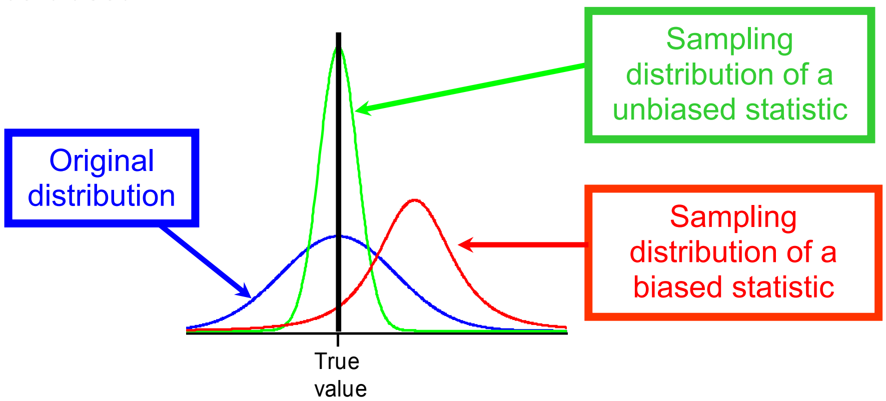
AP Statistics Notes – Unit Eight: Introduction to Inference

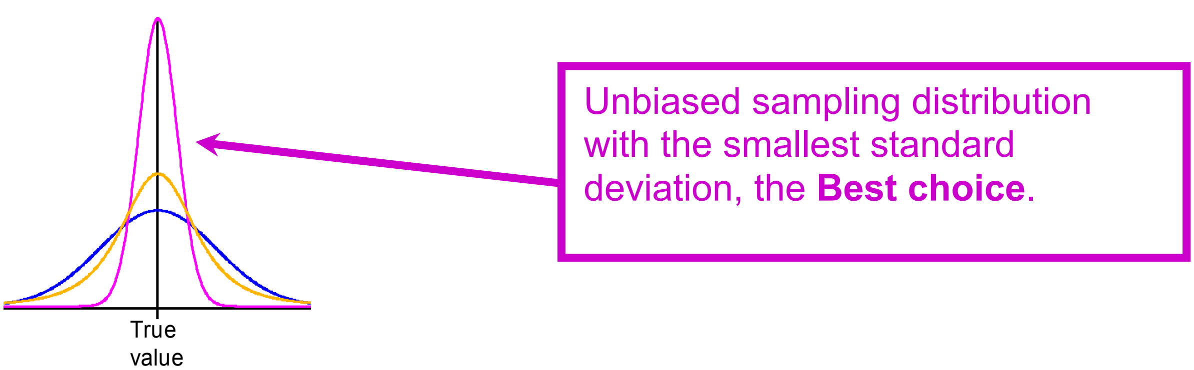
Syllabus Objectives: 4.1 – The student will estimate population parameters and margins of errors for means. 4.2 – The student will discuss the properties of point estimators, including biasedness and variability. 4.3 – The student will discuss the logic of confidence intervals, the meaning of a confidence level and an interval, and properties of confidence intervals.

When we select a sample, we want to infer some conclusion about the population that the sample represents. In this unit, we will be introduced to the two most common types of formal statistical inference: Confidence Intervals and Tests of Significance. Both types of inference are based on the sampling distributions of statistics. The purpose of this unit is to describe the reasoning used in inference – we will study specific procedures in later units.

- **Statistical Inference – provides methods for drawing conclusions about a population from sample data.**
  - **Point estimate** – A point estimate of a population characteristic is a single number that is based on sample data and represents a plausible value of the characteristic.
    - **Example:** A sample of weights of 34 male freshman students was obtained. If one wanted to estimate the true mean of all male freshman students, you might use the sample mean as a point estimate for the true mean.
  - **Bias** – A statistic with mean value equal to the value of the population characteristic being estimated is said to be an **unbiased** statistic. A statistic that is not unbiased is said to be **biased**.



  - **Variability** – Given a choice between several unbiased statistics that could be used for estimating a population characteristic, the best statistic to use is the one with the smallest variability – or the smallest standard deviation.

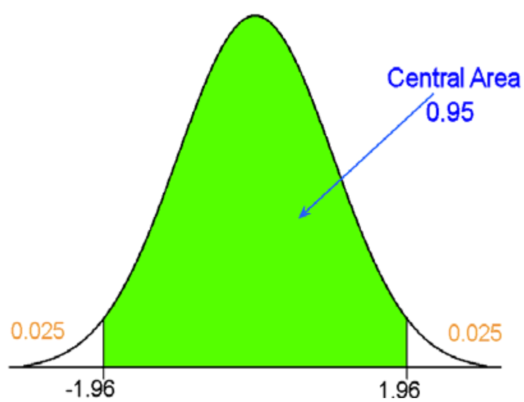


- **Confidence Intervals**

  - **Definition:** A confidence interval for a parameter is an interval of plausible values for the population characteristic. It is constructed so that, with a chosen degree of confidence, the value of the parameter will be captured inside the interval. We take the point estimate, $\bar{x}$, and add and subtract the same number to that, and that gives us an interval. (estimate $\pm$ margin of error)

  - **Confidence Level:** The confidence level associated with a confidence interval estimate specifies the success rate of the method used to construct the interval. The usual choices for levels are 90%, 95% and 99%, although others are possible. The level gives the probability that the interval will capture the true parameter value in repeated samples. For example, if the level is 95%, we are saying that in the long run, 95% of the resulting intervals would capture the true value of the parameter being estimated.

  - **Recall Sampling distributions** – For the sampling distribution of μ , $\mu_{\bar{x}} = \mu$ and and $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ and the sampling distribution of μ is approximately normal when the sample size is sufficiently large (*n* > 30). That means that approximately 95% of all large samples will result in a value of μ that is within $1.96\sigma_{\bar{x}} = 1.96\dfrac{\sigma}{\sqrt{n}}$ of the true population mean μ .

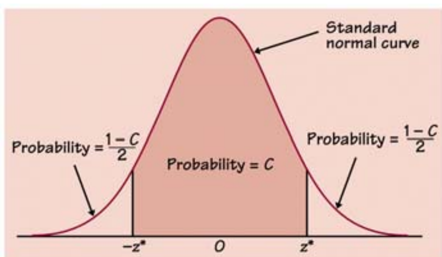


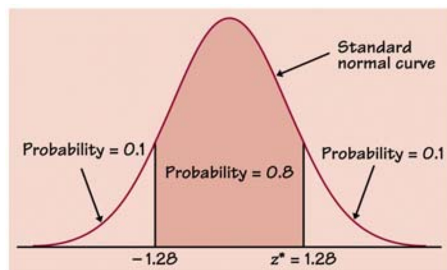
    Equivalently, this means that for 95% of all possible samples, μ will be in the interval $\mu - 1.96\dfrac{\sigma}{\sqrt{n}}$ to $\mu + 1.96\dfrac{\sigma}{\sqrt{n}}$.

  - **Conditions for constructing a confidence interval for μ:**
    1. The data is from an SRS from the <u>population of interest.</u>
    2. The sampling distribution of $\bar{x}$ is approximately normal. That means, either the population is known to be normal, or the sample size is large enough for us to use the Central Limit Theorem (CLT).

Syllabus Objective: 4.6 – The student will calculate the confidence interval for a mean.

- **Constructing a confidence interval for a sample mean**
  - Choose an SRS of size *n* from a population having unknown mean μ and known standard deviation σ. A level C confidence interval for μ is: $\bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}}$.
  - **Margin of error:** $z \cdot \frac{\sigma}{\sqrt{n}}$ is known as the margin of error (*m*). *z* is the critical value that we look up on the normal probability table and *n* is the sample size.
  - **Critical value:** Once you pick a confidence level, you must find the critical value (z-score) associated with that level. For example, if we want to be 80% confident (C = 0.80), then we must find the z-score that separates the middle 80% from the rest of the data. Subtract 0.8 from 1, divide that by 2 and we get 0.10. Using the normal probability table, find the z-score with a probability of 0.1 to the left of it. (It is 1.28 – see below).





  - **Most common critical values:**

C = 90%	Tail area = 0.05	z = 1.645
C = 95%	Tail area = 0.025	z = 1.96
C = 99%	Tail area = 0.005	z = 2.575

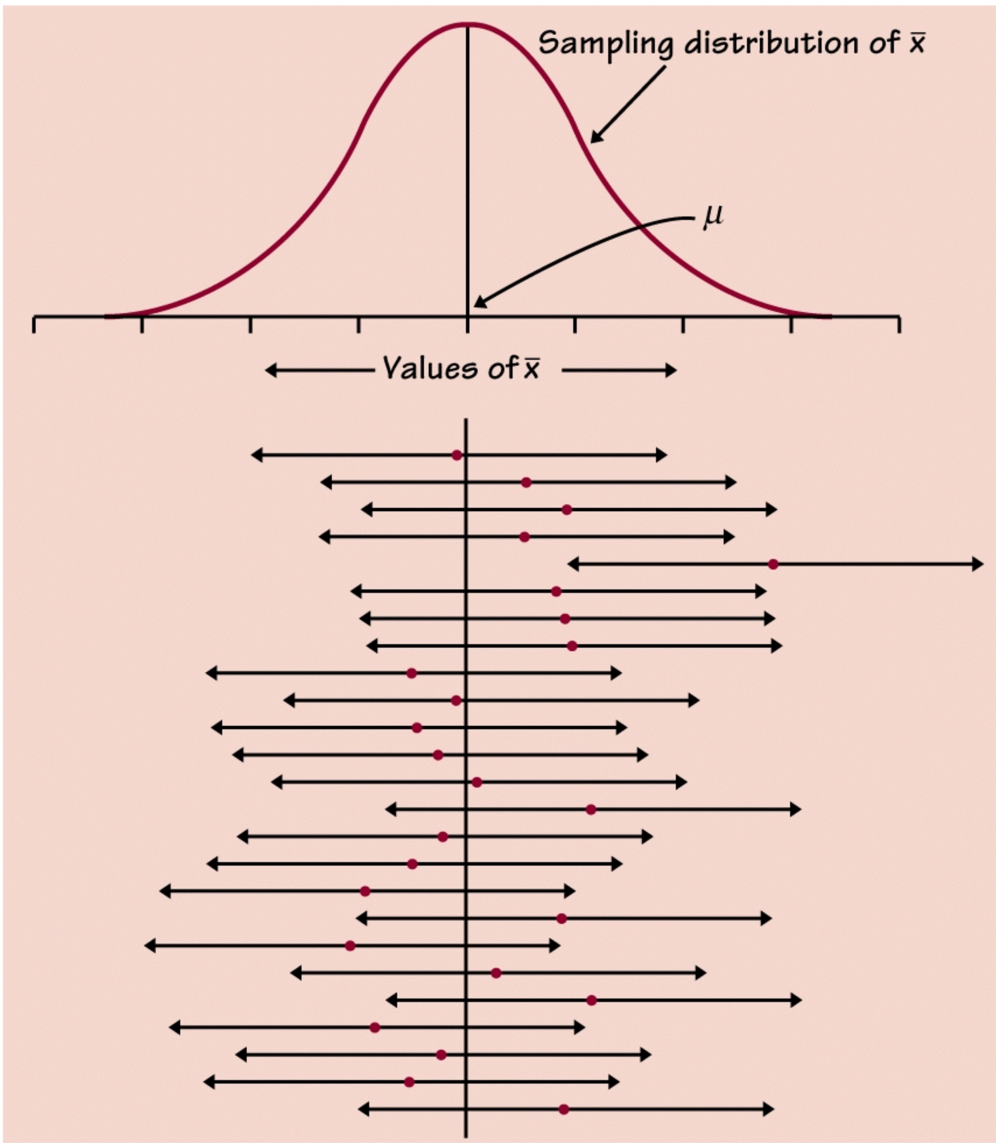
- **General construction of a C.I.:**
  1. Identify the population of interest and the parameter you want to draw conclusions about.
  2. Choose the appropriate inference procedure. Verify the conditions for using the selected procedure.
  3. If the conditions are met, carry out the inference procedure. CI = estimate $\pm$ margin of error.
  4. Interpret your results in the context of the problem.

***Once you choose, C, the confidence level, the margin of error follows from this choice. So, the margin of error gets smaller (your interval gets narrower) when: 1) *z* is smaller, or you pick a low C, 2) σ gets smaller or 3) *n* gets larger.

- **Sample size formula**

  If we rework the C.I. formula from the previous page, we can use the formula to solve for *n*, the sample size - $n \geq \left(\frac{z \cdot \sigma}{m} \right)^2$. If we know the level of confidence and margin of error we want, we can find the sample size needed for that confidence interval.

- **Graphical representation of confidence intervals:**



Twenty-five samples from the same population gave these 95% confidence intervals. In the long run, 95% of all samples give an interval that contains the true population mean, μ. Note above that 24 of the 25 intervals captured the true mean (96%).

- **Confidence Interval for a sample mean examples:**
  - **Example 1** – A certain filling machine has a true population standard deviation of $\sigma = 0.228$ ounces when used to fill catsup bottles. A random sample of 36 "6 ounce" bottles of catsup was selected from the output from this machine and the sample mean was 6.018 ounces. Find a 90% confidence interval estimate for the true mean fills of catsup from this machine.
    - **Solution:** We will follow the four steps previously stated. First, we must state the population parameter and the population of interest: $\mu =$ the true mean amount of catsup in the 6 ounce bottles. Secondly, we must verify that we can perform a 1-sample confidence interval. We know σ, the population standard deviation, we assume our sample is from an SRS and our sample size of 36 is greater than 30 so the CLT applies and the sampling distribution will be approximately normal. Now we can move to step 3 and find the confidence interval.
    - $\bar{x} = 6.018,\ \sigma = 0.228,\ n = 36$
    - To be 90% confident, the *z* critical value is 1.645.
    - $\bar{x} \pm (z \text{ critical value})\dfrac{\sigma}{\sqrt{n}} = 6.018 \pm (1.645)\dfrac{0.228}{\sqrt{36}} = 6.018 \pm 0.063$
    - The 90% confidence interval is: (5.955, 6.081)
    - The last step is that we must interpret it correctly. Here we would say that we are 90% confident that the true mean amount of catsup in 6 ounce bottles is between 5.955 ounces and 6.081 ounces.

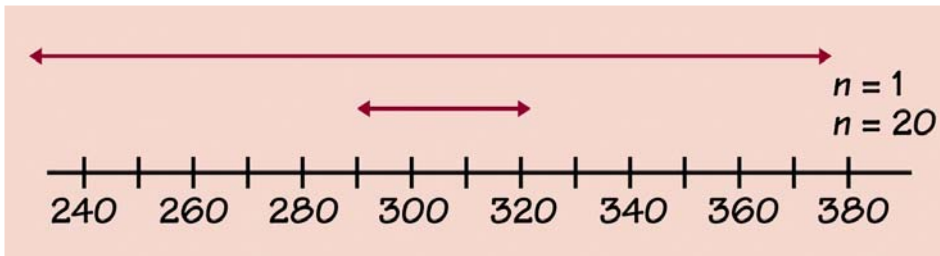
  - **Example 2** – A manufacturer of high-resolution video terminals must control the tension on the mesh of fine wires that lies behind the surface of the viewing screen. The tension is measured by an electrical device with output readings in millivolts (mV). Some variation is inherent in the production process. Careful study has shown that when the process is operating properly, the standard deviation of the tension readings is $\sigma = 43$ mV and that the distribution is approximately normal. Here are the tension readings from an SRS of 20 screens from a single day's production:

269.5	**297.0**	**269.6**	**283.3**	**304.8**	**280.4**	**233.5**	**257.4**	**317.5**	**327.4**
264.7	**307.7**	**310.0**	**343.3**	**328.1**	**342.6**	**338.8**	**340.1**	**374.6**	**336.1**

  Construct a 95% confidence interval for the mean tension of all the screens produced on this day.

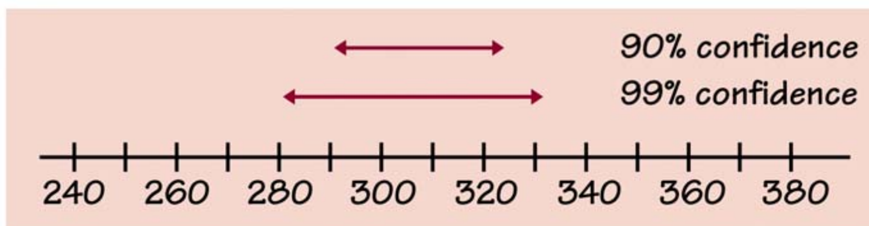
    - **Solution:** Step 1 – we identify our parameter and population of interest. The population of interest is all of the video terminals produced on the day in question. We want to estimate μ, the mean tension for all of these screens. Step 2 – We are told that we have an SRS of the population of interest and that the population distribution is approximately normal, so even though the sample size is less than 30, we may proceed. All conditions are satisfied to find the confidence interval.
    - Step 3: $\bar{x} = 306.3,\ \sigma = 43,\ n = 20$
    - To be 95% confident, the *z* critical value is 1.96.

- $\bar{x} \pm (z \text{ critical value})\dfrac{\sigma}{\sqrt{n}} = 306.32 \pm (1.96)\dfrac{43}{\sqrt{20}} = 306.32 \pm 9.615$
- The 95% confidence interval is (296.705, 315.935)
- Based on this sample, I am 95% confident that the true mean tension in the entire batch of video terminals produced that is day is between 296.705 mV and 315.935 mV.
- Suppose a single computer screen (*n* = 1) with the same mean of 306.32 had been used to find the confidence interval instead of a sample of 20. Notice how the two intervals compare below.



****Larger samples give shorter intervals.**

- Let's also compare the lengths of the 90% and 99% intervals for this same data.



****Smaller confidence gives shorter intervals.**

- **Example using the sample size formula** – Company management wants a report of the mean screen tension for the day's production accurate to within ± 5 mV with 95% confidence. How large a sample of video monitors must be measured to comply with this request?
  - **Solution:** For 95% confidence, the z critical value is 1.96. We also know that $\sigma = 43$ and the margin of error is 5.
  - $n \geq \left(\dfrac{z \cdot \sigma}{m}\right)^2 \rightarrow n \geq \left(\dfrac{(1.96)(43)}{5}\right)^2 \rightarrow n \geq 284.125, \text{ so take } n = 285$
  - Because *n*, sample size, must be a whole number, the company must measure the tension of 285 video screens to meet management's demand.

Syllabus Objectives: 4.9 – The student will discuss the logic of significance testing, null and alternative hypotheses, p-values, and one-and two-sided tests. 4.13 – The student will perform a test for a mean.

- **Tests of significance**
  - **Overview –** We will be studying two types of statistical inference. Use a confidence interval when your goal is to estimate a population parameter. The second type of inference, called **tests of significance**, has a different goal: to assess the evidence provided by data about some claim concerning a population. The test asks: Is the result we received reasonable or could we have gotten this result simply by chance?
  - **Hypotheses** – The test consists of two hypothesis statements.
    - The **null hypothesis,** H_0**,** says that there is *no effect* or *no change* in the population. If the null is true, the sample result is just chance at work. The null hypothesis is always a statement of equality. It contains the population characteristic (μ) and the hypothesized value, which is a specific number determined by the problem context. $H_0 : \mu = \text{ hypothesized value}$
    - The **alternative hypothesis,** H_a **,** is the effect we suspect is true and is the alternative to no effect. The alternative can be less than, greater than or not equal to the hypothesized value. The inequality forms are called one-sided alternatives and the not equal to is called the two-sided alternative.

      $H_a : \mu <$ $\qquad$ $H_a : \mu >$ $\qquad$ $H_a : \mu \neq$

      - ◊ **Hypothesis example 1:** You would like to determine if the diameters of the ball bearings you produce have a mean of 6.5 cm.

        $H_0 :\ \mu = 6.5$

        $H_a :\ \mu \neq 6.5$

        This is an example of the two-sided alternative.
      - ◊ **Hypothesis example 2:** The students entering into the math program used to have a mean SAT quantitative score of 525. Are the current students poorer (as measured by the SAT quantitative score)?

        $H_0 :\ \mu = 525$

        $H_a :\ \mu < 525$

        This is an example of the one-sided alternative.
  - **Test Statistic –** This is the function of sample data on which a conclusion to reject or fail to reject H_0 is based. The ***z* test statistic** for a sample mean is $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$.
  - **P-Value –** A **P-Value**, also called the **observed significance level**, is a measure of inconsistency between the hypothesized value for a population characteristic and the observed sample. It is the probability, computed assuming that H_0 is true, that the observed outcome would take a value as extreme or more extreme than that actually observed.

- The smaller the P-Value is, the stronger is the evidence against H_0 provided by the data.
- The decisive value of P is called the **significance level** and we write it as α, the Greek letter alpha. If the P-value is as small as or smaller than alpha, we say that the data are **statistically significant at level** α. "Significant" does not mean "important". It just simply means that it is "not likely to happen just by chance". We get to choose the value of alpha. The most common values of alpha are 0.01, 0.05 and 0.10.
  - ◊ H_0 should be rejected if P-value $\leq \alpha$.
  - ◊ H_0 should not be rejected if P-value $> \alpha$.

- **Steps for the Hypothesis test**

  - STEP 1: Identify the population of interest and the parameter you want to draw conclusions about. You can do this in the statement $\mu = \ldots$
  - STEP 2: State the null and alternative hypotheses in symbols. You can also state this in words to combine Steps 1 and Step 2.
  - STEP 3: Choose the appropriate inference procedure. Verify the conditions for using the selected procedure. That means, name the procedure you are doing and check the assumptions.
  - STEP 4: If the conditions are met, carry out the inference procedure.

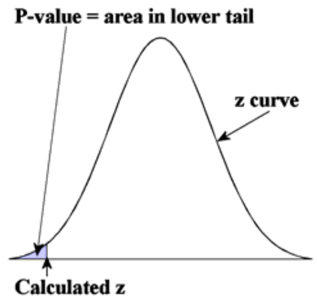
    -Calculate the test statistic – the mechanics part!
    -Find the P-value
  - STEP 5: Make your decision – Compare your p-value to α and decide to Reject or Fail to Reject H_0
  - STEP 6: Interpret your results in the context of the problem.
  - **Tests of significance assess the evidence against H_o. If the evidence is strong, we can confidently reject H_o in favor of the alternative. Failing to find evidence against H_o means only that the data are consistent with H_o, not that we have clear evidence that H_o is true. We either reject of fail to reject H_o. We never use the word "accept".

- **Cautions before we begin**
  1. The data MUST be an SRS from the population of interest.
  2. Because $\bar{x}$ is strongly influenced by a few extreme observations, outliers can have a big effect.
  3. If the sample size is small (n < 30) and the population is not normal, the inference cannot be accurately carried out.
  4. To carry out these procedures, you MUST know the standard deviation, σ, of the population. This unrealistic requirement renders our formulas of little use in statistical practice; however, we will learn in the next unit what to do when σ is not known.
  5. Remember there is more error associated with our inference procedures then just the confidence level or significance level we choose, which takes care of sampling errors. Practical difficulties, such as undercoverage and nonresponse in a sample survey, can cause additional errors.

- **Finding the significance level in the test**
  - If a significance level is not given, we will use $\alpha = 0.05$. Technically, we do not need a significance level. We can interpret our p-value without it and state whether we think it is statistically significant or not (small or large).
  - To find the p-value, we use the *z* test statistic and then the normal distribution table to look it up. You must multiply your p-value by two if you are doing a two-sided test.
  - The name of our Hypothesis test – single sample test of a population mean. Our test will take one of the three forms below. Finding the associated p-value is shown.

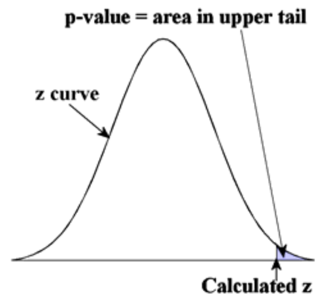
H_0: μ = hypothesized mean
H_A: μ < hypothesized mean

$$\text{P-value} = P\left(Z < \frac{\bar{x} - \text{hypothesized mean}}{\sigma / \sqrt{n}}\right)$$



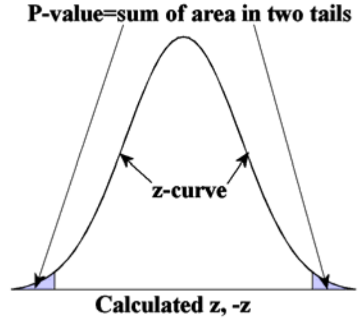
H_0: μ = hypothesized mean
H_A: μ > hypothesized mean

$$\text{P-value} = P\left(Z > \frac{\bar{x} - \text{hypothesized mean}}{\sigma / \sqrt{n}}\right)$$



H_0: μ = hypothesized mean
H_A: μ ≠ hypothesized mean

$$\text{P-value} = 2P\left(Z > \left|\frac{\bar{x} - \text{hypothesized mean}}{\sigma / \sqrt{n}}\right|\right)$$



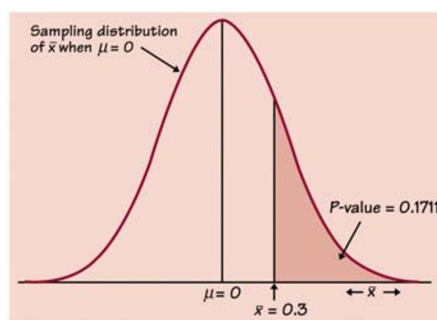
- **Tests of Significance for a sample mean examples**

  - **Example 1 -** Diet colas use artificial sweeteners to avoid sugar. These sweeteners gradually lose their sweetness over time and manufacturers test for loss of sweetness before marketing them. The tasters score the cola on a "sweetness score" of 1 to 10. The cola is then stored and after four months, the tasters rate the sweetness again using the same scale. The bigger the differences, the bigger the loss of sweetness. A sample of 10 trained tasters has a sample mean of 0.30. It is known that the individual tasters' scores vary according to a normal distribution with $\sigma = 1$. Is this good evidence that the cola lost sweetness in storage?

    - **Solution:** Step 1: Identify the population of interest and the parameter you want to draw conclusions about. $\mu =$ the true mean loss of sweetness in the diet cola.
    - Step 2: Write the hypotheses for the test. This will be a one-sided test. If there is NO sweetness loss, then our hypothesized value would be zero. If there is a loss, there would be a positive number. $H_0 : \mu = 0 \quad H_a : \mu > 0$
    - Step 3: Choose the appropriate inference procedure and verify the conditions for using the procedure. We are doing a one-sample mean hypothesis *z* test. We will assume we have an SRS of the population of diet cola. We also are told that the population has a normal distribution and the population standard deviation is given.
    - Step 4: Carry out the inference procedure. Find your variables, the test statistic and P-Value. $n = 10,\ \bar{x} = 0.3,\ \sigma = 1$

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{0.3 - 0}{1/\sqrt{10}} = \frac{0.3}{0.316} = 0.95$$

$$P\left(z > 0.95\right) = 0.1711$$



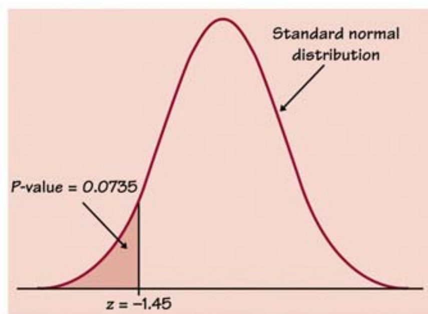
    - Step 5: Make your decision. Our P-Value is 0.1711. Since an alpha level is not given in the problem, we will use $\alpha = 0.05$. Since 0.1711 > 0.05, our decision is to Fail to Reject H_0.
    - Step 6: Interpret your results in the context of the problem. What our results show is that this could easily happen just by chance. The sample mean is not far from our expected value. That is, 17% of all samples would give a mean score as large or larger than 0.3 just by chance when the true population mean is 0. An outcome this likely to occur just by chance is not good evidence against the null hypothesis. These results are NOT statistically significant at the 0.05 level.

- **Example 2 -** In a discussion of the education level of the American workforce, someone says, "The average young person can't even balance a checkbook." The NAEP survey says that a score of 275 or higher on its quantitative test reflects the skill needed to balance a checkbook. The NAEP random sample of 840 young Americans had a mean score of $\bar{x} = 272,$ a bit below that level. Is this sample result good evidence that the mean for all young men is less than 275? The population standard deviation is 60.

  - **Solution:** Step 1: $\mu =$ the true mean NAEP score of all young American men.
  - Step 2: $H_0 : \mu = 275$ and $H_0 : \mu < 275$
  - Step 3: We will use a one-sample *z* test. We assume the data come from an SRS from all young American men. Since *n* = 840, the central limit theorem tells us that the sampling distribution of $\bar{x}$ will be approximately normal. Also, σ is known, so all of the conditions are met.
  - Step 4: $n = 840,\ \bar{x} = 272,\ \sigma = 60$

    $$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{272 - 275}{60 / \sqrt{840}} = \frac{-3}{2.07} = -1.45$$

    $$P - Value = P(z < -1.45) = 0.0735$$



  - Step 5: Let $\alpha = 0.05$. Then our P-Value is larger than the significance level. (0.0735 > 0.05). This means that we will Fail to Reject H_0. However, note that if we had used a different significant level, our decision would change. If $\alpha = 0.10$, then 0.0735 < 0.10 and we would Reject H_0 and say that our data WAS statistically significant at the 0.10 level and believe the alternative hypothesis to be true.
  - Step 6: A mean score as low as 272 would occur about 7 times in 100 samples if the population mean were 275. This is modest evidence that the mean NAEP score for all young Americans is less than 275, but is not significant at the 0.05 level. There is NOT significant evidence to suggest the NAEP score is less than 275 at the 0.05 level. There is significant evidence to suggest the NAEP score is less than 275 at the 0.10 level.

- **Example 3:** Let us revisit the manufacturing example used in Confidence Intervals Example 2. The manufacturer knows from careful study that the proper tension of the mesh in a video terminal is 275 mV. Is there significant evidence at the 1% level that $\mu \neq 275$, give our sample of 20 provides a mean of 306.32?

- **Solution:** Step 1: $\mu =$ the true mean tension of the screens produced that day.
- Step 2: This test is a 2-sided test. $H_0 : \mu = 275$ and $H_0 : \mu \neq 275$
- Step 3: We will use a one-sample mean *z* test. We assume the data come from an SRS from the population of interest. We were previously told that the population follows a normal distribution. Also, σ is known, so all of the conditions are met.
- Step 4: $n = 20,\ \bar{x} = 306.32,\ \sigma = 43$

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{306.32 - 275}{43 / \sqrt{20}} = \frac{31.32}{9.615} = 3.26$$

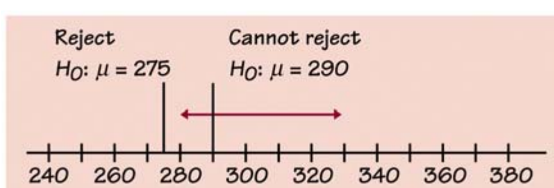
$$P - Value = 2P(z > 3.26) = 2(0.0006) = 0.0012$$

- Step 5: The problem states that $\alpha = 0.01$. On this problem, our P-Value is smaller than the significance level. (0.0012 < 0.01). This means that we will Reject H_0.
- Step 6: Based on this sample, there IS significant evidence to show that the screen tension for the day's population is not at the desired 275 mV level.

- **The similarities between confidence intervals and significance tests**
  - A level α two-sided significance test rejects a hypothesis $H_0 : \mu = \mu_0$ exactly when the value μ_0 falls outside a level $1 - \alpha$ confidence interval for μ.
  - A 99% confidence interval would be related to a 2-sided test with a significance level of 0.01. A 95% confidence interval would be related to a 2-sided test with a significance level of 0.05. A 90% confidence interval would be related to a 2-sided test with a significance level of 0.10.
  - **Example:** Again we will revisit the manufacturing problem. We will find the 99% confidence interval for the mean screen tension μ.

$$\bar{x} \pm (z\ critical\ value)\frac{\sigma}{\sqrt{n}} = 306.32 \pm (2.576)\frac{43}{\sqrt{20}} = 306.32 \pm 24.768 = (281.55, 331.09)$$

  - We are 99% confident that this interval captures the true population mean, μ. But our hypothesized population mean of 275 is not in the interval. So, we conclude that our null hypothesis, $\mu = 275$, is implausible. Thus, we conclude that μ is different from 275. Note that this is consistent with our conclusion in Example 3 above. If our null hypothesis had been 290, then that value would have been consistent with our 99% confidence interval, we would have "captured" it and we would not have been able to reject H_0.



Syllabus Objective: 4.10 – The student will discuss the concepts of Type I and Type II errors and the concept of power.

- **Type I and Type II errors**
  - When you set up a hypothesis test, the result is either strong support for the alternate hypothesis (if the null hypothesis is rejected) or there is not sufficient evidence to refute the claim of the null hypothesis (you are stuck with it, because there is a lack of strong evidence against the null hypothesis).
  - A **Type I error** is to reject the null hypothesis when it is actually true. A **Type II error** is to fail to reject the null hypothesis (believe it is true) when actually it is false.
  -

	Null Hypothesis	
Decision	**True**	**False**
Accept H_0	No Error	Type II Error β
Reject H_0	Type I Error α	No Error

  - The probability of a Type I error is denoted by α and is called the **level of significance** of the test. The probability of a Type II error is denoted by β.
  - **Relationship between the two errors:** Generally, with everything else held constant, decreasing one type of error causes the other to increase. As $\alpha\uparrow,\ \beta\downarrow$ and vice versa. The only way to decrease both types of error simultaneously is to increase the sample size. No matter what decision is reached, there is always the risk of one of these errors.
  - **Which is worse?** Look at the consequences of Type I and Type II errors and then identify the largest α that is tolerable for the problem. Employ a test procedure that uses this maximum acceptable value of α (rather than anything smaller) as the level of significance (because using a smaller α increases β).
  - **Examples of Type I and Type II errors**
    - **Example 1:** Consider a medical test where the hypotheses are equivalent to H_0 : the patient has a specific disease and H_a : the patient doesn't have the disease. Then, a Type I error is equivalent to a false negative (i.e., saying the patient does not have the disease when in fact, he does.) A Type II error is equivalent to a false positive. (i.e., saying the patient has the disease when, in fact, he does not.)

    - **Example 2:** When a law firm represents a group of people in a class action lawsuit and wins that lawsuit, the firm receives a percentage of the group's monetary settlement. That settlement amount is based on the total number of people in the group—the larger the group and the larger the settlement, the more money the firm will receive.

A law firm is trying to decide whether to represent car owners in a class action lawsuit against the manufacturer of a certain make and model for a particular defect. If 5% or less of the cars of this make and model have the defect, the firm will not recover its expenses. Therefore, the firm will handle the lawsuit only if it is convinced that more than 5 percent of cars of this make and model have the defect. The firm plans to take a random sample of 1,000 people who bought this car and ask them if they experienced this defect in their cars.

(a) Define the parameter of interest and state the null and alternative hypothesis that the law firm should test.

Solution: *p* = the true proportion of all cars of the specified make and model that have the defect. $H_0 : p = 0.05$ and $H_0 : p > 0.05$

(b) In the context of the situation, describe Type I and Type II errors <u>and</u> describe the consequences of each of these for the law firm.

Solution: Type I error: The law firm believes that the proportion of cars that have the defect is greater than 0.05, when in fact it is not. Consequence of the Type I error: The firm will not recover its expenses, resulting in a loss to the firm. Type II error: The law firm is not convinced that the proportion of cars that have the defect is greater than 0.05, when in fact it is. Consequence of the Type II error: The firm will miss an opportunity to make money on this case.

- **Power**
    - **Definition:** The probability that a fixed level α significance test will reject H_0 when a particular alternative value of the parameter is true is called the **power** of the test against that alternative.
    - The power of a test against any alternative is 1 – P(Type II error) or 1 – P(β).
    - If the probability of a Type II error is 10% (0.10), then the Power of the test is .9 or 90%.
    - A high power is desirable. A P-Value describes what would happen supposing the null hypothesis is true. Power describes what would happen supposing that a particular alternative is true.
    - There are four ways to increase the power of a test.
        1. **Increase** α . If we increase α , β will decrease and the power will increase.
        2. Increase the sample size. More data provides more information and makes the test more powerful.
        3. Decrease σ . Decreasing variability will affect the shape of the curve.
        4. Consider an alternative hypothesis value that is farther away from what you believe the true parameter to be. Choose one as far as possible from your hypothesized value.
    - Using a significance test with low power makes it unlikely that you will find a significant effect even if the truth is far from the null hypothesis.

- **Examples from previous AP exams**

  - **Example 1:** A safety group claims that the mean speed of drivers on a highway exceeds the posted speed limit of 65 miles per hour (mph). To investigate the safety group's claim, which of the following statements is appropriate?

    (A) The null hypothesis is that the mean speed of drivers on this highway is less than 65 mph.
    (B) The null hypothesis is that the mean speed of drivers on this highway is greater than 65 mph.
    (C) The alternative hypothesis is that the mean speed of drivers on this highway is greater than 65 mph.
    (D) The alternative hypothesis is that the mean speed of drivers on this highway is less than 65 mph.
    (E) The alternative hypothesis is that the mean speed of drivers on this highway is greater than or equal to 65 mph.

    - **Solution:** The answer is C. Answers A and B cannot be correct because the null hypothesis is always a statement of equality. Since the question states we believe it EXCEEDS the limit, greater than is the symbol we are looking for. E is not correct because it also includes "equal to" 65, which is not part of the alternative hypothesis.

  - **Example 2:** A random sample has been taken from a population. A statistician, using this sample, needs to decide whether to construct a 90 percent confidence interval for the population mean or a 95 percent confidence interval for the population mean. How will these intervals differ?

    (A) The 90 percent confidence interval will not be as wide as the 95 percent confidence interval.
    (B) The 90 percent confidence interval will be wider than the 95 percent confidence interval.
    (C) Which interval is wider will depend on how large the sample is.
    (D) Which interval is wider will depend on whether the sample is unbiased.
    (E) Which interval is wider will depend on whether a *z*-statistic or a *t*-statistic is used.

    - **Solution:** The answer is A. The confidence level C helps determine the width of the interval. If all other things remain constant and the level is increased, to be that confident, we must increase our interval making it wider. A lower confidence level results in a narrower interval.

  - **Example 3:** A consulting statistician reported the results from a learning experiment to a psychologist. The report stated that on one particular phase of the experiment a statistical test result yielded a p-value of 0.24. Based on this p-value, which of the following conclusions should the psychologist make?

    (A) The test was statistically significant because a p-value of 0.24 is greater than a significance level of 0.05.

(B) The test was statistically significant because p = 1 – 0.24 = 0.76 and this is greater than a significance level of 0.05.
(C) The test was not statistically significant because 2 times 0.24 = 0.48 and that is less than 0.5.
(D) The test was not statistically significant because, if the null hypothesis is true, one could expect to get a test statistic at least as extreme as that observed 24% of the time.
(E) The test was not statistically significant because, if the null hypothesis is true, one could expect to get a test statistic at least as extreme as that observed 76% of the time.

- **Solution:** The answer is D. The P-Value of 0.24 is large (we usually compare ot 0.05) and we would fail to reject the null hypothesis, meaning the test was not statistically significant.

- **Example 4:** A quality control inspector must verify whether a machine that packages snack foods is working correctly. The inspector will randomly select a sample of packages and weigh the amount of snack food in each. Assume that the weights of food in packages filled by the machine have a standard deviation of 0.30 ounce. An estimate of the mean amount of snack food in each package must be reported with 99.6 percent confidence and a margin of error of no more than 0.12 ounce. What would be the minimum sample size for the number of packages the inspector must select?

  (A) 8
  (B) 15
  (C) 25
  (D) 52
  (E) 60

  - **Solution:** The answer is D. Using the sample size formula, $m = 0.12,\ \sigma = 0.30$ and the *z* critical value for 99.6% confidence is 2.88.
    $n \ge \left(\frac{(2.88)(0.30)}{0.12}\right)^2 \rightarrow n \ge (7.2)^2 \rightarrow n \ge 51.84$. Rounding this to the largest whole number, we get 52.

- **Example 5:** In a test of the hypothesis H_o: $\mu = 100$ versus H_a: $\mu > 100$, the power of the test when $\mu = 101$ would be greatest for which of the following choices of sample size *n* and significance level α?

  (A) $n = 10$, $\alpha = 0.05$
  (B) $n = 10$, $\alpha = 0.01$
  (C) $n = 20$, $\alpha = 0.05$
  (D) $n = 20$, $\alpha = 0.01$
  (E) It cannot be determined from the information given.

  - **Solution:** The answer is C. Two of the ways to increase power is to increase the sample size and increase the significance level. Answer C has the largest sample size and largest significance level.

- **Example 6:** The analysis of a random sample of 500 households in a suburb of a large city indicates that a 98 percent confidence interval for the mean family income is ($41,300, $58,630). Could this information be used to conduct a test of the null hypothesis n a test of the hypothesis H_o: $\mu = 40{,}000$ against the alternative hypothesis H_a: $\mu \neq 40{,}000$ at the $\alpha = 0.02$ level of significance?

  (A) No, because the value of σ is not known.
  (B) No, because it is not known whether the data are normally distributed.
  (C) No, because the entire data set is needed to do this test.
  (D) Yes, since $40,000 is not contained in the 98 percent interval, the null hypothesis would be rejected in favor of the alternative, and it could be concluded that the mean family income is significantly different from $40,000 at the $\alpha = 0.02$ level.
  (E) Yes, since $40,000 is not contained in the 98 percent interval, the null hypothesis would not be rejected, and it could be concluded that the mean family income is significantly different from $40,000 at the $\alpha = 0.02$ level.

  - **Solution:** The answer is E. We can make a judgment without doing a hypothesis test by looking at the interval. The interval does NOT contain the hypothesized value, so it is not a plausible value. Therefore we would reject and say the results were statistically significant.

- **Example 7:** Ten students were randomly selected from a high school to take part in a program designed to raise their reading comprehension. Each student took a test before and after completing the program. The mean of the differences between the score after the program and the score before the program is 16. It was decided that all students in the school would take part in this program during the next school year. Let μ_A denote the mean score after the program and μ_B denote the mean score before the program for all students in the school. The 95 percent confidence interval estimate of the true mean difference for all students is (9, 23). Which of the following statements is a correct interpretation of this confidence interval?

  (A) $\mu_A > \mu_B$ with probability 0.95.
  (B) $\mu_A < \mu_B$ with probability 0.95.
  (C) μ_A is around 23 and μ_B is around 9.
  (D) For any μ_A and μ_B with $(\mu_A - \mu_B) \geq 14$, the sample result is quite likely.
  (E) For any μ_A and μ_B with $9 < (\mu_A - \mu_B) < 23$, the sample result is quite likely.

  - **Solution:** The answer is E. The confidence level does not refer to a probability so choices A and B are incorrect. The interval is just stating that we are 95% confident that the true parameter is contained in this range, so it would be very likely to obtain a sample result in this interval.