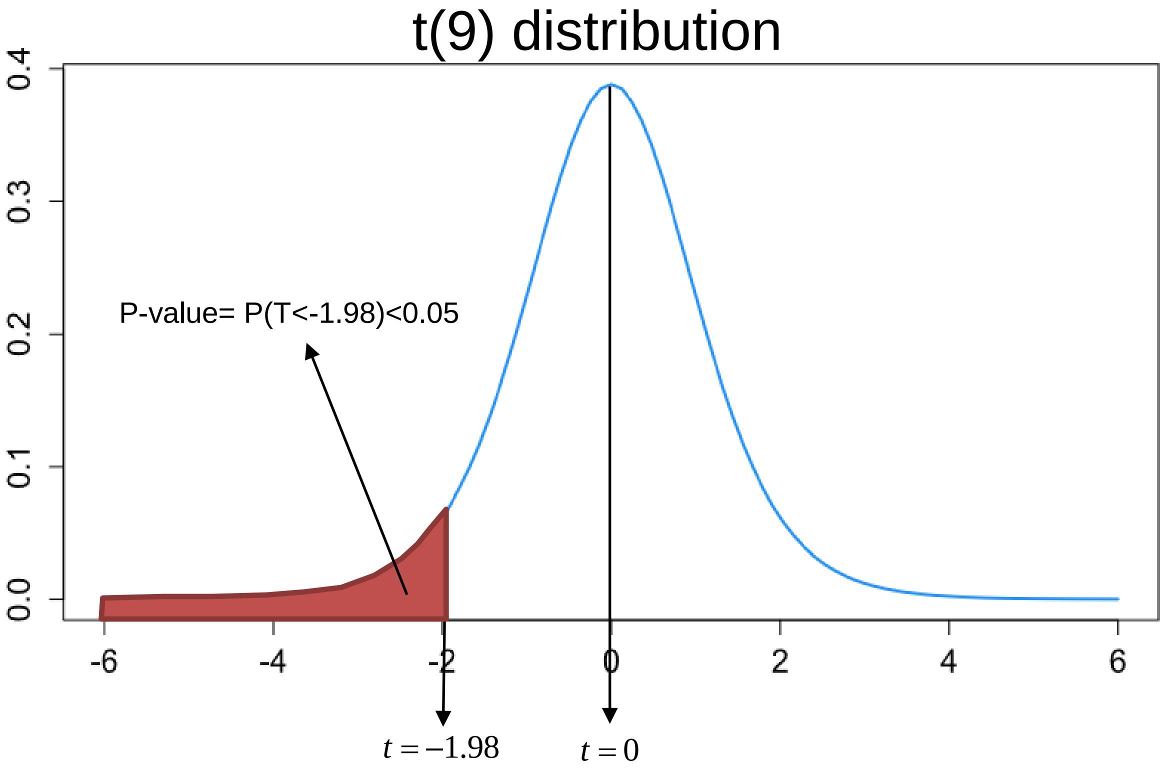
3

4

Using the t table to get a P-value



	Upper tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.30	[illegible]	[illegible]	[illegible]	[illegible]	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.18	[illegible]	[illegible]	[illegible]	[illegible]	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	[illegible]	[illegible]	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	[illegible]	[illegible]	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	[illegible]	[illegible]	2.303	2.681	3.055	3.428	3.930	4.318


t(9) distribution
P-value= P(T<-1.98)<0.05
0.0
0.1
0.2
0.3
0.4
-6
-4
-2
0
2
4
6
t = −1.98
t = 0

Can we use this CI for the HT?

Radon Detectors Again…

- We also built a confidence interval for this problem in lecture 11…

1-sample mean test and CI by hand and with SPSS

Is there convincing evidence that the mean reading of all detectors of this type differs from the true value of 105? Use α = 0.10 for a hypothesis test. Carry out a test in detail and write a brief conclusion.

The mean and standard deviation for this sample of 12 detectors are **104.13** and **9.40**, respectively.

91.9	97.8	111.4	122.3	105.4	95.0
103.8	99.6	119.3	104.8	101.7	96.6

Remember:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

1

2

3

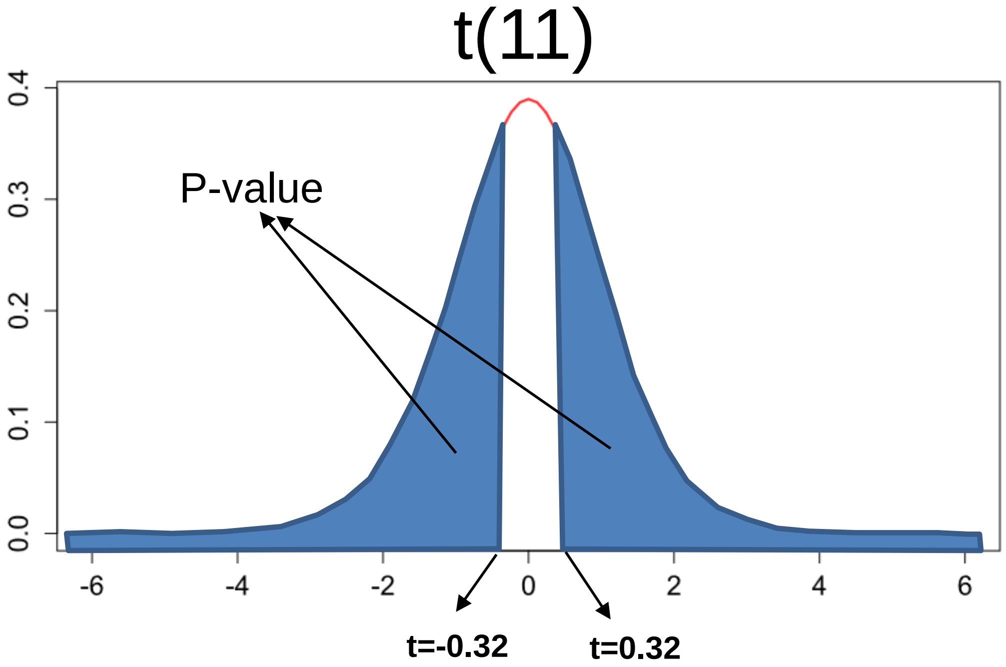
4

Using the t table for a P-value

P(t > 0.32) > 0.25

t = 0.32

	Upper tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221


t(11)
P-value
0.0
0.1
0.2
0.3
0.4
-6
-4
-2
0
2
4
6
t=-0.32
t=0.32

Can we use this CI for the HT?

Guidelines for One Sample t-Test
When is it ok to use the t procedures?

Sample size	Use t procedures
n < 15	if your data looks like a Normal distribution (you can check this with a Normal Quantile Plot)
15 ≤ n < 40	except in the presence of outliers or strong skewness. Some skewness is ok!
n ≥ 40	except in the presence of outliers, ok even if data are very skewed.

Use different procedure

Investigate the cause: data wrongly recorded, equipment malfunction, Response bias