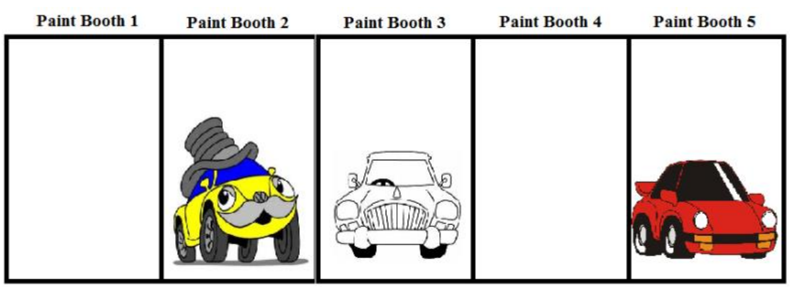
SPC

TOPICS: Rational Subgrouping, Capability Analysis

Problem 1. A car company was asked by its customer to run Xbar and R charts on the process characteristic "paint thickness." The car company has 5 different paint booths. The car company's engineering team proposed ***2 different sampling plans***.

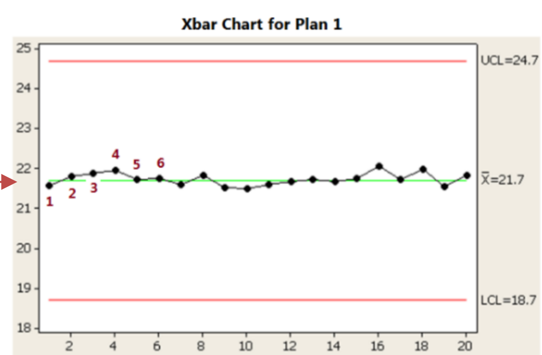


Plan 1:

- 5 cars will be randomly selected each day.
- ***1 car*** will be selected from ***each of the 5 paint booths***.
- ***Paint thicknesses for the 5 cars are averaged together*** *and the* ***average is plotted on the Xbar chart*** for that day.

The data and Xbar chart for this sampling plan are:

Day	Booth 1	Booth 2	Booth 3	Booth 4	Booth 5	Avg	Range
1	20.34	20.64	19.94	24.42	24.02	21.87	4.48
2	19.22	19.05	20.79	25.39	23.78	21.65	6.34
3	18.53	22.23	19.92	25.04	23.49	21.84	6.51
4	20.22	20.07	19.79	22.3	24.02	21.28	4.23
5	21.15	19.72	21.53	23.1	25.19	22.14	5.47
6	19.39	20.8	19.21	24.03	23.62	21.41	4.82
...	...	...	...	...	...	...	...



1(a) Does the **Xbar chart** indicate the *differences in paint thicknesses between the booths or the differences in paint thicknesses between the days?*

Between Booths **Between Days**

Solution: Each Xbar point on the chart represents a day. This chart displays day-to-day differences.

1(b) The **R chart** is not drawn. Would the **R chart** be used to indicate the *differences in paint thicknesses between the booths* or the *differences in paint thicknesses between days?*

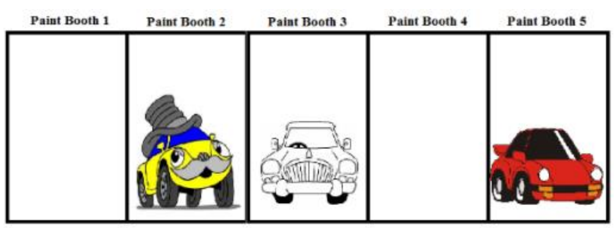
Between Booths Between Days

Solution: Each R point is the range between the paint thicknesses from the 5 booths. Also, since the Xbar chart is tracking between days, then the R chart must track within a day. Within a day, the subgroup is one car from each booth.

1(c) Does the **Xbar chart** appear to be "in control" considering all the Type I rules that we've talked about in class? If no, what is the issue?

ANSWER: No. It is in control with respect to all points being between the LCL and UCL on the control chart that we call special cause Rule 1. It DOES break the rule of having 15 points or more points within 1 standard deviation of the center line. Student's answer should say something about the points "hugging" the centerline or that this phenomenon represents the "too good to be true" scenario.

Problem 1 continued.



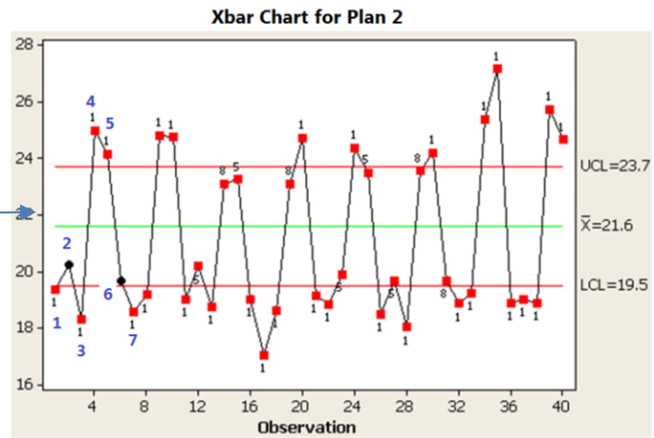
Plan 2:

- 5 cars will be sampled each day.
- The ***5 cars will be selected from one paint booth. Paint thickness for the 5 cars are averaged and the average is plotted on the Xbar chart for that day.***
- On the following day, the ***next paint booth is selected and the 5 cars are selected from it and averaged.***

For example, 5 cars from Booth 1 are sampled on Monday; 5 cars from Booth 2 are sampled on Tuesday; 5 cars from Booth 3 are sampled on Wednesday, etc.

The data and Xbar chart for this sampling plan are:

Day	Booth	Car 1	Car 2	Car 3	Car 4	Car 5	Avg	Range
1	1	19.41	19.11	21.26	20.61	20.09	20.10	2.15
2	2	20.2	20.78	20.76	19.41	19.55	20.14	1.37
3	3	18.75	19.01	20.18	17.81	20	19.15	2.37
4	4	24.59	24.11	24.53	26	23.79	24.60	2.21
5	5	22.59	23.71	23.28	23.52	25.86	23.79	3.27
6	1	20.16	20.97	20.25	19.98	19.36	20.14	1.61
7	2	20.56	20.29	20.55	21.85	19.74	20.60	2.11
...	...	...	...	...	...	...	...	...



1(d) Is the **Xbar chart** being used to indicate the *differences in paint thicknesses between the booths* or the *differences in paint thicknesses within a booth?*

Between Booths Within a Booth

Solution: The Xbar chart is displaying the large spread in thicknesses that different booths are producing.

1(e) The **R chart** is not drawn. Would the **R chart** be used to indicate the *differences in paint thicknesses between the booths* or the *differences in paint thicknesses within a booth?*

Between Booths **Within a Booth**

Solution: By looking at the table above, we can see by the Range value that there is not a lot of variation or spread WITHIN a booth.

1(f) Is there more variation between booths or within booths considering the data from Plan 1 and Plan 2?

Between Booths Within a Booth

Solution: You can see this by comparing the range values on the two different tables. The range values for the spread between the booths is averaging ~5, while the range values of the spread within a booth is averaging ~2.25.

1(g) Let's say all the cars that were painted in a given month are sitting in a back lot. In order to test paint thicknesses for that day, you take a random sample of 5 cars from the lot and measure their paint thicknesses. You average the 5 data points to put on the Xbar chart and their ranges are plotted on an R chart. What's one problem with this sampling plan?

Possible ANSWERS:

- If a point is out of control, you will never know which paint booth that it came from in order to correct the process.
- If the distribution of results from each booth are different, then this measurement scheme will never allow you to determine what these distributions are.

- You could randomly select your sample of 5 for cars that all came from the same booth.
- Other answers are possible if you explain another “reasonable” problem that potentially *could* happen, even if only rarely.

Problem 2. Capability Analysis Experiment. Form a team of 2, 3, or 4 members. Each team will need a tape measure or ruler and an experimental object: Pig Popper, leap frog, flying monkey, air mail, chicken or dragon flingers, etc.

The object is to **collect some type of distance data**; e.g., distance an object lands from a target, where objects may be:

- Ball popped from a pig’s mouth,
- Leaping frog,
- Flying monkey,
- Air mail,
- Flinging chicken or dragon, or
- Ball from nerf-like gun.

(a) Being as precise as possible (i.e. more decimal places = better!), one (or two) team members will shoot/fling/throw the object while the other team members will measure the distances. Write down what data your team is collecting and ***how specifically the distance is measured*** (e.g., from target to center of ball, or left arm of monkey to target, or center of frog’s head to target, etc.). Drawings are appropriate in providing clear explanations!

(b) After ~10 trial runs, set lower and upper specifications for the “ideal” distance after determining realistic values for the distances. List the specification limits below:

LSL: **Realistic value dependent on item used to collect data** USL: **Realistic value dependent on item**

(c) Collect 30 consecutive data points. Enter the data points in a column in Minitab and name the column appropriately. Again, make your measurements as “tight” (or resolute) as possible. Don’t just simply round your measurements to the nearest unit (recall our Normality vs. Rounding discussion!).

(d) **[+1]** Before conducting a capability analysis, check these necessary assumptions and **attach a copy of the appropriate graphics: 1. Process Control, 2. Normality, 3. IID.**

(e)-(g) Based on your graphics, answer the following.

(e) Are the distances “in control?” If not, what is the main issue affecting process control? (e.g., points beyond the LCL or UCL, too many points hugging the centerline, trend in the mean, etc.)

(f) Is your team’s data normally distributed? Record the p-value here for the normality test: ***p*-value:**

(g) Is your team’s data IID? If it is not, what lag value is significant and is it positive or negative?

(h) Regardless of whether your process met the appropriate assumptions to move forward (e.g., IID data, process in control) with a Capability Analysis, do so.

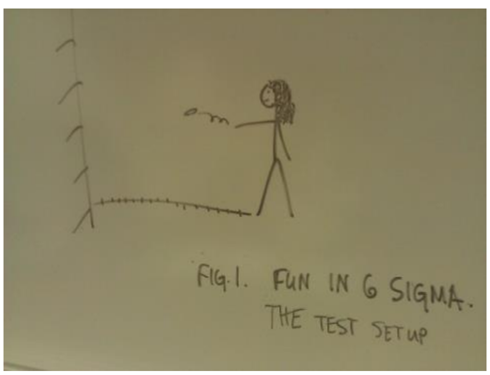
Using the specification limits that you set in part (b), determine the “capability” of the team members’ distances to meet these specifications. Perform the appropriate capability analysis listed below according to the normality of your process and record the capability indices: C_p and C_{pk} (if they exist), and P_p and P_{pk}.

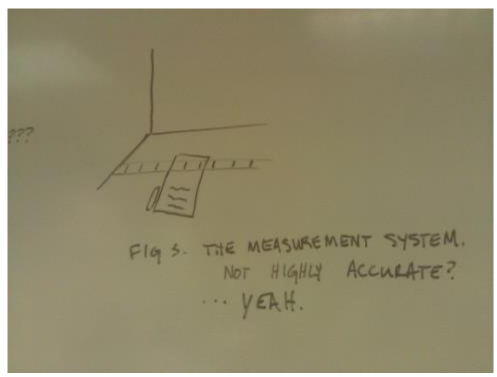
Recall: Data normally distributed: **Stat** > **Quality Tools** > **Capability Analysis** > **Normal;** your subgroup size is 1.

Recall: Data not normally distributed: **Stat** > **Quality Tools** > **Capability Analysis** > **Normal;** your subgroup size is 1; Select **Transform** and choose the **Box-Cox transformation**.

C_p= C_{pk}= P_p= P_{pk}=

Problem 3. Coin Toss Exercise: This is an exercise in which team member Mr. Potato head flipped a coin near a classroom wall to get as close to it as possible. The distance (in cm) from the wall to the center of the coin is collected for his 30 tosses. Data for Mr. Potato Head is in: **HmwkSet6_CapabilityAnalysisDATA**.

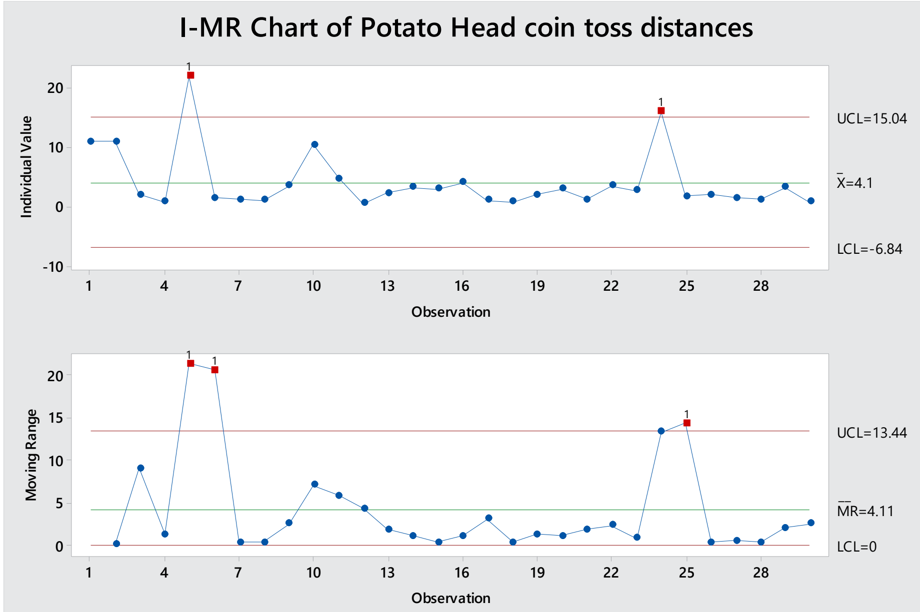




(a) **Yes or No.** Construct an I-MR chart to determine if his coin tosses are in control. Turn on all tests for Type I Error. Is his data in control based on the I-MR chart?

Yes **No**

Solution: I would say the process is not stable because there are three aspects of the graphs that suggest this. There is a string of points very near the centerline between trials 12 and 23 on the Individuals chart. Naturally, there are no “low” points below 0 since the process yields positive data. There are out of control points on the moving range chart, and although this is feasible because of the nature of the process, there are 3 such points (out of 30). Also, the two out of control points on the Individuals chart makes me curious if there is some type of pattern over an interval of every 20 points (although we know there isn’t). But, could there be a long toss every 20+ trials (do to fatigue or regrouping)?

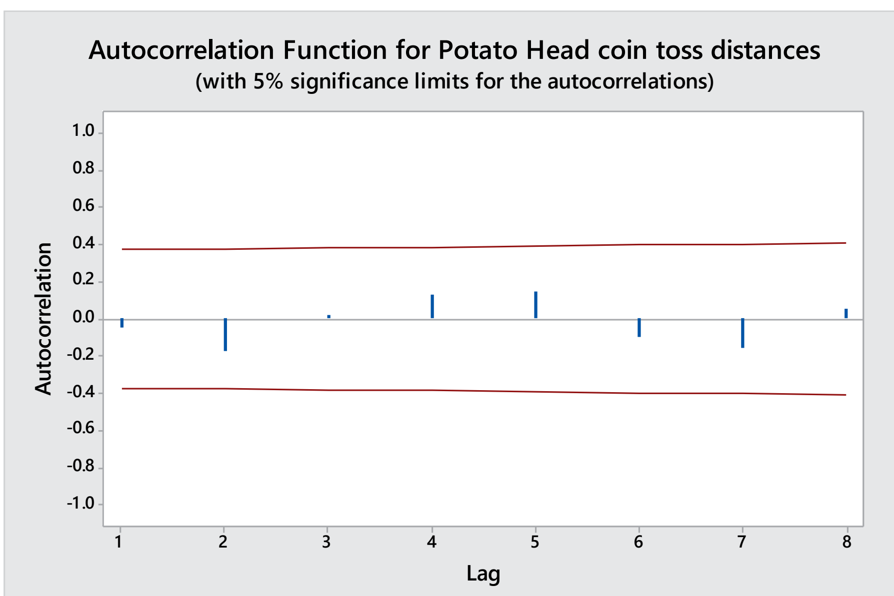


(b) True or False. Before performing a capability analysis on process data, it is necessary to make sure the data is in control first.

True False

(c) Check normality for Potato Head's process. Based on the p-value of the normality test, can you assume his data is normally distributed?

Yes **No**



(d) Check for dependency in Pareto Head's tosses. Is there dependency in his coin tosses?

Yes **No**

(e) Perform a Box-Cox Transformation on Potato Head's data to transform it into normally distributed data. Store the transformed data in another column in Minitab. What transformation (e.g., ln(X), X2, $\sqrt{X}$, etc.) are you applying to the data?

Stat > **Control Charts** > **Box-Cox Transformation**

Solution: Transformation $\mathbf{Y = ln(X)}$

(f) Run a normality test on the transformed data. Is the transformed data normally distributed according to the normality test's p-value?

Yes No

(g) Perform a capability analysis on Potato Head's data with lower and upper specifications set at 1 and 15 cm. Report the Cp and Cpk values. Make sure to apply the appropriate transformation as determined in part (e).

Stat > **Quality Tools** > **Capability Analysis** > **Normal;** your subgroup size is 1; Select Transform and choose the Box-Cox transformation

Solution: Cp = **0.51, Cpk** = **0.34.**

(h) Given the Cp value from part (g), which is larger – Voice of Customer (specification spread) or Voice of Process (process spread)?

VOC **VOP**

Since the ***Cp value is less than 1***, this means the numerator of Cp is less than the denominator of Cp. So, this means the Voice of Process ($6\hat{\sigma}$) is larger than VOC: (USL – LSL).

(i) Use Minitab's distribution identification abilities to write down just one distribution that "fits" Potato Head's ORIGINAL data (not transformed data!), where "fits" means the distribution's p-value is at least greater than 0.05. Make sure the chosen distribution is NOT a transformation (e.g., Box-Cox or Johnson transformation). Report the distribution's scale, threshold, spread, etc. parameter values as well.

Stat > **Quality Tools** > **Individual Distribution Identification**

Distribution: **Parameter values** (such as scale, spread, threshold):

Goodness of Fit Test

Distribution	AD	P	
Lognormal	0.534	0.158	**Location: 0.913, Scale: 0.957**
3-Parameter Weibull	0.683	0.081	**Shape: 0.836, Scale: 3.283, Threshold: 0.438**
Loglogistic	0.436	0.236	**Location: 0.843, Scale: 0.535**

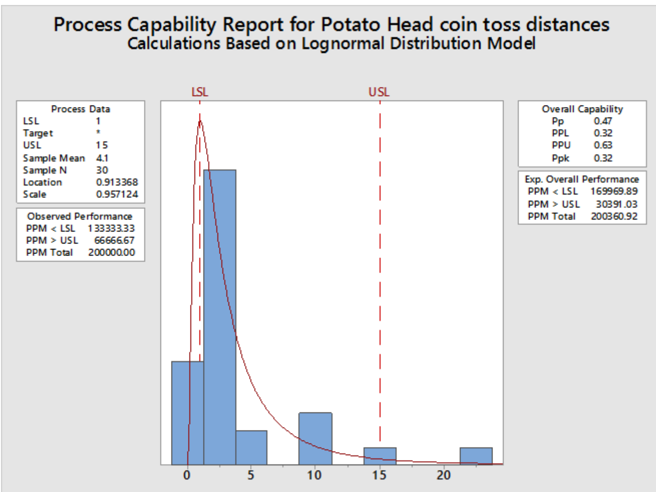
Solution: *Although the Box-Cox and Johnson Transformations show good fits, they are* ***transformations*** *and not fitting a distribution to the original data set.* We want to determine what distribution (e.g., Weibull, Lognormal, etc) fits our original data without a transformation. The "good" distributions for fitting the data WITHOUT the transformation are highlighted above: Lognormal, 3-Parameter Weibull (not great), Loglogistic.

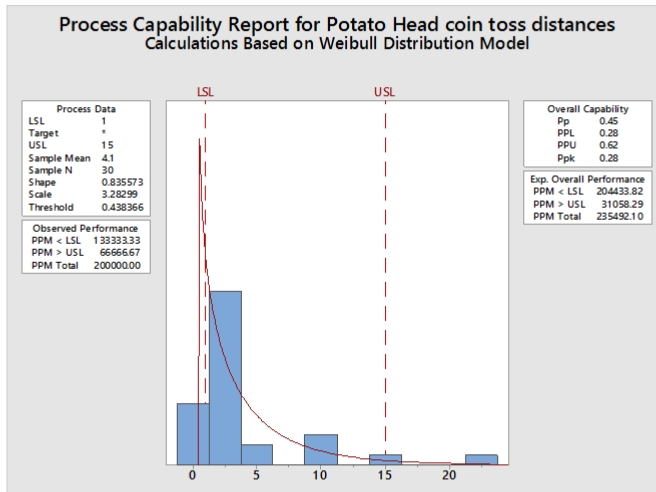
(j) Perform a Capability Analysis using the distribution you reported from part (i). The instructions for doing this are in the Lesson 16 notes.

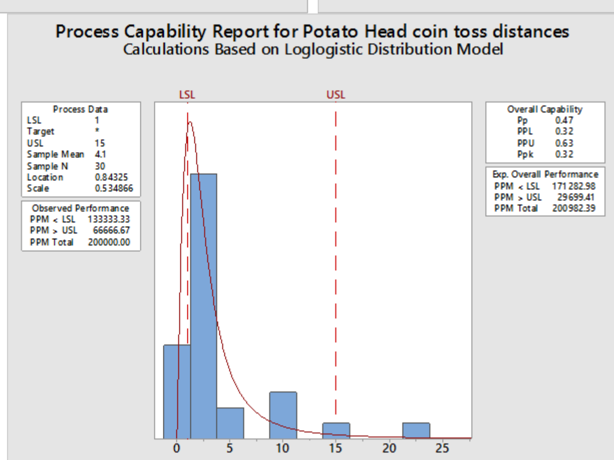
Stat > **Quality Tools** > **Capability Analysis** > **Non-normal**, and select the distribution you want from the drop down menu list. Report the Pp and Ppk values.

Pp = Ppk =

Solution:
Lognormal: Pp = 0.47, Ppk = 0.32
3-Parameter Weibull: Pp = 0.45, Ppk = 0.28
Loglogistic: Pp = 0.47, Ppk = 0.32



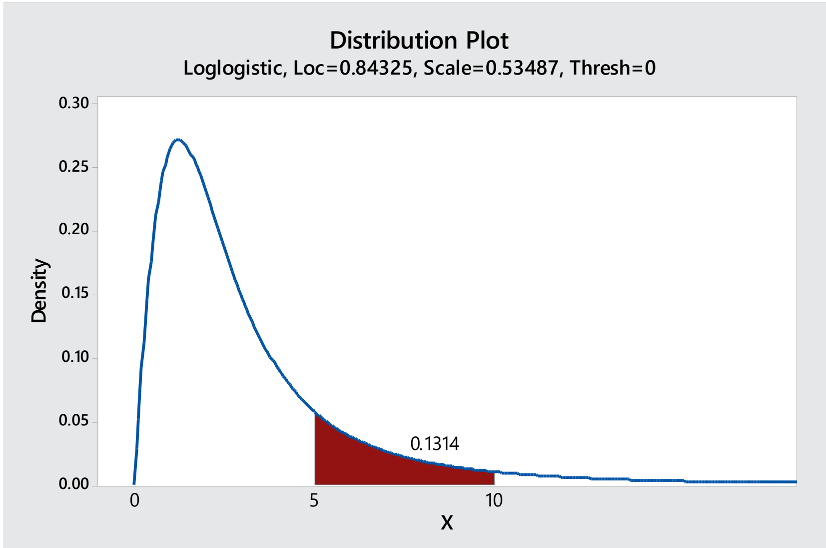


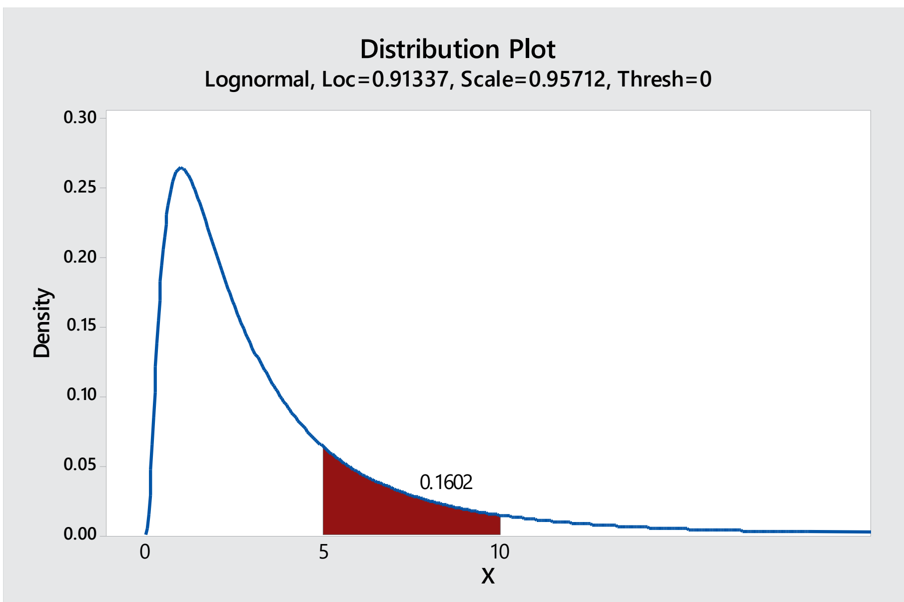


(k) Use the distribution from part (i) to determine the proportion of his throws that would land between 5 and 10 cm. You can use either of the following methods:

- Write out the expression you are integrating below and then evaluate it, or
- Use Minitab's Probability Distribution Plot to fit the desired distribution function with parameters reported from part (i). Sketch a graph below of the distribution and shade the area corresponding to the appropriate probability.

Solution: Loglogistic: 0.1314, Lognormal: 0.1602, Weibull: 0.1812





(l) Use the transformed data to determine the proportion of his throws that would land between 5 and 10 cm. Recall that the transformed data does follow a normal distribution and that the values 5 and 10 have not been transformed.

Determine the proportion correct to at least 2 decimal places.

Show your Minitab work (or sketch graph and shade proportion) to support your answer.

Solution: It makes sense that the probability matches the Lognormal distribution (above) since applying a log transformation to the normal is what a Lognormal is. Make sure that you applied the log transformation to the specification limits; i.e., LSL = ln(5.0), USL = ln(10.0)