Fundamental of Hypothesis Testing

- There two types of statistical inferences, Estimation and Hypothesis Testing

- Hypothesis Testing: A hypothesis is a **claim** (assumption) about one or more **population** parameters.
  - Average price of a six-pack in the U.S. is $\mu = \$4.90$
  - The population mean monthly cell phone bill of this city is: $\mu = \$42$
  - The average number of TV sets in U.S. Homes is equal to three; $\mu = 3$

- It Is always about a population parameter, not about a sample statistic
- Sample evidence is used to assess the probability that the claim about the population parameter is true

A. It starts with Null Hypothesis, H_0

$$H_0: \mu = 3 \quad \text{and} \quad \bar{X} = 2.79$$

1. We begin with the assumption that H_0 is true and any difference between the sample statistic and true population parameter is due to chance and not a real (systematic) difference.
2. Similar to the notion of “innocent until proven guilty”
3. That is, “innocence” is a null hypothesis.

Null Hypo, Continued

4. Refers to the status quo
5. Always contains "=" , "≤" or "≥" sign
6. May or may not be rejected

B. Next we state the Alternative Hypothesis, $\mathbf{H_1}$

1. Is the opposite of the null hypothesis
   1. e.g., The average number of TV sets in U.S. homes is not equal to 3 ($H_1: \mu \neq 3$)
2. Challenges the status quo
3. Never contains the "=" , "≤" or "≥" sign
4. May or may not be proven
5. Is generally the hypothesis that the researcher is trying to prove. Evidence is always examined with respect to $\mathbf{H_1}$, never with respect to $\mathbf{H_0}$.
6. We never "accept" $\mathbf{H_0}$, we either "reject" or "not reject" it

Summary:

- In the process of hypothesis testing, the null hypothesis initially is assumed to be true
- Data are gathered and examined to determine whether the evidence is strong enough with respect to the alternative hypothesis to reject the assumption.
- In another words, the burden is placed on the researcher to show, using sample information, that the null hypothesis is false.
- If the sample information is sufficient enough in favor of the alternative hypothesis, then the null hypothesis is rejected. This is the same as saying if the persecutor has enough evidence of guilt, the “innocence is rejected.
- Of course, erroneous conclusions are possible, type I and type II errors.

Reason for Rejecting H_0

- Illustration: Let say, we **assume** that average age in the US is 50 years (H0=50). If in fact this is the true (unknown) population mean, it is unlikely that we get a sample mean of 20. So, if we have a sample that produces an average of 20, them we **reject** that the null hypothesis that average age is 50. (note that we are rejecting our assumption or claim). (would we get 20 if the true population mean was 50? NO. That is why we reject 50)

How Is the Test done?

- We use the distribution of a Test Statistic, such as Z or t as the criteria.

A. Rejection Region Method:

- Divide the distribution into rejection and non-rejection regions
- Defines the unlikely values of the sample statistic if the null hypothesis is true, the critical value(s)
  - Defines rejection region of the sampling distribution
- Rejection region(s) is designated by α , (level of significance)
  - Typical values are .01, .05, or .10
- α is selected by the researcher at the beginning
- α provides the critical value(s) of the test

Rejection Region or Critical Value Approach:

Level of significance = α

H_0: $\mu = 12$
H_1: $\mu \neq 12$

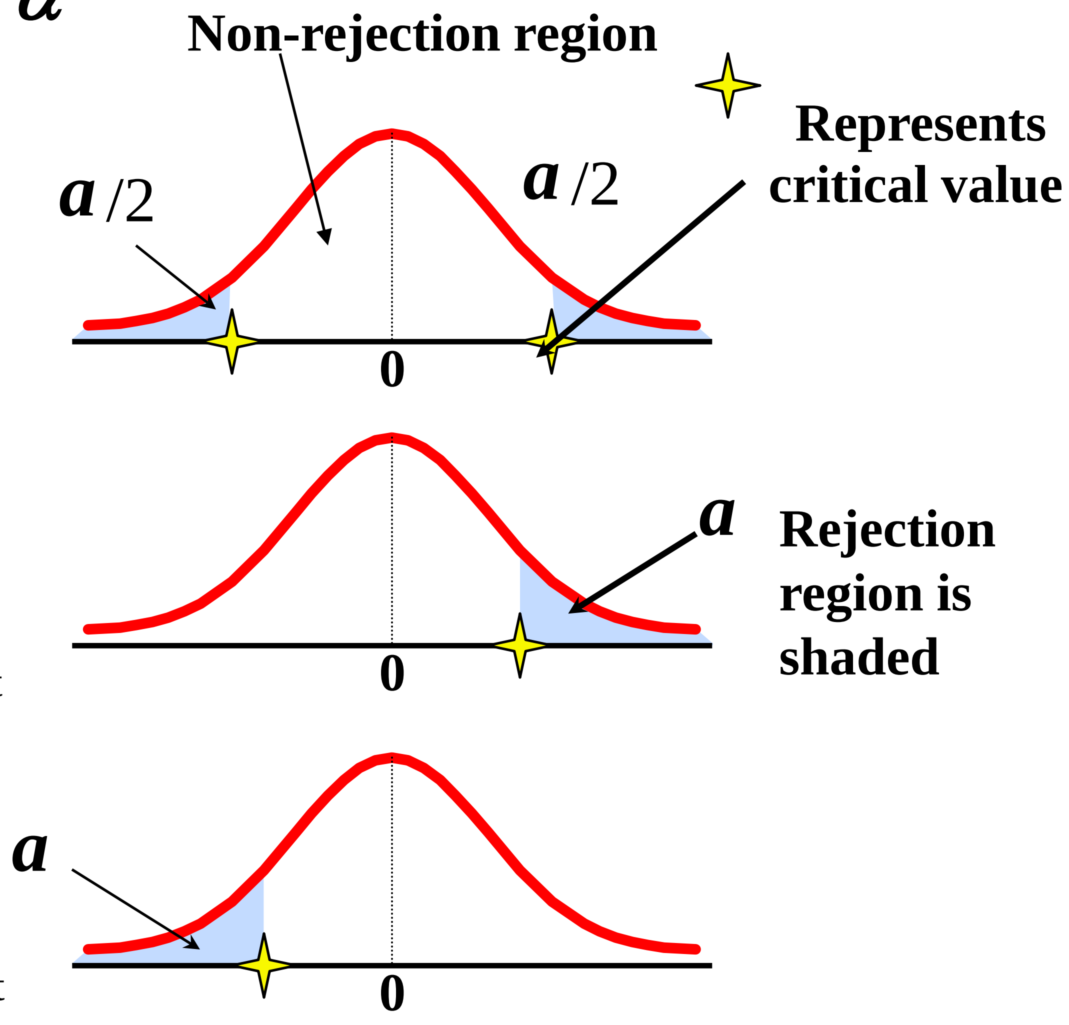
Two-tail test

H_0: $\mu \leq 12$
H_1: $\mu > 12$

Upper-tail test

H_0: $\mu \geq 12$
H_1: $\mu < 12$

Lower-tail test

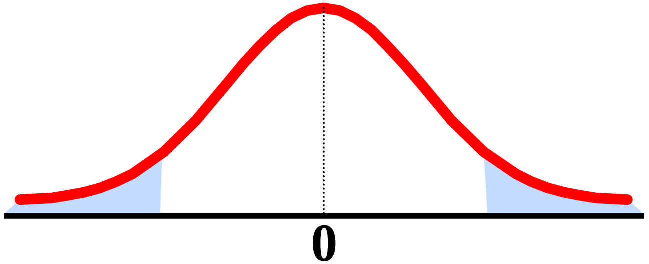


- P-Value Approach –
  - P-value=Max. Probability of (Type I Error), calculated from the sample.
- Given the sample information what is the size of blue are?

$H_0: \mu = 12$

$H_1: \mu \neq 12$

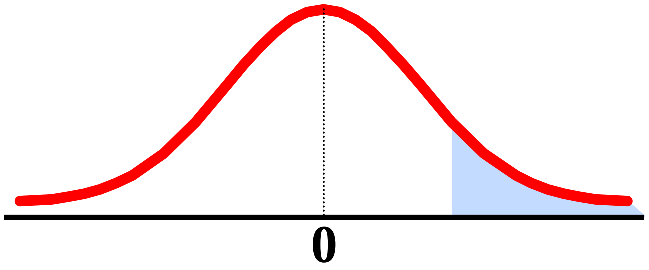
Two-tail test



$H_0: \mu \leq 12$

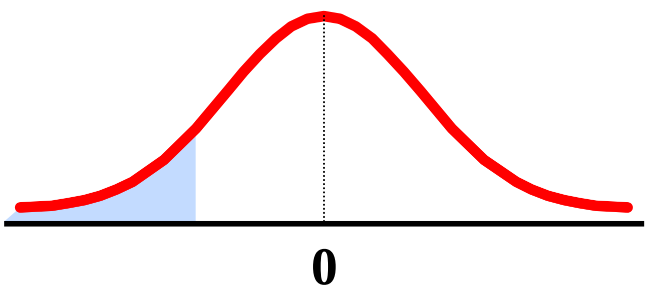
$H_1: \mu > 12$

Upper-tail test



$H_0: \mu \geq 12$

$H_1: \mu < 12$



Type I and II Errors:

- The size of α, the rejection region, affects the risk of making different types of incorrect decisions.

Type I Error

- Rejecting a true null hypothesis when it should NOT be rejected
- Considered a serious type of error
- The probability of Type I Error is α
- It is also called level of significance of the test

Type II Error

- Fail to reject a false null hypothesis that should have been rejected
- The probability of Type II Error is β

Decision	Actual Situation			
	Hypothesis Testing		Legal System	
	H0 True	H0 False	Innocence	Not innocence
Do Not Reject H_0	**No Error** $(1-\alpha)$	**Type II Error** (β)	No Error (not guilty, found not guilty) $(1-\alpha)$	Type II Error (guilty, found not guilty) (β)
Reject H_0	**Type I Error** (α)	**No Error** $(1-\beta)$	Type I Error (Not guilty, found guilty) (α)	No Error (guilty, found guilty) $(1-\beta)$

- Type I and Type II errors cannot happen at the same time

1. Type I error can only occur if H_0 is true
2. Type II error can only occur if H_0 is false
3. There is a tradeoff between type I and II errors. If the probability of type I error (α) increased, then the probability of type II error (β) declines.
4. When the difference between the hypothesized parameter and the actual true value is small, the probability of type two error (the non-rejection region) is larger.
5. Increasing the sample size, n, for a given level of α**,** reduces β

- **B. P-Value approach to Hypothesis Testing:**

1. The rejection region approach allows you to examine evidence but restrict you to not more than a certain probability (say α = 5%) of rejecting a true H_0 by mistake.
2. The P-value approach allows you to use the information from the sample and then calculate the maximum probability of rejecting a true H_0 by mistake.
3. Another way of looking at P-value is the probability of observing a sample information of "A=11.5" when the true population parameter is "12=B". The P-value is the maximum probability of such mistake taking place.

4. That is to say that P-value is the smallest value of α for which H_0 can be rejected based on the sample information
5. Convert Sample Statistic (e.g., sample mean) to Test Statistic (e.g., Z statistic)
6. Obtain the p-value from a table or computer
7. Compare the p-value with α
   - If p-value $<$ α , reject H_0
   - If p-value $\geq$ α , do not reject H_0

Test of Hypothesis for the Mean

σ known

The test statistic is:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

σ Unknown

The test statistic is:

$$t_{n-1} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

Steps to Hypothesis Testing

1. State the H_0 and H_1 clearly
2. Identify the test statistic (two-tail, one-tail, and Z or t distribution
3. Depending on the type of risk you are willing to take, specify the level of significance,
4. Find the decision rule, critical values, and rejection regions. If –CV<actual value (sample statistic) <+CV, then do not reject the H_0
5. Collect the data and do the calculation for the actual values of the test statistic from the sample

Steps to Hypothesis testing, continued

Make statistical decision

Do not Reject H_0

Reject H_0

Conclude H_0 may be true

Conclude H_1 is "true"

(There is sufficient evidence of H1)

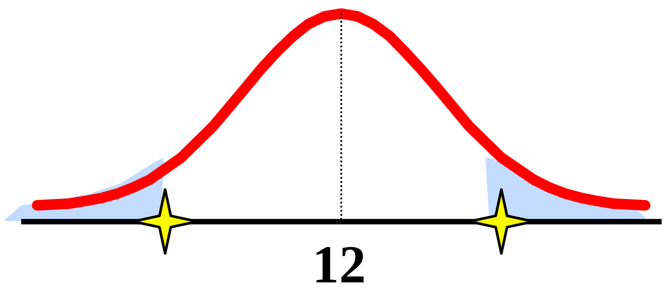
Make management/business/ administrative decision

When do we use a two-tail test?
when do we use a one-tail test?

- The answer depends on the question you are trying to answer.
- A two-tail is used when the researcher has no idea which direction the study will go, interested in both direction.
  (example: testing a new technique, a new product, a new theory and we don't know the direction)
- A new machine is producing 12 fluid once can of soft drink. The quality control manager is concern with cans containing too much or too little. Then, the test is a two-tailed test. That is the two rejection regions in tails is most likely (higher probability) to provide evidence of H_1.

$$H_0 : \mu = 12 \text{ oz}$$

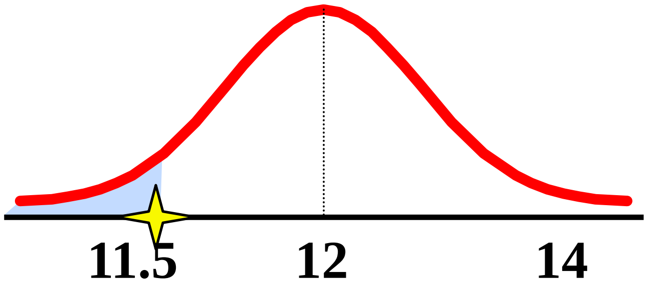
$$H_1 : \mu \neq 12 \text{ oz}$$



- One-tail test is used when the researcher is interested in the direction.
- Example: The soft-drink company puts a label on cans claiming they contain 12 oz. A consumer advocate desires to test this statement. She would assume that each can contains at least 12 oz and tries to find evidence to the contrary. That is, she examines the evidence for less than 12 0z.
- What tail of the distribution is the most logical (higher probability) to find that evidence? The only way to reject the claim is to get evidence of less than 12 oz, left tail.

$$H_0 : \mu \geq 12 \text{ oz}$$

$$H_1 : \mu < 12 \text{ oz}$$



Review of Hypo. Testing

- What is HT?
- Probability of making erroneous conclusions
    - Type I – only when Null Hypo is true
    - Type II – only when Null Hypo is false
- Two Approaches
    - The Rejection or Critical Value Approach
    - The P-value Approach (we calculate the observed level of significance)
- Test Statistics
    - Z- distribution if Population Std. Dev. is Know
    - t-distribution if the Population Std. Dev. is unknown

Rejection Region or Critical Value Approach:

The given level of significance = α

H_0: $\mu = 12$

H_1: $\mu \neq 12$

Two-tail test

H_0: $\mu \leq 12$

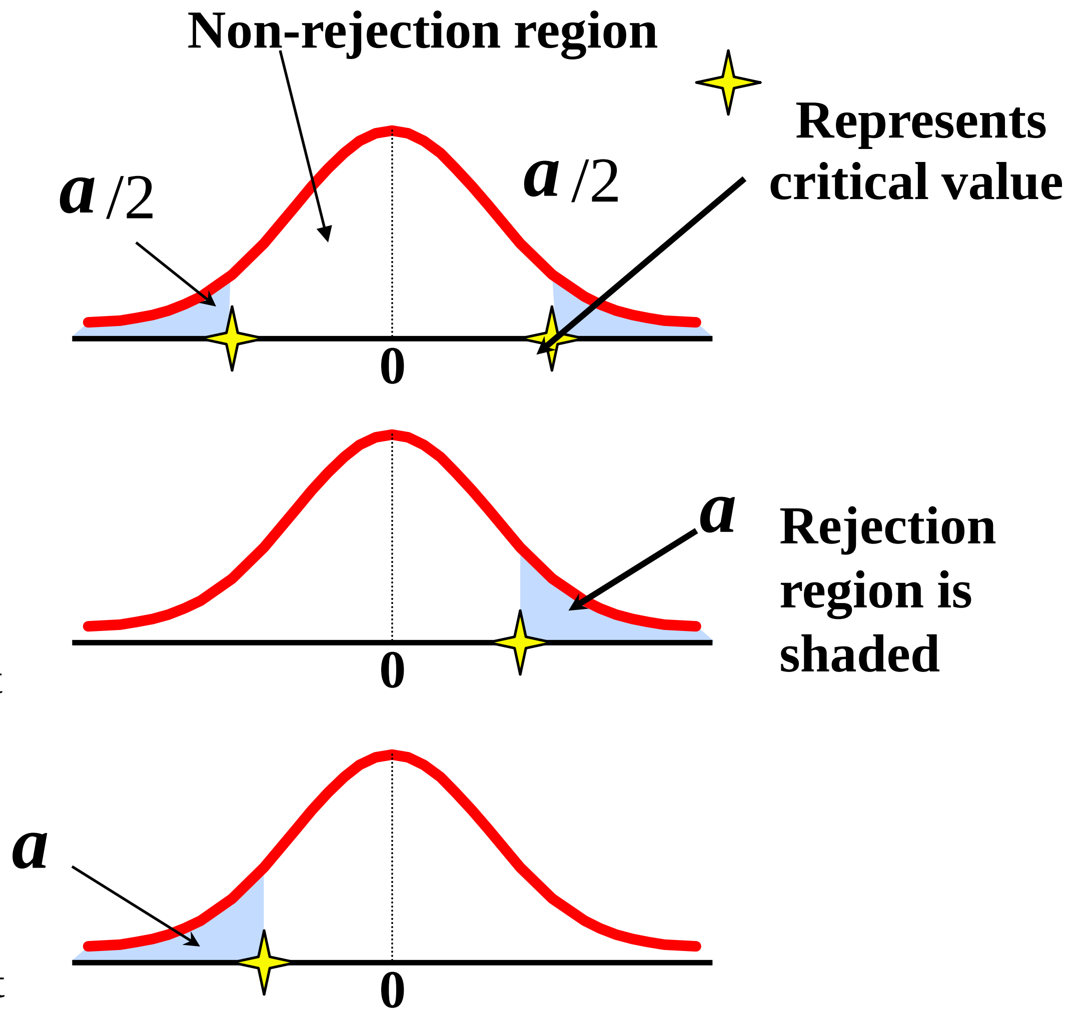
H_1: $\mu > 12$

Upper-tail test

H_0: $\mu \geq 12$

H_1: $\mu < 12$

Lower-tail test

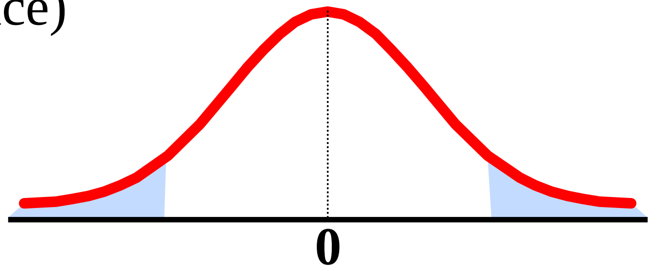


- P-Value Approach –
  - P-value=Max. Probability of (Type I Error), calculated from the sample.
- Given the sample information what is the size of the blue areas? (The observed level of significance)

H_0: $\mu = 12$

H_1: $\mu \neq 12$

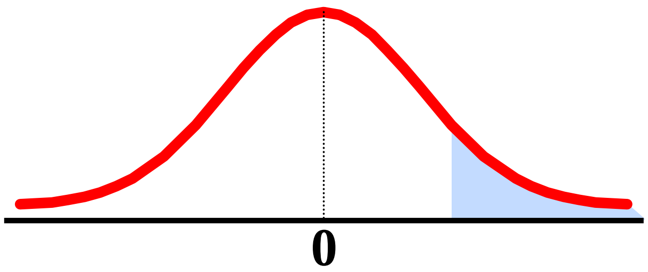
Two-tail test



H_0: $\mu \leq 12$

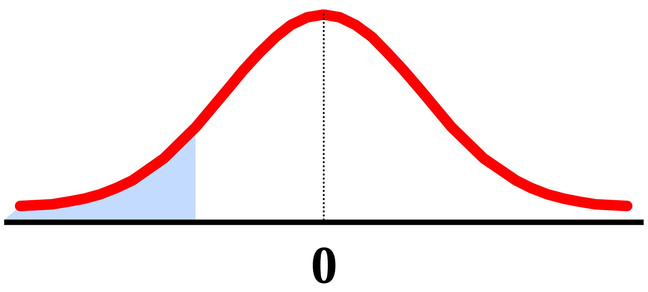
H_1: $\mu > 12$

Upper-tail test



H_0: $\mu \geq 12$

H_1: $\mu < 12$



- Example 1:
- Let's assume a sample of 25 cans produced a sample mean of 11.5 0z and the population std dev=1 0z.

- Question 1:
  - At a 5% level of significance (that is allowing for a maximum of 5% prob. of rejecting a true null hypo), is there evidence that the population mean is different from 12 oz?
    - Null Hypo is:?
    - Alternative Hypo is?

Can both approaches be used to answer this question?

A: Rejection region approach: calculate the actual test statistics and compare it with the critical values

B: P-value approach: calculate the actual probability of type I error given the sample information. Then compare it with 1%, 5%, or 10% level of significance.

– Interpretation of Critical Value/Rejection Region Approach:

– Interpretation of P-value Approach:

-
  - At a 5% level of significance (that is allowing for a maximum of 5% prob. of rejecting a true null hypo), is the evidence that the population mean is less than 12 oz?
    - Null Hypo is:?
    - Alternative Hypo is?

Can both approaches be used to answer this question?

- Interpretation of Critical Value Approach:

- Interpretation of P-value Approach:

- Question 3:
  - If in fact the pop. mean is 12 oz, <u>what is the probability</u> of obtaining a sample mean of 11.5 or less oz (sample size 25)? Null
    - Null Hypo is:?
    - Alternative Hypo is?
- Question 4:
  - If in fact the pop. mean is 12 oz, and the sample mean is 11.5 (or less), <u>what is the probability</u> of erroneously rejecting the null hypo that the pop. mean is 12 oz?
    - Null Hypo is:?
    - Alternative Hypo is?

Can both approaches be used to answer these question?

Example2

An automated production line system took an average of 90 minutes to create the finished product. Management has changed several key components of the production line recently, and wishes to know if the production time is still 90 minutes. A sample of 20 production runs took, on average, 85 minutes to create the finished product. Assuming that production times are normally distributed with $\sigma=7$ minutes, test the claim that the average production time has remained at 90 minutes. Use a significance level of 0.05.

Hypothesis Test:

H0: $\mu=90$ (claim)
H1: $\mu\neq 90$
Rejection region on either side beyond the critical value.
Critical value $z=\pm 1.96$ (two sided with $\alpha=0.05$).
Test Statistic: $z = (85-90)/(7/\sqrt{20}) = -3.19$
Test statistic is in the rejection region, hence we reject H0.
At $\alpha=0.05$ there is significant evidence that the production time has changed.

Example 3

A health club claims that its members lose an average of at least 10 pounds withi n the first month of joining. A consumer agency wishes to verify the claim and selects a sample of 36 first-time members. They lost an average of 9.2 pounds in the first month. If $\sigma=2.4$ pounds, test the club's claim at a 0.01 significance level.

Hypothesis Test:

H0: $\mu \geq 10$ (claim)
H1: $\mu<10$
Rejection region on the left side only, beyond the critical value of $z=-2.33$. (one sided with $\alpha=0.01$)
Test Statistic: $z= (9.2-10)/(2.4/\sqrt{36}) = -2.0$
We are not in the rejection region, hence do not reject H0.
At $\alpha=0.01$ we do not have sufficient evidence to reject the health club's claim.

Example 4

A retailer claims that the average consumer spends at most \$30 at their store. A tax auditor doubts that value and selects a random sample of 42 consumers. They spend an average of \$36.12 at the store. If consumer spending has an population standard deviation of \$14.30, calculate the p-value and test the retailer's claim. Use $\alpha=0.1$.

Hypothesis Test, with p-value:

H0: $\mu \leq 30$ (claim)
H1: $\mu > 30$
Rejection region on the right side only, beyond the critical value of
$z=+1.28$ (one tail, $\alpha=0.1$)
Test statistic: $z=(36.12-30)/(14.30/\sqrt{42})=2.77$
P-value is $p=1-.9972=0.0028$. (no need to double, as one-tailed test) We reject H0.
Note: Since the p-value is much lower (0.28%) than our signifcance level (10%) this is a very strong rejection!
Based on the collected sample, we have very significant evidence to reject the retailer's claimed value.

Example 5

In 1996 a population census revealed that the mean income level in a northern community of 5,000 inhabitants was \$19,368. A researcher believes that in 2006 this number had decreased. A 2006 random sample of 25 inhabitants has mean income \$18,610 with standard deviation \$2520. Test the researcher's claim using $\alpha=0.025$.

This is a hypothesis test, with a small sample and unknown σ, hence t-values must be used. We must also assume that the income levels are normally distributed.

Hypotheses are H0: $\mu \geq \$19368$, H1: $\mu < \$19368$ (claim)

Critical value, with d.f.=24 and a one-tailed test with 0.025 significance is t=-2.064. The rejection region is anything beyond the critical value (to the left).

Our test statistic is $t = (\bar{x}-\mu) / s/\sqrt{n} = -1.50$

We do not reject H0.

At $\alpha=0.025$ there is insufficient evidence to support the claim that average income has changed from 1996 levels.

Example 6

A published report states the mean first frost date in Southern Saskatchewan is date number 252 (September 10^{th}). A researcher doubts this value and has collected data on the first frost for the past 15 years. Her sample mean date is date number 256 (September 14^{th}) with a sample standard deviation of 7 days. Test the claim at the 0.05 level.

This is a hypothesis test, with a small sample and unknown σ, hence t-values must be used. We must also assume that the distribution of all first frost dates is normal.

Hypotheses are $H0: \mu=252$ (claim), $H1: \mu \neq 252$

Critical value, for a two-tailed test with d.f.=14 and $\alpha=0.05$ is $t=\pm 2.145$. We reject if our test statistic lies beyond (on either tail).

The test statistic is $t = (\bar{x}-\mu) / s/\sqrt{n} \;= 2.213$

Reject H0. At $\alpha=0.05$ there is sufficient evidence to doubt the stated first frost date of September 10^{th}.

Connection to Confidence Intervals

- While the confidence interval estimation and hypothesis testing serve different purposes, they are based on same concept and conclusions reached by two methods are consistent for a two-tail test.

- In CI method we estimate an interval for the population mean with a degree of confidence. If the estimated interval contains the hypothesized value under the hypothesis testing, then this is equivalent of not rejecting the null hypothesis. For example: for the beer sample with mean 5.20, the confidence interval is:

  $$P(4.61 \le \mu \le 5.78)=95\%$$

- Since this interval contains the Hypothesized mean ($4.90), we do not (did not) reject the null hypothesis at $\alpha = .05$

- Did not reject and within the interval, thus consistent results.