Inference for Means: t-Distribution

Sections 6.4 - 6.6, 6.10 - 6.13
- Single Mean (6.4)
- Intervals and tests (6.5, 6.6)
- Difference in means
- Distribution, intervals and tests (6.10 - 6.12)
- Matched pairs (6.13)
- Intervals and tests (6.8, 6.9)
- Correlation

Comments
- Re-grade requests due today. Hand them to me in person or slide them under my office door.
- My Monday office hours are being permanently switched to Tuesday, 1 - 2:30. This change will be updated on the website.
- Both Heather and Sam have office hours Monday
- If you can't make my office hours, I am always available to talk after class.

Inference Using N(0,1)

If the distribution of the sample statistic is normal:

A confidence interval can be calculated by

\[ \text{sample statistic} \pm z^* \times SE \]

where \( z^* \) is a N(0,1) percentile depending on the level of confidence.

A p-value is the area in the tail(s) of a N(0,1) beyond

\[ z = \frac{\text{sample statistic} - \text{null value}}{SE} \]

CLT for a Mean

Population

Distribution of Sample Data

Distribution of Sample Means

SE of a Mean

The standard error for a sample mean can be calculated by

\[ SE = \frac{\sigma}{\sqrt{n}} \]

Standard Deviation

The standard deviation of the population is

a) \( \sigma \)
b) \( s \)
c) \( \frac{\sigma}{\sqrt{n}} \)
**Standard Deviation**

The standard deviation of the sample is

a) $\sigma$

b) $s$

c) $\frac{\sigma}{\sqrt{n}}$

**Standard Deviation**

The standard deviation of the sample mean is

a) $\sigma$

b) $s$

c) $\frac{\sigma}{\sqrt{n}}$

*The standard error is the standard deviation of the statistic.*

---

**CLT for a Mean**

If $n \geq 30^*$, then

$$\bar{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

*Smaller sample sizes may be sufficient for symmetric distributions, and 30 may not be sufficient for very skewed distributions or distributions with high outliers.

---

**Standard Error**

$$SE = \frac{\sigma}{\sqrt{n}}$$

- We don’t know the population standard deviation $\sigma$, so estimate it with the sample standard deviation, $s$

$$SE = \frac{s}{\sqrt{n}}$$

---

**t-distribution**

- Replacing $\sigma$ with $s$ changes the distribution of the z-statistic from normal to $t$

- The $t$ distribution is very similar to the standard normal, but with slightly fatter tails to reflect this added uncertainty

---

**Degrees of Freedom**

- The $t$-distribution is characterized by its *degrees of freedom (df)*

- Degrees of freedom are calculated based on the sample size

- The higher the degrees of freedom, the closer the $t$-distribution is to the standard normal
Normality Assumption

- Using the t-distribution requires an extra assumption: the data comes from a normal distribution.
- Note: this assumption is about the original data, not the distribution of the statistic.
- For large sample sizes we do not need to worry about this, because s will be a very good estimate of \( \sigma \) and \( t \) will be very close to N(0,1).
- For small sample sizes (\( n < 30 \)), we can only use the t-distribution if the distribution of the data is approximately normal.

Small Samples

- If sample sizes are small, only use the t-distribution if the data looks reasonably symmetric and does not have any extreme outliers.
- Even then, remember that it is just an approximation.
- In practice/life, if sample sizes are small, you should just use simulation methods (bootstrapping and randomization).

Confidence Intervals

\[
\text{sample statistic} \pm t^* \times SE
\]

\[
\bar{X} \pm t^* \times \frac{s}{\sqrt{n}}
\]

\( df = n - 1 \)

\( t^* \) is found as the appropriate percentile on a t-distribution with \( n - 1 \) degrees of freedom.

IF \( n \) is large or the data is normal.
Hypothesis Testing

\[ t = \frac{\text{sample statistic} - \text{null value}}{\text{SE}} \]

\[ H_0: \mu = \mu_0 \]

\[ df = n - 1 \]

The p-value is the area in the tail(s) beyond \( t \) in a \( t \)-distribution with \( n - 1 \) degrees of freedom, IF \( n \) is large or the data is normal.

Chips Ahoy!

A group of Air Force cadets bought bags of Chips Ahoy! cookies from all over the country to verify this claim. They hand counted the number of chips in 42 bags.

\[ \bar{X} = 1261.6, \ s = 117.6 \]


Can we use hypothesis testing to prove that there are 1000 chips in every bag? ("prove" = find statistically significant)

(a) Yes
(b) No

Can we use hypothesis testing to prove that the average number of chips per bag is 1000? ("prove" = find statistically significant)

(a) Yes
(b) No

Statements of equality are always in the null hypothesis, and we can only reject or fail to reject the null, never accept the null. Therefore, we cannot use hypothesis testing to prove equality.

Can we use hypothesis testing to prove that the average number of chips per bag is more than 1000? ("prove" = find statistically significant)

(a) Yes
(b) No

Chips Ahoy!

Can we use hypothesis testing to prove that there are more than 1000 chips in each bag, on average?

(a) Yes
(b) No
(c) Cannot tell from this data

2) Give a 99% confidence interval for the average number of chips in each bag.

\[ \bar{X} = 1261.6, \ s = 117.6, \ n = 42 \]
### Chips Ahoy!

1. State hypotheses:
   - $H_0 : \mu = 1000$
   - $H_a : \mu > 1000$

2. Check conditions:
   - $n = 42 \geq 30$ ✔️

3. Calculate test statistic:
   $$ t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1261.6 - 1000}{117.6/\sqrt{42}} = 14.4 $$

4. Compute p-value:
   - $p - value = 0$

4. Interpret in context:
   - This provides extremely strong evidence that the average number of chips per bag of Chips Ahoy! cookies is significantly greater than 1000.

This is 99% confident that the average number of chips per bag of Chips Ahoy! cookies is between 1212.6 and 1310.6 chips.

---

### SE for Difference in Means

$$ SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} $$

$df = \text{smaller of } n_1 - 1 \text{ and } n_2 - 1$

---

### t-distribution

For a difference in means, the degrees of freedom for the $t$-distribution is the smaller of $n_1 - 1$ and $n_2 - 1$

The test for a difference in means using a $t$-distribution is commonly called a **t-test**
### The Pygmalion Effect

Teachers were told that certain children (chosen randomly) were expected to be "growth spurters," based on the Harvard Test of Inflected Acquisition (a test that didn't actually exist). These children were selected randomly.

The response variable is change in IQ over the course of one year.


---

### Pygmalion Effect

1. State hypotheses:
   
   \[ H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_A : \mu_1 - \mu_2 > 0 \]

   - Mean IQ change for "growth spurters"
   - Mean IQ change for control students

2. Check conditions:
   
   \[ 255 \geq 30,65 \geq 30 \]

3. Calculate t-statistic:
   
   \[ t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{12.22 - 8.42}{3.8 \sqrt{\frac{1}{65} + \frac{1}{255}}} = 2.096 \]

   \[ s_p \approx \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(255 - 1)12.0^2 + (65 - 1)13.3^2}{255 + 65 - 2}} \]

4. Compute p-value:
   
   \[ p \approx 0.02 \]

5. Interpret in context:
   
   We have evidence that positive teacher expectations significantly increase IQ scores in elementary school children.

---

### Matched Pairs

- A matched pairs experiment compares units to themselves or another similar unit, rather than just compare group averages
- Data is paired (two measurements on one unit, twin studies, etc.)
- Look at the difference in responses for each pair

---

### From the paper:

"The difference in gains could be ascribed to chance about 2 in 100 times"

---

* \( s_1 \) and \( s_2 \) were not given, so I set them to give the correct p-value

---

We are 95% confident that telling teachers a student will be an intellectual "growth spurter" increases IQ scores by between 0.17 and 7.43 points on average, after 1 year.
**Pheromones in Tears**

- Do pheromones (subconscious chemical signals) in female tears affect testosterone levels in men?
- Cotton pads had either real female tears or a salt solution that had been dripped down the same female’s face
- 50 men had a pad attached to their upper lip twice, once with tears and once without, order randomized.
- Response variable: testosterone level


**Matched Pairs**

Why do a matched pairs experiment?

- Increase the power of the test
- Decrease the margin of error for intervals
- All of the above

Using matched pairs decreases the standard deviation of the response, which increases the power of tests and decreases the margin of error for intervals.

**Matched Pairs**

- Matched pairs experiments are particularly useful when responses vary a lot from unit to unit
- We can decrease standard deviation of the response (and so decrease standard error of the statistic) by comparing each unit to a matched unit

**Matched Pairs**

- For a matched pairs experiment, we look at the difference between responses for each unit, rather than just the average difference between treatment groups
- Get a new variable of the differences, and do inference for the difference as you would for a single mean

---

Pheromones in Tears

- The average difference in testosterone levels between tears and no tears was -2.17 pg/ml.
  - "pg" = picogram = 0.001 nanogram = 10^-12 gram
- The standard deviation of these differences was 46.5
- Average level before sniffing was 155 pg/ml.
- The sample size was 50 men
- Do female tears lower male testosterone levels?
  - (a) Yes  (b) No  (c) ???
- By how much? Give a 95% confidence interval.

---

**Pheromones in Tears**

1. State hypotheses: $H_0: \mu_d = 0$ vs $H_1: \mu_d < 0$

2. Check conditions: $n = 50 \geq 30$

3. Calculate test statistic: $t = \frac{\bar{x}_d - \mu_0}{SE} = \frac{-21.7}{6.58} = -3.3$

4. Compute p-value: `p = pt(-3.3, df=49, tail="lower")` $= 0.000354$

5. Interpret in context:
   - This provides strong evidence that female tears decrease testosterone levels in men, on average.
Pheromones in Tears

1. Check conditions: $n = 50 \geq 30$ 
   
   $\bar{X} = -21.7$
   
   $SE = 6.58$

2. Find the $t^*$ value 
   
   $t^* = 2$

3. Compute the confidence interval: 
   
   $-21.7 \pm 2(6.58) = (-34.86, -8.54)$

4. Interpret in context: 
   
   We are 95% confident that female tears on a cotton pad on a man’s upper lip decrease testosterone levels between 8.54 and 34.86 pg/ml, on average.

Social Networks and the Brain

- Is the size of certain regions of your brain correlated with the size of your social network?
- Social network size measured by many different variables, one of which was number of Facebook friends. Brain size measured by MRI.
- The sample correlation between number of Facebook friends and grey matter density of a certain region of the brain (left middle temporal gyrus), based on 125 people, is $r = 0.354$. Is this significant? 
  
  (a) Yes  
  
  (b) No


Correlation

$SE = \sqrt{\frac{1 - r^2}{n - 2}}$

$t$-distribution

$df = n - 2$

Social Networks and the Brain

1. State hypotheses: 
   
   $H_0: \rho = 0$
   
   $H_a: \rho \neq 0$

2. Check conditions: $n = 125 \geq 30$

3. Calculate test statistic: 
   
   $t = \frac{r}{SE} = \frac{0.354}{0.084} = 4.20$

4. Compute p-value: 
   
   $p-value = 0.00005$

5. Interpret in context: 
   
   This provides strong evidence that the size of the left middle temporal gyrus and number of Facebook friends are positively correlated.

To Do

- Read Chapter 6
- Do Homework 5 (due Tuesday, 10/30)