1.

QUESTION #1 and #2

CAN NOT BE SHARED SORRY

2.

$n = 3$

3. A sample of size 3 was drawn from a population and the resulting observations are reported below.

$$2.2, 2, 10.3$$

good idea to order them: 2.0, 2.2, 10.3

Using these observed values, report the following:

(a) (2 points) the mean

$$\bar{X} = \frac{2.2 + 2.0 + 10.3}{3} = \frac{14.5}{3} = \boxed{\frac{29}{6} \text{ or } 4.8\bar{3}}$$

(b) (2 points) the median

$$\text{median} = Q(0.5) = X_i + [(n \cdot p + 0.5) - i] \cdot (X_{i+1} - X_i) = X_2 + [(2) - (2)](X_3 - X_2) = X_2 - (0)$$

$$i = \lfloor n \cdot p + 0.5 \rfloor = \lfloor 3 \cdot (0.5) + 0.5 \rfloor = \lfloor 2.0 \rfloor = 2$$

$$= X_2 = \boxed{2.2}$$

(c) (2 points) the variance

Using sample variance:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{1}{3-1}\left[\left(2.0 - \frac{29}{6}\right)^2 + \left(2.2 - \frac{29}{6}\right)^2 + \left(10.3 - \frac{29}{6}\right)^2\right]$$

$$= \frac{1}{2}[6.93 + 8.03 + 29.9] = \boxed{22.42}$$

(d) (2 points) the standard deviation

Using sample standard deviation:

$$s = \sqrt{s^2} = \sqrt{22.42} \approx \boxed{4.74}$$

5. A company specializing in the installation and maintanence of "infinity pools" records the number of service requests they receive each month for two years. The number of requests are presented in the table below:

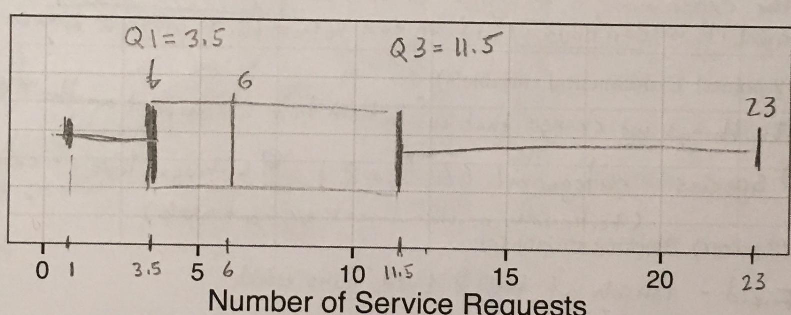
```
The decimal point is 1 digit(s) to the right of the |

0 | 122333444      -> 1, 2, 2, 2, ...
0 | 556669999         5, 5, 6, 6, ...
1 | 4                 14
1 | 77                17, 17
2 | 013               20, 21, 23
```

Note that 0 | 4 represents 4 and 1 | 2 represents 12. In this case, the first quartile is $Q(.25) = 3.5$, the median is 6, and the third quartile is $Q(.75) = 11.5$.

(a) (10 points) Using the axes below, create a box plot to summarize the data. Label all important values. Draw a star over unusual observations.

$Q3 = 11.5$
$Q1 = 3.5$ $\Rightarrow$ $IQR = Q3 - Q1 = 11.5 - 3.5 = 8$

Whisker length at most $1.5 \cdot IQR = 12$



Finding Upper Whisker: $Q3 + 1.5 \cdot IQR = 23.5$

biggest observation inside 23.5 is 23

So whisker extends to 23

Finding Lower Whisker: $Q1 - 1.5 \cdot IQR = 3.5 - 12 = -8.5$

smallest point larger than -8.5 is 1

So whisker extends to 1.

The whiskers cover all the observations so none are unusual

4. An agriculturist is attempting to determine which of three species of corn (A, B, and C) yield the most grain per acre. Since the yield may depend on the fertilizer used, the researcher intends to use fertilizers with different concentrations of Nitrogen as well - low Nitrogen, medium-low Nitrogen, medium-high Nitrogen, and high Nitrogen. There are 8 fields (scattered around Iowa) available to perform this expiriment. Each field is divided into 24 single acre plots and the combinations of species and fertilizer are randomly assigned so that within each field every combination is used exactly twice. At harvest time, the amount of grain each plot yields is recorded and the combination of corn species and fertilizer that gives the highest average yield is chosen.

   (a) (2 points) Is this an experiment or an observational study? Explain.

   The agriculturist is actively manipulating the scenarios from which the observations are collected - so this is an experiment.

   (b) Identify the following (if there was not one, simply put "not used"). Additionally, label each as continuous or discrete.

      i. (2 points) Response variable(s):

      We are collecting and measuring the grain each acre yields - this is what the experiment is designed to do, so $\boxed{\text{Yield per acre}}$ is our response and it is $\boxed{\text{continuous}}$ (if we can believe that yield is measured in weight)

      ii. (2 points) Experimental variable(s):

      The things we change that we believe have an impact on the result:

      ① Species - categorical (A, B, or C) ② Nitrogen levels - categorical low / low-med / med / med-high / high

      (technically neither discrete or continuous)

      iii. (2 points) Blocking variable(s):

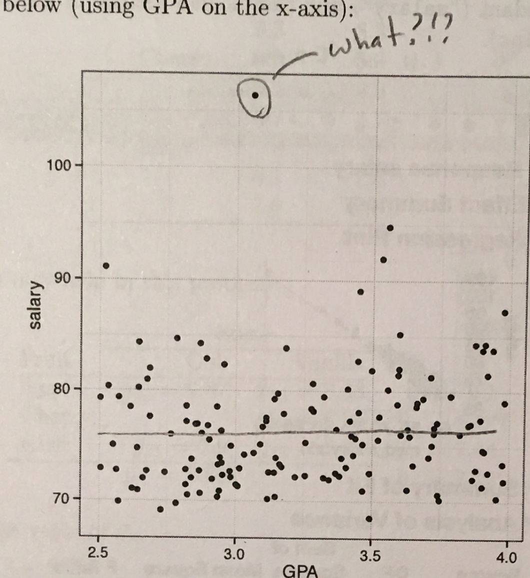
      Field - which of the 8 fields was used.

   (c) (2 points) Was replication used in this study? If so, where was it applied? If not, how could we have applied it?

   Yes - in each field, the combinations are used twice each.
   Repeating the same combinations within a block = replication.

6. A survey given to members of a national engineering honor society who have recently graduated is attempting to determine the relationship between salary and GPA. The graph below displays 150 responses.

The results are depicted below (using GPA on the x-axis):



A line of best fit should have a slope of about 0 since there isn't much pattern here

Here are some summaries of the data (again using the actual score as the x-value and the person's evaluation of their score as the y-value):

$$b_1 = \frac{\Sigma x_i y_i - n\cdot\bar{x}\bar{y}}{\Sigma x_i^2 - n\cdot\bar{x}^2} = \frac{37299 - 150\cdot(3.25)(76.5)}{882126 - 150\cdot(3.25)^2} = 0.00004517$$

$$\sum_{i=1}^{150} x_i = 487 \qquad \bar{x} = \frac{487}{150} = 3.25 \qquad \sum_{i=1}^{150} x_i^2 = 1609$$

$$\sum_{i=1}^{150} y_i = 11474 \qquad \bar{y} = \frac{11474}{150} = 76.5 \qquad \sum_{i=1}^{150} y_i^2 = 882126$$

$$\sum_{i=1}^{150} x_i y_i = 37299$$

$$b_0 = \bar{y} - b_1\bar{x} = 76.5 - b_1\cdot 3.25 = 76.496$$

(a) Using the summaries above, the survey workers fit a linear relationship between **GPA** (x) and **salary** (y).

i. (5 points) Write the equation of the fitted linear relationship.

$$\hat{y} = b_0 + b_1 x \Rightarrow \boxed{\hat{y} = 76.496 - 0.00004517\cdot x}$$

where x is GPA and $\hat{y}$ is predicted salary

ii. (5 points) Using the fitted line, what do we suppose the salary will be for an engineer with a GPA of 3.0?

$$\hat{y} = 76.496 - 0.00004517\cdot(3.0) = 76.49624$$

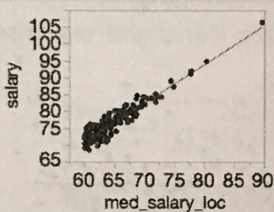
(b) Discouraged by the relationship between salary and GPA, the surveyors remember that they know the address of each respondant and are able to determine the median income of the area in which the respondant lives. The JMP output below comes from fitting a linear relationship using the annual salary of the respondant ("salary") and the median income of the area in which the respondant lives (med_salary_loc).

eng_sal: Fit Least Squares

Response salary

Effect Summary

Regression Plot



Summary of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	3884.3322	3884.33	1173.720
Error	148	489.7940	3.31	Prob > F
C. Total	149	4374.1263		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.2289215	2.172865	1.03	0.3067
med_salary_loc	1.1486576	0.033528	34.26	<.0001*

b_0

b_1

Effect Tests

Effect Details

i. (5 points) Write the equation of the fitted linear relationship.

$\hat{Y} = b_0 + b_1 \cdot X \Rightarrow$ $\boxed{\hat{Y} = 2.2289215 + 1.1486576 \cdot X}$ where X is median salary by location and $\hat{Y}$ is predicted salary

ii. (5 points) Find and interpret the value of R^2 for the fitted quadratic relationship.

$$R^2 = \frac{SSTO - SSE}{SSTO} = \frac{4374.1263 - 489.7940}{4374.1263} = .8880$$

So 88.8% of the variability in salary is explained by the median of the location

7. A winery is experimenting with blending a small amount of non-grape tastes into its current harvest of grapes (to add "notes"). They are considering three fruit additions (apple, cherry, and kiwi) and two spice additions (oak and vanilla). Three wine experts working for the company test the fruit/spice combinations and provide a rating from 0 to 10 (with 10 being the highest).

The results are recorded below.

	Spice	
Fruit	Oak	Vanilla
Apple	9.8	8.4
	9.9	8.9
	9.3	8.3
Cherry	~~8.3~~ 5.3	~~5.3~~ 4.1
	5.5	4.9
	5.7	4.7
Kiwi	8	7
	8.1	7.3
	7.6	6.6

typos

$\bar{Y}_{21} = \frac{5.3+5.5+5.7}{3} = 5.5$

$\bar{Y}_{22} = \frac{4.1+4.9+4.7}{3} = \frac{13.7}{3} = 4.57$

The following summaries may help in this problem:

	Spice		
Fruit	Oak	Vanilla	
Apple	$\bar{y}_{11} = 9.67$	$\bar{y}_{12} = 8.53$	$\bar{y}_{1\cdot} = 9.1$
Cherry		~~$\bar{y}_{2\cdot} = 5.04$~~	$\bar{Y}_{2\cdot} = 5.04$
Kiwi	$\bar{y}_{31} = 7.9$	$\bar{y}_{32} = 6.97$	$\bar{y}_{3\cdot} = 7.44$
	$\bar{y}_{\cdot 1} = 7.69$	$\bar{Y}_{\cdot 2}$	$\bar{y}_{\cdot\cdot} = 7.19$

(a) (2 points) Report the value of $\bar{y}_{21}$

$\bar{Y}_{21} = \frac{5.3+5.5+5.7}{3} = 5.5$

(b) (2 points) Report the value of $\bar{y}_{\cdot 2}$

$\bar{Y}_{\cdot 2} = \frac{1}{3}(\bar{Y}_{12} + \bar{Y}_{22} + \bar{Y}_{32}) = \frac{1}{3}(8.53 + 4.57 + 6.97)$
$= 6.69$

(c) (3 points) Find the fitted main effect of fruit, a_1, a_2, and a_3, that you would get from factorial model that ignores interactions.

$a_1 = \bar{Y}_{1\cdot} - \bar{Y}_{\cdot\cdot} = 9.1 - 7.19 = 1.911$ (makes taste 1.911 pts. higher on average)

$a_2 = \bar{Y}_{2\cdot} - \bar{Y}_{\cdot\cdot} = 5.04 - 7.19 = -2.16$ (makes taste 2.16 pts lower on average)

$a_3 = \bar{Y}_{3\cdot} - \bar{Y}_{\cdot\cdot} = 7.44 - 7.19 = 0.244$ (makes taste 0.244 pts higher on average)

(d) (3 points) Ignoring possible interactions, give the estimated values $\hat{y}_{22}$ and $\hat{y}_{23}$.

$\hat{Y}_{22} = \bar{Y}_{\cdot\cdot} + a_2 + b_2$
$= 7.19 + (-2.16) + (-0.50)$
$= 4.53$

$b_1 = \bar{Y}_{\cdot 1} - \bar{Y}_{\cdot\cdot} = 7.69 - 7.19$
$= 0.50$
$b_2 = \bar{Y}_{\cdot 2} - \bar{Y}_{\cdot\cdot} = 6.69 - 7.19$
$= -0.50$

$\hat{Y}_{23}$ does not exist, $\hat{Y}_{32} = 7.19 + 0.244 + (-.5) = 6.934$

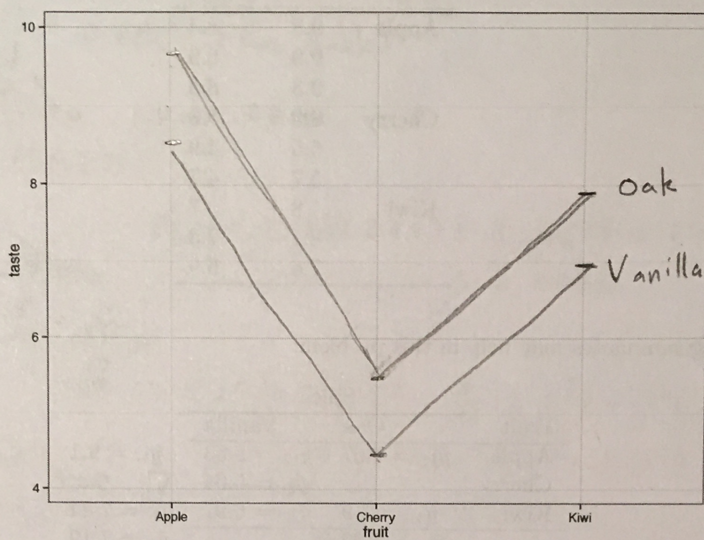
(e) (2 points) How do the estimated values computed above compare to the average for the same combinations seen in the data? Does it appear that ignoring interactions was a good choice?

$\hat{Y}_{22} = 4.53$ and $\bar{Y}_{22} = 4.57$ (pretty close)

$\hat{Y}_{32} = 6.934$ and $\bar{Y}_{32} = 6.97$ (pretty close)

both results are pretty close ⇒ ignoring interactions may have been good

(f) (5 points) Using the template below, create a profile plot for this data:



(g) (2 points) Using the plot does it appear that there are interactions between fruit and spice type? Which combination would you recommend?

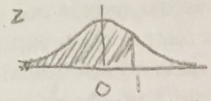
The segments imply that there are no interactions (all segments parallel) So we are probably justified to ignore interactions.

Oak and Apple has the highest average rating.

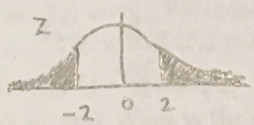
8. Let X be a normal random variable with a mean of 1 and a varaince of 16 (i.e., $X \sim N(1,16)$) and let Z be a random variable following a standard normal distribution. Find the following probabilities (note: the attached standard normal probability table may be helpful):

(a) (2 points) $P(Z \leq 1)$

$P(Z \leq 1) = 0.8413$ (from table)

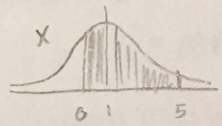


(b) (2 points) $P(|Z| \geq 2)$



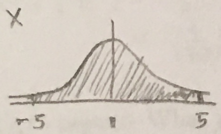
$P(Z \leq -2) + P(Z \geq 2) = 0.0228 + (1 - 0.9772)$
$= 0.0456$

(c) (2 points) $P(0 \leq X < 5)$



$P(0 \leq X \leq 5) = P\left(\frac{0-1}{4} \leq \frac{X-1}{4} \leq \frac{5-1}{4}\right) = P(-0.25 \leq Z \leq 1)$
$= P(Z \leq 1) - P(Z \leq -0.25)$
$= 0.8413 - 0.4013$
$= 0.44$

(d) (2 points) $P(|X| \leq 5)$



$P(|X| \leq 5) = P(-5 \leq X \leq 5) = P\left(\frac{-5-1}{4} \leq Z \leq \frac{5-1}{4}\right) = P(-1.5 \leq Z \leq 1)$
$= P(Z \leq 1) - P(Z \leq -1.5)$
$= 0.8413 - 0.0668 = 0.7745$

9. Suppose that X is a continuous random variable with cumulative density function (cdf):

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-3x} & x \geq 0 \end{cases}$$

$F(x) = P(X \leq x)$

(a) (2 points) What is the probability that X takes a value less than 1?

$P(X \leq 1) = F(1) = 1 - e^{-3(1)} = \boxed{1 - e^{-3}}$

(b) (2 points) What is the probability that X takes a value greater than 2?

$P(X > 2) = 1 - P(X \leq 2) = 1 - F(2) = 1 - (1 - e^{-3(2)}) = \boxed{e^{-6}}$

(c) (2 points) Derive f(x), the probability density function.

the derivative of F(x) is f(x)

$F'(x) = (1 - e^{-3x})' = 0 - (-3)e^{-3x} = 3e^{-3x}$ if $x \geq 0$

$F'(x) = (0)' = 0$ if $x < 0$

So

$$f(x) = \begin{cases} 3e^{-3x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

10. Suppose we have a bag containing five tiles, three of which are labelled 1 and two of which are labelled 2. Assume that each of the five tiles has an equal chance of being drawn. The number on the tile tells us how many times we will roll a fair die. For instance, if we draw a tile with the number 2 on it, we will roll a die twice but if we draw a tile with a 1 on it, we will only roll a die once. For this problem,

- let X be the number on the tile
- let Y be the sum of our rolls

(a) (2 points) Find $f_X(x)$.

X can be 1 or 2. 3/5 tiles say "1" and $\frac{2}{5}$ tiles say "2".

So $\boxed{f_X(x) = \begin{cases} 3/5 & \text{if } x=1 \\ 2/5 & \text{if } x=2 \\ 0 & \text{otherwise} \end{cases}}$ since $P(X=1) = 3/5$ and $P(X=2) = 2/5$

(b) (2 points) Find $f_{Y|X}(6|1)$.

$f_{Y|X}(6|1) = P(Y=6|X=1) = P(\text{"sum of rolls is 6"} \mid \text{"roll the die once"})$
$= \frac{1}{6}$

(c) (2 points) Find the joint probability $f(1,6)$.

$f_{XY}(1,6) = f_{Y|X}(6|1) \cdot f_X(1) = \frac{1}{6} \cdot \frac{3}{5} = \frac{3}{30} = \boxed{\frac{1}{10}}$

(d) (2 points) Find the joint probability $f(2,6)$.

$f(6|2) =$ Probability of sum of rolls $= 6$ given 2 rolls $= 5/36$
$f_X(2) = \frac{2}{5}$

$f_{XY}(2,6) = f_{Y|X}(6|2) \cdot f_X(2)$
$= \frac{5}{36} \cdot \frac{2}{5}$
$= \frac{2}{36}$
$= \boxed{\frac{1}{18}}$

(e) (3 points) Find $f_Y(6)$.

$f_Y(6) = \sum_{\text{all } x} f_{XY}(x,6) = f_{XY}(1,6) + f_{XY}(2,6)$
$= \frac{1}{10} + \frac{1}{18} = \frac{18}{180} + \frac{10}{180} = \boxed{\frac{28}{180}}$

(f) (2 points) Find $f_{X|Y}(2|6)$.

$f_{X|Y}(2|6) = \frac{f_{XY}(2,6)}{f_Y(6)} = \frac{1/18}{28/180} = \boxed{\frac{10}{28}}$

11. O-rings are elastomer loops designed to create a seal between the interface of two parts of a mechanical device. Because the elasticity of the material used to make them can be impacted by temperature (which can lead to the seal being broken) it is important to make sure that the O-ring is functional at the temperatures the part they are used in will be exposed to. Two composites (Composite X and Composite Y) are being tested in an O-ring that will be used in a part of a satellite that will be exposed to very low temperatures. A sample of 50 O-rings from each composite are placed in a chamber, where the temperature is gradually reduced until the seal is broken. Suppose that each composite has some mean failure temperature, μ_X for Composite X and μ_Y for Composite Y, and some variance in failure temperature, σ_X^2 for Composite X and σ_Y^2 for Composite Y. Before any observations are recorded, we can consider the sampled values from Composite X to be random variables $X_1, X_2, \ldots, X_{50}$ with $\mathbb{E}(X_i) = \mu_X$ and $Var(X_i) = \sigma_X^2$. We can also consider the sampled values from Composite Y to be random variables $Y_1, Y_2, \ldots, Y_{50}$ with $\mathbb{E}(Y_i) = \mu_Y$ and $Var(Y_i) = \sigma_Y^2$.

Let $\bar{X} = \frac{1}{50}X_1 + \frac{1}{50}X_2 + \ldots + \frac{1}{50}X_{50}$ and let $\bar{Y} = \frac{1}{50}Y_1 + \frac{1}{50}Y_2 + \ldots + \frac{1}{50}Y_{50}$.

(a) (3 points) What is the expected value of $\bar{X}$ (use appropriate symbols if needed).

These are just properties of sample means

$$E(\bar{X}) = \tfrac{1}{50}E(X_1) + \tfrac{1}{50}\cdot E(X_2) + \ldots + \tfrac{1}{50}E(X_{50})$$
$$= \tfrac{1}{50}\cdot\mu_X + \tfrac{1}{50}\cdot\mu_X + \ldots + \tfrac{1}{50}\cdot\mu_X = 50\left(\tfrac{1}{50}\mu_X\right) = \boxed{\mu_X}$$

(b) (3 points) What is the variance of $\bar{X}$ (use appropriate symbols if needed).

$$Var(\bar{X}) = \left(\tfrac{1}{50}\right)^2 Var(X_1) + \left(\tfrac{1}{50}\right)^2 Var(X_2) + \ldots + \left(\tfrac{1}{50}\right)^2 Var(X_{50})$$
$$= \left(\tfrac{1}{50}\right)^2\sigma_X^2 + \left(\tfrac{1}{50}\right)^2\sigma^2 + \ldots + \left(\tfrac{1}{50}\right)^2\sigma_X^2 = 50\left[\left(\tfrac{1}{50}\right)^2\sigma_X^2\right] = \sigma_X^2/50$$

(c) (3 points) What is the distribution of $\bar{X}$ (use appropriate symbols if needed).

by CLT since $n = 50$, $\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{50}\right)$

(d) (3 points) What is the expected value of $\bar{Y}$ (use appropriate symbols if needed).

as in (a) we get $E(\bar{Y}) = \mu_Y$

(e) (3 points) What is the variance of $\bar{Y}$ (use appropriate symbols if needed).

as in (b) $Var(\bar{Y}) = \sigma_Y^2/50$

(f) (3 points) What is the distribution of $\bar{Y}$ (use appropriate symbols if needed).

by CLT $\bar{Y} \sim N(\mu_Y, \sigma_Y^2/50)$ since $n \geq 25$

(g) (6 points) Let $\bar{D} = \bar{X} - \bar{Y}$. What is the distribution of $\bar{D}$ (use appropriate symbols if needed).

$E(\bar{D}) = E(\bar{X}) - E(\bar{Y}) = \mu_X - \mu_Y$

$Var(\bar{D}) = Var(\bar{X}) + Var(\bar{Y}) = \sigma_X^2/50 + \sigma_Y^2/50$

Since $\bar{X}$ and $\bar{Y}$ are both normal,

$\bar{D} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{50} + \frac{\sigma_Y^2}{50}\right)$

12. After running the O-ring experiment, the researchers found $\bar{x} = 50$ K and $\bar{y} = 53$ K. Suppose that $\sigma_X^2 = 10$ and $\sigma_Y = 20$.

(a) (4 points) Provide a 90% confidence interval for μ_X.

$\alpha = 0.10$

for 90% conf. interval, we use $Z_{.95} = 1.645$

$\bar{X} \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_X^2}{n}}$

$50 \pm 1.645\sqrt{\frac{10}{50}} \Rightarrow (49.26, 50.73)$

(b) (4 points) Provide a 99% confidence interval for μ_X.

$\alpha = 0.01$, for 99% confidence we use $Z_{1-\frac{\alpha}{2}} = Z_{.995} = 2.58$

$\bar{X} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_X^2}{n}} \Rightarrow 50 \pm 2.58\sqrt{\frac{10}{50}}$

$\Rightarrow (48.85, 51.15)$

(c) (4 points) Provide a 95% confidence interval for μ_Y.

$\alpha = 0.05$, for 95% confidence interval, we use $Z_{1-\alpha/2} = 1.96$

$\bar{Y} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_Y^2}{n}} \Rightarrow 53 \pm 1.96\sqrt{\frac{20}{50}} \Rightarrow (51.76, 54.24)$

(d) ~~(6 points)~~ Provide a 95% confidence interval for $\mu_X - \mu_Y$ (hint: you can use the distribution of $\bar{D}$). Does this provide ~~any evidence~~ that one O-ring is better than the other?

Not covered

(e) (2 points) Is ~~there any~~ evidence that one O-ring is better than the other?

Not covered

13. A company recently did a major overhaul to their server system hardware and is checking to make sure that there have been no changes in the download speed. The previous download speed had an average of 63.4 Mbps. A systems analyst took 10 readings on the download speeds during the course of a day to check. Her results are below (in Mbps):

63.63, 63.4, 63.51, 63.14, 63.38, 63.35, 63.53, 63.37, 63.53, 63.71

The sample average is 63.45 and the sample variance is 0.026.

(a) (5 points) Provide a 90% confidence interval for the mean download speed.

not covered

(b) (5 points) Provide a 95% lower confidence bound for the mean download speed.

not covered

(c) (10 points) Conduct a hypothesis test at the 95% confidence level for the null hypothesis $\mu = 63.4$ against the alternative $\mu \neq 63.4$. Include your hypothesis statement, the test statistic, the p-value, your decision rule, and your conclusion.

not covered