Exam Title: Statistical Thinking for Industrial Problem Solving

Sample Questions

The following sample questions are not inclusive and do not necessarily represent all of the types of questions that comprise the exams. The questions are not designed to assess an individual's readiness to take a certification exam.

Question 1

Which tool do you use to establish a baseline level of performance for the process you are trying to improve?

A. control chart
B. process map
C. scatterplot
D. cause and effect diagram

correct answer= "A"

Question 2

You are leading a problem-solving team. Which tool helps you define the problem, identify variables you'll need to study, and set the project scope?

A. Gage R&R
B. Process Map
C. Design of Experiments
D. Hypothesis Test

correct answer= "B"

Question 3

Which two tools are useful for identifying potential root causes in a problem-solving investigation? (Choose two.)

A. The 5 Whys
B. Measurement System Analysis (MSA)
C. Ishiskawa (Cause-and-Effect) Diagram
D. Regression

correct answer= "A, C"

Question 4

Which graph shows the three characteristics of the distribution of a variable (centering, shape, and spread)?

A. Pareto plot
B. Scatterplot
C. Histogram
D. Run chart

correct answer= “C”

Question 5

In a dataset, an analyst has determined that two quantitative variables are highly correlated. Which kind of graph would effectively tell this story?

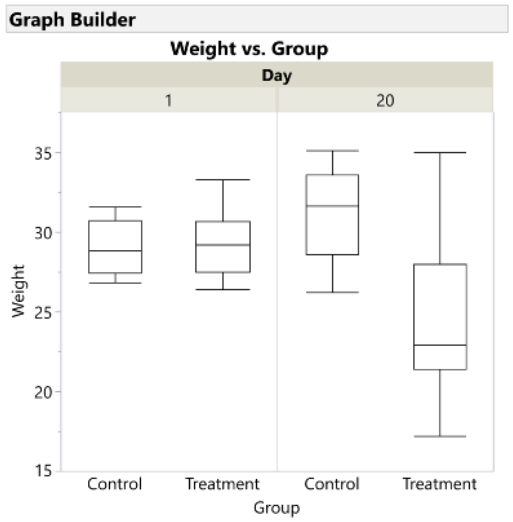
A. Histogram
B. Box plot
C. Bar chart
D. Scatterplot

correct answer= “D”

Question 6

A scientist at a biotech company is investigating the effects of a drug on the weight of mice. Mice in the treatment group received the drug, mice in the control group did not. All mice were weighed twice, once on Day 1 and again on Day 20.

Interpret this graph of the study results and determine which conclusion is correct.



A. Mice in the treatment group generally lost weight over time, but the variation in their weights increased.
B. The control group had a decrease in variation over time, and those mice gained weight on average.
C. All mice in both groups gained weight over time and the variation in their weights increased.
D. All treatment mice lost weight, while all control mice gained a small amount of weight.

correct answer= “A”

Question 7

You have generated some graphs you need to share with partners outside of your organization. The underlying data used to create the visualizations contains sensitive information and cannot be shared outside of your organization. What format should you use for sharing these results?

A. an interactive dashboard that is viewed within JMP or another application
B. static image files such as .jpg or .png
C. an interactive dashboard that is viewed in a web browser
D. a script or program that can be used to generate the graphs when applied to the data

correct answer= “B”

Question 8

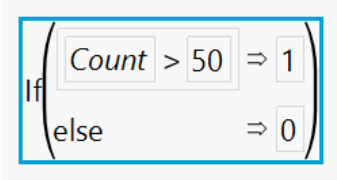
What does a row represent in a data set that is ready for analysis?

A. a variable
B. an experiment
C. an observation
D. a process

correct answer= “C”

Question 9

When you run the formula shown below, which statement is TRUE?



A. rows with a count value of more than 50 will have the value 1
B. rows with a count value of more than 50 will have a value of 0
C. rows with no count value will be 0
D. rows with a count value of 50 will be 1

correct answer= “A”

Question 10

Control limits on an X-bar control chart are calculated using which two numbers? (Choose two.)

A. an estimate of the overall standard deviation
B. an estimate of the customer specifications
C. an estimate of the grand mean
D. an estimate of the within subgroup standard deviation

correct answer= “C,D”

Question 11

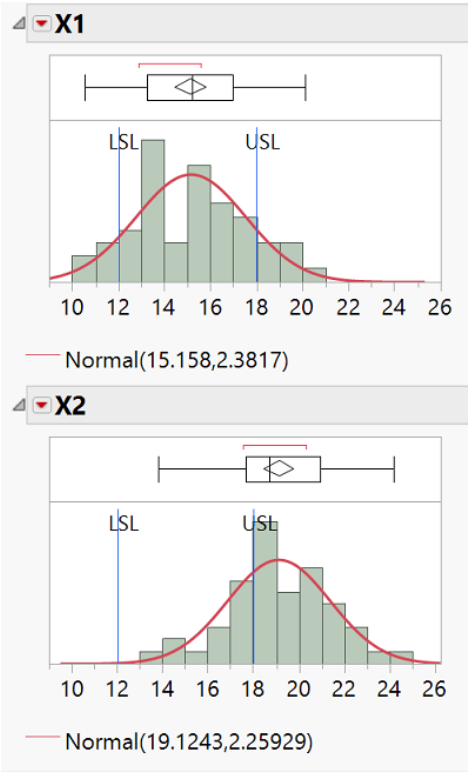
An operator randomly selected and measured the diameters of shafts. Six consecutive parts were measured per day. This was repeated for 30 days. Which control chart should be created with these data?

A. Run chart
B. I and MR chart
C. p chart
D. X-bar and R chart

correct answer= “D”

Question 12

Below is the process capability output from two processes, X1 and X2. Which statement is TRUE?

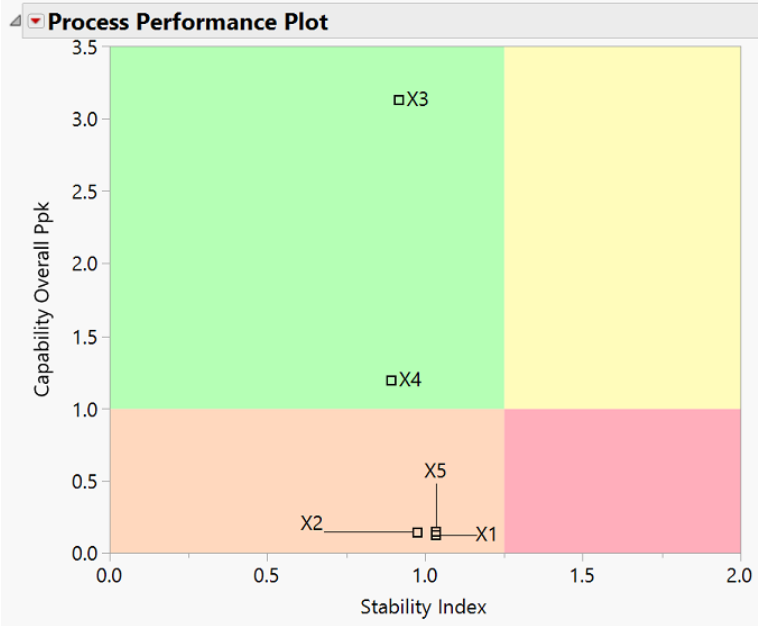


A. Cpk is better for X1.
B. Cpk is better for X2.
C. Cp is better for X1.
D. Cp is better for X2.

correct answer= “A”

Question 13

Based on the graph below, which two processes are both capable and stable? (Choose two.)



A. X1
B. X2
C. X3
D. X4
E. X5

correct answer= “C,D”

Question 14

What are two characteristics that can be evaluated in a measurement system analysis? (Choose two.)

A. repeatability
B. homogeneity
C. bias
D. Sampling error

correct answer= “A,C”

Question 15

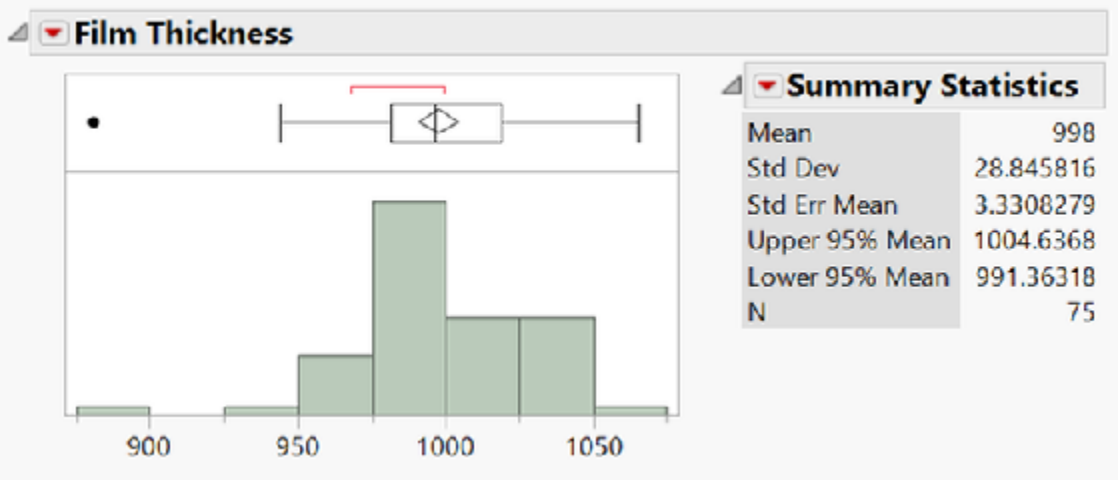
For measurement system analysis, match the following terms with their descriptions.

Term		Definition
A. Stability	_____	There is an increase in measurement variation across the operational range of the parts under study.
B. Nonlinearity	_____	The ability to obtain consistent measurements when testing is performed by different operators on the same parts using the same gauge.
C. Reproducibility	_____	The difference between gauge measurements and the true (or standard) value.
D. Bias	_____	The gauge yields consistent measurements when the same parts are measured over time.

correct answer= “B, C, D, A”

Question 16

An engineer measures the film thickness of 75 wafers. Based on the distribution and the summary statistics shown below, what can you say about the true population mean of film thickness values?



A. The mean of the population is 998.
B. You are 95% confident the true mean of the population is between 991 and 1004.
C. You are 95% confident the true mean of the population is between 918 and 1082.
D. There is not enough data to estimate the mean of the population.

correct answer= “B”

Question 17

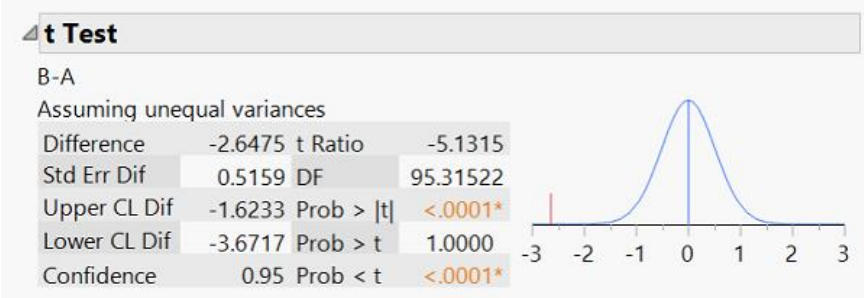
Which of the statements is true about p-values when testing for the difference between two means at a significance level of 0.05?

A. If the p-value is less than 0.05, there is a statistical difference in the means.
B. If the p-value is greater than 0.05, there is a statistical difference in the means.
C. If the p-value is less than 0.05, there is not a statistical difference in the means.
D. If the p-value is less than 0.05, the means are statistically equal.

correct answer= “A”

Question 18

The team performed a two sample t-test on the shipping time from two vendors. Based on this analysis, what can you conclude?



A. The means are equal.
B. The means are significantly different.
C. The means are not significantly different.
D. Vendor A is better.

correct answer= “B”

Question 19

A scientist creates two new formulations of a storage buffer and would like to determine which one is most effective at increasing the optical density (OD) measured in a bioassay.

Given:

- The estimated standard deviation for the bioassay is 0.10 OD units.
- The scientist only wants to run 6 replicates per formulation.
- The acceptable alpha risk is 0.05 and the power is 0.80.

Note:

- You may use Sample Size and Power platform under DOE > Design Diagnostics in JMP to answer this question.

How large does the difference in ODs between the bioassays need to be for the formulations to be considered different from one another?

A. 0.307 OD units
B. 0.180 OD units
C. 0.071 OD units
D. 0.117 OD units

correct answer= “B”

Question 20

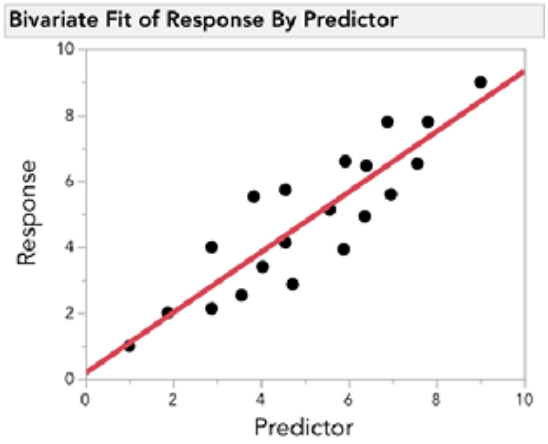
Which of the statements is true?

A. If you find correlations, then you understand cause and effect.
B. Correlation is when one outcome affects another outcome.
C. Correlation and causation are two words for the same concept.
D. Correlation between two variables does not imply causation.

correct answer= “D”

Question 21

You fit a simple linear regression model. Which statement is true?



Linear Fit

Response = 0.1776578 + 0.9152991*Predictor

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	69.361119	69.3611	64.1016
Error	18	19.476881	1.0820	Prob > F
C. Total	19	88.838000		<.0001*

Parameter Estimates

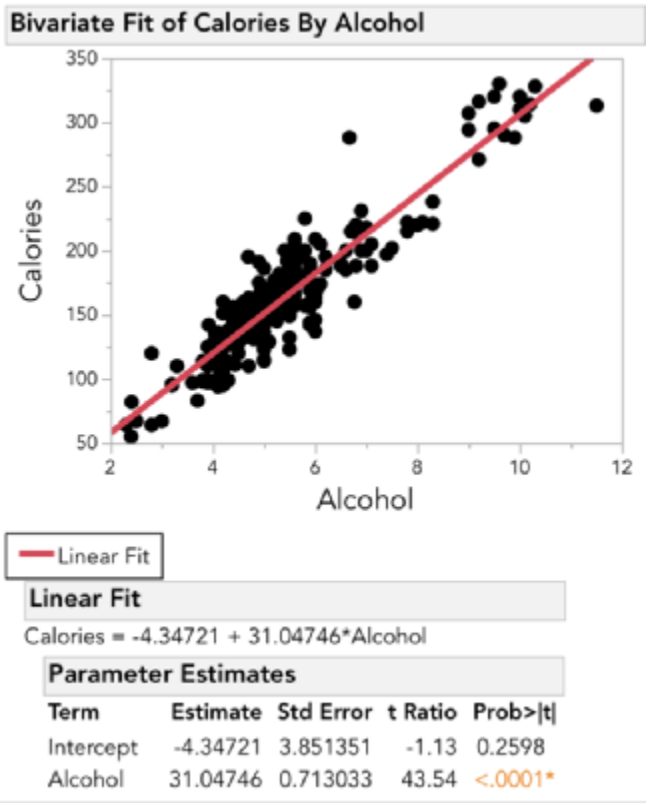
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.1776578	0.628999	0.28	0.7808
Predictor	0.9152991	0.114322	8.01	<.0001*

A. There is no relationship between the response and the predictor.
B. There is curvature in the relationship between the response and the predictor.
C. There is a significant linear relationship between the response and the predictor.
D. As the predictor increases, the response decreases.

correct answer= “C”

Question 22

You collect data on the percent alcohol and calories for 340 commercial beers (at a standard volume). Use the regression formula below to calculate the number of calories for a beer with 8.2% alcohol. ***NOTE:*** *You may use a calculator for this question.*



A. 150
B. 250
C. 300
D. 325

correct answer= “B”

Question 23

When do you use multiple linear regression?

A. when you want to model the relationship between a categorical response and more than one predictor
B. when you want to understand correlations between a collection of variables
C. when you want to model the relationship between a continuous response and more than one predictor
D. when you have retrospective data and want to establish causation

correct answer= “C”

Question 24

You fit a linear regression model for Yield as a function of several predictors. What can you conclude from this model?

Summary of Fit

RSquare	0.65262
RSquare Adj	0.560232
Root Mean Square Error	2.296903
Mean of Response	91.8378
Observations (or Sum Wgts)	120

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	25	931.6842	37.2674	7.0639
Error	94	495.9218	5.2758	Prob > F
C. Total	119	1427.6060		<.0001*

A. The model does not explain a significant amount of the variation in Yield.
B. The model explains 91.8% of the variation in **Yield**.
C. You have identified the root cause of the variation in **Yield**.
D. There is a significant relationship between at least one of the predictors and **Yield**.

correct answer= “D”

Question 25

Which statement about logistic regression models is TRUE?

A. Logistic regression models can have only a continuous variable as the response.
B. Logistic regression models can have only categorical variables as predictors.
C. Logistic regression models can have only a single predictor.
D. Logistic regression models can have only a categorical variable as the response.

correct answer= “D”

Question 26

What are two advantages of using statistically designed experiments (DoE) over One Factor at a Time (OFAT) experiments? (Choose two.)

A. DoE allows you to change many factors at a time.
B. DoE is less efficient than OFAT.
C. DoE allows you to study interaction effects.
D. DoE allows you to study fewer factors than OFAT.

correct answer= “A,C”

Question 27

You are the Operations Manager of a production facility. You would like to run a DoE on your process to study 3 factors at two levels each using a full factorial experiment. How many runs will you execute?

A. 6
B. 8
C. 9
D. 12

correct answer= “B”

Question 28

What is the goal of predictive modeling?

A. to identify the most important predictors
B. to describe what has happened in the past
C. to predict what might happen next
D. to estimate model parameters

correct answer= “C”

Question 29

A model is developed to accurately classify defective parts. Match the following terms with their definitions.

	Term		Definition
A.	False Negative	_____	A part is defective, and it is classified as defective.
B.	True Positive	_____	A part is good, and it is classified as good.
C.	True Negative	_____	A part is good, and it is classified as defective.
D.	False Positive	_____	A part is defective, but it is classified as good.

correct answer= “B, C, D, A”

Sample Scenario

The Statistical Thinking for Industrial Problem Solving exam has a practical section on the exam where you will be asked to work with sample data and work in JMP to answer the test questions. Below is a sample of the type of scenario you could see on the exam.

The data table used in this sample scenario, ***forming-yield.jmp****, can be downloaded using this* ***link****. If you are enrolled in the Statistical Thinking for Industrial Problem Solving (STIPS) course, you can also open the file in the Virtual Lab in the course.*

Question 30

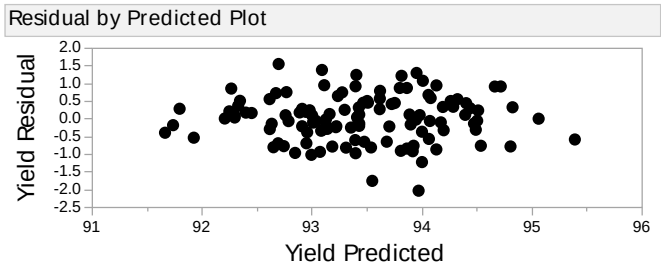
You are studying the yield of a forming process.

In JMP, open the file **forming-yield.jmp**. Fit a multiple linear regression model to **Yield**, with all of the factors and their two-way interactions.

1. Describe the pattern in the Residual by Predicted plot?
   a. There is curvature in the residuals
   b. There is heteroskedasticity in the residual plot
   c. There is a severe outlier that is tilting the regression line
   d. The residuals look randomly scattered around the center line of zero

   Correct answer = d

   (Rationale: There is no obvious pattern in the residual plot. The points are randomly scattered around the center line.)

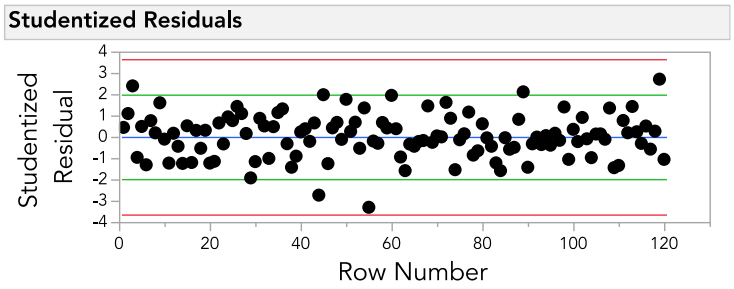


2. Look at the Studentized Residual plot. What do you learn from this plot?

   a. There are several outliers
   b. There are no outliers
   c. The response is normally distributed
   d. There is evidence of multicollinearity

   Correct answer = b

   (Rationale: None of the points are outside the red lines in the plot, so there are no outliers.)



Externally studentized residuals with 95% simultaneous limits (Bonferroni) in red, individual limits in green.

3. Right-click over the Parameter Estimates table and select Columns > VIF. What do you learn from the VIF values?

   a. There is a serious problem with multicollinearity
   b. Multicollinearity is not a serious problem
   c. There is a highly influential observation
   d. There is heteroskedasticity in the residuals

   Correct answer = b

   (Rationale: None of the VIFs are large (they are all less than 5), so multicollinearity is not a serious problem.)

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	98.483102	1.380285	71.35	<.0001*	.
Stirrer Speed	-0.013934	0.009589	-1.45	0.1492	1.091029
Modifer %	0.6585682	0.069337	9.50	<.0001*	1.0918066
Filler %	-0.346483	0.073012	-4.75	<.0001*	1.1571967
Viscosity	0.0115598	0.031612	0.37	0.7153	1.2159451
Catalyst[Catalyst A]	-0.001283	0.065071	-0.02	0.9843	1.0572626
(Stirrer Speed-64.8192)*(Modifer %-1.63383)	-0.022892	0.011162	-2.05	0.0428*	1.1890087
(Stirrer Speed-64.8192)*(Filler %-15.5217)	0.0251179	0.011077	2.27	0.0254*	1.3146062
(Stirrer Speed-64.8192)*(Viscosity-11.8938)	0.014594	0.004832	3.02	0.0032*	1.3443541
(Stirrer Speed-64.8192)*Catalyst[Catalyst A]	0.0066039	0.01034	0.64	0.5244	1.2500169
(Modifer %-1.63383)*(Filler %-15.5217)	0.0104746	0.081437	0.13	0.8979	1.2481704
(Modifer %-1.63383)*(Viscosity-11.8938)	0.0324526	0.034034	0.95	0.3425	1.1673591
(Modifer %-1.63383)*Catalyst[Catalyst A]	-0.002186	0.06978	-0.03	0.9751	1.1058097
(Filler %-15.5217)*(Viscosity-11.8938)	-0.008385	0.036419	-0.23	0.8184	1.3706597
(Filler %-15.5217)*Catalyst[Catalyst A]	-0.069765	0.076384	-0.91	0.3632	1.266315
(Viscosity-11.8938)*Catalyst[Catalyst A]	-0.020434	0.032595	-0.63	0.5321	1.2844769

4. Look at the Effect Summary table. Which term is the most significant (given the other terms in the model)?

   a. Modifier %
   b. Stirrer Speed
   c. The intercept
   d. Viscosity

   Correct answer = a

(Rationale: Modifier % has the lowest p-value (0.00000). Filler % also has a p-value of 0.00000, but Modifier % has the largest LogWorth 15.051.)

Effect Summary

Source	LogWorth	PValue
Modifer %	15.051	0.00000
Filler %	5.177	0.00001
Stirrer Speed*Viscosity	2.498	0.00318
Stirrer Speed*Filler %	1.595	0.02542
Stirrer Speed*Modifer %	1.369	0.04279
Stirrer Speed	0.826	0.14923 ^
Modifer %*Viscosity	0.465	0.34252
Filler %*Catalyst	0.440	0.36317
Stirrer Speed*Catalyst	0.280	0.52444
Viscosity*Catalyst	0.274	0.53209
Viscosity	0.145	0.71535 ^
Filler %*Viscosity	0.087	0.81837
Modifer %*Filler %	0.047	0.89790
Modifer %*Catalyst	0.011	0.97506
Catalyst	0.007	0.98431 ^

5. Which interaction term is significant (given the other terms in the model)?

   a. Filler %*Viscosity
   b. Modifier %*Viscosity
   c. Stirrer Speed * Filler %
   d. Modifier % * Filler %

   Correct answer = c

   (Rationale: Stirrer Speed*Filler % has the lowest p-value (0.02542) of the interactions listed.)

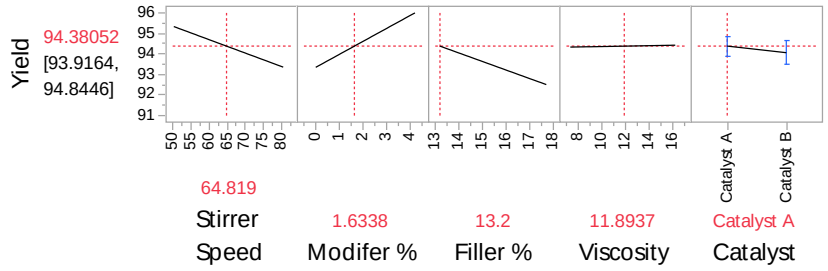
6. Use the Prediction Profiler to explore and interpret the model coefficients. Change the value of Filler % from the low value (13.2) to the high value (17.7). What do you observe?

   a. The slope of the line for Stirrer Speed changes from negative to positive
   b. The slope of the line for Modifier % changes from positive to negative
   c. The slope of the line for Viscosity changes from negative to positive
   d. There are no interactions between Filler % and the other predictors

   Correct answer = a

   (Rationale: The interaction between Stirrer Speed and Filler % is significant, so as the value of Filler % is change the slope of the line for Stirrer Speed changes.)

Prediction Profiler



Prediction Profiler