

Unit 4: Introduction to Inference

Statistics S-100 Teaching Team

Summer 2024

Sampling variability

Confidence intervals

Hypothesis testing

The relationship between tests and confidence intervals

Sampling variability

STATISTICAL INFERENCE

The goal of **statistical inference** is to learn about a population based on the information in a single sample:

- Estimation is the framework for using a **sample statistic** to estimate a **population parameter**.
- Conducting a **hypothesis test** allows us to quantify the strength of evidence for a specific conjecture made about a population parameter.

This unit illustrates the principles of inference in the setting of estimating a population mean.

- Future units will extend inference to other settings.
- We will also discuss important caveats and common misunderstandings in future units.

YOUTH RISK FACTOR BEHAVIOR SURVEILLANCE SYSTEM (YRBSS)

- The YRBSS is a survey conducted by the US CDC to measure health-related activity in high school aged youth.
- 2.6 million high school students participated between 1991 and 2013.
- The yrbss data in the oibiostat package contains responses from 13,583 participants in 2013.

```
#load the data
```

```
library(oibiostat)
```

```
data("yrbss")
```

```
yrbss[c(1:3, 13582, 13583), c(1:3, 6:8)]
```

##	age	gender	grade	height	weight	helmet.12m
## 1	14	female	9	NA	NA	never
## 2	14	female	9	NA	NA	never
## 3	15	female	9	1.73	84.37	never
## 13582	17	female	12	1.60	77.11	sometimes
## 13583	17	female	12	1.57	52.16	did not ride

POPULATION PARAMETER VERSUS SAMPLE STATISTIC

Population parameter

The CDC was interested in estimating the health behaviors of the *target population*; i.e., all high school students in the US in 2013.

The population mean weight among all high school students in the US in 2013 is an example of a **population parameter**.

- Typically, the value of a population parameter is unknown.
- The symbol μ is used to denote a population mean.

Sample statistic

The descriptive measures for a sample are referred to as **sample statistics**.

The sample mean weight among the 13,572 students in yrbss is an example of a sample statistic.

- The value of a sample statistic is known.
- The symbol \bar{x} is used to denote a sample mean.

The sample mean is a **point estimate** of the population mean; i.e., a single value estimate of the population mean.

SAMPLING VARIABILITY

In nearly all studies, there is one target population and one sample.

- The CDC took a random sample of 13,572 high school students from the population of 21.2 million high school students in the US.

Suppose that a different random sample (of the same size) were taken from the same population:

- The sample could consist of different participants, thus. . .
- The sampled weight values could be different, producing a different sample mean weight than in the initial sample.

Sampling variability refers to the idea that the value of an estimate (i.e., a sample statistic) varies from sample to sample.

- The value of a sample statistic is not fixed until a specific sample is observed.
- The sample mean \bar{X} is a random variable. When we observe a specific sample mean \bar{x} , we are observing one possible value that \bar{X} can take on.

SAMPLING FROM A POPULATION

Typically, the exact values of population parameters are unknown.¹

But, we can observe the effect of sampling variability in a **hypothetical** setting where μ is known. This hypothetical setting allows us to understand how the distribution of \bar{X} behaves.

- Suppose that the 13,572 individuals in `yrbss` represent our target population.
 - ◊ Let mean weight in `yrbss` be the population parameter, μ_{weight} .
- Take a random sample from `yrbss` (e.g., $n = 30$) and calculate \bar{x}_{weight} , the mean weight among the sampled individuals.
 - ◊ How well does \bar{x}_{weight} estimate μ_{weight} ?
- Repeatedly take a random sample and compute the sample mean in order to construct a **sampling distribution of \bar{X}** .

¹Hence, the need for inference...

TAKING ONE RANDOM SAMPLE OF SIZE 30 FROM yrbss

```
#load the dataset  
library(oibiostat)  
data("yrbss")
```

```
#remove rows with missing values  
yrbss_complete <- yrbss[complete.cases(yrbss$weight), ]
```

```
#set parameters  
sample_size <- 30
```

```
#obtain random sample of row numbers  
set.seed(5011)  
sample_rows <- sample(1:nrow(yrbss_complete), sample_size)
```

```
#calculate point estimates from sampled rows  
mean(yrbss_complete$weight[sample_rows]); sd(yrbss_complete$weight[sample_rows])
```

```
## [1] 69.628
```

```
## [1] 18.26424
```

THE SAMPLE MEAN AS A RANDOM VARIABLE

Any sample statistic is a random variable, due to the nature of random sampling from a population.²

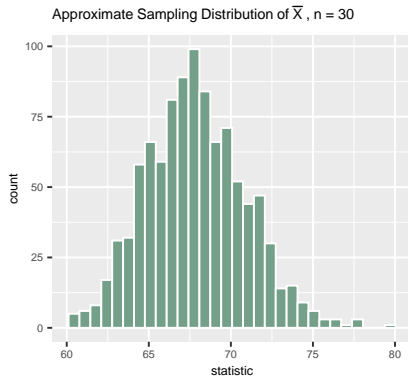
- The value of a sample statistic varies from sample to sample.
- The value of a sample statistic is not fixed until a specific sample is observed.

²Some additional formalism: Each observation X_1, X_2, \dots, X_n in a sample of size n can be thought of as a random variable. A function of random variables is itself a random variable, hence why \bar{X} is a random variable.

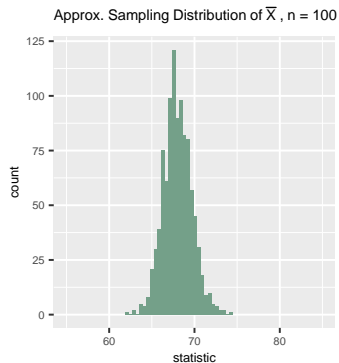
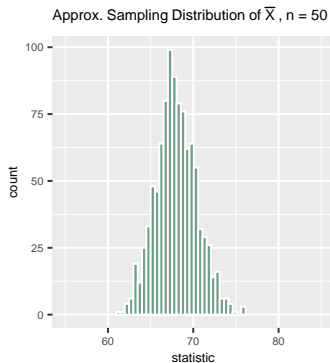
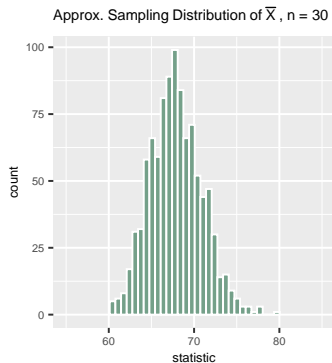
THE SAMPLE MEAN AS A RANDOM VARIABLE...

The **sampling distribution of \bar{X}** conveys the sample-to-sample variability in \bar{X} .

- The sampling distribution of \bar{X} is centered at μ ; i.e., $E(\bar{X}) = \mu$.
- The standard deviation of the sampling distribution, $SD(\bar{X})$, is referred to as the **standard error of the sample statistic**.
- The standard error is influenced by:
 - ◇ the underlying variation of the data
 - ◇ the sample size n



SAMPLING DISTRIBUTION OF A SAMPLE STATISTIC...



As n increases, the sampling distribution becomes more symmetric and bell-shaped, in addition to becoming more narrow (i.e., the variability decreases).

CENTRAL LIMIT THEOREM (CLT) FOR THE SAMPLE MEAN

If a sufficiently large sample of n independent observations are collected from a population with mean μ and standard deviation σ , the sampling distribution of \bar{X} is well approximated by a normal distribution with the following parameters:

$$E(\bar{X}) = \mu \quad \text{SD}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

- The underlying distribution of individual observations does not need to be normally distributed for the CLT to apply, but the distribution should not be strongly skewed.
- One rule of thumb: the sample size should be at least $n = 30$.
- Larger sample sizes are recommended if it is known that the underlying distribution of individual observations exhibits skew.
- By the Law of Large Numbers (LLN), $E(\bar{X}) \rightarrow \mu$ as $n \rightarrow \infty$.

A NOTE ABOUT HANDLING MISSINGNESS...

Earlier, we based our analysis only on individuals with data for the `weight` variable.

- This is an example of doing a **complete case analysis (CCA)**, an analysis restricted to individuals with complete information.
- Our conclusions were based on information from 12,579 cases out of the 13,583 survey participants. The rate of missingness for `weight` was 7.39%.
- Why might it be important to be transparent about how we handled the missing data?

Imputation is a common approach for dealing with missingness that involves using information on other variables to fill in the missing values with plausible ones. The details are beyond the level of this course.³

³Check out [this article from the BMJ](#) if you are interested in learning more about the topic.

Confidence intervals

CONFIDENCE INTERVALS

A **confidence interval** provides an estimate for a population parameter along with a margin of error that gives a plausible range of values for the population parameter.

A confidence interval for a population mean μ has the general form

$$\bar{x} \pm m \rightarrow (\bar{x} - m, \bar{x} + m),$$

where m is the margin of error.

To calculate m , we use what is known about the *sampling variability* of \bar{X} .

USING s AS AN APPROXIMATION OF σ

The **standard error** of the sample mean, $SE(\bar{X})$, measures the sample-to-sample variability of \bar{X} ; i.e., the extent to which sample means oscillate around the population mean.

- From theory, the standard error equals $\frac{\sigma}{\sqrt{n}}$.
- In practice, however, σ is typically unknown.
- The sample standard deviation s is a reasonably good estimate of σ .

Thus, $SE(\bar{X})$ is approximately equal to $\frac{s}{\sqrt{n}}$.

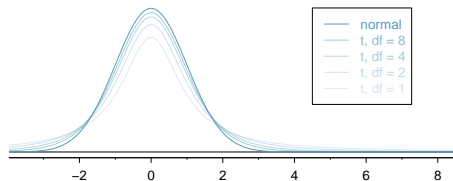
- Using s to estimate σ introduces another source of variability.
- Can adjust for this by modeling the sampling distribution of \bar{X} with the t distribution rather than the normal.

THE t DISTRIBUTION

The t distribution is symmetric, bell-shaped, and centered at 0.

- Its shape is very close to the (standard) normal distribution, but the tails of a t distribution are thicker. This adjusts for the variability introduced by using s as an estimate of σ .
- It has one parameter, *degrees of freedom* (df).^a
- When df is large ($df \geq 30$), the t and z distributions are virtually identical.

^aDegrees of freedom can be represented as ν .



GENERAL FORM FOR A CONFIDENCE INTERVAL

A $(1 - \alpha)(100)\%$ CI for μ is given by

$$\bar{x} \pm t^* \times \frac{s}{\sqrt{n}},$$

where t^* , the critical t -value, is the point on a t distribution with degrees of freedom $n - 1$ that has area $(1 - \alpha/2)$ to the left (and area $\alpha/2$ to the right).

- The **margin of error** $t^* \frac{s}{\sqrt{n}}$ consists of the **standard error** $\frac{s}{\sqrt{n}}$ and the critical t -value t^* .
- The standard error is computed from features of the sample: s and n .
- The critical t -value t^* changes based on the desired level of confidence (and sample size).

Common confidence levels: 90%, 95%, 99%

CALCULATING THE CRITICAL t -VALUE, t^*

The function `qt()` identifies the point on a t distribution with df degrees of freedom that has area p to the left.

- For $t_{df=n-1}$, `qt(p, df)` calculates t such that $p = P(T \leq t)$.
- For a 95% confidence interval, find the critical t -value such that 95% of the distribution is between $-t^*$ and t^* .
- The critical t -value for a 95% confidence interval where $n = 30$ is 2.045.

```
qt(0.975, df = 29)
```

```
## [1] 2.04523
```

CALCULATING A CONFIDENCE INTERVAL

The confidence interval for population mean weight, from the earlier sample of 30 individuals:

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$
$$69.628 \pm (2.045) \frac{18.264}{\sqrt{30}}$$
$$(62.81, 76.45) \text{ kg}$$

Interpretation: With 95% confidence, the interval (62.81, 76.45) kg captures the population mean weight of high school students (who responded to the YRBSS survey in 2013).

```
#letting R do the work  
t.test(yrbss_complete$weight[sample_rows])$conf.int  
## [1] 62.80802 76.44798  
## attr("conf.level")  
## [1] 0.95
```

Full explanation of `t.test()` coming in the labs.

INTERPRETING A CONFIDENCE INTERVAL

Suppose that from a random sample of 60 American adults, a 95% confidence interval for the population mean weight is (160.89, 185.71) lbs.

It is tempting to conclude μ is within the interval (160.89, 185.71) lbs with probability 0.95. . .

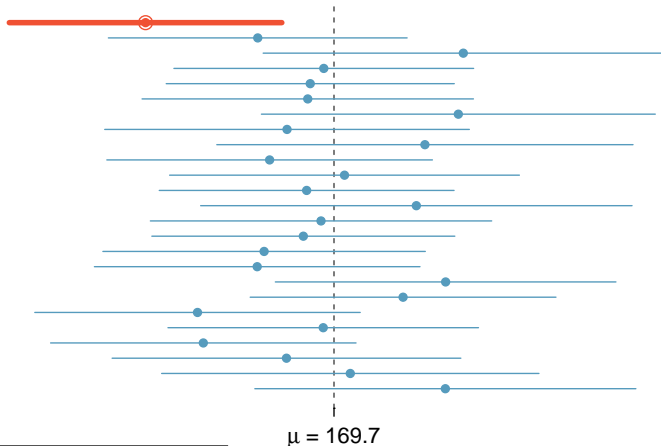
- However, it would be **incorrect** to claim $P(160.89 \leq \mu \leq 185.71) = 0.95$.
- Before the sample is observed, it is valid to state that the *random interval* $(\bar{X} - t_{0.975}^* \frac{s}{\sqrt{n}}, \bar{X} + t_{0.975}^* \frac{s}{\sqrt{n}})$ contains μ with probability 0.95.
- Once the sample is observed, the interval is a *fixed interval*, and either does or does not contain μ .

The correct interpretation relies on the theoretical construct of repeated sampling. If many samples of $n = 60$ were taken and a CI calculated for each one, approximately 95% of the intervals would contain μ .

- Intuitive explanation: a CI expresses values of μ consistent with the observed data.

INTERPRETING A CONFIDENCE INTERVAL...

Twenty-five samples of size $n = 60$ were taken from the 'artificial' population, then a 95% CI for the population mean adult weight calculated based on each sample. Only 1 of these 25 intervals did not contain the population mean, $\mu = 169.7$ lbs.⁴



⁴In a realistic setting, μ is not known. Here, the complete dataset is treated as the population.

HIDDEN ASSUMPTIONS

1. The data used to calculate the confidence interval are from a random sample taken from the target population.
2. While the population mean from the target population is not known, the target population is well-defined.

Both conditions are true in this classroom example of sampling from yrbss, but may be difficult to verify in practice.

Hypothesis testing

INTUITION BEHIND HYPOTHESIS TESTING

Let's first think about a simple coin tossing example. Suppose a friend gives you a coin and asks you to investigate whether it is fair or biased.

- First, make a hypothesis. Let's assume that the coin is fair and that the probability of seeing heads is equal to 0.50.⁵
- Suppose you plan to flip the coin 20 times and record the outcomes.
 - ◊ If the coin is actually fair, what should the sampling distribution of the number of heads (out of 20 tosses) look like?
- You carry out the 20 tosses...
 - ◊ Suppose you see 12 heads. Do you think the coin is biased?
 - ◊ Suppose you see 19 heads. Do you think the coin is biased?

⁵This is an example of a hypothesis test for a single *proportion*, rather than a *mean*. We'll return to this in Unit 9, but the logic of hypothesis testing is the same regardless of the population parameter of interest.

EXAMPLE: CHERRY BLOSSOM TEN MILE RUN

The Cherry Blossom Ten Mile Run is an annual road race in Washington, D.C. The race takes place on the first Sunday in April so as to coincide with the bloom of the 3,000 cherry trees planted as a gift from the Japanese government in 1912. In recent years about 15,000 amateur runners participated.

In 2006, the mean time for all runners who finished the race was 93.3 minutes.

Is the typical US runner getting faster or slower over time?

We will address this question using a random sample of 100 runners who finished the 2017 Cherry Blossom Ten Mile Run.

EXAMPLE: CHERRY BLOSSOM TEN MILE RUN...

Let's apply the logic of hypothesis testing:

- First, make a hypothesis about the population parameter of interest.
 - ◊ Hypothesize that the population mean run time in 2017 is equal to the 2006 time.
 - ◊ $H_0: \mu = 93.3$ minutes
- If it were true that $\mu = 93.3$, what would the sampling distribution of \bar{X} look like when $n = 100$?
 - ◊ We can use simulation to help us visualize this distribution.⁶
- Where does the observed sample mean run time in 2017 fall on the sampling distribution?
 - ◊ Does the evidence support the initial hypothesis that $\mu = 93.3$ minutes?
 - ◊ Or, does the evidence suggest that the population mean run time in 2017 was different from 93.3 minutes?

⁶We'll have to make some assumptions to visualize the null distribution. More on this in the lab...

LOGIC OF HYPOTHESIS TESTING

Observations come from either of two competing distributions:

- The *null* distribution: a usual distribution that has been true in the past
- The *alternative* distribution: new distribution induced by an intervention or a changing condition

We conclude that observations come from the null distribution, unless...

- The value of an **observed statistic** is so extreme that it would be unlikely to occur under the null distribution.

In other words, we compare the observed statistic and hypothesized null value to determine which of the following conclusions is more plausible:

- The difference is small enough that the observed data come from the null distribution. Any discrepancy is simply due to sampling variability.
- The difference is large enough that the observed data do not come from the null distribution. The discrepancy seems too large to be solely due to sampling variability.

FORMAL APPROACH TO HYPOTHESIS TESTING

1. Formulate null and alternative hypotheses.
2. Specify a significance level, α .
3. Calculate a test statistic.
4. Calculate a p -value.
5. Draw conclusions in the context of the original problem.

1. NULL AND ALTERNATIVE HYPOTHESES

Start with a conjecture about the population parameter of interest, then formulate statistical null and alternative hypotheses.

The **null hypothesis** (H_0) posits a value for the population parameter of interest and represents a claim to be tested.

- H_0 can also be thought of as representing the status quo; i.e., no change from the past.
- $H_0 : \mu = \mu_0$

The **alternative hypothesis** (H_A) is an alternative claim and is often represented by a range of parameter values.

- $H_A : \mu \neq \mu_0$, or $H_A : \mu < \mu_0$, or $H_A : \mu > \mu_0$

Generally, an investigator suspects that the null hypothesis is not true and performs a hypothesis test in order to evaluate the strength of evidence against the null hypothesis.

1. NULL AND ALTERNATIVE HYPOTHESES...

We are interested in assessing whether run times seem different in 2017 versus 2006.

- $H_0 : \mu = 93.3 \text{ min}$ ⁷
- $H_A : \mu \neq 93.3 \text{ min}$

This form of H_A is called a two-sided alternative.

- $H_A : \mu < 93.3 \text{ min}$ would be a one-sided alternative.

The choice of one- or two-sided alternative is context-dependent.

⁷A null hypothesis has the general format $H_0 : \mu = \mu_0$. The symbol μ_0 represents the numeric value that μ is hypothesized to equal under the null. In this case, the mean run time in 2006 is chosen as μ_0 .

2. SPECIFYING A SIGNIFICANCE LEVEL α

The significance level α can be thought of as the value that quantifies how rare or unlikely an event must be in order to represent sufficient evidence against H_0 .

- i.e., if the observed statistic occurs with probability smaller than α under the assumption that H_0 is true, then it is considered an extreme observation
- Typically, α is chosen to be a small value such as 0.10, 0.05, or 0.01.
- Always specify α **before** looking at the data!

In the context of decision errors, α is the probability of making a Type I error.

- Type I error refers to incorrectly rejecting the null hypothesis.⁸

⁸More on decision errors coming in Unit 6.

3. CALCULATE A TEST STATISTIC

The test statistic measures the discrepancy between the observed data and what would be expected if the null hypothesis were true.

- When considering the sampling distribution of \bar{X} , how many standard deviations is the observed sample mean from the hypothesized population mean, μ_0 ?
- Recall that the standard deviation of \bar{X} is well-approximated by $\frac{s}{\sqrt{n}}$.

When testing hypotheses about a population mean, the test statistic is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}},$$

where the test statistic t follows a t distribution with $n - 1$ degrees of freedom.

- A larger test statistic indicates greater departure from μ_0 and stronger evidence against H_0 (and in favor of H_A).

4. CALCULATE A p -VALUE

If the null hypothesis were true, what is the probability that we would observe a result **as or more extreme** than the observed sample value?

- The p -value is a conditional probability computed *assuming that H_0 is true*.
- The p -value is **not** the probability that H_0 is true (nor is it the probability that H_A is false).

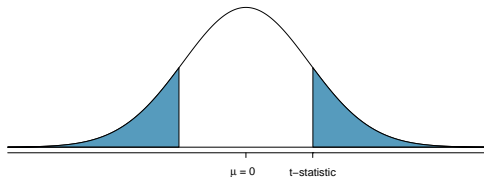
Using what is known about the distribution of the test statistic, calculate the p -value associated with the test statistic then compare it to the significance level α .

- A result is considered unusual if its p -value is less than α .

4. CALCULATE A p -VALUE...

For a two-sided alternative, $H_A : \mu \neq \mu_0$, the p -value is the total area from both tails of the null distribution that are beyond the absolute value of the test statistic.

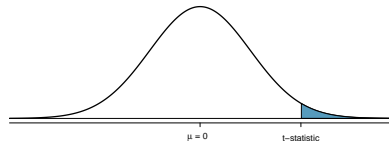
$$p\text{-value} = 2P(T \geq |t|) = P(T \leq -|t|) + P(T \geq |t|)$$



For a one-sided alternative, the p -value is the area in the tail of the null distribution that matches the direction of the alternative.

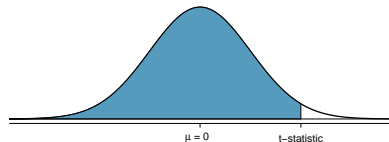
For $H_A : \mu > \mu_0$:

$$p\text{-value} = P(T \geq t)$$



For $H_A : \mu < \mu_0$:

$$p\text{-value} = P(T \leq t)$$



4. CALCULATE A p -VALUE...

```
t.test(cherry.sample$net_sec/60, mu = 93.3, alternative = "two.sided")
```

```
##  
## One Sample t-test  
##  
## data:  cherry.sample$net_sec/60  
## t = 3.0486, df = 99, p-value = 0.002949  
## alternative hypothesis: true mean is not equal to 93.3  
## 95 percent confidence interval:  
##  95.10571 101.83796  
## sample estimates:  
## mean of x  
##  98.47183
```

Alternatively, compute the test statistic and use `pt()`.

```
2*pt(3.0486, df = 100 - 1, lower.tail = FALSE) #two-sided p-value
```

```
## [1] 0.002948914
```

```
pt(3.0486, df = 100 - 1, lower.tail = FALSE) #p-value for HA: mu > 93.3
```

```
## [1] 0.001474457
```

5. DRAW CONCLUSIONS

The smaller the p -value, the stronger the evidence against the null hypothesis.

- If the p -value is as small or smaller than α , we **reject** the null hypothesis. The result is statistically significant at level α and there is evidence to accept the alternative hypothesis.
- If the p -value is larger than α , we **fail to reject** the null hypothesis. The result is not statistically significant at level α . In other words, the evidence does not contradict the null hypothesis.

A subtle but important point: not rejecting H_0 is not the same as proving that H_0 is true. We simply do not have sufficient evidence that H_0 is not true!

5. DRAW CONCLUSIONS...

A p -value represents the probability of observing a test statistic **as extreme or more extreme** than what was observed, under the **assumption that the null hypothesis is true**.

Interpretation in context:

- Let's use the p -value for $H_A : \mu > 93.3$ min.
- The chance of seeing a sample of 100 runners with mean completion time of 98.5 minutes or greater if the actual mean completion time for all runners in 2017 were actually 93.3 minutes equals 0.0015.

Conclusion:

- Since 0.0015 is very small (relative to $\alpha = 0.05$), these data suggest evidence against the null hypothesis that the mean completion time in 2017 is 93.3 minutes. There is evidence at the $\alpha = 0.05$ significance level to reject H_0 and accept $H_A : \mu > 93.3$ min.
- It would be very unusual to see a sample mean of 98.5 minutes or greater if the population mean were actually 93.3 minutes, so we can conclude that the mean completion time in 2017 is greater than 93.3 minutes.

5. DRAW CONCLUSIONS...

What if the sample mean had actually been 94.5 minutes, with a one-sided p -value of 0.241?

Conclusion:

- Since 0.241 is not small (relative to $\alpha = 0.05$), these data *do not suggest evidence against the null hypothesis* that the mean completion time in 2017 is 93.3 minutes.
- Instead, these data are consistent with the sample-to-sample variation that we would expect when drawing a random sample of 100 completion times when the average completion time is actually 93.3 minutes.
- Since $p > \alpha$, there is insufficient evidence to reject H_0 ; i.e., we fail to reject H_0 .
- **Note:** In statistics, we do not “accept” H_0 . We simply have not seen evidence against H_0 .

GUIDANCE ON INTERPRETING p -VALUES

A p -value represents the probability of observing results *as or more extreme than what was observed*, under the *assumption that the null hypothesis is true*.

- A “small” p -value suggests that the null hypothesis is not true; i.e., represents evidence that we can reject the null hypothesis.
 - ◇ Conceptual thinking: We observed results that are so unlikely to happen if the null is true that it makes more sense to conclude that the null is false.
 - ◇ Example: We flip a coin 100 times and see heads 95 times. This is so unlikely to happen with a fair coin that it seems reasonable to conclude the coin is biased.
- A “large” p -value indicates that there is insufficient evidence against the null hypothesis; i.e., the results are consistent with the null hypothesis (being true).
 - ◇ This is **not** the same as claiming that we have found evidence that the null is true! We simply have not yet observed evidence indicating the null hypothesis is false.
 - ◇ Example: We flip a coin 100 times and see heads 45 times. This is not that unlikely to happen with a fair coin, so we can't conclude the coin is biased. We also have not proven that the coin is fair.

GUIDANCE ON INTERPRETING p -VALUES...

What is considered extreme enough evidence to be convincing?

- A common practice is to consider p -values smaller than 0.05 to be “statistically significant”.
 - ◊ Intuition: If we see results that would only happen less than 5% of the time when the null hypothesis is true, then we have sufficient evidence against the null hypothesis.
- In scientific literature, this benchmark is referred to as α ; i.e., the significance level.
- The value of α is pre-specified before an analysis is done. Common values of α include 0.10, 0.05, 0.01, etc.

However, we should avoid thinking of $\alpha = 0.05$ as a definitive cutoff.

- How should we interpret a p -value of 0.08?
- How should we interpret a p -value of 0.045?

ASSUMPTIONS

The same assumptions behind the CLT are behind hypothesis testing:

- Observations are independent of each other.
- The sample size is reasonably large.

The same hidden assumptions behind confidence intervals also apply:

- The data used for the hypothesis test are representative of the target population.
- The target population is well-defined.

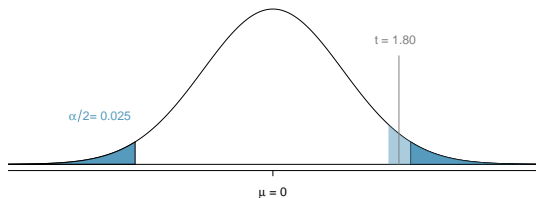
TWO-SIDED VS. ONE-SIDED HYPOTHESIS TESTS

Make this choice **before** looking at the data!

In practice, two-sided tests are seen as more rigorous and are typically expected by most journals and regulatory authorities.

It is “easier” to reject H_0 for a one-sided test than a two-sided test, when conducted at the same α level.^a

^aMore about choosing between one-sided and two-sided tests in *OI Biostat* Section 4.3.5



The relationship between tests and confidence intervals

TWO-SIDED TESTS AND CONFIDENCE INTERVALS

The relationship between a hypothesis test and the corresponding confidence interval is defined by the significance level, α .

- Hypothesis test: Is the sample statistic far enough away from the hypothesized null value to be considered extreme?
- Confidence interval: Is the hypothesized null value close enough to the sample statistic to be plausible?

The “far enough” and “close enough” are defined by α .

- If a 95% CI for a parameter does not contain the hypothesized null value, then the data contradict the null hypothesis at significance level $\alpha = 0.05$.
- If a 95% CI for a parameter does contain the hypothesized null value, then the data do not contradict the null hypothesis at significance level $\alpha = 0.05$.

TWO-SIDED TESTS AND CONFIDENCE INTERVALS...

If a 95% confidence interval for a population mean does not contain a hypothesized value μ_0 , then:

- The data contradict the the null hypothesis
 $H_0 : \mu = \mu_0$ at significance level $\alpha = 0.05$
- The implied two-sided alternative hypothesis is
 $H_A : \mu \neq \mu_0$

Based on the following 95% confidence interval computed from the 2017 sample of run times, is there sufficient evidence to reject $H_0 : \mu = 93.3$ minutes at the $\alpha = 0.05$ significance level?

```
t.test(cherry.sample$net_sec/60,  
       conf.level = 0.95)$conf.int  
  
## [1] 95.10571 101.83796  
## attr("conf.level")  
## [1] 0.95
```

IN PRACTICE...

A confidence interval provides a range of plausible values for a parameter.

A hypothesis test measures the strength of evidence against a null hypothesis.

Both are important tools for conducting inference.

- A lot more information/guidance about interpreting confidence intervals and hypothesis tests coming up in future units.