Introduction to Statistics I

Textbook: Elementary Statistics (4th Edition, by Navidi and Monk), and Workshop Statistics (4th Edition, by Rossman and Chance).

Previous Lecture

- Intuition for confidence intervals for quantitative vars
- Estimated *se* when we don't know σ: $\overline{se} = \frac{s}{\sqrt{n}}$
- t-distr, degrees of freedom (df), critical values t^*_{n-1}
- Conf Interval is: $\overline{x} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right)$ (if tech conds met)

Class Notices

- Exam2 Grades on Canvas/GradeScope
- Tutoring at QRC?.

---

§9.1 - 9.3: Hypothesis Tests for Population Mean

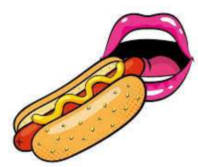
So you've got a hypothesis that challenges conventional wisdom, and you've sampled a population to prove it.

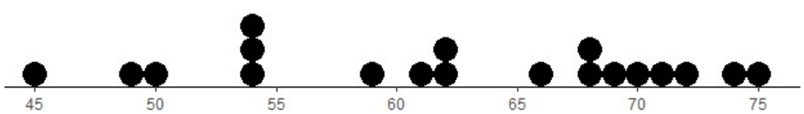
Let's put conventional wisdom on trial!

A **hypothesis test** is a procedure which:

- Presents a hypothesis,
- Systematically examines the evidence,
- And makes a decision about the hypothesis using the evidence.

Example: Nathan's Hot Dog eating contest has competitors eat as many hot dogs as possible in ten minutes. A dot plot below shows how many hot dogs the winner ate for 19 years from 2002-2020.





hot dogs eaten in 10 mins by the winner

We can view this as a **sample** of the greater population of winners of Nathan's hot dog contest.

Example RQ: Is the average # of hot dogs the contest winners eat in 10 mins greater than 60?

Hypothesis Test

A **hypothesis test** is a six-step procedure to test the likelihood of a hypothesis.

It Puts a hypothesis on trial, where the sample & it's statistic are the evidence.

Is the evidence strong enough to support the hypothesis?

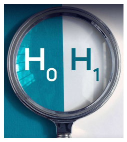
We will go over...

Steps of a Hypothesis Test for Quantitative Data

1. **Lay out the problem and check technical conditions**. 

   State parameter, variable, and population.

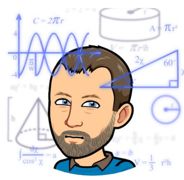
   Check SRS and if population normal or sample size at least 30.

2. **State hypotheses**. 

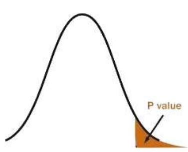
   What's conventional wisdom? What are you trying to show? Null & Alternative hypotheses.

3. **Set Significance-level** 

   How sure do you want to be that you're right? $\alpha = 0.05$?

4. **Calculate test statistic** (t/z-score). 

   Sample the population, then check how many SDs your sample mean is from the assumed parameter.

5. **Find *P*-value**. 

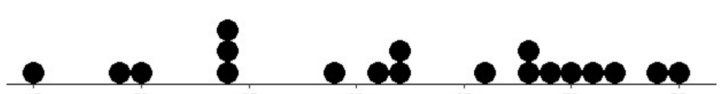
   The *P*-value is the probability of observing your sample mean or something more extreme (assuming the null is true).

6. **State Conclusion**: Evidence, Decision, In-Context Summary. 

Step 1: Lay out the problem and check tech conditions. 

Define the: **Parameter** of interest (what symbol?). **Variable** to collect. **Population** to sample from.

Example RQ: Is the average # of hot dogs the contest winners eat in 10 mins greater than 60?



To examine this, let's do a hypothesis test. First, lay out the problem:

Parameter (Symbol?), Variable (Variable Type?), Population

Parameter: Average # of hot dogs competition winners can eat in 10 mins. **Symbol**: μ.

Variable: # hot dogs eaten in 10 mins. **Variable Type**: Quantitative.

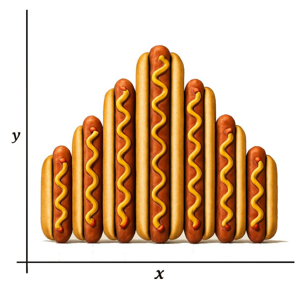
Population: All competition winners (past and future).

Check Technical Conditions

Hypothesis testing relies on CLT, so we must do a tech check.

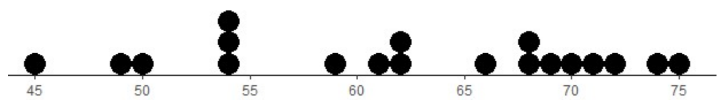
Recall: CLT (for means) results hold if sampling is SRS, and also:

Population data is normal, *OR* Sample size is at least 30.



Back to Example: sample size is 19, and SRS isn't mentioned.

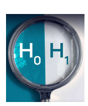
$19 \ngtr 30$, so this test is only valid if # of hot dogs eaten follows normal distr.



Based on the dot plot above, it's *not inconceivable* that this sample came from a normal distr.

We need to make it **explicit** that we're making this **assumption** as we go forward.
"We assume the pop. is normal."
Also, "Since this wasn't a SRS, we'll just assume the sampling doesn't generate bias." 

Step 2: State Hypotheses



In a hypothesis test, there are two hypotheses!

Null Hypothesis (H_0): current (conventional) view of parameter (null is innocent/true). 

Alternative Hypothesis (H_1): alternative view of parameter we're hoping to show (null is guilty/false). 

For null hypothesis (H_0), parameter μ is equal to some value μ_0, denoted: $H_0 : \mu = \mu_0$.

For alternative hypothesis (H_1), parameter μ is either less (<), more than (>), or not equal to ($\neq$) **that same value** (μ_0).

Back to Example: Can we use the hot dog data to conclude the average # of hot dogs that competition winners eat in 10 mins is greater than 60?

Null Hypothesis. $H_0 : \mu = 60$.

Alternative Hypothesis (H_1). Asserts μ is either:

- Less than μ_0,
- More than μ_0,
- Or not equal to μ_0.

Alternative hypothesis looks like one (and only one) of the following:

$$H_1 : \mu > \mu_0, \qquad H_1 : \mu < \mu_0, \qquad H_1 : \mu \neq \mu_0.$$

Back to Example:

$$H_1 : \mu > 60$$

One Sided vs Two Sided Tests

When hypothesis tests use an alternative hypothesis of:

$H_1 : \mu > \mu_0$ OR $H_1 : \mu < \mu_0$, it's called a **one-sided test**.

If hyp. test uses $H_1 : \mu \neq \mu_0$, it's called a **two-sided test**.

Step 3: Set Sig-Level

Before beginning the experiment, we need to decide how small the P-value would need to be for us to reject H_0.

The P-value below which we reject H_0 is called the **significance level** α.

The default is $\alpha = 0.05$. Assume this, unless told otherwise.

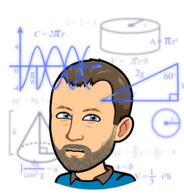
If P-value $< \alpha$, **reject null** H_0 and conclude alternative H_1.

If P-value $> \alpha$, **fail to reject** H_0 and fail to conclude H_1. (but don't conclude H_0 is true!)

In some areas of research, higher or lower significance levels are used.

In Our Example: we'll use $\alpha = 0.05$.

Step 4: Calculate Test Statistic (t/z-score)



The test statistic measures how far the observed statistic $\bar{x}$ is from the hypothesized parameter value μ_0.

If $\bar{x}$ is "**far**" from μ_0, then it's **likely** H_0 is **false**.

If $\bar{x}$ is "**near**" to μ_0 then it's **possible** H_0 is **true**.

Test Statistic Derivation

What's "far" or "near" is context dependent. However, we've seen how to create a consistent measure of distance by measuring it **in terms of SDs**. That is, the z-score.

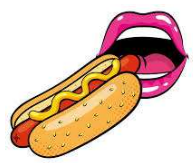
Note $\bar{x} - \mu_0$ is how far statistic $\bar{x}$ is from μ_0.

So, ideally the test statistic *should* be $z = \frac{\bar{x}-\mu_0}{\text{SD}} = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$. The distance fom $\bar{x}$ to μ_0 **in terms of SDs**.

But σ is unknown! So we replace σ w/the sample SD s.

So, we rewrite the test statistic as: $t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$. Also called a t-**score**.

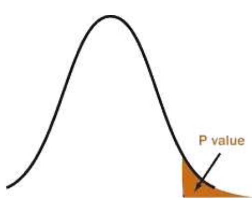
We use "t" instead of "z" to indicate we've estimated the SD σ w/the sample SD s.



Back to Example: $n = 19, \quad \bar{x} = 62.3, \quad s = 9.12$.

So: $t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} = \frac{62.3-60}{\frac{9.12}{\sqrt{19}}} \approx 1.099$.

Step 5: Find *P*-value



Instead of just looking at distance, we can ask:

If H_0 is correct, what's the probability P of observing this $\bar{x}$ (or something more surprising)?

This probability P is called the P-value.

We **can't** use the normal distr to calculate prob here because we estimated the SD with s.

Instead, we use the t-distr w/$n - 1$ degrees of freedom.

Back to Example: $t = 1.099$. Our sample size is 19.

What are the degrees of freedom?

bit.ly/introstatsdata

t-distr Calculator

$df = 19 - 1 = 18$.

What's probability of something greater?

We find: $P = 0.143$.

To find P-value:

- ♦ Input t-score and degrees of freedom.

- ♦ If H_1 <, then find "left tail" prob.

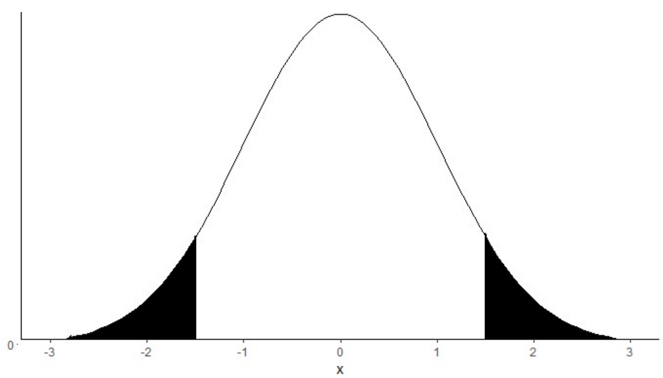
- ♦ If H_1 >, then find "right tail" prob.

- ♦ If $H_1 \neq$, then find "two-tail" prob.

P-value Derivation

The P-value that we calculate depends on the alternative hypothesis H_1.

- ▸ If $H_1 : \mu < \mu_0$, then we want probability of observing $\bar{x}$ or something lower.
- ▸ If $H_1 : \mu > \mu_0$, then we want probability of observing $\bar{x}$ or something greater.
- ▸ If $H_1 : \mu \neq \mu_0$, then we want probability of observing something at least as far from μ_0 as $\bar{x}$ is, but in either direction (more extreme than $\bar{x}$).



2-sided P-value

Interpretation. The P-value is the:

- ♦ Probability of observing your statistic or something more extreme (less than, greater than, or both),
- ♦ assuming the null hypothesis is true.

Back to Example: The prob of observing a sample mean of 62.3 or higher, assuming the population mean is 60, is 14.3%.

Step 6: State Conclusion 

We need to present **evidence** (P-value), make a **decision** (reject/fail to reject), and give an **in-context summary**. 

Have we enough evidence that the alternative H_1 is true (null H_0 is guilty/false )?

Or do we not have enough evidence (null H_0 *might* not be guilty ).

❗ We can never *conclude* the null (H_0) is true based on our data.

This is because $H_0 : \mu = \mu_0$ is an equality statement,
and statistics can't prove things are equal. (!?!)

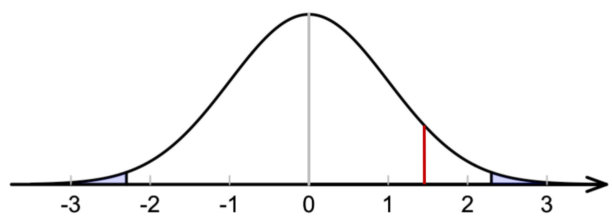
Decision: Given a significance level α:

if $P < \alpha$ - Reject null.

if $P > \alpha$ - Fail to reject null. (but don't say the null is true!)

- Large t leads to small P-value, indicating $\bar{x}$ is far away from μ_0.

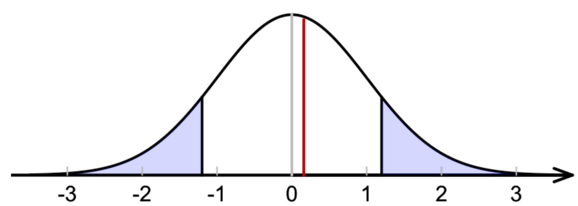
  Stat $\bar{x}$ is unlikely to be observed if H_0 is true.
  So, it's unlikely H_0 is true, and we **reject the null hypothesis**.



Large t (black line), μ_0 is grey line, $\bar{x}$ is red line, small P (purple).

- Small t leads to large P-value, indicating $\bar{x}$ is relatively close to μ_0.

  Stat $\bar{x}$ is frequently observed if H_0 is true. So $\bar{x}$ is consistent w/H_0.
  Thus H_0 *may* be true, so we **fail to reject the null hypothesis**.



Small t, large P.

- Small P: reject H_0 and conclude alternative H_1.
- Large P: fail to reject H_0 and fail to conclude H_1. (but **don't** conclude H_0 is true)

For our example the significance is $\alpha = 0.05$. Recall that $P = 0.143$. Decision?

"We **fail to reject** the null hypothesis." (it *could* be true!)

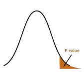
In-Context Summary

Write summary based on original RQ w/no statistical language.

Back to Example: Since $P > \alpha$, we **fail to reject the null hypothesis** and summarize that
"There's not enough evidence to conclude that the average # of hot dogs eaten by competition winners in 10 mins is more than 60."

If, instead P had been less than α, we would **reject the null hypothesis** and summarize that
"the average # of hot dogs eaten by competition winners in 10 mins *is* more than 60."

Recap: The Steps of a Hypothesis Test for Quantitative vars

1.  Lay out problem (param, var, pop) and check technical conditions.
2.  State hypotheses H_0, H_1
3.  Set Sig-level α
4.  Calculate test statistic (t/z-score)
5.  Calculate P-value
6.  State Conclusion: evidence, decision, in-context summary.

Activity: **9.3a**

Confidence Intervals and Hypothesis Testing

How do CIs relate to a hypothesis tests? Let's create a 95% CI for this hot dog example.

Recall: $n = 19$, $\bar{x} = 62.3$, $s = 9.12$, $df = ?$

$t^*_{18} = 2.101.$ (from t-distr calculator)

So, $62.3 \pm 2.101\left(\frac{9.12}{\sqrt{19}}\right)$.

And our 95% CI is: $(57.85,\ 66.74)$.

The CI contains 60. This is evidence that 60 is near the value of μ!

Do we fail to reject null hypothesis: $H_0 : \mu = 60$? (it depends upon H_1 and α)

CIs can predict the outcome of *two sided* hyp. tests: $H_1 : \mu \neq \mu_0$.

If the null hypothesized value μ_0 is in the interval, we fail to reject null ($H_0 : \mu = \mu_0$).

If μ_0 is outside the interval, we reject H_0.

CIs predict the outcome of a hyp. test if two things hold:

- Hyp. test is a "$\neq$" alternative hypothesis.
- Sig level α must "match" confidence level of CI.

E.g., a 95% CI predicts a test w/$\alpha = 0.05$.

A 99% CI predicts a test w/$\alpha = 0.01$.

A 90% CI predicts a test w/$\alpha = 0.10$.

Back to Example: Our 95% CI was: $(57.85,\ 66.74)$.

Assuming: $H_0 : \mu = 55$ and $H_1 : \mu \neq 55$. At $\alpha = 0.05$, what's the decision?

Because 55 not in CI, we **reject** null.

For: $H_0 : \mu = 65$, and $H_1 : \mu \neq 65$. At $\alpha = 0.10$, decision?

We need a 90% CI, not a 95% CI to know.

So, we can't make a decision. We must create a 90% CI.

What did we learn?

- Hypothesis test for quantitative vars
- CIs for two-sided hypothesis tests