AP Statistics Notes – Unit Nine: Inference for the Mean of a Population

Syllabus Objectives: 3.22 – The student will describe the properties of student's t-distributions. 3.23 – The student will solve problems using tables of student's t-distributions.

In the last unit, we learned the logic behind inferential procedures. In this unit, we will apply that logic to inference involving means. We learn how to build confidence intervals and perform significance tests for one mean as well as comparisons between two means. Further, we will be introduced to a new distribution that we can use when we do not know the population standard deviation.

- **Conditions for inference about a mean**
  - The population standard deviation, σ, is unknown. We must estimate σ from the data.
  - Our data are from a **simple random sample** of size *n* from the population of interest. This condition is very important.
  - Observations from the population have a **normal distribution** with mean μ and standard deviation σ. In practice, it is enough that the distribution be symmetric and single-peaked unless the sample is very small.

- **Standard Error**
  - **Definition:** When the standard deviation of a statistic is estimated from the data, the result is called the **standard error** of the statistic.
  - **Formula:** $s/\sqrt{n}$

- **The t distributions**
  - **Definition:** An Irish mathematician/statistician W.S. Gosset developed the techniques and derived the Student's t distributions that describe the behavior of

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

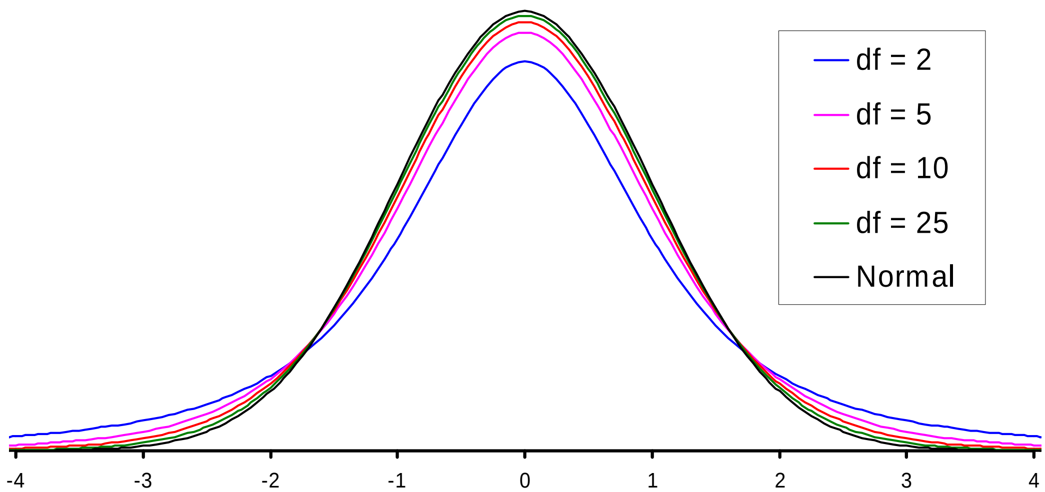
  - If X is a normally distributed random variable, the statistic follows a **t distribution** with df = *n* -1 (degrees of freedom). $$t=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$$

  - **Robustness of *t* procedures:** A confidence interval or test is called **robust** if the confidence level or p-value does not change very much when the assumptions of the procedure are violated. *t* procedures are quite robust against nonnormality of the population when there are no outliers, especially, when the distribution is roughly symmetric.
    - The condition of SRS is more important than the distribution being approximately normal. This condition should not be violated.
    - If the sample size is small (less than 15), use t procedures if the data is close to normal. If it is clearly nonnormal, then do not use t procedures.
    - If the sample size is medium (between 15 and 30), use t procedures as long as there are no outliers or strong skewness.

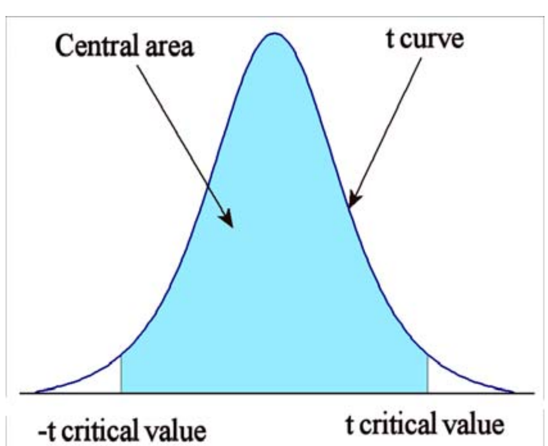
- For large samples (greater than 30), t procedures can be used for even skewed distributions due to the Central Limit Theorem (CLT).
- Always make a plot to check for skewness and outliers before using the t procedures. This can be done with a histogram, boxplot, dotplot, stemplot or normal probability plot.

- **Characteristics of the t distribution:**
  1. It is similar in shape to the standard normal curve. It is symmetric, centered at zero, single-peaked and bell-shaped.
  2. The spread of the *t* distribution is LARGER than the normal distribution. This is because we are estimating σ with *s,* so there is **more variability**, or spread.
  3. As n increases, the degrees of freedom increases and the *t* curve approaches the normal curve very closely. This happens because *s* estimates σ better as the sample size gets bigger.
  4. The *t* values are always a bit larger than the *z* values.
  5. We must use a new table – the *t* table instead of the *z* table.

- Comparison of normal and t distibutions



- **Notice: As df increase, *t* distributions approach the standard normal distribution. Since each *t* distribution would require a table similar to the standard normal table, we usually only create a table of critical values for the *t* distributions.**



- **The *t* table**

Central area captured:	0.80	0.90	0.95	0.98	0.99	0.998	0.999
Confidence level:	80%	90%	95%	98%	99%	99.8%	99.9%
Degrees of freedom							
1	3.08	6.31	12.71	31.82	63.66	318.29	636.58
2	1.89	2.92	4.30	6.96	9.92	22.33	31.60
3	1.64	2.35	3.18	4.54	5.84	10.21	12.92
4	1.53	2.13	2.78	3.75	4.60	7.17	8.61
5	1.48	2.02	2.57	3.36	4.03	5.89	6.87
6	1.44	1.94	2.45	3.14	3.71	5.21	5.96
7	1.41	1.89	2.36	3.00	3.50	4.79	5.41
8	1.40	1.86	2.31	2.90	3.36	4.50	5.04
9	1.38	1.83	2.26	2.82	3.25	4.30	4.78
10	1.37	1.81	2.23	2.76	3.17	4.14	4.59
11	1.36	1.80	2.20	2.72	3.11	4.02	4.44
12	1.36	1.78	2.18	2.68	3.05	3.93	4.32
13	1.35	1.77	2.16	2.65	3.01	3.85	4.22
14	1.35	1.76	2.14	2.62	2.98	3.79	4.14
15	1.34	1.75	2.13	2.60	2.95	3.73	4.07
16	1.34	1.75	2.12	2.58	2.92	3.69	4.01
17	1.33	1.74	2.11	2.57	2.90	3.65	3.97
18	1.33	1.73	2.10	2.55	2.88	3.61	3.92
19	1.33	1.73	2.09	2.54	2.86	3.58	3.88
20	1.33	1.72	2.09	2.53	2.85	3.55	3.85
21	1.32	1.72	2.08	2.52	2.83	3.53	3.82
22	1.32	1.72	2.07	2.51	2.82	3.50	3.79
23	1.32	1.71	2.07	2.50	2.81	3.48	3.77
24	1.32	1.71	2.06	2.49	2.80	3.47	3.75
25	1.32	1.71	2.06	2.49	2.79	3.45	3.73
26	1.31	1.71	2.06	2.48	2.78	3.43	3.71
27	1.31	1.70	2.05	2.47	2.77	3.42	3.69
28	1.31	1.70	2.05	2.47	2.76	3.41	3.67
29	1.31	1.70	2.05	2.46	2.76	3.40	3.66
30	1.31	1.70	2.04	2.46	2.75	3.39	3.65
40	1.30	1.68	2.02	2.42	2.70	3.31	3.55
60	1.30	1.67	2.00	2.39	2.66	3.23	3.46
120	1.29	1.66	1.98	2.36	2.62	3.16	3.37
z critical values	1.28	1.645	1.96	2.33	2.58	3.09	3.29

*****To use this table with confidence intervals, find your confidence level at the top and go down that column until you find the row containing your degrees of freedom. If your df is not on the table, use the closest most conservative one. For example, if df = 34, use df = 30. If df = 80, use df = 60.**

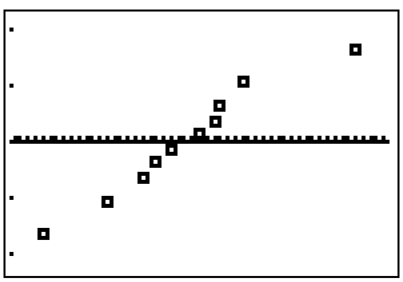
Syllabus Objectives: 4.6 – The student will calculate the confidence interval for a mean. 4.13 – The student will perform a test for a mean.

- **One-sample *t* Confidence Interval procedures**
  - Identify the population of interest and the parameter you want to draw conclusions about.
  - Choose the appropriate inference procedure. Verify the conditions for using the selected procedure. Make sure to plot the sample data to guarantee the sample data is symmetric.
  - If the conditions are met, carry out the inference procedure. CI = estimate $\pm$ margin of error.

$$\bar{x} \pm (\text{t critical value}) \frac{s}{\sqrt{n}}$$

  - Interpret your results in the context of the problem.

- **Confidence Interval for a sample mean example:**
  - **Example** – Ten randomly selected shut-ins were each asked to list how many hours of television they watched per week. The results are: 82, 66, 90, 84, 75, 88, 80, 94, 110 and 91. Find a 90% confidence interval estimate for the true mean number of hours of television watched per week by shut-ins.
    - **Solution:** We will follow the four steps previously stated in the last unit. First, we must state the population parameter and the population of interest: $\mu =$ the true mean number of hours of television watched per week by shut-ins. Secondly, we must verify that we can perform a 1-sample *t* confidence interval. We must first assume the 10 results are an SRS from the population of shut-ins. We do not know σ, the population standard deviation and our sample is small (less than 15). We must plot the sample data and check for normality. Plugging the data into the TI-84 and running a 1-Var Stats gives us the needed statistics for our C.I. and we can graph the data.



****Notice:** the normal probability plot looks reasonably linear, so it is **reasonable** to assume that the number of hours of television watched per week by shut-ins is normally distributed.

    - Now we can move to step 3 and find the confidence interval.
    - $\bar{x} = 86,\ s = 11.842,\ n = 10$
    - We find the critical *t* value of 1.833 by looking on the *t* table in the row corresponding to df = 9, in the column with top label 90%.
    - $\bar{x} \pm (t \text{ critical value}) \frac{s}{\sqrt{n}} = 86 \pm (1.833) \frac{11.842}{\sqrt{10}} = 86 \pm 6.86$
    - The 90% confidence interval is: (79.14, 92.86)
    - The last step is that we must interpret it correctly. Based on this sample, we are 90% confident that the true mean number of hours of television watched per week by shut-ins is between 79.14 and 92.86 hours.

- **One-sample *t* test procedures**
  - To test the null hypothesis, we use the following *t* statistic:

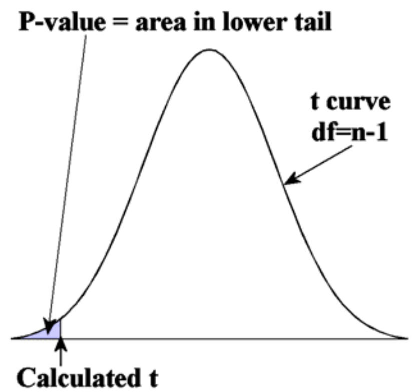
$$t = \frac{\bar{x} - \text{hypothesized mean}}{s/\sqrt{n}}$$

  - The approximate p-value for this test statistic is found using a *t* random variable with *n* -1 degrees of freedom.

  - **One-sided:**

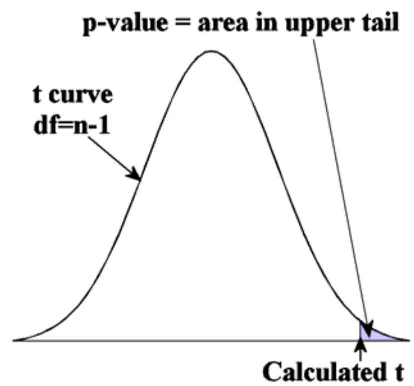
H_0: μ = hypothesized mean
H_A: μ < hypothesized mean

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



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

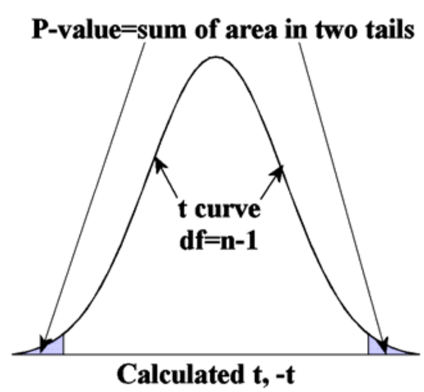
$$\text{P-value} = P\left(t > \frac{\bar{x} - \text{hypothesized mean}}{s/\sqrt{n}}\right)$$



  - **Two-sided:**

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

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



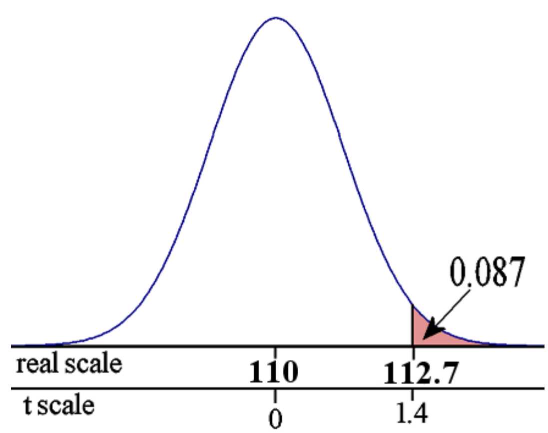
- **Tests of Significance for a sample mean examples**
  - **Example 1 -** A manufacturer of a special bolt requires that this type of bolt have a mean shearing strength in excess of 110 lb. To determine if the manufacturer's bolts meet the required standards a sample of 25 bolts was obtained and tested. The sample mean was 112.7 lb and the sample standard deviation was 9.62 lb. Use this information to perform an appropriate hypothesis test with a significance level of 0.05.
    - **Solution:** Step 1: Identify the population of interest and the parameter you want to draw conclusions about. $\mu =$ the true mean shearing strength of this specific type of bolt
    - Step 2: Write the hypotheses for the test. This will be a one-sided test. $H_0 : \mu = 110 \quad H_a : \mu > 110$
    - Step 3: Choose the appropriate inference procedure and verify the conditions for using the procedure. We are doing a one-sample mean hypothesis *t* test. We will assume we have an SRS of the population of bolts. Since our sample size is not over 30 and we do not have the sample data to plot and check for normality, we will have to assume our population is normally distributed.
    - Step 4: Carry out the inference procedure.

The test statistic is $\mathbf{t} = \dfrac{\bar{\mathbf{x}} - \mathbf{110}}{\mathbf{s}/\sqrt{\mathbf{n}}}$

$$t = \frac{\bar{x} - 110}{s/\sqrt{n}} = \frac{112.7 - 110}{9.62/\sqrt{25}} = \frac{2.7}{1.924} = 1.40$$

$$P\left(t > 1.40\right) = 0.087$$

See the *t* table on the next page as to how the p-value was found.



t \ df	13	14	15	16	17	18	19	20	21	22	23	24
0.0	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
0.1	0.461	0.461	0.461	0.461	0.461	0.461	0.461	0.461	0.461	0.461	0.461	0.461
0.2	0.422	0.422	0.422	0.422	0.422	0.422	0.422	0.422	0.422	0.422	0.422	0.422
0.3	0.384	0.384	0.384	0.384	0.384	0.384	0.384	0.384	0.384	0.383	0.383	0.383
0.4	0.348	0.348	0.347	0.347	0.347	0.347	0.347	0.347	0.347	0.347	0.346	0.346
0.5	0.313	0.312	0.312	0.312	0.312	0.312	0.311	0.311	0.311	0.311	0.311	0.311
0.6	0.279	0.279	0.279	0.278	0.278	0.278	0.278	0.278	0.277	0.277	0.277	0.277
0.7	0.248	0.248	0.247	0.247	0.247	0.246	0.246	0.246	0.246	0.246	0.245	0.245
0.8	0.219	0.219	0.218	0.218	0.217	0.217	0.217	0.217	0.216	0.216	0.216	0.216
0.9	0.192	0.192	0.191	0.191	0.190	0.190	0.190	0.189	0.189	0.189	0.189	0.189
1.0	0.168	0.167	0.167	0.166	0.166	0.165	0.165	0.165	0.164	0.164	0.164	0.164
1.1	0.146	0.145	0.144	0.144	0.143	0.143	0.143	0.142	0.142	0.142	0.141	0.141
1.2	0.126	0.125	0.124	0.124	0.123	0.123	0.122	0.122	0.122	0.121	0.121	0.121
1.3	0.108	0.107	0.107	0.106	0.105	0.105	0.105	0.104	0.104	0.104	0.103	0.103
1.4	0.092	0.092	0.091	0.090	0.090	0.089	0.089	0.088	0.088	0.088	0.087	0.087
1.5	0.079	0.078	0.077	0.077	0.076	0.075	0.075	0.075	0.074	0.074	0.074	0.073
1.6	0.067	0.066	0.065	0.065	0.064	0.064	0.063	0.063	0.062	0.062	0.062	0.061
1.7	0.056	0.056	0.055	0.054	0.054	0.053	0.053	0.052	0.052	0.052	0.051	0.051
1.8	0.048	0.047	0.046	0.045	0.045	0.044	0.044	0.043	0.043	0.043	0.042	0.042
1.9	0.040	0.039	0.038	0.038	0.037	0.037	0.036	0.036	0.036	0.035	0.035	0.035
2.0	0.033	0.033	0.032	0.031	0.031	0.030	0.030	0.030	0.029	0.029	0.029	0.028
2.1	0.028	0.027	0.027	0.026	0.025	0.025	0.025	0.024	0.024	0.024	0.023	0.023
2.2	0.023	0.023	0.022	0.021	0.021	0.021	0.020	0.020	0.020	0.019	0.019	0.019
2.3	0.019	0.019	0.018	0.018	0.017	0.017	0.016	0.016	0.016	0.016	0.015	0.015
2.4	0.016	0.015	0.015	0.014	0.014	0.014	0.013	0.013	0.013	0.013	0.012	0.012
2.5	0.013	0.013	0.012	0.012	0.011	0.011	0.011	0.011	0.010	0.010	0.010	0.010
2.6	0.011	0.010	0.010	0.010	0.009	0.009	0.009	0.009	0.008	0.008	0.008	0.008
2.7	0.009	0.009	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007	0.006	0.006
2.8	0.008	0.007	0.007	0.006	0.006	0.006	0.006	0.006	0.005	0.005	0.005	0.005
2.9	0.006	0.006	0.005	0.005	0.005	0.005	0.005	0.004	0.004	0.004	0.004	0.004
3.0	0.005	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.003	0.003	0.003	0.003
3.1	0.004	0.004	0.004	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.002
3.2	0.003	0.003	0.003	0.003	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002
3.3	0.003	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
3.4	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.001
3.5	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
3.6	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
3.7	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
3.8	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.000	0.000
3.9	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
4.0	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

t = 1.4
n = 25
df = 24
Tail area = 0.087

- Step 5: Make your decision. Our p-value is 0.087. The significance level given in the problem was $\alpha = 0.05$. Since 0.087 > 0.05, our decision is to Fail to Reject H_0.
- Step 6: Interpret your results in the context of the problem. At a level of significance of 0.05, there is insufficient evidence to conclude that the mean shearing strength of this brand of bolt exceeds 110 pounds.

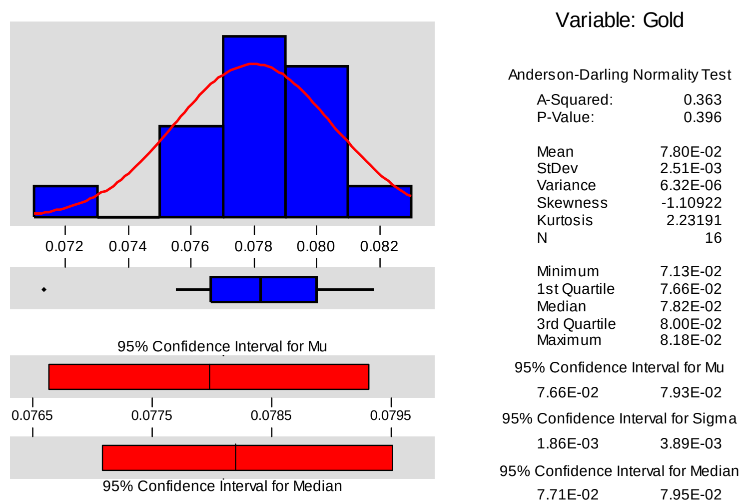
o **Example 2 –** A jeweler is planning on manufacturing gold charms. His design calls for a particular piece to contain 0.08 ounces of gold. The jeweler would like to know if the pieces that he makes contain (on the average) 0.08 ounces of gold. To test to see if the pieces contain 0.08 ounces of gold, he made a sample of 16 of these particular pieces and obtained the following data.

0.0773 0.0779 0.0756 0.0792 0.0777 0.0713 0.0818 0.0802
0.0802 0.0785 0.0764 0.0806 0.0786 0.0776 0.0793 0.0755

Use a level of significance of 0.01 to perform an appropriate hypothesis test.

- **Solution:** Step 1: $\mu =$ the true mean gold content for this particular type of charm.
- Step 2: $H_0 : \mu = 0.08$ and $H_a : \mu \neq 0.08$
- Step 3: We will use a one-sample mean *t* test. We assume the data come from an SRS of gold for this particular charm. We will plot the data to make sure the assumption of normality is valid.

Descriptive Statistics



We can see that with the exception of one outlier, the data is reasonably symmetric and mound shaped in shape, indicating that the assumption that the population of amounts of gold for this particular charm can reasonably be expected to be normally distributed.

- Step 4: Doing a 1-Var stats on the TI-84:
  $n = 16,\ \bar{x} = 0.077981,\ s = 0.0025143$

$$t = \frac{\bar{x} - 0.08}{s/\sqrt{n}} = \frac{0.077981 - 0.08}{0.0025143/\sqrt{16}} = \frac{-0.002019}{0.000628575} = -3.20$$

This is a two-tailed test. Looking up in the table of tail areas for *t* curves, with df = 15, we see the table entry is 0.003, so

$$P - Value = 2(0.003) = 0.006$$

- Step 5: The problem states that $\alpha = 0.01$. Our p-value is smaller than the significance level (0.006 < 0.01). This means that we will Reject H_0 at the 0.01 level of significance.
- Step 6: At the 0.01 level of significance, there is convincing evidence that the true mean gold content of this type of charm is not 0.08 ounces. Actually, when there is convincing evidence of rejecting a null hypothesis for the two-sided alternative, a one-tailed claim is also supported. In this case, at the 0.01 level of significance, there is convincing evidence that the true mean gold content of this type of charm is less than 0.08 ounces.

Syllabus Objectives: 3.20 – The student will analyze sampling distributions between two independent sample means. 4.7 – The student will calculate the confidence interval for a difference between two means (paired and unpaired). 4.14 – The student will perform a test for a difference between two means (paired and unpaired).

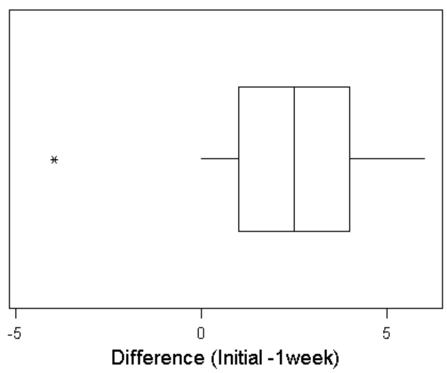
- **Comparing two population or treatment means – Paired *t* test**
  - Comparative studies are more convincing that single-sample investigations. For that reason, one-sample inference is less common than comparative inference. One common design to compare two treatments makes use of ONE sample procedures. In a **matched pairs design**, subjects are matched in pairs and each treatment is given to one subject in each pair.
  - Assumptions:
    1. The samples are **paired**.
    2. The *n* sample differences can be viewed as a random sample for a population of differences.
    3. The number of sample differences is large (generally at least 30) OR the population distribution of differences is approximately normal.
  - Alternate hypothesis and finding the p-value:
    1. $H_a : \mu_d >$ hypothesized value
       P-value = Area under the appropriate t curve to the right of the calculated t.
    2. $H_a : \mu_d <$ hypothesized value
       P-value = Area under the appropriate t curve to the left of the calculated t.
    3. $H_a : \mu_d \neq$ hypothesized value
       P-value = $2 \cdot$(Area to the right of t) if it is positive and
       $2 \cdot$(Area to the left of t) if it is negative.

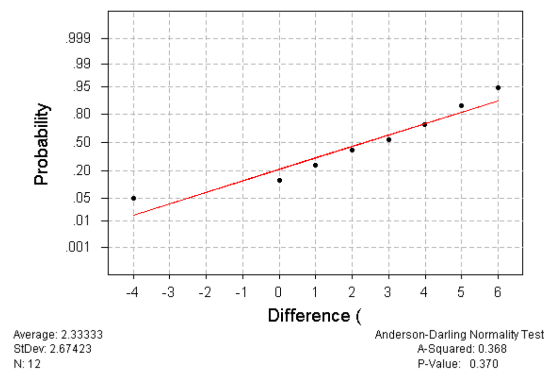
  - **Matched pair Example 1 –** A weight reduction center advertises that participants in its program lose an average of at least 5 pounds during the first week of the participation. Because of numerous complaints, the state's consumer protection agency doubts this claim. To test the claim at the 0.05 level of significance, 12 participants were randomly selected. Their initial weights and their weights after 1 week in the program appear below. Set up and perform an appropriate hypothesis test with a significance level of 0.05.

Member	Initial Weight	One Week Weight
1	195	195
2	153	151
3	174	170
4	125	123
5	149	144
6	152	149
7	135	131
8	143	147
9	139	138
10	198	192
11	215	211
12	153	152

Member	Initial Weight	One Week Weight	Difference Initial -1week
1	195	195	0
2	153	151	2
3	174	170	4
4	125	123	2
5	149	144	5
6	152	149	3
7	135	131	4
8	143	147	-4
9	139	138	1
10	198	192	6
11	215	211	4
12	153	152	1

- **Solution:** Step 1: $\mu_d =$ the true mean of the individual weight changes (initial weight – weight after one week). We are looking at the **mean differences.** $\mu_d = \mu_1 - \mu_2 = \mu_{\text{initial weight}} - \mu_{\text{week 1 weight}}$
- Step 2: $H_0 : \mu_d = 5$ and $H_a : \mu_d < 5$
- Step 3: We will use a matched pairs *t* test. According to the statement of the example, we can assume that the sampling is random. The sample size (12) is small, so from the boxplot of the differences, we see that there is one outlier but never the less, the distribution is reasonably symmetric and the normal plot confirms that is it is reasonable to assume that the population of differences (weight losses) is normally distributed.





- Step 4: Doing a 1-Var stats on the list of differences in the TI-84: $n = 12,\ \bar{x}_d = 2.333,\ s_d = 2.674$

We use the test statistic:
$$t = \frac{\bar{x}_d - \text{hypothesized value}}{s_d / \sqrt{n}} = \frac{\bar{x}_d - 5}{s_d / \sqrt{n}}$$

$$t = \frac{\bar{x}_d - 5}{s_d / \sqrt{n}} = \frac{2.333 - 5}{2.674 / \sqrt{12}} = \frac{-2.667}{0.77192} = -3.45$$

This is a one-tailed test, so looking up the *t* value of 3.45 under df = 11 in the table of tail areas for *t* curves, we find the p-value = 0.002.

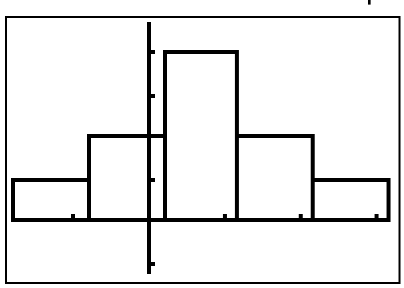
- Step 5: The problem states that $\alpha = 0.05$. Our p-value is smaller than the significance level (0.002< 0.05). This means that we will Reject H_0 at the 0.01 level of significance.
- Step 6: At the 0.05 level of significance, there is strong evidence that the mean weight loss for those who took the program for one week is less than 5 pounds.

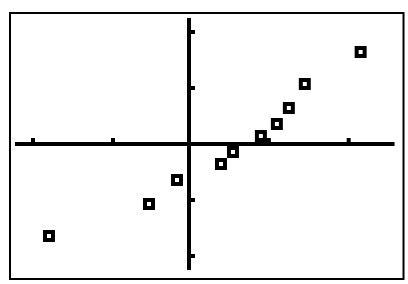
- o **Matched pair Example 2 –** Investigators at the U.S. Department of Agriculture wished to compare methods of determining the level of *E. coli* bacteria contamination in beef. Two different methods (A and B) of determining the level of contamination were used on each of ten randomly selected specimens of a certain type of beef. The data obtained, in millimicrobes/liter of ground beef, for each of the methods are shown in the table below.

		Specimen									
		1	2	3	4	5	6	7	8	9	10
Method	A	22.7	23.6	24.0	27.1	27.4	27.8	34.4	35.2	40.4	46.8
	B	23.0	23.1	23.7	26.5	26.6	27.1	33.2	35.0	40.5	47.8

Is there a significant difference in the mean amount of *E. coli* bacteria detected by the two methods for this type of beef? Provide a statistical justification to support your answer.

- **Solution:** Step 1: $\mu_d =$ the true mean difference (method A – method B) in the level of *E. coli* bacteria contamination in beef detected by the two methods. We are looking at the **mean differences.** $\mu_d = \mu_1 - \mu_2 = \mu_A - \mu_B$
- Step 2: $H_0 : \mu_d = 0$ and $H_a : \mu_d \neq 0$. If there is **no** difference in the two methods, we would expect their difference to be zero. Most matched pairs tests will have hypothesized values of zero.
- Step 3: Since the observations are obtained on 10 randomly selected specimens, it is reasonable to assume that the 10 data pairs are independent of one another. Looking at a histogram of the differences (-0.3, 0.5, 0.3, 0.6, 0.8, 0.7, 1.2, 0.2, -0.1, -1.0), it is symmetric with no apparent outliers. The normal probability plot has a linear pattern. It appears that the normal distribution is a reasonable option in this case.





- Step 4: Doing a 1-Var stats on the list of differences in the TI-84: $n = 10,\ \bar{x}_d = 0.29,\ s_d = 0.629727$

$$t = \frac{\bar{x}_d - 0}{s_d / \sqrt{n}} = \frac{0.29}{0.629727 / \sqrt{10}} = \frac{0.29}{0.199137} = 1.46$$

This is a one-tailed test, so looking up the *t* value of 1.46 under df = 9 in the table of tail areas for *t* curves, we find the p-value = 0.179.

- Step 5: Since the problem does not state a significance level, we will use $\alpha = 0.05$. Since the p-value is greater than 0.05, we will Fail to Reject H_0 at the 0.05 level of significance.
- Step 6: At the 0.05 level of significance, we do not have statistically significant evidence to conclude that there is a difference in the mean amount of *E. coli* bacteria detected by the two methods for this type of beef.

- **Comparing two population or treatment means - independent**
  - The goal of inference is to compare the responses to two treatments or to compare the characteristics of two populations.
  - We have a separate sample from each treatment or each population.
  - Conditions for comparing two means
    - We have **two SRSs,** from two distinct populations. The samples are **independent**. That is, one sample has no influence on the other. With matched pairs, the samples were dependent.
    - Both populations are **normally distributed**. The means and standard deviations of the populations are unknown.

  - **Notation – Comparing Two Means:**

	Mean Value	Variance	Standard Deviation
Population or Treatment 1	μ_1	σ_1^2	σ_1
Population or Treatment 2	μ_2	σ_2^2	σ_2

There are four unknown parameters, the two means and the two standard deviations. We use the sample means and standard deviations to estimate the unknown parameters.

  - **Notation for the samples:**

	Sample Size	Mean	Variance	Standard Deviation
Population or Treatment 1	n_1	$\bar{x}_1$	s_1^2	s_1
Population or Treatment 2	n_2	$\bar{x}_2$	s_2^2	s_2

Note the subscripts. They remind us which sample a statistic comes from.

- **Sampling distribution for comparing two means –** If the random samples on which $\bar{x}_1$ and $\bar{x}_2$ are based are selected **independently** of one another, then

- 1. $\mu_{\bar{x}_1-\bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$

- 2. $\sigma^2_{\bar{x}_1-\bar{x}_2} = \sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2} = \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$ $\qquad \sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

- If n_1 and n_2 are both large or the population distributions are (at least approximately) normal, then $\bar{x}_1$ and $\bar{x}_2$ each have (at least approximately) a normal distribution. This implies that the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is also normal or approximately normal.
- Because $\bar{x}_1 - \bar{x}_2$ has a normal distribution, we can standardize it to obtain a standard normal *z* statistic. This is the two-sample *z* statistic used:

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

. It has a normal distribution and we use the *z* table.

- Because we would not know the population standard deviations, we estimate them by the sample standard deviations from our two samples. The result is a two-sample *t* statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

It has approximately a *t* distribution with degrees of freedom df given by

$$\text{df} = \frac{(V_1 + V_2)^2}{\dfrac{V_1^2}{n_1 - 1} + \dfrac{V_2^2}{n_2 - 1}}$$

where $V_1 = \dfrac{s_1^2}{n_1}$ and $V_2 = \dfrac{s_2^2}{n_2}$

You truncate df to an integer.

- Avoiding the df formula: We do not need to be quite that accurate. We can use procedures based on the statistic *t* with critical values from the *t* distribution with degrees of freedom equal to the smaller of $n_1 - 1$ and $n_2 - 1$. These procedures are always conservative for any two normal populations.

- **The Two-Sample *t* procedures**
  - Draw an SRS of size n_1 from a normal population with unknown mean μ_1, and draw an independent SRS of size n_2 from another normal population with unknown mean μ_2. The confidence interval for $\mu_1 - \mu_2$ is given by:

$$(\bar{x}_1 - \bar{x}_2) \pm t \text{ critical value } \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

  - To test the hypothesis $H_0 : \mu_1 = \mu_2$, compute the two-sample *t* statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

    and use p-values or critical values for the *t* distribution.

- **Examples of two-sample procedures**
  - **Example 1 – Confidence interval:** Two kinds of thread are being compared for strength. Fifty pieces of each type of thread are tested under similar conditions. The sample data is given in the following table. Construct a 98% confidence interval for the difference of the population means.

	Sample Size	Sample mean	Sample Standard Deviation
Thread A	50	78.3	5.62
Thread B	50	87.2	6.31

    - **Solution:** Step 1: First identify the parameter we are interested in. $\mu_1 - \mu_2 =$ the true mean difference of the tensile strengths of these two threads. Step 2: We must check our assumptions. We will assume that we have two independent SRSs of each type of thread. Since we do not have the data, we cannot check the sample plots; however, both samples sizes are greater than 30, therefore the CLT guarantees normality. Step 3: Calculate the confidence interval. We have the sample means and standard deviations, we need the critical value. Look on the table of *t* critical values under 98% for df=49. Since 49 is not listed, we will use the conservative df=40. This gives us a critical value of 2.423.

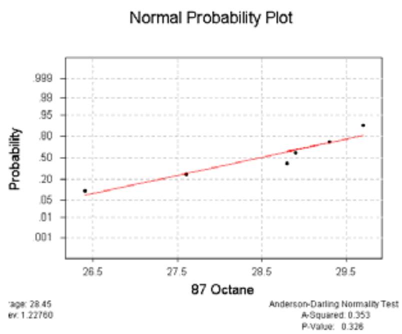
$$(78.3-87.2)\pm 2.423\sqrt{\frac{5.62^2}{50}+\frac{6.31^2}{50}} = -8.9 \pm 2.895 = (-11.885, -6.005)$$

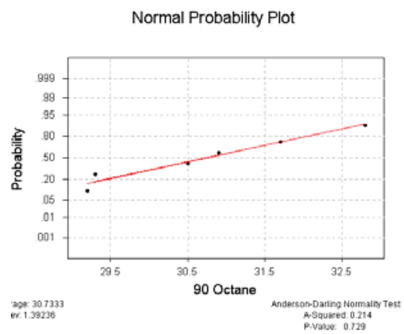
- Step 4: Based on this sample, I am 98% confident that the true difference of the means of the tensile strengths is between -11.885 and -6.005.

- o **Example 2 – Confidence Interval:** A student recorded the mileage he obtained while commuting to school in his car. He kept track of the mileage for twelve different tankfuls of fuel, involving gasoline of two different octane ratings. Compute the 95% confidence interval for the difference of mean mileages. His data follow:

87 Octane	90 Octane
26.4, 27.6, 29.7	30.5, 30.9, 29.2
28.9, 29.3, 28.8	31.7, 32.8, 29.3

- **Solution:** Step 1: First identify the parameter we are interested in. $\mu_1 - \mu_2 =$ the true mean difference of the mileages. Step 2: We must check our assumptions. We have to assume that the samples were independent and random and that the underlying populations were normally distributed since the sample sizes were small. By looking at the following normality plots, we see that the assumption of normality for each of the two populations of mileages appears reasonable. Given the small sample sizes, the assumption of normality is very important, so one would be a bit careful utilizing this result.





- Step 3: Calculate the confidence interval. We have the sample means and standard deviations, we need the critical value. Look on the table of *t* critical values under 95% for df=5. This gives us a critical value of 2.571.

$$n_1 = n_2 = 6,\ \bar{x}_1 = 28.45,\ \bar{x}_2 = 30.73,\ s_1 = 1.228,\ s_2 = 1.392$$

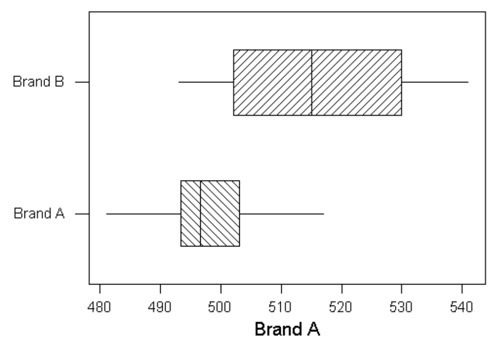
$$(28.45-30.73)\pm 2.571\sqrt{\frac{1.228^2}{6}+\frac{1.392^2}{6}} = -2.28 \pm 1.4765 = (-3.7565, -0.8035)$$

- Step 4: Based on this sample, I am 95% confident that the true difference of the mean mileages is between -3.7565 and -0.8035.

- o **Example 3 - Test:** In an attempt to determine if two competing brands of cold medicine contain, on the average, the same amount of acetaminophen, twelve different tablets from each of the two competing brands were randomly selected and tested for the amount of acetaminophen each contains. The results (in milligrams) follow. State and perform an appropriate hypothesis test. Use a significance level of 0.01.

Brand A	Brand B
517, 495, 503, 491	493, 508, 513, 521
503, 493, 505, 495	541, 533, 500, 515
498, 481, 499, 494	536, 498, 515, 515

- **Solution:** Step 1: $\mu_1 =$ the true mean amount of acetaminophen in cold tablet Brand A. $\mu_2 =$ the true mean amount of acetaminophen in cold tablet Brand B.
- Step 2: $H_0 : \mu_1 = \mu_2$ and $H_a : \mu_1 \neq \mu_2$
- Step 3: The samples were selected independently and randomly. Since the samples are not large, we need to be able to assume that the populations of amounts of acetaminophen are both normally distributed. As we can see from the boxplots below, the assumption that the underlying distributions are normally distributed appears to be quite reasonable.



- Step 4: Doing a 1-Var stats on sample data for the two samples on the TI-84: $n_1 = n_2 = 12,\ \bar{x}_1 = 497.83,\ \bar{x}_2 = 515.67,\ \mathrm{s}_1 = 8.83,\ \mathrm{s}_2 = 15.144$

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{497.83 - 515.67 - 0}{\sqrt{\frac{8.83^2}{12} + \frac{15.144^2}{12}}} = \frac{-17.84}{5.0605} = -3.53$$

This is a two-tailed test, so looking up the *t* value of 3.53 under df = 11 in the table of tail areas for *t* curves, we find the p-value to be between 0.001<p<0.0025. Multiplying this by two because it is a two-tailed test, we have a p value between 0.002 and 0.005.

- Step 5: Since the problem states a significance level of $\alpha = 0.01$, we will Reject H_0 because the p-value is much smaller than the significance level.
- Step 6: At the 0.01 level of significance, the data provides strong evidence that the mean amount of acetaminophen is not the same for both brands. Specifically, there is strong evidence that the average amount per table for Brand A is less than that for Brand B.

- **Example 4 - Test:** We would like to compare the mean fill of 32 ounce cans of beer from two adjacent filling machines. A sample of 35 cans from machine 1 gave a mean of 16.031 and a standard deviation of 0.043. A sample of 31 cans from machine 2 gave a mean of 16.009 and a sample standard deviation of 0.052. Perform a significance test at the 0.05 significance level.

  - **Solution:** Step 1: $\mu_1 =$ the true mean amount of beer in machine 1. $\mu_2 =$ the true mean amount of beer in machine 2.
  - Step 2: $H_0 : \mu_1 = \mu_2$ and $H_a : \mu_1 \neq \mu_2$
  - Step 3: The samples were selected independently and randomly. Since the samples are large $n_1 = 35 > 30$ and $n_2 = 31 > 30$, the Central Limit Theorem states that the sampling distributions of both will be approximately normal.
  - Step 4:

$$t = \frac{\overline{x}_1 - \overline{x}_2 - \mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{16.031 - 16.009 - 0}{\sqrt{\frac{0.043^2}{35} + \frac{0.052^2}{31}}} = \frac{0.022}{0.0118345} = 1.86$$

    This is a two-tailed test, so looking up the *t* value of 1.86 under df = 30 in the table of tail areas for *t* curves, we find the p-value to be between 0.025<p<0.05. Multiplying this by two because it is a two-tailed test, we have a p value between 0.05 and 0.10.
  - Step 5: Since the problem states a significance level of $\alpha = 0.05$, we will Fail to Reject H_0 because the p-value is larger than the significance level.
  - Step 6: At the 0.05 level of significance, there is insufficient evidence to support a claim that the two machines produces bottles with different mean fills.