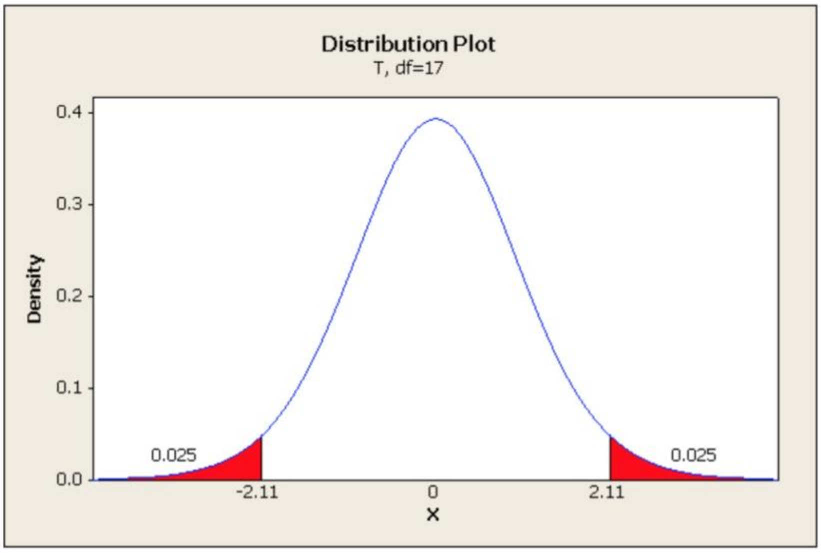
ST 305 HW 8 Solution
(Total points: 48)

Ch 7.1: 16, 18, 19, 21, 22, 25, 26, 33, 41

16. (1 point)



From the table of t distribution critical values, the 2.5 %(0.025) critical value is 2.110.

This leads to the rejection of the null hypothesis at the 5% level for a two sided alternative when $t < -2.110$ or $t > 2.110$

18. (1 point)

The null and alternative hypotheses are,

$$H_0 : \mu = 0$$
$$H_a : \mu \neq 0$$

The computer software reports that the sample mean is $\bar{x} = 15.3$ and the two-tailed P-value is 0.074.

The P-value for testing the alternative hypothesis, $H_a : \mu > 0$ is,

$$\begin{aligned} p &= 2 \times [\text{Two-tailed P-value}] \\ &= 2(0.074) \\ &= 0.148 \end{aligned}$$

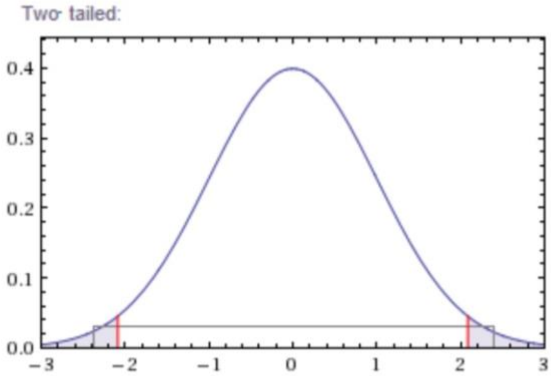
Therefore, the P-value for the significant test is 0.148.

19. (1 point)

Given that $\bar{x} = -15.3$ and $P = 0.074$ for a t test with hypotheses

$$H_0 : \mu = 0$$
$$H_a : \mu \neq 0$$

The *P*-value would be 0.037 for $\mu < 0$ and 0.963 for $\mu > 0$.



Under the null hypothesis, the true mean is somewhere in the white area. However, if $\bar{x}$ is negative, then we are solely concerned with the shaded area on the left side.

21. (4 points)

(a)

The degrees of freedom are calculated as follows:

Determine the degrees of freedom using the following formula,

$$\begin{aligned} df &= n - 1 \\ &= 27 - 1 \\ &= \boxed{26} \end{aligned}$$

(b)

The critical values which bracket t^* for 26 degrees of freedom and $t = 2.01$ are 1.706 and 2.056.

(c)

The one-sided *P*-value falls between the values 0.05 and 0.025. However, because this is a two sided t test, the probabilities are $2 \times 0.05 = 0.10$ and $2 \times 0.025 = 0.05$.

Therefore, the P-value is lies between $\boxed{\text{0.05 and 0.10.}}$

(d)

From the *t*-distribution tables, the critical values at 0.05 level for 26 degrees of freedom about two tailed is ± 2.056.

Since, $t = 2.01 < 2.056$. So, there is not significant at the 5% level.

From the *t*-distribution tables, the critical values at 0.01for 26 degrees of freedom about two tailed is ± 2.779.

Since, $t = 2.01 < 2.779$. So, there is not significant at the 1% level.

The value of $t = 2.01$ is not significant at the 5% level and is also not significant at the 1% level.

22. (3 points)

Given that the one-sample *t* statistic has the value *t* = –2.55 for sample size n = 14 and hypotheses

$$H_0 : \mu = 20$$
$$H_a : \mu < 20$$

(a) The degrees of freedom are calculated as follows:

The degree of freedom for a one sample t test is *n*–1.

$$n - 1 = 14 - 1$$
$$= \boxed{13}$$

(b) From statistical table, we can observe that the test statistic value of *t* (–2.55) for 13 degrees of freedom falls between the one-sided *P*-value falls between the values 0.02 and 0.01.

(c) The exact *P*-value is calculated using EXCEL software to be 0.0121.

25. (2 points)

(a) Hypotheses are as follows:

$H_0 : \mu = 10$ The average number of uses that a person can produce in 5 minutes is 10.

$H_a : \mu < 10$ There is a decrease in the performance, or the average number of uses is less than 10.

(b) We wish to perform a one-sided *t* test of significance. The sample mean is 7.88 with a sample standard deviation of 2.35. There are 34 participants.

The *t* statistic is calculated as follows:

$$t = \frac{\overline{x} - \mu}{s / \sqrt{n}}$$
$$= \frac{7.88 - 10}{2.35 / \sqrt{34}}$$
$$= -5.26$$

With 33 degrees of freedom, this *t* statistic gives a probability that is much smaller than the significance level of 0.05.

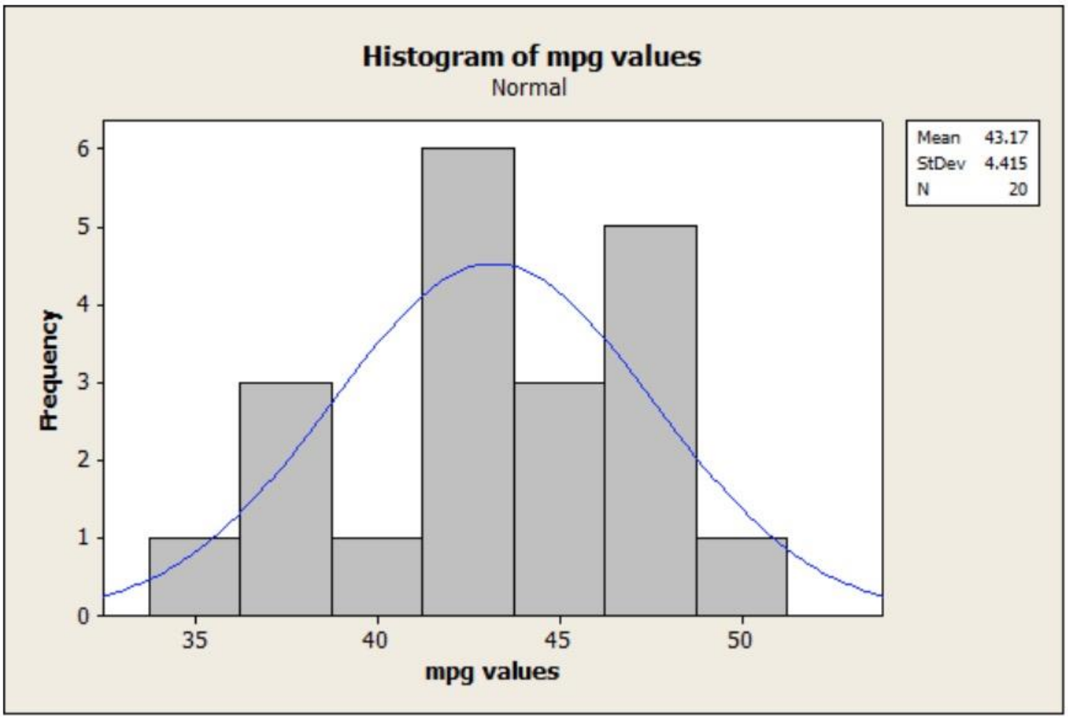
The *P*-value is 3.948×10^{-6}. As this *p*-value is less than the significance level of 0.05, we have sufficient evidence to reject the null hypothesis.

Therefore, we conclude that there is a statistically significant effect of observed rudeness on task performance.

26. (3 points)

a)

The graphical summary of the given 20 observations is as follows:



From the graph, it can be observed that the data is approximately symmetric and there are no outliers in the data.

Hence, it is appropriate to analyze these data using methods based on Normal / *t*-distribution.

(b)

The sample mean of mpg values is,

$$\bar{x} = \frac{\sum x_i}{n} = \frac{863.4}{20} = \boxed{43.17}$$

The sample standard deviation of mpg values is,

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{370.342}{20-1}} = \boxed{4.415}$$

The standard error is,

$$se = \frac{s}{\sqrt{n}} = \frac{4.415}{\sqrt{20}} = \boxed{0.9872}$$

Since the sample standard deviation is unknown, use *t*-distribution to find the margin of error at 95% confidence level.

The critical value of '*t*' at 19 degrees of freedom and at 0.05significance level,

$$t^*_{0.025,19} = 2.093$$

Then, Margin of error, M.E. is,

$$M.E. = t^* * \frac{s}{\sqrt{n}} = 2.093 \times \frac{0.9872}{\sqrt{20}} = \boxed{2.066}$$

c)

Then, the 95% confidence interval for μ are given as,

$$\bar{x} \pm M.E = 43.17 \pm 2.066 = \boxed{(41.104,\ 45.236)}$$

The 95% confidence interval the mean mpg is $(41.104, 45.236)$.

33. (2 points)

(a)

Null hypothesis is

$H_0 : \mu = 0$

Alternative hypothesis is

$H_a : \mu \neq 0$

Test statistic is

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \sim t(n-1)$$

$$= \frac{328 - 0}{256 / \sqrt{16}}$$

$$t = 5.13 \sim t(15)$$

P – Value is

$$P = 2P(T \geq |5.13|) = 2P(T \geq 5.13)$$

Where T has the $t(15)$ distribution from 'TABLE D' we see that $P(T \geq 4.073) = 0.0005.$

Therefore, we conclude that the P – value is less than $2 \times 0.0005 = 0.001$

Therefore we reject the null hypothesis H_0.

Hence we conclude that there was change in 'NEAT'.

(b)

95 % confidence interval for the change in 'NEAT' is

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

$$= 328 \pm \left(2.131 \times \frac{256}{\sqrt{16}} \right)$$

$$= 328 \pm 136.384$$

$$= (191.616, 464.384)$$ (Calories)

The mean 'NEAT' increase is between 192 and 464 calories.

41. (4 points)

(a)

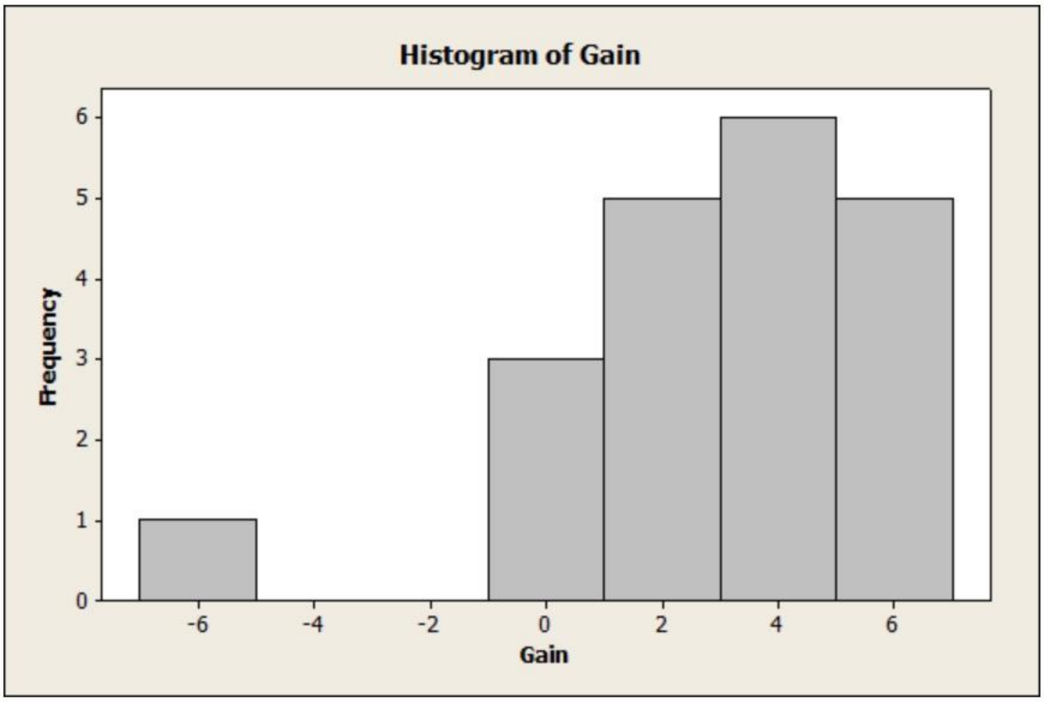
The null hypothesis is that the pretest and post test has the same mean scores.

i.e., $H_0 : \mu = 0$

The alternative hypothesis is that the mean score of posttest is greater than mean score of pretest

i.e., $H_a : \mu > 0$

(b)



From the above histogram we can observe that the distribution is skewed to left. There is an outlier. The spread is from – 6 to 6.

Mean of the distribution of gains is

$$\bar{x} = \frac{2+0+6+.......+2+3+3}{20}$$

$$= 2.5$$

Standard deviation of the distribution of gains is

$$s = \sqrt{\frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2}$$

$$= \sqrt{\frac{1}{20-1}\left[(2-2.5)^2 + (0.2.5)^2 + ... + (3-2.5)^2\right]}$$

$$= 2.893$$

(c)

Test statistic is

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad \square\, t(n-1)$$

$$= \frac{2.5 - 0}{\frac{2.893}{\sqrt{20}}}$$

$$t = 3.865 \quad \square\, t(19)$$

Degrees of freedom = n – 1 = 20 -1 = 19

P – Value is

$$P = P(T \geq t) = P(T \geq 3.86) = 0.000527$$

Since $P = 0.000527,$ therefore the test is statistically significant.

Hence we reject the null hypothesis.

(d)

95 % confidence interval for the mean improvement is

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

$$= 2.5 \pm \left(2.093 \times \frac{2.893}{\sqrt{20}}\right)$$

$$= 2.5 \pm 1.354$$

$$= (1.146, 3.854)$$

Ch 7.2: 61, 62, 63, 64, 65, 75, 76abc, 80, 81, 83

61. (4 points)

(a) The given statement is **wrong**.

In order to test the hypotheses should contain the population mean instead of the sample means.

(b) The given statement is **wrong**.

The samples are not independent.

(c) The given statement is **wrong**.

Here to reject the null hypothesis, we need a small *P*-value.

(d) The given statement is **wrong**.

When the test is left-tailed test ($H_a : \mu_1 < \mu_2$), it results **negative *t* value**.

62. (2 points)

(a)

The known 95% confidence interval for the difference of two means is (0.8, 2.3).

The confidence interval does not include 0. It can be concluded that there is a sufficient

evidence to reject the null hypothesis that the means are equal.

Hence, at the 5% level, the difference in means is statistically significant.

(b)

Larger samples will generally give a smaller margin of error for the difference between

two sample means. In the equation for two-sample *t* confidence interval, sample sizes are

in the denominator of the margin of error and so maximizing sample size will minimize

margin of error.

63 (2 points)

(a) For a *t*-statistic of $t = -2.18$ with 10 degrees of freedom, the *P*-value is 0.05425. Thus, we cannot reject the null hypothesis in favor of the alternative at the 5% significance level.

(b) For the same *t*-statistic and degrees of freedom as part (a), but for a one-sided alternative that the difference in means is negative, the *P*-value is 0.02712 and so we may reject the null hypothesis in favor of the alternative.

64 (1 point)

Given $\overline{x}_1 = 100$, $\overline{x}_2 = 110$, $s_1 = 18$, $s_2 = 15$, $n_1 = 50$, and $n_2 = 40$, we wish to find a 95% confidence interval for the difference in corresponding values of μ using the second approximation for degrees of freedom.

The equation we use is

$$(\overline{x}_1 - \overline{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

For 40 – 1 = 39 degrees of freedom, the 95% confidence level two-sided t^* value is approximately 2.021.

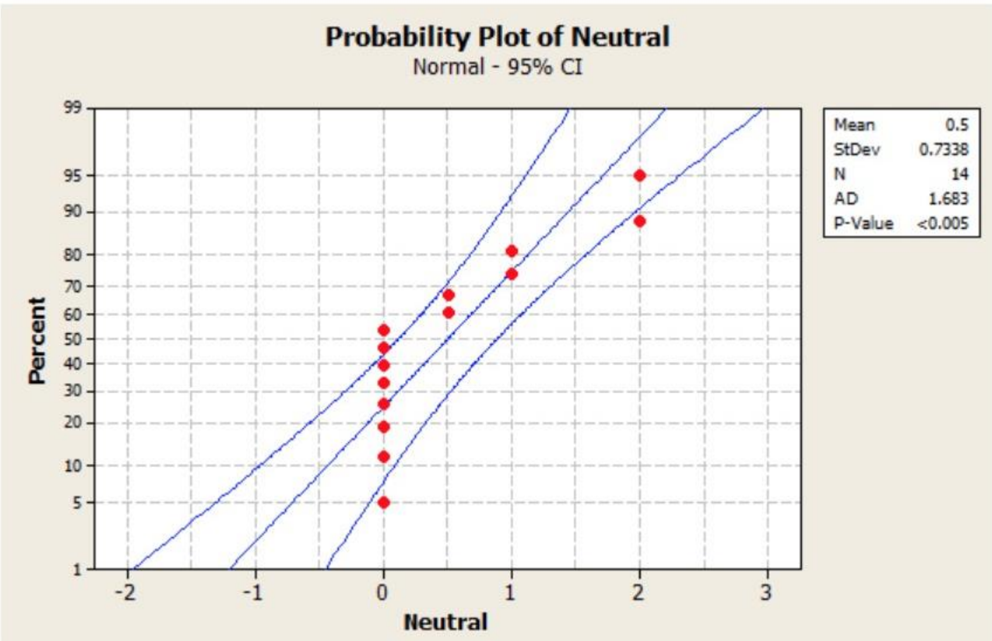
$$(100 - 110) \pm 2.021\sqrt{\frac{18^2}{50} + \frac{15^2}{40}} = -10 \pm 7.032$$

The 95% confidence interval is $\boxed{(-17.03, \ -2.97)}$.

Intuitively, we can conclude that this 95% confidence interval holds fewer values than a 99% confidence interval. This is because a confidence interval with higher confidence needs a wider range in order to increase confidence.

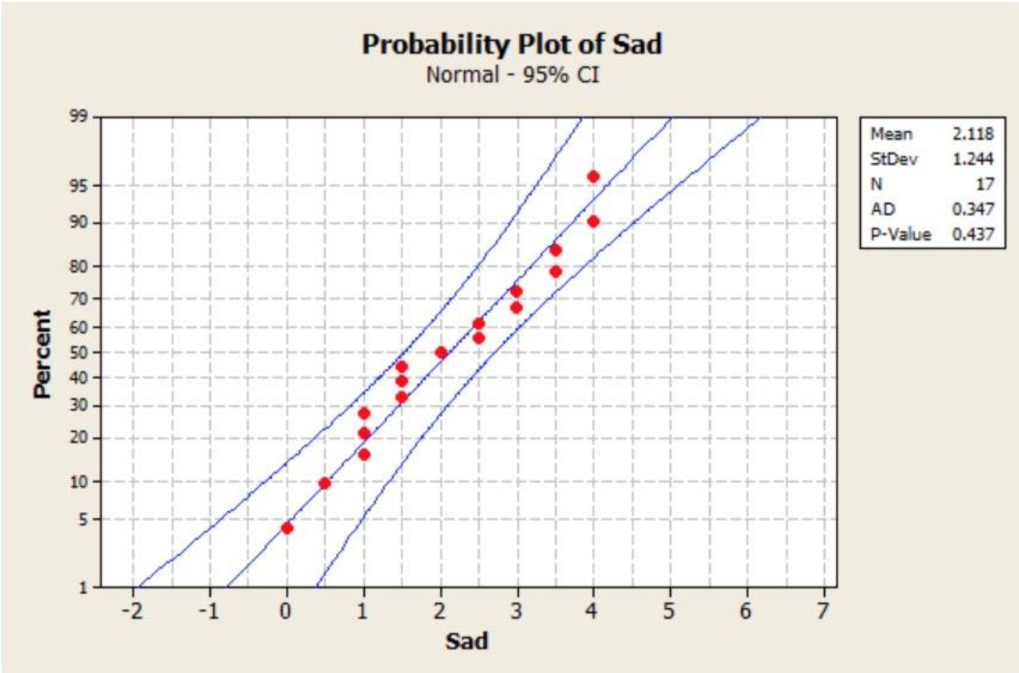
65 (4 points)

(a)



-+

From the normal probability plot, we can observe that some of data points are just beyond the control lines, enen though all data points are equally balancing from the straight line, so the data set of the price of purchasing insulated water bottle for the Neutral group follows approximately Normal distribution.



From the normal probability plot, we can observe that all data points lies within the control limits and close to central line, so the data set of the price of purchasing insulated water bottle for the sad group follows Normal distribution.

(b)

Descriptive Statistics: Neutral, Sad

Variable	Total Count	Mean	StDev
Neutral	14	0.571	0.730
Sad	17	2.118	1.244

(c)

It is required to test whether there is any sufficient significant difference of the mean price of purchasing insulated water bottles for the Neutral group and sad group.

The null hypothesis:

H_0 : There is no sufficient significant difference of the mean price of purchasing insulated water bottles for the Neutral group and sad group.

That is, $H_0 : \mu_1 = \mu_2$

H_1 : There is sufficient significant difference of the mean price of purchasing insulated water bottles for the Neutral group and sad group.

That is, $H_1 : \mu_1 \neq \mu_2$

(d)

Let us consider, $\bar{x}_1$ represents the average price of purchasing insulated water bottles for the Neutral group and the $\bar{x}_2$ represents the average price of purchasing insulated water bottles for the sad group.

The information regarding problem is as follows,

$n_1 = 14 \quad n_2 = 17$
$\bar{x}_1 = 0.571 \quad \bar{x}_2 = 2.118$
$s_1 = 0.730 \quad s_2 = 1.244$

Here, we can analyze this study by using two sample t-tests is good approach, since the population standard deviations are unknown and the size of the sample is small $(n < 30)$.

Here, the level of significance is $\alpha = 0.05$

75 (1 point)

From the information, observe that a study of food portion sizes reported that over a 17 year period.

The average size of a soft drink consumed by Americans aged 2 years and older increase from 13.1 ounce to 19.9

The additional information needed to compute a confidence interval for the increase is needed to either sample sizes and standard deviations or degrees of freedom and a more accurate value for the *P – value*.

The confidence interval will give us useful information about the magnitude of the difference.

76 (3 points)

(a)

It is not reasonable to assume that the given data is normally distributed, because the given sample size is small.

(b)

The null hypothesis is that the two groups of children consume equal amounts of sweetened drinks.

$$H_0 : \mu_1 = \mu_2$$

The alternative hypothesis is that the two groups of children do not consume equal amounts of sweetened drinks.

$$H_a : \mu_1 \neq \mu_2$$

Test statistic is $t = \dfrac{\overline{x_1} - \overrightarrow{x_2}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

$$= \frac{8.2 - 14.5}{\sqrt{\dfrac{10.7^2}{20} + \dfrac{8.2^2}{5}}}$$

$$= -1.44$$

Degree of freedom is smaller of 19 and 4 .That is 4.

P – Value is

$$P = 2P(T \geq 1.44)$$

$$= 2 \times 0.111644$$

$$= 0.223287$$

Since $P = 0.223287,$ therefore we accept the null hypothesis.

(c)

95 % confidence interval for the difference in means is

$$\left(\overline{x_1} - \overline{x_2}\right) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$= (8.2 - 14.5) \pm \left((2.776) \sqrt{\frac{10.7^2}{20} + \frac{8.2^2}{5}} \right)$$

$$= -6.3 \pm 12.15511$$

$$= (-18.4551\ oz, 5.855108\ oz)$$

80 (3 points)

(a)

Consider Null and Alternative hypothesis.

Null hypothesis, H_0: There is no significant difference between the average ads from sources of Wall Street Journal and National Enquirer.

That is, $H_0 : \mu_1 = \mu_2$

Alternative hypothesis, H_a: There is significant difference between the average ads from sources of Wall Street Journal and National Enquirer.

That is, $H_a : \mu_1 \neq \mu_2$

Calculate the test statistic.

$$
\begin{aligned}
t &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\
&= \frac{4.77 - 2.43}{\sqrt{\frac{(1.50)^2}{66} + \frac{(1.64)^2}{61}}} \\
&= \frac{2.34}{0.28} \\
&= 8.37
\end{aligned}
$$

$$
\begin{aligned}
df &= n_1 + n_2 - 2 \\
&= 66 + 61 - 2 \\
&= 125
\end{aligned}
$$

$$p - value = 0.0000 \qquad \left(\text{Use Excel's } \left(= T.DIST.2T(8.37, 125)\right)\right)$$

The level of significance is 0.05

Compare the *p-value* with the level of significance.

Here, the *p-value* is lesser than the level of significance then reject the Null hypothesis.

Hence, conclude that there is significant difference between the average ads from sources of Wall Street Journal and National Enquirer.

(b)

Calculate 95% confidence interval for the difference.

The level of significance is 0.05

From the standard *t* table values, observe that the critical value of *t* for the two tail test at 5% level of significance and 125 degrees of freedom is 1.979

Therefore,

$$CI = (\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\left(\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)$$

$$= (4.77 - 2.43) \pm (1.979)\left(\sqrt{\frac{(1.50)^2}{66} + \frac{(1.64)^2}{61}}\right)$$

$$= 2.34 \pm (1.979)(0.280)$$
$$= 2.34 \pm 0.55$$
$$= (1.79, 2.89)$$

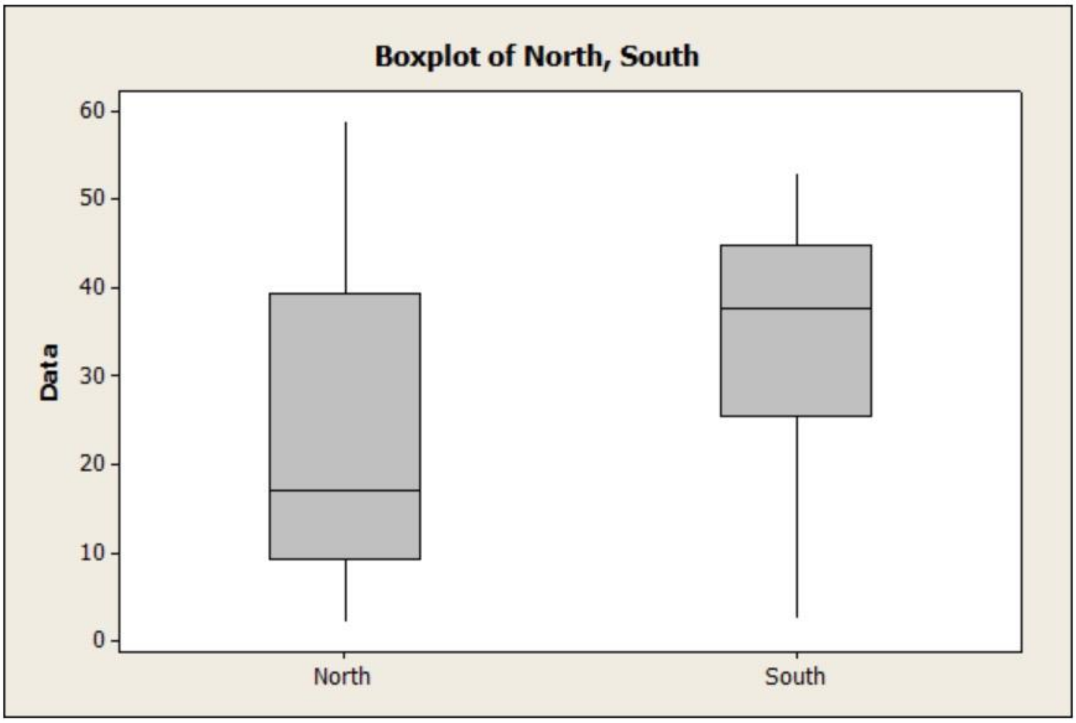
(c)

From the 95% confidence interval, observe that the 0 value is not included in these intervals then the Null hypothesis is rejected.

Hence, conclude that there is significant difference between the average ads from sources of Wall Street Journal and National Enquirer.

81 (5 points)

(a)



From the above box-plot we can observe that the North distribution is right skewed. The five-number summary is 2.2, 10.2, 17.05, 39.1, 58.8 cm.

While the south distribution is left skewed. The five-number summary is 2.6, 26.1, 37.70, 44.6, 52.9 cm.

There are no clear outliers in the above two plots.

(b)

The methods of this section seem to be appropriate, because there are no outliers in the given data.

(c)

The null hypothesis for comparing the two samples of Tree DBH's is,

$$H_0 : \mu_1 = \mu_2$$

The alternative hypothesis for comparing the two samples of tree DBH's is,

$$H_0 : \mu_1 \neq \mu_2$$

Since we do not know that the mean DBH of the trees in the North half of the Tract is greater than or lesser than the mean DBH of trees in the south half, therefore it is better to use the two-sided alternative.

(d)

Test statistic is,

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

$$= \frac{23.70 - 34.53}{\sqrt{\frac{17.50^2}{30} + \frac{14.26^2}{30}}}$$

$$= -2.63$$

Degree of freedom is 29.

P-value is,

$$P = 2P(T \geq 1 - 2.631)$$

$$= 2P(T \geq 2.63)$$

$$= 2 \times 0.006762$$

$$= 0.013524$$

Since P = 0.014, we will accept the null hypothesis below 1% significance level and we reject the null hypothesis above 1% significance level.

(e)

95% confidence interval for the difference in Mean DBH's is,

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

$$= (23.70 - 34.53) \pm (2.045)\sqrt{\frac{17.50^2}{30} + \frac{14.26^2}{30}}$$

$$= -10.83 \pm 8.428424$$

$$= (-19.2584\,\text{cm},\ -2.40158\,\text{cm})$$

We are 95% confident that the mean difference in DBH's is between -19.2584 cm and -2.40158 cm.

83 (2 points)

The following table is the summary of sales of present and previous year of a product.

sales	n	$\bar{x}$	s
present year	70	53	15
last year	55	50	18

Using t-sample t procedures, the 95% confidence interval for the difference in mean number of units sold at all retail stores is given by

$$\left(\overline{x_1} - \overline{x_2}\right) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Here,

t^* is the critical value of t at 54 degrees of freedom and at0.05 significance level= 2.009

$$(53 - 50) + 2.009 * \sqrt{\frac{15^2}{70} + \frac{18^2}{55}}$$

3 6.0621

(-3.06, 9.0621)

b) The difference in sample means may be due to random fluctuations which need not be accounted.