AP statistics Chapter 9 Note-Taking Guide
Testing a Claim

Name Settle
Per ___ Date ___

9.1

Significance test

A formal procedure for comparing observed data with a claim (hypotheses) whose truth we want to assess. We express the results of a significance test in terms of a probability that measures how well the data and the claim agree.

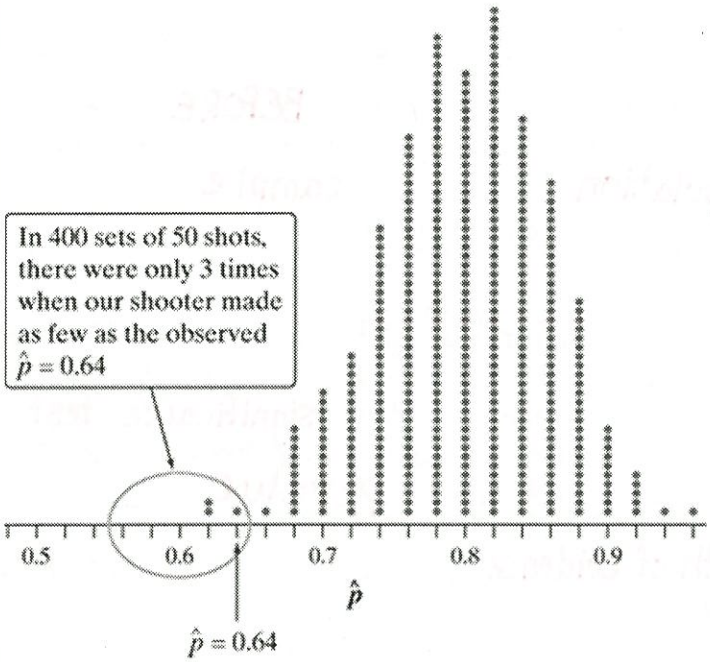
Significance tests

- Deal with claims about a population
- Ask if sample data give good evidence against a claim
- "If we took many random samples and the claim were true, we would get a result like this ___% of the time"
- BASIC IDEA: An outcome that would rarely happen if a claim were true is good evidence that the claim is NOT TRUE!

Sample: A basketball player claims to make 80% of the free throws he attempts. We think he might be exaggerating. To test this claim, we'll ask him to shoot some free throws. Suppose he shoots 50 free throws and makes 32 of them.

A) What is his sample proportion ($\hat{p}$)? $\hat{p} = \frac{32}{50} = 0.64$

B) What can we conclude about the player's claim based on the sample data? We can perform a simulation to find out. We used Fathom software to simulate 400 sets of 50 shots assuming that the player really is an 80% free throw shooter. The dotplot of the results is below. Each dot is the proportion of shots made for each group of 50 attempts.



You can say how strong the evidence against the player's claim is by giving the probability that he would make as few as 32 out of 50 free throws if he really makes 80% in the long run.

Based on the simulation, our estimate of this probability is 3/400 = 0.0075. The observed statistic, $\hat{p} = 0.64$, is so unlikely that it gives convincing evidence that the player's claim is not true

Be sure that you understand why this evidence is convincing. There are two possible explanations of the fact that our virtual player made only $\hat{p} = 0.64$ of his free throws:

1. The player's claim is correct ($p = 0.8$), and by bad luck, a very unlikely outcome occurred.

2. The population proportion is actually less than 0.8, so the sample result is not an unlikely outcome.

9.1A

Stating a hypotheses

- **Null hypotheses** (H_0) a statement of no difference (make 80% shots)
- **Alternative hypotheses** (H_a) the claim that we hope or suspect is true instead of the null hypotheses. (makes less than 80%)
- **One sided** alternative hypotheses will claim that the actual parameter is either less than or greater than the null hypotheses $p<$ or $p>$
- **Two sided** alternative hypotheses will claim that the actual parameter is Not the parameter stated in the null hypotheses. $p\neq$

Sample: Mike is an avid golfer who would like to improve his play. A friend suggests getting new clubs and lets Mike try out his 7-iron. Based on years of experience, Mike has established that the mean distance that balls travel when hit with his old 7-iron is μ = 175 yards with a standard deviation of σ = 15 yards. He is hoping that this new club will make his shots with a 7-iron more consistent (less variable), so he goes to the driving range and hits 50 shots with the new 7-iron.

Problem:

(a) Describe the parameter of interest in this setting.

Mike wants to be more consistent, so the parameter is the standard deviation (σ) of the distance he hits the ball with the new 7 iron.

(b) State appropriate hypotheses for performing a significance test.

More consistent = less variation so less than 15 yards

$H_0: \sigma = 15$
$H_a: \sigma < 15$

9.1A

Notes about hypotheses:

- The hypotheses should express the hopes or suspicions we have BEFORE we see the data. It is cheating to look at the data first and then frame the hypotheses to fit what the data show.
- Hypotheses always refer to a population, not to a sample.
- If unsure, use a 2-sided alternative hypotheses!

9.1A

Logic of significance tests

Analogy of a trial

For a defendant (H_0) to be brought to trial, there must be some preliminary evidence that he is guilty. It is the job of the jury to decide whether the evidence is convincing (no other plausible explanations for how the crime was committed). We do a significance test when the sample statistic provides preliminary evidence that the alternative hypotheses is true. To determine whether the evidence is convincing, we calculate a p-value.

9.1A

Interpreting P-values

The probability that measures the strength of evidence against a null hypotheses is called a p-value.

- The smaller the p-value, the stronger the evidence against the null hypotheses provided by the data.
- The p-value is the conditional probability $P(H_a|H_0)$

In the **free-throw shooter example**, the estimated p-value of 0.0075 is strong evidence against the null hypotheses H_0: $p = 0.80$. For that reason, we would reject H_0 in favor of the alternative H_a: $p < 0.80$. It appears that the virtual player makes fewer than 80% of his free throws.

Later, we will learn how to calculate the p-values.

Sample: When Mike was testing a new 7-iron, the hypotheses were

$$H_0: \sigma = 15$$

$$H_a: \sigma < 15$$

where σ = the true standard deviation of the distances Mike hits golf balls using the new 7-iron. Based on 50 shots with the new 7-iron, the standard deviation was s_x = 10.9 yards.

Problem: A significance test using the sample data produced a *P*-value of 0.002.

(a) Interpret the *P*-value in this context.

If the true standard deviation is 15, then the probability of getting a 10.9 standard deviation by chance is 0.002 (2 in 1000)

(b) Do the data provide convincing evidence against the null hypothesis? Explain.

Yes, the p-value is really small so random chance explanation is not valid. So there is convincing evidence that the true standard deviation with the new 7 iron is less than 15.

9.1A

Statistical significance

The final step in performing a significance test is to draw a conclusion about the competing claims you were testing. We will make one of two decisions based on the strength of the evidence:

- Reject Ho
- Fail to reject Ho – Does not guarantee that H_0 is true just that there is not enough evidence to reject.

***DO NOT ACCEPT Ho, you will lose credit on AP exam!

9.1A

Significance Level

- To determine what P-value is considered small (Reject Ho) or large (Fail to reject Ho) we compare it to a significance level (α).
- Significance level requires evidence against Ho to be so strong that is would happen less than _____% of the time by chance when H_0 is true.
- When our P-value is less than our chosen α, we say that the result is **statistically significant**.
- Significant doesn't mean important, it means not likely to happen by chance.
- If you are going to draw a conclusion based on statistical significance, the significance level should be determined BEFORE the data are produced.
- Significance levels should range from 0.1 to 0.01. Courts will accept 0.05 or lower.
- When in doubt or if not explicitly stated in a problem, use 0.05

Sample: For his second semester project in AP Statistics, Zenon decided to investigate if students at his school prefer name-brand potato chips to generic potato chips. He randomly selected 50 students and had each student try both types of chips, in random order. Overall, 34 of the 50 students preferred the name-brand chips. Zenon performed a significance test using the hypotheses:

$$H_0: p = 0.5$$

$$H_a: p > 0.5$$

where p = the true proportion of students at his school that prefer name-brand chips. The resulting *P*-value was 0.0055.

Problem: What conclusion would you make at each of the following significance levels?

(a) $\alpha = 0.01$ 0.0055 < 0.01

Reject Ho, There is convincing evidence that more than 50% of the students prefer name-brand chips

(b) $\alpha = 0.001$ 0.0055 ≮ 0.01

Fail to reject Ho There is not convincing evidence that more than 50% of the students prefer name-brand chips.

9.1B

Errors

When we draw conclusions from a significance test, we hope our conclusions will be correct. Sometimes though, our conclusions will be incorrect. There are two types of mistakes we can make:

1. Type 1 error: we reject H_0 when H_0 is true
2. Type 2 error: we fail to reject H_0 when H_0 is false.

		Truth about the population: H_0 true	Truth about the population: H_0 false (H_a true)
Conclusion based on sample	Reject H_0	Type I error	Correct conclusion
	Fail to reject H_0	Correct conclusion	Type II error

Sample: The manager of a fast-food restaurant want to reduce the proportion of drive-through customers who have to wait more than 2 minutes to receive their food once their order is placed. Based on store records, the proportion of customers who had to wait at least 2 minutes was $p = 0.63$. To reduce this proportion, the manager assigns an additional employee to assist with drive-through orders. During the next month the manager will collect a random sample of drive-through times and test the following hypotheses:

$H_0: p = 0.63$

$H_a: p < 0.63$

where p = the true proportion of drive-through customers who have to wait more than 2 minutes after their order is placed to receive their food.

Problem: Describe a Type I and a Type II error in this setting and explain the consequences of each.

Type 1: Manager decides that the true proportion of customers who has to wait at least 2 minutes has reduced when it really hasn't. Manager would think they are keeping up when they aren't. Lose customers.

Type 2: Manager decides that the true proportion of customers who has to wait at least 2 minutes has not reduced, when it really has. He hires another employee to keep up.

9.1B

Error probabilities

We can assess the performance of a significance test by looking at the probabilities of making these two errors.

1) Type 1 error probabilities P(type 1 error) = P(reject H_0 | H_0 is true)

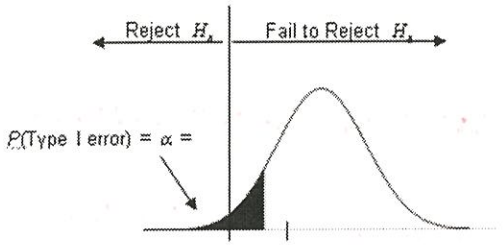
In the fast food alternate example, we were testing the following hypotheses:

** convict an innocent person

$H_0: p = 0.63$

$H_a: p < 0.63$

where p = the true proportion of drive-through customers who have to wait more than 2 minutes after their order is placed to receive their food. Suppose that the manager decided to carry out this test using a random sample of 250 orders and a significance level of $\alpha = 0.10$. What is the probability of a making a Type I error?



To make a Type I error means that we reject H_0 when H_0 is actually true. In this case, a Type I error occurs when the true proportion of customers that have to wait at least 2 minutes remains $p = 0.63$, but we get a value of $\hat{p}$ small enough that the P-value is less than 0.10. When H_0 is true, this will happen 10% of the time.

In other words, P(type 1 error) = α

Type 2 error probabilities P(Fail to reject H_0 | H_0 is false) ** guilty person goes free

A significance test makes a Type II error when it fails to reject a null hypothesis that really is false. There are many values of the parameter that satisfy the alternative hypothesis, so instead we concentrate on the other option. We can calculate the probability that a test does reject H_0 when an alternative is true. This probability is called the **power** of the test against that specific alternative.

		Truth about the population	
		H_0 true	H_0 false (H_a true)
Conclusion based on sample	Reject H_0	Type I error	Power
	Fail to reject H_0	Correct conclusion	Type II error

The probability of making a Type II error is 1 - power.

**We will not ask you to calculate power or probability of a type 2 error, but you will be asked to estimate using a simulation OR identify a power based on a diagram.

9.1B

Planning studies: The power of a statistical test

How large should a sample be when we plan to carry out a significance test?

Here are the questions we must answer to decide how many observations we need:

1. **Significance level**? If you insist on a smaller significance level (such as 1% rather than 5%), you have to take a larger sample. A smaller significance level requires stronger evidence to reject the null hypothesis.
2. **Practical Importance**? At any significance level and desired power, detecting a small difference requires a larger sample than detecting a large difference.
3. **Power**? If you insist on higher power (such as 99% rather than 90%), you will need a larger sample. Higher power gives a better chance of detecting a difference when it is really there.

↓ α → small difference → larger sample
higher power →

Sample: Suppose that the manager of the fast food restaurant wants to change some aspects of his study about the proportion of drive-through customers that have to wait at least 2 minutes to receive their food after they place their order.

α

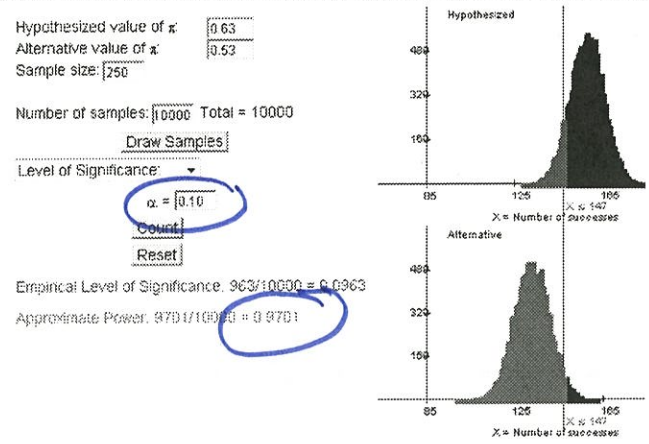
1. Significance level: To reduce the possibility of a Type I error and avoid the possibility of unnecessarily paying an extra employee, the manager reduces the significance level from 0.10 to 0.01.
2. Practical importance: To justify the additional cost of the extra employee, the manager decides that the true proportion must be reduced to at most 0.53.
3. To get faster results, the manager reduces the sample size from 250 to 100.

How will these changes affect the power of the test of $H_0: p = 0.63$ versus $H_a: p < 0.63$?

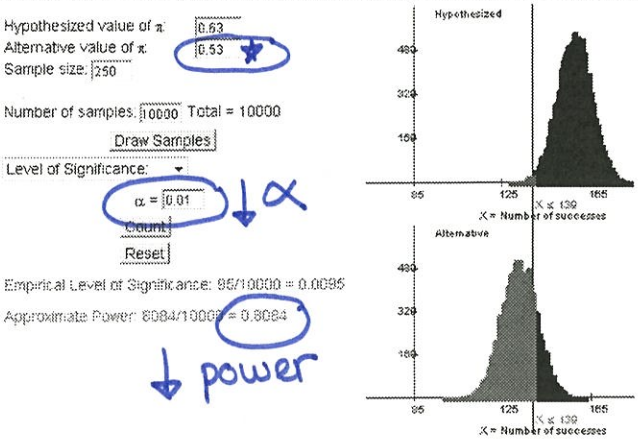
↓ α → ↓ power ↓ sample → ↓ power

↓ p_0 for H_a ↑ power

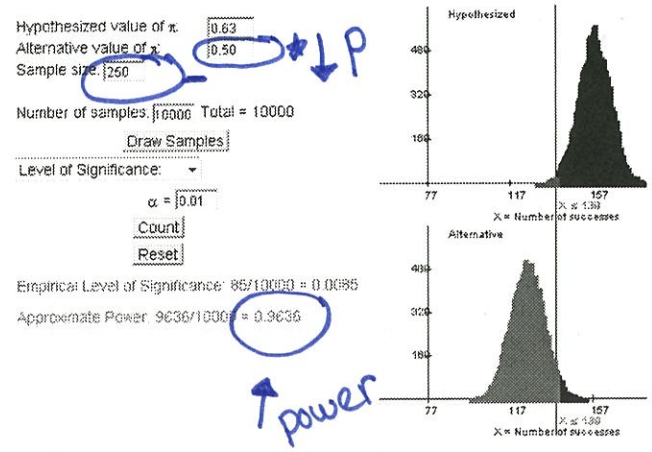
Power Simulation



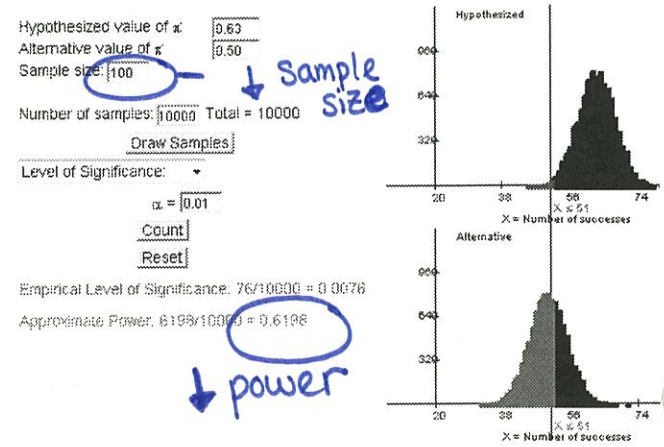
Power Simulation



Power Simulation



Power Simulation



9.2A

Significance tests for population proportions

Conditions must be met:

1. **Random**: Data should come from a well-designed (SRS) random sample or randomized experiment. Otherwise we can't infer to the population or establish cause and effect.

2. **Normal:** sampling distribution of the statistic is approximately normal
   - Normal condition for proportions: $np_0 \geq 10$ $n(1-p_0) \geq 10$

   P will be replaced with $\mathbf{p_0}$ which is the population proportion.

3. **Independent:** sampling with replacement for the population allows us to use standard deviation formulas, or if sampling without replacement, we meet the 10% condition for independence $n \leq 0.10N$. or more than 10n in the population

PROBLEM: Check the conditions for carrying out a significance test of the virtual basketball player's claim.

SOLUTION: The three required conditions are

- Random: SRS of the 50 shots
- Normal: Assume H_0 is true so $p_0 = 0.80$ so $50(0.8) = 40 \geq 10$ and $50(.2) = 10 \geq 10$
- Independent: player will shoot more than 500 free throws in lifetime

9.2A

Calculations: Test statistic and P-value

A significance test uses sample data to measure the strength of evidence against H_0. Here are some principles that apply to most tests:

- The test compares a Statistic calculated from sample data with the value of the parameter stated by the null hypotheses.
- Values of the statistic far from the null parameter in the direction specified by the alternative hypothesis give evidence against H_0.
- To assess how far the statistic is from the parameter, standardize the statistic.
- This value is called the **test statistic**:

$$z = \frac{\text{Statistic} - \text{parameter}}{\text{Standard deviation of statistic}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

- The test statistic measures how far the sample result is from the null parameter value, in what direction, on a standardized scale.
- You can use the test statistic to find the P-value of the test

Sample:. A basketball player claims to make 80% of the free throws he attempts. We think he might be exaggerating. To test this claim, we'll ask him to shoot some free throws. In an SRS of 50 free throws, the player made 32 free throws

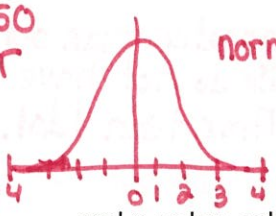
a) Calculate the test statistic

$\hat{p} = \frac{32}{50} = 0.64$ $\quad z = \frac{0.64 - 0.80}{\sqrt{\frac{0.8(0.2)}{50}}} = \frac{-0.16}{0.0566} = \boxed{-2.83}$

b) Find and interpret the p-value.

look up -2.83 in table or normalcdf (-10, -2.83, 0, 1) = 0.0233

$H_a: p < 0.80$



If H_0 (0.8) is true, there is a 2.33% chance he would shoot this bad.

**both the p-value from a simulation and p-value calculated using the normal distribution are estimates of the p-value.

9.2A

Four step process for significance testing

State: What hypotheses do you want to test, and at what significance level? Define any parameters you use. State H_0 and H_a.

Plan: Name procedure you are using. Check conditions.

Do: If conditions are met, perform calculations

- Compute the test statistic
- Find the P-value

Conclude: Interpret the results of your test in the context of the problem.

9.2A

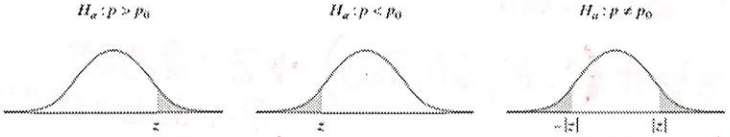
One sample z test for a proportion

A way to compute the test statistic is the one-sample z test.

- Choose an SRS of size n from a large population that contains an unknown proportion of p successes.
- To test the hypotheses $H_0: p = p_0$, compute the **z statistic**

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

- Find the P-value by calculating the probability of getting a z statistic this large or larger in the direction specified by the alternative hypothesis H_a



- Must meet Normal and Independent conditions to use this test

Sample: On shows like American Idol, contestants often wonder if there is an advantage to performing last. To investigate this, a random sample of 600 American Idol fans is selected and they are shown the audition tapes of 12 never-before-seen contestants. For each fan, the order of the 12 videos is randomly determined. Thus, if the order of performance doesn't matter, we would expect approximately 1/12 of the fans to prefer the last contestant they view. In this study, 59 of the 600 fans preferred the last contestant they viewed. Does this data provide convincing evidence that there is an advantage to going last?

State: Test the hypotheses at $\alpha = 0.05$ significance level

$H_0: p = 1/12 = 0.0833$ $H_a: p > 1/12 = 0.0833$

where p = true proportion of American Idol fans who prefer the last performance.

Plan: Use one sample z-test for proportions

- Random: SRS of 600 fans ✓
- Normal: $600(0.0833) = 50 \geq 10$ and $600(0.9167) = 550 \geq 10$ ✓
- Independent: there are more than 6000 American Idol fans.

Do: $\hat{p} = 59/600 = 0.0983$

test statistic $z = \frac{0.0983 - 0.0833}{\sqrt{\frac{0.0833(0.9167)}{600}}} = \frac{0.015}{0.0113} = 1.33$

P-value: normalcdf(1.33, 10, 0, 1) = 0.092 → 9.2%

Conclude:

Since our p-value is greater than our significance level (0.092 > 0.05) we fail to reject H_0. We do not have sufficient evidence to support performing last in American Idol.

9.2A

What happens when the data don't support Ha?

There is no need to continue with the significance test. The conclusions is clear, we fail to reject H_0. If you aren't paying attention, you may end up performing the test. The test will give you the same conclusion, fail to reject H_0. A lot more work with the same answer!

9.2A

One proportion z test on the calculator

- Press stat
- Arrow to tests
- Choose 5:1-propZTest
- Enter Po → from Ho
- Enter x (number of successes)
- Enter n (number of trials)
- Enter alternative hypotheses
- Calculate

Use 4 step process

Sample: According to the National Campaign to Prevent Teen and Unplanned Pregnancy, 20% of teens aged 13 to 19 say that they have electronically sent or posted sexually suggestive images of themselves. The counselor at a large high school worries that the actual figure might be higher at her school. To find out, she gives an anonymous survey to a random sample of 250 of the school's 2800 students. All 250 respond, and 63 admit to sending or posting sexual images. Carry out a significance test at the α = 0.05 significance level. What conclusion should the counselor draw?

Do step: $p_0 = 0.20$ $\hat{p} = \frac{63}{250} = 0.252$ $H_0: p = 0.20$ $H_a: p > 0.20$

$z = \frac{0.252 - 0.20}{\sqrt{\frac{0.2(0.8)}{250}}}$ = 1 prop z test (0.20, 63, 250) → $z = 2.055$, P-value = 0.0199

P-value is smaller than significance level (0.0199 < 0.05) so reject H_0. We have convincing evidence that more than 20% teens at the school "sext"

9.2B

More on Two sided tests

We perform a two sided test when looking for convincing evidence that the true parameter is different from the hypothesized value of the parameter, $H_0: p \neq p_0$

Sample: According to the Centers for Disease Control and Prevention (CDC) Web site, 50% of high school students have never smoked a cigarette. Taya wonders whether this national result hold true in her large, urban high school. For her statistics class, Taya takes an SRS of 150 students from her school. She gets responses from all 150 students, and 90 say that they have never smoked a cigarette. What should Taya conclude? Give appropriate evidencd to support your answer.

State: We will test at the $\alpha = 0.05$ significance level

$H_0: p = 0.50$ $H_a: p \neq 0.50$

where p = proportion of students that have never smoked a cigarette.

Plan: We will use a 1-sample z test for proportions, Check conditions:

Random: SRS of 150 students

Normal: $150(0.5) = 75 \geq 10$ and $150(0.5) = 75 \geq 10$

Independent: more than 1500 students at the large urban high school.

Do: $\hat{p} = \frac{90}{150} = 0.6$ $z = \frac{0.6 - 0.5}{\sqrt{0.5(0.5)/150}} =$ stat → test → 1 prop z test $z = 2.4495$

$p_0 = 0.5$, x: 90, n: 150, $\neq p_0$

p-value: 0.00715

$0.00715 < 0.05$

Conclude: Reject H_0, We have convincing evidence that the true proportion of students that have never smoked a cigarette is not 0.50.

→ so what is the proportion?

9.2B

confidence intervals and significance tests

Significance tests can tell us if the smaller specific population proportion differs from the entire population proportion.

Confidence intervals give us an idea of what the actual proportion may be. Therefore, a confidence interval can be more informative.

Sample: Taya found that 90 of an SRS of 150 students said that they had never smoked a cigarette. We checked the conditions for performing the significance test earlier. Before we construct a confidence interval for the population proportion p, we should check that both $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$. Since the number of successes and the number of failures in the sample are 90 and 60, we can proceed with our calculations. Calculate the 95% confidence interval for the true proportion of students at the school that never smoked a cigarette.

Estimate the actual proportion of students at the high school that have never smoked a cigarette with 95% confidence by using a 1 sample z interval. Conditions were checked in previous problem and above.

$z = \text{invNorm}(.05/2, 0, 1) = 1.96$

$\text{interval} = 0.6 \pm 1.96\sqrt{\frac{0.6(0.4)}{150}} =$ stat → tests → 1 prop z interval

x: 90, n = 150, C-level: 0.95

(0.5216, 0.6784)

We are 95% that the interval from 0.52 to 0.68 contains the actual proportion of students who have never smoked a cigarette at Taya's High School.

9.2B

Connection between confidence intervals and two sided significance tests

There is a link between confidence intervals and two-sided significance tests:

- The confidence interval gives an approximate range of values p_o that would Not be rejected by a two sided test at the 1 - Confidence Level significance level.
- However, with proportions the link is not perfect (approximation) because of the standard error used for the confidence interval is based on the sample proportion $\hat{p}$, while the denominator statistic is based on the value p_o from the null hypotheses

Test statistic: $z = \dfrac{\hat{p} - p_o}{\sqrt{\dfrac{p_o(1-p_o)}{n}}}$ Confidence interval: $\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

- So on the AP exam it is acceptable to use a confidence interval rather than the test statistic to address a one-sided alternative hypotheses but not 2-sided alternative hypotheses.

9.3A

Population Means and significance testing

- When working with population proportions we use z-test since they are based from the z-values from a standard normal distribution.
- When working with population mean, we use t-values from a t distribution with n – 1 degrees of freedom.
- To keep it straight we have "**zap tax**" for z for a proportion/t for a mean ($\bar{x}$)

9.3A

Conditions for population means using a t distribution

- **Random:** The data come from a random sample from a sample of size n from the population of interest or a randomized experiment. This condition is very important.
- **Normal**:
  - **Sample size less than 15**: use t procedures if the data appear close to normal. If it is **skewed** or has **outliers** DO Not use t procedures.
  - **Sample size of at least 15**: the t procedures can be used except if the data has outliers or strong skewness
  - **Large samples**: The t procedures can be used even for clearly skewed distributions when the sample is large ($n \geq 30$)

we will use boxplots on calculator to evaluate!

- **Independent**: individual observations are independent so if we sample without replacement the 10% must be met population is larger than 10n

Sample: A classic rock radio station claims to play an average of 50 minutes of music every hour. However, every time you turn to this station it seems like there is a commercial playing. To investigate their claim, you randomly select 12 different hours during the next week and record what the radio station plays in each of the 12 hours. Here are the number of minutes of music in each of these hours:

44 49 45 51 49 53 49 44 47 50 46 48

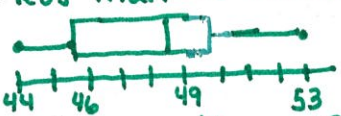
Problem: Check the conditions for carrying out a significance test of the company's claim that it plays an average of at least 50 minutes of music per hour.

$L_1 \rightarrow$ statplot → ▭ ..

L_1 F:1 Zoom 9 Trace

Solution:

- Random: SRS of 12 hours
- Normal: less than 15 check for skewness & outliers



No outliers, Approximately normal

- Independent: more than 120 hours of radio play time

9.3A

Calculations: Test statistic and P-value

- Perform calculations assuming the H_0 is true.
- The test statistic measures how far n differs from the parameter value specified by H_0 in standardized units.
- Due to our t distritbution our test statistic is $t = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$

When the normal condition is met, this statistic has a t-distribution with n-1 degrees of freedom.

- Once we have calculated the test statistic we can use table B or a calculator to find the p-value for a significance test about μ.

Sample: A classic rock radio station claims to play an average of 50 minutes of music every hour. However, every time you turn to this station it seems like there is a commercial playing. To investigate their claim, you randomly select 12 different hours during the next week and record what the radio station plays in each of the 12 hours. Here are the number of minutes of music in each of these hours:

44 49 45 51 49 53 49 44 47 50 46 48

You want to test $H_0: \mu = 50$ versus $H_a: \mu < 50$.

Problem: Compute the test statistic and *P*-value for these data. $\bar{x} = 47.917$, $s_x = 2.81$, $n = 12$

Solution: $t = \frac{47.917 - 50}{2.81/\sqrt{12}} = -2.57$ $df = 12 - 1 = 11$

P-value = table between 0.02 and 0.01 ≈ 0.015

.3A

Using to calculator to compute P-values from a t distribution

- Go the the distribution menu (2nd vars)
- Choose tcdf(
- Enter lower bound for t value (if t ≥ #, use the number; if t ≤ # use -100)
- Enter upper bound for t value (if t ≤ #, use the number; if t ≥ # use 100)
- Enter degrees of freedom (n-1)
- If you are using a two sided test, multiply answer by 2

] 2 sided, you choose

Sample: For a job satisfaction study, the hypotheses are Ho: μ = 0 and Ha:μ ≠ 0
Where μ is the mean difference in job satisfaction scores (self-paced – machine-paced) in the population of assembly line workers at the company. Data from a random sample of 18 workers gave $\bar{x} = 17$ and $S_x = 60$.

a) Calculate the test statistic by hand. Show your work.

$t = \frac{17 - 0}{60/\sqrt{18}} = 1.202$

b) Now use your calculator to find the p-value. What conclusion can you draw?

2 sided so 2tcdf(1.202, 100, 17) = 0.246
(lower bound, upper bound, degrees of freedom)

0.246 ≮ 0.05, we fail to reject Ho, Not enough evidence to conclude that job satisfaction differs when self paced or machine-paced.

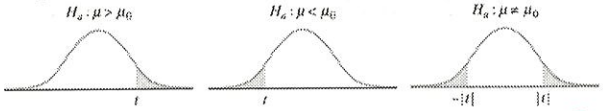
9.3A

One sample t test

- Choose an SRS of sizen from a large population with an unknown mean μ
- To test the hypotheses Ho: $\mu = \mu_0$, compute the one-sample t statistic

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

- Find the p-value by calculating the probability of getting a t statistic this large or larger in the direction specified by the alternative hypotheses H_a in a t-distribution
  df = n-1



- Population distribution is normal or meets normal requirements
- Independence condition is met

Sample: Every road has one at some point—construction zones that have much lower speed limits. To see if drivers obey these lower speed limits, a police officer used a radar gun to measure the speed (in miles per hour, or mph) of a random sample of 10 drivers in a 25 mph construction zone. Here are the results:

27 33 32 21 30 30 29 25 27 34

Problem:

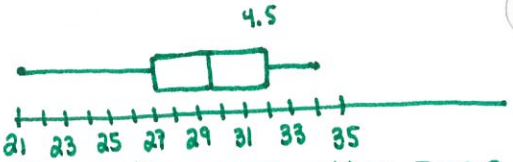
(a) Can we conclude that the average speed of drivers in this construction zone is greater than the posted 25 mph speed limit?

(b) Given your conclusion in part (a), which kind of mistake—a Type I or a Type II error—could you have made? Explain what this mistake means in this context.

Solution:

State: We want to test hypotheses at the 0.05 significance level.
$H_0: \mu = 25$ $H_a: \mu > 25$ μ = true mean speed of drivers in construction zone.

Plan: Use one sample t-test for significance
Random: SRS 10 drivers
Normal: n < 15 so check for outliers & skewness
Slightly skewed left, no outliers ok to continue



Independent: more than 100 drivers that will pass through the construction zone

Do: $\bar{x} = 28.8$ $s_x = 3.94$ $n = 10$

$$t = \frac{28.8 - 25}{3.94/\sqrt{10}} = 3.05$$

P-value: tcd(3.05, 100, 9) = 0.007
(lower bound, upper bound, degrees of freedom)

Conclude: 0.007 < 0.05, reject Ho,
we have convincing evidence that the mean speed of drivers through a construction zone is greater than 25 mph.

(b) Since we rejected the null hypothesis, it is possible we made a type I error. It is possible we rejected it when it is actually true. We determied the mean speed was greater than 25 mph when it truly is not.

9.3A

Technology

Because t procedures are so common, all statistical software packages will do the calculations (but not checking conditions) for you.

Calculator instructions

Raw data entry	Summary statistics entry
• Enter data into L1 • Press STAT • Arrow over to TESTS • Choose 2:T-test • Inpt: Data • μ_o: • list: L1 • Freq:1 • μ: choose based on Ha • calculate	• Press STAT • Arrow over to TESTS • Choose 2:T-test • Inpt: Stats • μ_o: • $\bar{x}$: • S_x: • n: sample size • μ: choose based on Ha • calculate

Sample:

A college professor suspects that students at his school are getting less than 8 hours of sleep a night, on average. To test his belief, the professor asks a random sample of 28 students, "how much sleep did you get last night?" Here are the data (in hours):

9 6 8 6 8 8 6 6.5 6 7 9 4 3 4 5 6 11 6 3 6 6 10 7 8 4.5 9 7 7

Do these data provide convincing evidence in support of the professors suspicion? Carry out a significance test at the α = 0.05 level to help answer this question.

Usually do all four steps, but lets just do the last 2

Stat → tests → t-test → data μ_o: 8, List: L_1, Freq: 1, $< \mu_o$

$t = -3.63$ $p = 5.9 \times 10^{-4}$ 0.00059 $\bar{x} = 6.64$ $Sx = 1.98$

0.00059 < 0.05, reject H_o, Convincing evidence that students are getting less than 8 hours of sleep.

9.3A

Two sided significance tests

Collection 1

	29.4874 mm
	50
Sx →	0.0934676 mm
Se →	0.0132183 mm
Width	29.2717 mm
	29.4225 mm
	29.4821 mm
	29.5544 mm
	29.7148 mm

S1 = mean ( )
S2 = count ( )
S3 = stdDev ( )
S4 = stdError ( )
S5 = min ( )
S6 = Q1 ( )
S7 = median ( )
S8 = Q3 ( )
S9 = max ( )

In the children's game Don't Break the Ice, small plastic ice cubes are squeezed into a square frame. Each child takes turns tapping out a cube of "ice" with a plastic hammer hoping that the remaining cubes don't collapse. For the game to work correctly, the cubes must be big enough so that they hold each other in place in the plastic frame but not so big that they are too difficult to tap out. The machine that produces the plastic ice cubes is designed to make cubes that are 29.5 millimeters (mm) wide, but the actual width varies a little. To make sure the machine is working well, a supervisor inspects a random sample of 50 cubes every hour and measures their width. The Fathom output summarizes the data from a sample taken during one hour.

Problem:

(a) Interpret the standard deviation and the standard error provided by the computer output.

(b) Do these data give convincing evidence that the mean width of cubes produced this hour is not 29.5 mm?

Solution:

(a) Standard deviation: The widths of the cubes are about 0.093mm from the mean width

In random samples of size 50, the sample mean will vary from the true mean by 0.013

(b) *State:* We want to test the hypotheses at a 0.05 significance level.
$H_0: \mu = 29.5$ $H_a: \mu \neq 29.5$ where μ = true mean width of ice cubes.

Plan: Use a 1-sample t-test for significance
Random: SRS of 50 cubes
Normal: Sample is large (≥ 30) so approximately normal.
Independent: more than 500 ice cubes made by the machine.

Do: $\bar{x} = 29.4874$, $Sx = 0.0935$, $n = 50$

$t = \frac{29.4874 - 29.5}{0.0935/\sqrt{50}} =$ or stats → test → t-test stats → $\mu_0: 29.5, \bar{x} = 29.4874, Sx = 0.0935$ $n: 50$, $\neq \mu_0$

$t = -0.953$, p-value = 0.345

Conclude: 0.345 $\not< 0.05$, fail to reject H_0, There is not convincing evidence that the actual mean is not 29.5.

9.3A **Two sided significance tests and confidence intervals**

- Unfortunately the significance test doesn't tell us the actual value of μ, for that we need a confidence interval.
- The connection between significance tests and confidence intervals is even stronger for means than it was for proportions. That is because both the test statistic and the confidence interval use the standard error of $\bar{x}$ in the calculations.
- So when the two sided significance test at a level α, rejects H_0, the 100(1 – α)% confidence interval for μ will not contain the hypothesized value μ_0
- And when the test fails to reject Ho, the confidence interval will contain μ_0.

Estimate of Collection 1 (Estimate Mean)

Attribute (numeric): Width

Interval estimate for population mean of **Width**

Count:	50
Mean:	29.4874 mm
Std dev:	0.0934676 mm
Std error:	0.0132183 mm
Confidence level:	95.0 %
Estimate:	29.4874 mm +/- 0.0265632 mm
Range:	29.4609 mm to 29.514 mm

Sample: Don't break the ice
Here is Fathom output for a 95% confidence interval for the true mean width of plastic ice cubes produced this hour.

Problem:
(a) Interpret the confidence interval. Would you make the same conclusion with the confidence interval as you did with the significance test in the previous example?
(b) Interpret the confidence level.

Solution:
(a) We are 95% confident that the interval from 29.4609 to 29.514 contains the true mean width of plastic ice cubes.
Yes, 29.4874 lies within the interval so we fail to reject H_0

(b) If we were to take many random samples of size 50 and make 95% confidence intervals for each. 95% of those intervals will contain the actual mean width of the plastic ice cubes.

9.3B **Inference for means: Paired data**

- Comparative studies are more convincing than single sample investigations.
- Therefore, one-sample inference less common than comparative inference.
- Study designs that involve making 2 observations on the same individual, or one observation on each of 2 similar individuals, result in **paired data**. (this experiment design is called matched pairs)

- When paired data result from measuring the same variable twice, we can make comparisons by analyzing the differences in each pair. If the conditions for inferences are met, we can use one-sample t procedures to perform inference about the mean difference μ_d. These methods are called **paired t procedures**.

Sample: For their second semester project in AP Statistics, Libby and Kathryn decided to investigate which line was faster in the supermarket, the express lane or the regular lane. To collect their data, they randomly selected 15 times during a week, went to the same store, and bought the same item. However, one of them used the express lane and the other used a regular lane. To decide which lane each of them would use, they flipped a coin. If it was heads, Libby used the express lane and Kathryn used the regular lane. If it was tails, Libby used the regular lane and Kathryn used the express lane. They entered their randomly assigned lanes at the same time and each recorded the time in seconds it took them to complete the transaction.

Time in Express Lane E (seconds)	Time in Regular Lane R (seconds)	
337	342	5
226	472	246
502	456	-46
408	529	121
151	181	30
284	339	55
150	229	79
357	263	-94
349	332	-17
257	352	95
321	341	20
383	397	14
565	694	129
363	324	-39
85	127	42

Problem: Carry out a test to see if there is convincing evidence that the express lane is faster.

Solution: Since the data is paired, we will consider the differences in time (R-E) so a positive value means the express lane was faster.

State: We will test the hypotheses at a 0.05 significance level.

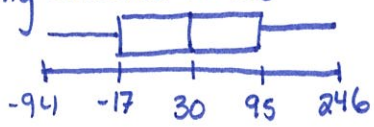
$H_0: \mu_d = 0$ $H_a: \mu_d > 0$ where μ = actual mean difference in time to checkout.

Plan: Use a paired t-test for means

Random: Randoms times at supermarket and randomly selected lanes

Normal: 15 < 30 so check for outliers/skewness

no outliers, no strong skew → approx. normal



Independent: Differences are independent

Do:

$t = \frac{42.67 - 0}{84.02/\sqrt{15}} = 1.967$ Stats → test → t-test → data

$\mu_0 = 0$, L3, $> \mu_0$

P-value = 0.03

0.03 < 0.05 Reject H_0

Conclude: We have convincing evidence that the express lane is faster than the regular lane.

9.3B

About paired data

- Individual scores are dependent
- Differences in scores are not dependent (independent)
- Be sure to report degrees of freedom when using calculator
- If subjects in an experiment were not randomly chosen, we can't generalize our findings to entire population.
- If subjects in an experiment were not randomly assigned a treatment, we can't make an inference about cause and effect.
- A confidence interval gives more information than a significance test.

9.3B

Using tests wisely

Significance tests are widely used in reporting the results of research in many fields.
New drugs require significant evidence of effectiveness and safety.
Courts ask about statistical significance in hearing discrimination cases.
In all cases, statistical significance is valued because it points to an effect that is unlikely to happen by chance.

- Statistical significance is not the same thing as practical importance. Pay attention to the actual data (plotting data - outliers) along with the statistical significance to avoid this.
- The foolish user of statistics who feeds the data to a calculator or computer without exploratory analysis will often be embarassed. If you were to actually graph the data, would it be worthwhile to calculate?
- Don't ignore the lack of significance. Usually more data will be needed to help us make some conclusions.
- When planning a study, verify that the test you plan to use has a high probability (power) of detecting a difference of the size you hope to find.
- Statistical inference is not valid for all sets of data. Badly designed surveys or experiments will yield invalid results. Always ask how the data was produced.
- Running tests multiple times to get the significance you want will have little meaning.