W&M CSCI 628: Design of Experiments
Homework 1

Megan Rose Bryant

September 2, 2014

1.2

Suppose that you want to investigate the factors that potentially affect cooking rice.

a.) *What would you use as a response variable in this experiment? How would you measure the response?*

The appropriate response variable would be the weight of the final product. Available process knowledge tells us that as rice cooks it absorbs the liquid in which it is cooked and that each type of rice has a certain weight when properly cooked. Therefore the response variable can be measured by weighing the finished product, given that the appropriate proper cook-weight is known.

b.)*List all of the potential sources of variability that could impact the response.*

A major source of variability is the type of rice (i.e. long or short, etc.). Each type of rice has a different ideal weight. A possible nuisance factor would be variations among brand and batch of both the rice, the cooking liquid, and tools used in the cooking method. Another source of variability would be the liquid in which the rice is cooked; water, broth, and salt water are some of the possible mediums. The cooking method would also be a significant source of variability as rice can be cooked in many different ways, the most common being fried, stove-top boiled, or via rice-cooker. The starting amount of rice is another important factor that affects the final product. Furthermore, cook temperature, cook time, altitude, humidity, temperature, and even day of the week are all factors which must be considered as sources of potential variability.

c.) *Complete the first three steps of the guidelines for designing experiments in section* 4*.*

Step 1: Recognition of and statement of the problem. The objective of this experiment is to determine which factors affect the cooking of rice. Therefore, we will be conducting a characterizing experiment.

Step 2: Selection of the Response Variable. For this experiment, we would use the weight of the cooked rice as a response variable. This would allow us to properly measure whether or not the rice has been correctly cooked and thereby allow us to characterize the factors.

Step 3: Choice of factors, levels, and range. We would select the design factors to be cooking liquid, cooking method, and cooking time, and starting amount. The type of rice should be a held constant factor and a suitable type would be either short or long grain rice. The tools used in the cooking method should also be a held constant factor (i.e. the same pan, pot, rice-cooker, etc. will be used in each trial). The controllable nuisance factors are brand name, batch, and day of the week. A factor that would be allowed-to-vary would be the specific number of rice grains; only the starting weight would be held constant. These can be controlled by selecting a particular brand and blocking within that selection and blocking the schedule. Uncontrollable nuisance factors include humidity and altitude, but, since they can be measured, their effects can be compensated for by using the analysis of covariance procedure.

Cooking liquid should have water, broth, and saltwater as levels. Cooking method should have frying pan, stove-top boil, and rice-cooker as levels. Cooking time should range from 0 minutes to 20 minutes. Cooking temperature should range from low-high.

2.2

The minitab output for a random sample of data is shown below. Some quantities are missing. Compute the values of the missing quantities.

Variable	N	Mean	SE Mean	Std. Dev.	Sum
Y	16	?	0.159	?	399.851

$$SE_{\bar{y}} = \frac{\sigma}{\sqrt{n}} \qquad \mu = \frac{\sum_{i=1}^{N} y_i}{n}$$
$$0.159 = \frac{\sigma}{\sqrt{16}} \qquad \mu = \frac{\sum_{i=1}^{16} y_i}{16}$$
$$0.159 = \frac{\sigma}{4} \qquad \mu = \frac{399.851}{16}$$
$$\sigma = 0.636 \qquad \mu = 24.991$$

Therefore, the completed table is as follows:

Variable	N	Mean	SE Mean	Std. Dev.	Sum
Y	16	24.991	0.159	0.636	399.851

2.5

Consider the Minitab output shown below.

One Sample Z
Test of $\mu = 30$ vs $\neq 30$
The assumed standard deviation= 1.2

N	Mean	SE Mean	95% CI	Z	P
16	31.2000	0.3000	(30.6120, 31.7880)	?	?

a.)*Fill in the missing values in the output. What conclusion would you draw?*

$$SE_{\bar{y}} = \frac{\sigma}{\sqrt{n}} \qquad Z = \frac{\bar{y}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \qquad P = 2[1-\phi(|Z|)]$$
$$0.3000 = \frac{\sigma}{\sqrt{16}} \qquad = \frac{31.200-30}{\frac{1.2}{\sqrt{16}}} \qquad = 2[1-\phi(|4|)]$$
$$0.3000 = \frac{\sigma}{4} \qquad = \frac{1.2}{\frac{1.2}{4}} = \frac{1.2}{.3} \qquad = 2[1-.999969]$$
$$\sigma = 1.2 \qquad = 4 \qquad = .000063$$

Therefore, our completed table would be as follows:

One Sample Z
Test of $\mu = 30$ vs $\neq 30$
The assumed standard deviation= 1.2

N	Mean	SE Mean	95% CI	Z	P
16	31.2000	0.3000	(30.6120, 31.7880)	4	.0001

Since the sample mean is in the 95% confidence interval, we can conclude that we would fail to reject the null hypothesis at a 5% significance level.

b.)*Is this a one-sided or two-sided test?*

We know that this is a two-sided test since the alternative hypothesis is $\mu \neq 30$.

Note: If the H_a was $\mu < 30$ or $\mu > 30$, we would have a one-sided test.*

c.) *Use the output and the normal table to find a 99 percent CI on the mean.*

A 99% confidence interval gives us an α of 0.01. Therefore, the 99% confidence interval on the true population of the mean is:

$$\begin{gathered}
\bar{y} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{y} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \\
31.2 - Z_{\frac{.01}{2}} \frac{1.2}{\sqrt{16}} \leq \mu \leq 31.2 + Z_{\frac{.01}{2}} \frac{1.2}{\sqrt{16}} \\
31.2 - Z_{.005} \frac{1.2}{4} \leq \mu \leq 31.2 + Z_{.005} \frac{1.2}{4} \\
31.2 - Z_{.005} * .3 \leq \mu \leq 31.2 + Z_{.005} * .3 \\
31.2 - 2.5667 * .3 \leq \mu \leq 31.2 + 2.5667 * .3 \\
31.2 - .77001 \leq \mu \leq 31.2 + .77001 \\
30.43 \leq \mu \leq 31.97
\end{gathered}$$

Therefore, the 99% confidence interval is: $(30.43, 31.97)$ i.e.
$(31.2 \pm .770)$.

Note: Our sample mean, $\bar{y} = 31.2$, is in the 99% confidence interval.*

d.) *What is the P-value if the alternative hypothesis is $H_1 : \mu > 30$?*

If our alternative hypothesis is modified so that $H_1 : \mu > 30$, we would now have a one-sided test (as opposed to our original two sided test). Therefore, the new P value would only be concerned with one tail of the distribution.

Since we have modified the alternative to a >, this means that the new P-value would be half of the original P-value. Therefore, we have:

$$P^* = \frac{1}{2}P = \frac{1}{2}0.000063 = .000032$$

2.8

Consider the Minitab output shown below.

One Sample T: Y

Test of $\mu = 91$ vs $\neq 91$

Variable	N	Mean	Std. Dev	SE Mean	95% CI	T	P
Y	25	92.5805	?	0.4673	(91.6160, ?)	3.38	0.002

a.)*Fill in the missing values in the output. Can the null hypothesis be rejected at the 0.05 level? Why?*

$$SE_{\bar{y}} = \frac{S}{\sqrt{n}}$$
$$0.4673 = \frac{S}{\sqrt{25}}$$
$$0.4673 = \frac{S}{\sqrt{25}}$$
$$S = 2.3365$$

$$\text{Lower Bound} = \bar{y} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}$$
$$91.6160 = 92.5805 - t_{\frac{.05}{2},25-1}\frac{2.3365}{\sqrt{25}}$$
$$91.6160 = 92.5805 - t_{.025,15}\frac{2.3365}{5}$$
$$91.6160 = 92.5805 - t_{.025,15}.4673$$
$$0.9645 = 0.4673t_{.025,15}$$
$$t_{.025,24} = 2.0640$$

$$\text{Upper Bound} = \bar{y} + t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}$$
$$\text{Upper Bound} = 92.5805 + t_{\frac{.05}{2},25-1}\frac{1.8692}{\sqrt{25}}$$
$$\text{Upper Bound} = 92.5805 + 2.0640\frac{2.3365}{5}$$
$$\text{Upper Bound} = 92.5805 + 2.0640 * 0.4673$$
$$\text{Upper Bound} = 93.5450$$

Therefore, our completed table would be as follows:

One Sample T: Y

Test of $\mu = 91$ vs $\neq 91$

Variable	N	Mean	Std. Dev	SE Mean	95% CI	T	P
Y	25	92.5805	2.3365	0.4673	(91.6160, 93.5450)	3.38	0.002

Since our sample mean is in the 95% confidence interval, we cannot reject the null hypothesis in favor of the alternative.

b.)*Is this a one-sided or two-sided test?*

We know that this is a two-sided test since the alternative hypothesis is $\mu \neq 91$.

Note: If the H_a was $\mu < 91$ or $\mu > 91$, we would have a one-sided test.*

c.) *If the hypothesis had been* $H_0 : \mu = 90$ *versus* $H_1 : \mu \neq 90$ *would you reject the null hypothesis at the* 0.05 *level?*

A change in the null hypothesis would result in a change in the test statistic, t_0. Therefore, the new test statistic, t_0^* would need to be calculated before any statment on rejection could be made.

$$t_0^* = \frac{\bar{y} - \mu_0^*}{\frac{S}{\sqrt{n}}}$$
$$t_0^* = \frac{92.5805 - 90}{\frac{1.8692}{\sqrt{25}}}$$
$$t_0^* = \frac{2.5805}{\frac{2.3365}{5}}$$
$$t_0^* = \frac{2.5805}{0.4673}$$
$$t_0^* = 5.52$$

We know that with a two-sided test, the fixed significance level criteria for rejection is $|t_0 *| > t_{\frac{\alpha}{2}, n-1}$.

We know that $t_{\frac{\alpha}{2}, n-1} = t_{.025,24} = 2.0640$.

Since our new test statistic, $t_0^* = 5.52 > t_{.025,15} = 2.0640$, we would again reject the null hypothesis in favor of the alternative.

**Note: These calculations could have been intuited by reasoning that the new mean,* $\mu_0^* = 90 < \mu_0 = 91$, *was naturally rejected by the previously calculated* 95% *confidence interval.*

d.) *Use the output and the t table to find* 99 *percent two-sided CI on the mean.*

$$\bar{y} - t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{y} + t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}$$
$$92.5805 - t_{\frac{.01}{2}, 25-1} \frac{2.3365}{\sqrt{25}} \leq \mu \leq 92.5805 + t_{\frac{.01}{2}, 25-1} \frac{2.3365}{\sqrt{25}}$$
$$92.5805 - t_{.005,24} \frac{2.3365}{5} \leq \mu \leq 92.5805 + t_{.005,24} \frac{2.3365}{5}$$
$$92.5805 - t_{.005,24} * 0.4673 \leq \mu \leq 92.5805 + t_{.005,24} * 0.4673$$
$$92.5805 - 2.797 * 0.4673 \leq \mu \leq 92.5805 + 2.797 * 0.4673$$
$$92.5805 - 1.3322 \leq \mu \leq 92.5805 + 1.3322$$
$$91.2738 \leq \mu \leq 93.8878$$

Therefore, a 99% confidence interval would be: $(91.2738, 93.8878)$ i.e (92.5808 ± 1.3322).

e.) *What is the P-value if the alternative hypothesis is $H_1 : \mu > 91$*

If the alternative hypothesis is modified to be $H_1 : \mu > 91$, then we would have a one-sided test as opposed to a two-sided test. This means that the new P-value, P^*, would be half of the original P-value, since we are only interested in the happening of one tail. Therefore, since the new alternative is $> \mu_0$, we have:

$$P^* = \frac{1}{2}P = \frac{1}{2} * .002 = 0.001.$$

2.9

Consider the Minitab output shown below.

One Sample T: Y
Test of $\mu = 25$ vs $\mu > 25$

Variable	N	Mean	Std. Dev	SE Mean	95% Lower Bound	T	P
Y	12	25.6818	?	0.3360	?	?	0.034

a.) *How many degrees of freedom are there on the t-test statistic?*

There are $n - 1 = 12 - 1 = 11$ degrees of freedom on the t-test statistic.

b.) *Fill in the missing information*

$$SE_{\bar{y}} = \frac{S}{\sqrt{n}}$$
$$0.3360 = \frac{S}{\sqrt{12}}$$
$$0.3360 * \sqrt{12} = S$$
$$1.1639 = S$$
$$S = 1.1639$$

$$LB = \bar{y} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}$$
$$LB = 25.6818 - t_{\frac{.05}{2},12-1}\frac{1.1639}{\sqrt{12}}$$
$$LB = 25.6818 - t_{.025,11} * 0.3360$$
$$LB = 25.6818 - 2.201 * 0.3360$$
$$LB = 23.1448$$

$$t_0 = \frac{\bar{y}-\mu_0}{\frac{S}{\sqrt{n}}}$$
$$t_0 = \frac{25.6818-25}{\frac{1.1639}{\sqrt{12}}}$$
$$t_0 = \frac{0.6818}{.3360}$$
$$t_0 = 2.029$$

Therefore, the completed table is as follows:

One Sample T: Y
Test of $\mu = 25$ vs $\mu > 25$

Variable	N	Mean	Std. Dev	SE Mean	95% Lower Bound	T	P
Y	12	25.6818	1.1639	0.3360	23.1448	2.029	0.034

2.10

Consider the Minitab output shown below.

Two-Sample T-Test and CI: Y1, Y2

Two-sample T for Y1 vs Y2

	N	Mean	Std. Dev	SE Mean
Y1	20	50.19	1.71	0.38
Y2	20	52.52	2.48	0.55

Difference = mu (X1) - mu (X2)
Estimate for difference: -2.33341
95% CI for difference: $(-3.69547, -0.97135)$
T-test of difference = 0 (vs not = 0): T-value = -3.47
P-Value = 0.001 DF = 38
Both use Pooled Std. Dev. = 2.1277

a.) *Can the null hypothesis be rejected at the 0.05 level? Why?*

Since the estimate for difference lies, -2.33341, lies within the 95% confidence interval of $(-3.69547, -097135)$, we fail to reject the null hypothesis at the 0.05 significance level.

Note:If the estimate for difference was outside this interval, then we would be able to reject the null hypothesis in favor of the alternate.*

b.) *Is this a one-sided or two-sided test?*

This is a two-sided test, since the alternative hypothesis is that the difference is not equal to 0.

Note: If the alternative was either* > 0 *or* < 0*, then we would have a one-sided test.*

c.) *If the hypothesis had been* $H_0 : \mu_1 - \mu_2 = 2$ *versus* $H_0 : \mu_1 - \mu_2 \neq 2$ *would you reject the null hypothesis at the 0.05 level?*

This change in the null hypothesis would result in a change in the test statistic, t_0. Therefore, the new test statistic, t_0^* would need to be calculated before any statment on rejection could be made.

$$t_0^* = \frac{\bar{y_1}-\bar{y_2}}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$
$$t_0^* = \frac{50.19-52.52}{2.1277\sqrt{\frac{1}{20}+\frac{1}{20}}}$$
$$t_0^* = \frac{-2.33}{2.1277\sqrt{\frac{1}{10}}}$$
$$t_0^* = -3.4629$$

We know that with a two-sided test, the fixed significance level criteria for rejection is $|t_0*| > t_{\frac{\alpha}{2},\nu}$.

We know that $t_{\frac{\alpha}{2},\nu} = t_{.025,38} = 2.0244$.

Since our new test statistic, $|t_0^*| = |-3.4269| = 3.4269 > t_{.025,38} = 2.0244$, we would reject the null hypothesis in favor of the alternative and the 0.05 significance level.

d.)*If the hypothesis had been $H_0 : \mu_1 - \mu_2 = 2$ versus $H_0 : \mu_1 - \mu_2 < 2$ would you reject the null hypothesis at the 0.05 level?Can you answer this question without doing additional calculations? Why?*

This change in the null hypothesis would result in a change in the test statistic, t_0. Therefore, the new test statistic, t_0^* would need to be calculated before any statment on rejection could be made.

$$t_0^* = \frac{\bar{y_1}-\bar{y_2}}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_1^2}{n_2}}}$$
$$t_0^* = \frac{50.19-52.52}{\sqrt{\frac{(1.71)^2}{20}+\frac{(2.48)^2}{20}}}$$
$$t_0^* = \frac{-2.33}{\sqrt{0.146205+0.30752}}$$
$$t_0^* = \frac{-2.33}{\sqrt{0.453725}}$$
$$t_0^* = \frac{-2.33}{.6736}$$
$$t_0^* = -3.4591$$

We know that for a left tail test, $-t_{\frac{\alpha}{2},\nu} = -t_{.05,38} = -1.6859$.

Since our new test statistic, $t_0^* = -3.4591 < t_{.05,38} = -1.6859$, we would reject the null hypothesis in favor of the alternative and the 0.05 significance level.

e.) *Use the output and the t-table to find a 95 percent upper confidence bound on the difference in means.*

$$\begin{aligned}
\text{UB} &= \bar{y_1} - \bar{y_1} + t_{\frac{\alpha}{2},\nu} S_p \sqrt{\tfrac{1}{n_1} + \tfrac{1}{n_2}} \\
\text{UB} &= 50.19 - 52.52 + t_{\frac{.05}{2},38} 2.1277 \sqrt{\tfrac{1}{20} + \tfrac{1}{20}} \\
\text{UB} &= -2.33 + 2.0244 * 2.1277 \sqrt{\tfrac{1}{10}} \\
\text{UB} &= -2.33 + 1.3621 \\
\text{UB} &= -.9679
\end{aligned}$$

Therefore, the 95% upper confidence bound on the difference of means is -0.9679.

2.12

The viscosity of a liquid detergent is supposed to average 800 centistokes at 25° *C. A random sample of 16 batches of detergent is collected and the average viscosity is 812. Suppose we know that the standard deviation of viscosity is* $\sigma = 25$ *centistokes.*

a.) *State the hypotheses that should be tested.*

We know from the problem description that the average should be 800 centistokes. We also know that the standard deviation of viscosity is $\sigma = 25$, we may assume normality. Based on this information, we would use a two-sided, one-variable Z test with the following hypotheses:

$$\begin{aligned}
H_0 &: \mu_0 = 800 \\
H_1 &: \mu_0 \neq 800
\end{aligned}$$

b.) *Test these hypotheses using* $\alpha = 0.05$*. What are your conclusions?*

First, we must caclulate the test statistic.

$$\begin{aligned}
Z_0 &= \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \\
Z_0 &= \frac{812 - 800}{\frac{25}{\sqrt{16}}} \\
Z_0 &= \frac{12}{\frac{25}{4}} \\
Z_0 &= 1.92
\end{aligned}$$

Since we are using a two-sided, one-variable Z test, the fixed significance level criteria for rejection is $|Z_0| > Z_{\frac{\alpha}{2}}$. We have a significance level of 5%, therefore $Z_{\frac{\alpha}{2}} = Z_{\frac{.05}{2}} = Z_{.025} = 1.96$

Combining this we have:

$|Z_0| = 1.92 \ngtr Z_{\frac{\alpha}{2}} = 1.96$

Therefore, at a 5% significance level, we fail to reject the null hypothesis in favor of the alternative and cannot conclude that the viscosity of the liquid detergent is not 800 centistokes at $25^{\circ}C$.

c.) *What is the P-Value for the test?*

$$
\begin{aligned}
P &= 2[1 - \phi(|Z_0|)] \\
P &= 2[1 - \phi(|1.92|)] \\
P &= 2[1 - \phi(1.92)] \\
P &= 2[1 - \phi(1.92)] \\
P &= 2[1 - .97257] \\
P &= 2[.02743] \\
P &= 0.05486
\end{aligned}
$$

d.) *Find a 95% confidence interval on the mean shaft diameter.*

$$
\begin{gathered}
\bar{y} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{y} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \\
812 - Z_{\frac{.05}{2}} \frac{25}{\sqrt{16}} \leq \mu \leq 812 + Z_{\frac{.05}{2}} \frac{25}{\sqrt{16}} \\
812 - Z_{.025} \frac{25}{4} \leq \mu \leq 812 + Z_{.025} \frac{25}{4} \\
812 - 1.96 * \frac{25}{4} \leq \mu \leq 812 + 1.96 * \frac{25}{4} \\
812 - 12.25 \leq \mu \leq 812 + 12.25 \\
799.75 \leq \mu \leq 824.25
\end{gathered}
$$

Therefore, we have a 95% confidence interval of $(799.75, 824.25)$ i.e. (812 ± 12.25).

2.14

A normally distributed random variable has an unknown mean μ *and a known variance* $\sigma^2 = 9$. *Find the sample size required to construct a 95 percent confidence interval on the mean that has total length of* 1.0.

Since the sample is normally distributed and the variance is known, we can conclude that the appropriate test is a two-sided, one-variable Z test.

The corresponding confidence interval is: $\bar{y} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{y} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

Therefore, the desired length of the interval can be represented by:

$$2(Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = 1.0$$

We know that we are using a significance level of $\alpha = 0.05$, therefore we compute the following test statistic:

$$Z_{\frac{\alpha}{2}} = Z_{\frac{.05}{2}} = Z_{0.025} = 1.96$$

We can now combine all of the known information and solve for n.

$$\begin{aligned} &2(Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = 1.0 \\ &2(1.96 \frac{\sqrt{9}}{\sqrt{n}}) = 1.0 \\ &(1.96 \frac{3}{\sqrt{n}}) = .5 \\ &5.88 = .5\sqrt{n} \\ &11.76 = \sqrt{n} \\ &n = 138.2976 \end{aligned}$$

Since, we are only interested in whole numbers, we shound round up such that $n = 139$.

Therefore, we need a sample size of 139 in order to construct a 95% confidence interval on the mean of a normally distributed random variable with a known variance and an unknown mean.

2.15

The shelf life of a carbonated beverage is of interest. Ten bottles are randomly selected and tested, and the following results are obtained:

Days	
108	138
124	163
124	159
106	134
115	139

a.) *We would like to demonstrate that the mean shelf life exceeds 120 days. Set up an appropriate hypotheses for investigating this claim.*

We are only given information about the sample population in the problem description. Since the population mean and variance are unknown, we can conclude that the one-variable t-test is most appropriate. Furthermore, since we are interested in demonstrating that the mean shelf life exceeds 120 days, we know that the test will be one-sided.

$H_0 : \mu_0 = 120$
$H_1 : \mu_0 > 120$

b.) *Test these hypotheses using $\alpha = 0.01$. What are your conclusions?*

The test statistic for a one-variable, one-sided t-test is as follows:

$$t_0 = \frac{\bar{y} - \mu_0}{\frac{S}{\sqrt{n}}}$$

Therefore we must first calculate the sample mean.

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$
$$\bar{y} = \frac{1310}{10}$$
$$\bar{y} = 131$$

Next, we must calculate the sample variance.

$$S^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}$$
$$S^2 = \frac{1}{10-1}(\sum_{i=1}^{n}(y_i-\bar{y})^2)$$
$$S^2 = \frac{((108-131)^2+(124-131)^2+(124-131)^2+(106-131)^2+(138-131)^2+(163-131)^2+(159-131)^2+(134-131)^2+(139-131)^2)}{10-1}$$
$$S^2 = \frac{3438}{9}$$
$$S^2 = 382$$

Which implies that the sample standard deviation $S = \sqrt{382} = 19.5448$.

We then use this information to find the test statistic.

$$t_0 = \frac{\bar{y}-\mu_0}{\frac{S}{\sqrt{n}}}$$
$$t_0 = \frac{131-120}{\frac{19.5448}{\sqrt{10}}}$$
$$t_0 = \frac{11}{6.1806}$$
$$t_0 = 1.7798$$

For a one-variable, one-sided t-test, the fixed significance level criteria for rejection is as follows:

$t_0 = 1.7798 \not> t_{\alpha,n-1} = t_{0.01,9} = 2.8214$

Therefore, at the 1% significance level, we fail to reject the null hypothesis in favor of the alternative and cannot conclude that the mean shelf life of the soft drink exceeds 120 days.

c.)*Find the P-value for the test in part (b).*

The P-value in a one-variable, one-sided t-test is the probability above t_0.

$$P = 1 - \phi(1.7798)$$
$$P = 1 - .94559$$
$$P = 0.05441$$

d.)*Construct a 99% confidence interval on the mean shelf life.*

The appropriate form of the 99% confidence interval on the mean is as follows:

$$\bar{y} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{y} + t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}$$
$$131 - t_{\frac{.01}{2},10-1}\frac{19.5448}{\sqrt{10}} \leq \mu \leq 131 + t_{\frac{.01}{2},10-1}\frac{19.5448}{\sqrt{10}}$$
$$131 - t_{.005,9}\frac{19.5448}{\sqrt{10}} \leq \mu \leq 131 + t_{.005,9}\frac{19.5448}{\sqrt{10}}$$
$$131 - 3.24983.2498 * (6.1806) \leq \mu \leq 131 + 3.2498 * (6.1806)$$
$$131 - 20.0857 \leq \mu \leq 131 + 20.0857$$
$$110.9143 \leq \mu \leq 151.0857$$

Therefore the 99% confidence interval on the mean would be $(110.9143, 151.0857)$ i.e. (131 ± 20.0857).

2.22

An article in Solid State Technology, "Orthogonal Design for Process Optimization and Its Application to Plasma Etching" by G.Z. Yin and D.W. Jillie (May 1987) describes an experiment to determine the effect of the C_2F_6 flow rate on the uniformity of the etch on a silicon wafer used in integrated circuit manufacturing. All of the runs were made in random order. Data for the two flow rates are as follows:

C_2F_6 **Flow (SCCM)**	**Uniformity Observation**					
	1	**2**	**3**	**4**	**5**	**6**
125	2.7	4.6	2.6	3.0	3.2	3.8
200	4.6	3.4	2.9	3.5	4.1	5.1

a.)*Does the C_2F_6 flow rate affect average etch uniformity? Use $\alpha = 0.05$.*

We know that we are interested in determining whether or not the two SCCM treatments effect average etch uniformity. Therefore, we would employ a two variable, two sided t-test to test whether or not the difference in the averages is 0 with the following hypotheses:

$$H_0 : \mu_1 = \mu_2$$
$$H_0 : \mu_1 = \mu_2$$

We must first calculate the sample means.

$$\bar{y}_1 = \frac{1}{n}\sum_{j=1}^{n} y_{1,j}$$
$$\bar{y}_1 = \frac{1}{6}((2.7) + (4.6) + (2.6) + (3.0) + (3.2) + (3.8))$$
$$\bar{y}_1 = \frac{1}{6}(19.9)$$
$$\bar{y}_1 = 3.3167$$

$$\bar{y}_2 = \frac{1}{n}\sum_{j=1}^{n} y_{2,j}$$
$$\bar{y}_2 = \frac{1}{6}((4.6) + (3.4) + (2.9) + (3.5) + (4.1) + (5.1))$$
$$\bar{y}_2 = \frac{1}{6}(23.6)$$
$$\bar{y}_2 = 3.933$$

Next, we must calculate the standard deviations.

$$S_1 = \left[\frac{\sum_{j=1}^{n}(y_{1,j} - \bar{y})^2}{n-1}\right]^{\frac{1}{2}}$$
$$S_1 = \left[\frac{2.8883}{5}\right]^{\frac{1}{2}}$$
$$S_1 = \sqrt{.57766}$$
$$S_1 = .7600$$

$$S_2 = \left[\frac{\sum_{j=1}^{n}(y_{2,j} - \bar{y})^2}{n-1}\right]^{\frac{1}{2}}$$
$$S_2 = \left[\frac{3.3733}{5}\right]^{\frac{1}{2}}$$
$$S_2 = \sqrt{.67466}$$
$$S_2 = .8214$$

Next, we must calculate the pooled standard deviations.

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$
$$S_p = \sqrt{\frac{(5)0.57766 + (5)0.67466}{10}}$$
$$S_p = \sqrt{0.6262}$$
$$S_p = 0.7913$$

We can now calculate the t-test statistc.

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$
$$t_0 = \frac{3.3167 - 3.933}{0.7913\sqrt{\frac{1}{6}+\frac{1}{6}}}$$
$$t_0 = \frac{-.6163}{0.7913\sqrt{\frac{1}{12}}}$$
$$t_0 = -1.3498$$

The fixed significance level criteria for rejection for the two-sample,two-sided t-test is:

$$|t_0| = |-1.3498| = 1.4262 \not> t_{\frac{\alpha}{2},n_1+n_2-1} = t_{.025,5} = 2.571$$

Therefore at the 5% significance level we fail to reject the null hypothesis in favor of the alternative and cannot conclude that the difference in flow rates results in a difference in the average etch uniformity.

b.) *What is the P-value for the test in part (a)?*

For the paired t-test, the P-value is the sum of the probability above t_0 and below $-t_0$.

$$P = 2[1 - \phi(-1.4262)]$$
$$P = 2[1 - .8934]$$
$$P = 2[.1066]$$
$$P = 0.2131$$

c.) *Does the C_2F_6 flow rate affect the wafer-to-wafer variability in etch uniformity? Use $\alpha = 0.05$.*

Since we are interested in comparing the sample variablity, we will employ the F test with the following hypotheses:

$$H_0 : \sigma_1 = \sigma_2$$
$$H_1 : \sigma_1 \neq \sigma_2$$

We begin by calculating the sample means.

$$\bar{y}_1 = \frac{\sum_{i=1}^{n} y_{1,i}}{n} \qquad \bar{y}_2 = \frac{\sum_{i=1}^{n} y_{2,i}}{n}$$

$$\bar{y}_1 = \frac{2.7+4.6+2.6+3.0+3.2+3.8}{6} \qquad \bar{y}_2 = \frac{4.6+3.4+2.9+3.5+4.1+5.1}{6}$$

$$\bar{y}_1 = \frac{19.9}{6} \qquad \bar{y}_2 = \frac{23.6}{6}$$

$$\bar{y}_1 = 3.3167 \qquad \bar{y}_2 = 3.9333$$

Next, we must calculate the sample variances.

$$S_1^2 = \frac{\sum_{i=1}^{n}(y_{1,i}-\bar{y}_1)^2}{n-1}$$

$$S_1^2 = \frac{(2.7-3.3167)^2+(4.6-3.3167)^2+(2.6-3.3167)^2+(3.0-3.3167)^2+(3.2-3.3167)^2+(3.8-3.3167)^2}{6-1}$$

$$S_1^2 = \frac{2.8883}{5}$$

$$S_1^2 = .5776$$

$$S_2^2 = \frac{\sum_{i=1}^{n}(y_{2,i}-\bar{y}_2)^2}{n-1}$$

$$S_2^2 = \frac{(4.6-3.9333)^2+(3.4-3.9333)^2+(2.9-3.9333)^2+(3.5-3.9333)^2+(4.1-3.9333)^2+(5.1-3.9333)^2}{6-1}$$

$$S_2^2 = \frac{3.3733}{5}$$

$$S_2^2 = .67467$$

We can now calculate the F-test statistc.

$$F_0 = \frac{S_1^2}{S_2^2}$$

$$F_0 = \frac{.5776}{.6747}$$

$$F_0 = .8562$$

The fixed level criteria for rejection for the two-sided F-test is:

$$F_0 \not> F_{\frac{\alpha}{2},n_1-1,n_2-1} \qquad F_0 \not< F_{1-\frac{\alpha}{2},n_1-1,n_2-1}$$

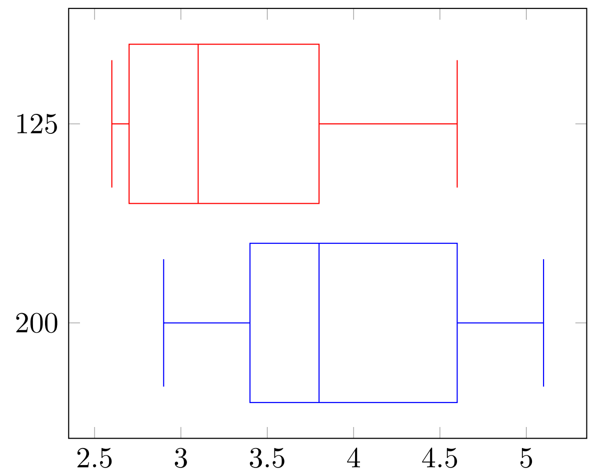
$$0.8562 \not> F_{\frac{.05}{2},6-1,6-1} \qquad 0.8562 \not< F_{1-\frac{.05}{2},6-1,6-1}$$

$$0.8562 \not> F_{.025,5,5} \qquad 0.8562 \not< F_{.975,5,5}$$

$$0.8562 \not> 7.1468 \qquad 0.8562 \not< 0.1399$$

Therefore, at the 5% significance level, we fail to reject the null hypothesis in favor of the alternative and cannot conclude that C_2F_6 flow rate affects the wafer-wafer variability in etch uniformity.

d.) *Draw box plots to assist in the interpretation of the data from this experiment.*



2.26

Twenty observations on etch uniformity on silicon wafers are taken during a qualification experiment for a plasma etcher. The data are as follows:

5.34	6.65	4.76	5.98	7.25
6.00	7.55	5.54	5.62	6.21
5.97	7.35	5.44	4.39	4.98
5.25	6.35	4.61	6.00	5.32

a.) *Construct a 95 percent confidence interval estimate of σ^2.*

Since we are interested in the variance of a single sample, we will use the χ^2 test as the basis for the appropriate confidence interval.

First, we must calculate the sample mean.

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$
$$\bar{y} = \frac{116.56}{20}$$
$$\bar{y} = 5.828$$

Next, we must caclulate the sample variance.

$$S^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$
$$S^2 = \frac{\sum_{i=1}^{n}(y_i - 5.828)^2}{20-1}$$
$$S^2 = \frac{15.01852}{19}$$
$$S^2 = .7904454$$

We are now ready to find the 95% confidence interval.

$$\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{(1-\frac{\alpha}{2}),n-1}}$$
$$\frac{(20-1).7904454}{\chi^2_{\frac{.05}{2},20-1}} \leq \sigma^2 \leq \frac{(20-1)7904454}{\chi^2_{(1-\frac{.05}{2}),20-1}}$$
$$\frac{(19).7904454}{\chi^2_{.025,19}} \leq \sigma^2 \leq \frac{(19).7904454}{\chi^2_{.975,19}}$$
$$\frac{(19).7904454}{\chi^2_{.025,19}} \leq \sigma^2 \leq \frac{(19).7904454}{\chi^2_{.975,19}}$$
$$\frac{15.01852}{32.8523} \leq \sigma^2 \leq \frac{15.01852}{8.9065}$$
$$0.4572 \leq \sigma^2 \leq 1.6862$$

Therefore, the 95% confidence interval on the variance is $(0.4572, 1.862)$ and (1.0717 ± 0.6145).

b.) *Test the hypothesis that $\sigma^2 = 1.0$. Use $\alpha = 0.05$. What are your conclusions?*

Since we are interested in testing whether the variance of population that is assumed to be normal equals a certain constant, we will use the χ^2 test with the following hypotheses:

$$H_0 : \sigma^2 = 1.0$$
$$H_1 : \sigma^2 \neq 1.0$$

We must first calculate the test statistic.

$$\chi^2_0 = \frac{(n-1)S^2}{\sigma^2_0}$$
$$\chi^2_0 = \frac{(20-1).7904454}{(1.0)^2}$$
$$\chi^2_0 = \frac{15.01852}{1}$$
$$\chi^2_0 = 15.01852$$

The fixed level criteria for rejection for the two-sided χ^2-test is:

$$
\begin{array}{r|l}
F_0 \not> \chi^2_{\frac{\alpha}{2},n-1} & F_0 \not< \chi^2_{1-\frac{\alpha}{2},n-1} \\
15.01852 \not> \chi^2_{\frac{.05}{2},20-1} & 15.01852 \not< \chi^2_{1-\frac{.05}{2},20-1} \\
15.01852 \not> \chi^2_{0.025,19} & 15.01852 \not< \chi^2_{0.975,19} \\
15.01852 \not> 32.8523 & 15.01852 \not< .89065
\end{array}
$$

Therefore, at the 5% significance level, we fail to reject the null hypothesis in favor of the alternative and cannot conclude that the variance in the etch uniformity of the silicon wafers is not 1.0

c.) *Discuss the normality assumption and its role in this problem.*

When we selected the χ^2 test as a basis for evaluating our hypothesis, we made an assumption regarding normality. That is to say, we assumed that the sample was drawn from a normaly distributed population. This was necessary for this problem since tests on variance are very sensitive to normality.

d.) *Check normality by constructing a normal probability plot. What are your conclusions?*

The following normal probability plot of the sample was created in minitab 17. Since the plotted points fall approximately on a straight line, we have no reason to question the normality assumption. If the plotted points varied significantly, then we would have reason to suspect that it is not normall distributed.