JMP Certified Associate:

Statistical Thinking for Industrial Problem Solving

Section 1: Statistical Thinking and Problem Solving: 10%

Define statistical thinking and problem solving

- Explain the benefits of statistical thinking
- List the principles of statistical thinking
- Describe the importance of following a structured approach to problem solving
- Explain the importance of a well-written problem definition
- Identify and define KPIs (Key Performance Indicators)
    - Define what success looks like and determine how to measure for success and improvement
    - Describe the importance of establishing a baseline for your KPIs using
    - Establish a baseline for your KPIs using control charts

Define the process for creation of a process map or SIPOC

- Interpret a process map
    - Explain the importance of a process map in understanding the process and scoping the problem
    - Define the parts of a process map
- Interpret a SIPOC model (Supplier, Input, Process, Output, Customer)
    - Explain why you would do a SIPOC

Identify root causes of a problem, compile and collect data

- Develop and interpret a cause and effect (C&E) diagram
    - Identify the components of a simple C&E diagram
    - Given a completed C&E diagram, determine action to take based on the completed diagram.
- Define rules for effective brainstorming
    - Explain methods for brainstorming and methods for gaining consensus.
    - Explain how Affinity diagrams are used for organizing output from a brainstorming session
- Define operational definitions and why they are important
    - Given a particular KPI, identify the appropriate operational definition for it
- Describe data types and modeling types

- Given a problem, be able to create a data collection plan
    - Determine the data you need to collect
    - Determine if there is historical data available that can be used
        - Determine locations/sources of existing data
    - Given a project, determine what other type of data are needed
        - Describe the benefits/risks of using different types of data (observational, retrospective/historical, experimental)
        - Describe different sampling strategies for collecting data

Section 2: Exploratory Data Analysis: 20%

Describe data using descriptive statistics and graphical summaries

- Describe measures and graphs for summarizing data
    - Interpret summary statistics
    - Generate and interpret: histograms, box plots, comparative box plots, run charts, and Pareto plots
        - Describe when to use each type of plot
- Distinguish between population vs sample, distinguish between parameter vs statistic
- Recognize the importance of probability distributions
    - Describe characteristics of the normal distribution.
    - Explain the central limit theorem
    - Use a normal quantile plot to determine if the underlying distribution is approximately normal.

Visualize and explore data

- View relationships between variables
    - using scatterplots
    - using box plots
    - using bar charts
- Create graphs to visualize many variables at a time
    - using heat maps
    - using multivari (variability) charts
    - using trellis plots and overlay variables
    - using geographic maps
- Describe the characteristics of a good data visualization
- Explore subsets of your data using data filtering
- Communicate the message in your data
- Identify your target audience
- Translate analytic results into practical/actionable results
    - Determine the best visualization to tell a story
    - Given a particular visualization, describe the story in the data in the simplest terms

Save and share results

- Produce high quality graphics
  - Edit and customize graphs
- Save work for reproducibility and reuse
  - Manage and organize your work (datasets and output)
  - Save scripts to data tables
- Select the appropriate method to share your work based on your intended audience
  - HTML w/data, PowerPoint, scripts, dashboards, copy and paste images with selection tool

Prepare data for analysis

- Structure data appropriately for analysis
  - Describe a data table (observations and variables)
  - Determine if data needs to be restructured for analysis
  - Describe restructuring techniques (split, stack, concatenate, sort, and join)
- Identify data quality issues such as missing or messy data, outliers, and redundant variables
  - Define the possible data quality issues
  - Recognize data quality issues
  - Determine and recommend actions to take to address data quality issues
- Group and bin data and create new variables
  - Describe need for derived variables
  - Demonstrate when recoding variables might be helpful
  - Create new columns with formulas

Section 3: Quality Methods: 20%

Use Statistical Process Control Charts

- Explain how control charts are used
  - Describe structure of control charts
    - Interpret the elements of a control chart
  - Explain how to use control charts for process improvement
  - Define special and common cause variation
  - Describe tampering
- Create simple charts, I and MR, X-bar and R or S, 3-way charts
  - Describe differences between I and MR charts and X-bar R or S charts
  - Describe how and why to use rational subgrouping
  - Use phase variables for exploratory analysis
- Use and interpret tests for special causes
  - Interpret control charts and determine if they are stable
  - Add tests for special causes to the charts, determine which tests to add
- Use control charts to determine if improvements have been sustained

Assess process capability

- Explain the difference between voice of the process to voice of the customer
  - Describe Cp, Cpk, Pp and Ppk
  - Define process capability
  - Compare and contrast control limits and spec limits
  - Compare and contrast capability to stability
- Interpret capability indices
  - Check the assumptions for interpreting capability indices
  - Relate capability measures to ppm defective
  - Interpret graphics and indices for non-normal data
  - Interpret graphics and indices for many variables at one time
    - Goal plot
    - Process performance graph
  - Use capability indices (including stability index and target index) and plots to determine if a problem is present and the nature of the problem
    - Instability
    - Off target
  - Determine actions to take based on capability measures

Conduct Measurement System Studies

- Define the goal/objective of (Measurement System Analysis), determine language and define terminology to be used:
    - Explain importance of evaluating your measurement system
    - Define terms associated with MSAs
        - repeatability, reproducibility, stability, bias, linearity
    - Describe the process of conducting an MSA
    - Design an MSA
- Utilize variability (multivari) charts to graphically evaluate the data from your MSA
    - Calculate common metrics for evaluating the measurement system
- Interpret variance components output to determine the next actions to take from your MSA
- Interpret Gauge R&R analysis results
- Examine EMP (Evaluate Measurement Process) output
    - Interpret the EMP output to determine if your measurement system is fit for use or if it needs improvement
        - Using average and range charts
        - Using parallelism plots

Section 4: Decision Making with Data: 15%

Define/describe statistical intervals

- Contrast the different types of intervals: confidence, tolerance, and prediction
- Explain the need for statistical intervals
    - Compare point estimation to statistical intervals
- Explain when to use each type of interval
- Interpret and explain each type of interval
- Describe the effect of sample size on statistical intervals

Perform statistical tests

- Compare confidence intervals with hypothesis tests
- Perform a test for one mean
- Perform a test for two or more means
    - Perform, interpret and explain multiple comparison tests
- Interpret and explain output from means tests
- Compare statistical significance with practical significance

Determine sample size

- Define statistical power and alpha risk
- Explain the factors that influence power and sample size calculations
- Determine sample size for a given power (and s, and delta and alpha risk)
    - Calculate sample size for one and two sample mean problems
- Determine the power for a given sample size (and s, delta, and alpha risk)

Section 5: Correlation and Regression: 20%

Be able to define and use correlation

- Use correlation to measure relationships between continuous variables
    - Describe the strength of a linear association using correlation. Provide a definition of correlation.
    - Interpret a correlation matrix and scatterplot matrix
    - Interpret a density ellipse
- Describe common errors in interpreting correlations
    - Explain the difference between correlation and causation.
        - Lurking variables
- Describe the impact of outliers on correlation.
- Explain why correlation does not apply when there is curvature.

Define simple linear regression

- Explain the difference between correlation and regression.
- Explain and interpret the coefficients in the regression model.
- Describe the method of least squares.
- Evaluate regression model assumptions.
    - List the model assumptions and use a residual analysis to determine if model assumptions are met.
    - Explain what might be causing particular patterns in residual plots (curvature, heteroscedasticity, outliers).
- Interpret regression analysis results.
    - Perform a regression analysis.
    - Determine whether the regression model is significant.
    - Interpret the slope coefficient.
    - Interpret the confidence interval for the slope coefficient.
- Fit regression models with curvature
    - Fit a model with a quadratic effect and interpret the results.

Define/Explain Multiple Linear Regression

- Fit multiple linear regression models
    - Describe when multiple linear regression is appropriate.
    - Explain the difference between SLR and MLR.
- Compare/contrast the goals of explanatory and predictive modeling.
- Interpret the multiple linear regression model
    - Interpret the coefficients in a regression model (continuous and categorical)
    - In JMP, use the prediction profiler to interpret model coefficients and interactions.

- Be able to interpret the output of a multiple regression model
- Verify model assumptions
    - Conduct a residual analysis and determine if regression assumptions have been met.
- Identify influential observations
    - Describe the impact of influential observations
    - Interpret Cook's D values
- Explain/describe/define collinearity
    - Interpret the VIF
    - Describe the impact of collinearity
- Perform variable selection
    - Define variable selection and explain the need for variable selection
    - List methods of model variable selection (forward, backward, mixed)
    - Use the Effect Summary table in JMP to reduce the model

Define/explain logistic regression

- Explain the difference between logistic regression and multiple linear regression.
    - Describe when logistic regression is appropriate.
    - Explain how logistic regression works.
- Fit a logistic regression model and interpret results.
    - Interpret the coefficients in a logistic regression model (continuous and categorical)
    - Interpret a logistic curve
    - Use the prediction profiler to interpret model coefficients and understand interaction effects.
- Identify potential problems with the logistic regression model such as:
    - Separation
    - Not enough data

Section 6: Design of Experiments: 10%

Define DOE and compare it to OFAT

- Determine the appropriate circumstances for using DOE.
    - Explain the difference between DOE and OFAT
- Explain why you need to do a DOE.
- Describe the importance of a well-designed experiment.

Discuss best practices and considerations to take when running an experiment

- Determine what you want to measure.
- Explain the importance of randomization and the consequences of not doing it
- Explain the trade-offs between budget and the complexity of the experiment
- Determine the goals of your DOE

Analyze the results of an experiment

- Analyze the results of a factorial experiment
    - Interpret analysis results
    - Identify significant model effects
    - Use the Prediction Profiler to explore your model

Design a simple experiment

- Define the following terms: factor, level, treatment, replicate
- Describe a full factorial experiment
- Design a 2^k full factorial experiment in JMP

Section 7: Predictive Modeling and Text Mining: 5%

Explain the difference between explanatory modeling versus predictive modeling

- Explain overfitting
- Describe model validation

Interpret results from a predictive model with validation

Interpret results from a multiple linear model with validation**.**

- Interpret RMSE (RASE)

Interpret results from a logistic model with validation.

- Interpret the misclassification rate (error)
- Interpret the confusion matrix
- Define the following terms: False positive, false negative, true positive, true negative

Interpret decision tree models.

- Describe a decision (partition) tree model
- Describe the difference between a classification tree and a regression tree.

Describe neural network models.

- Describe the basic structure of a neural network (input layer, hidden layer, output layer).