JMP Instructional data set 033

Semiconductor Clean Room Study

STATISTICAL DISCOVERY

JMP platforms and statistical methods

Distribution:	Histogram; box plots; summary statistics
Graph Builder:	Scatter plots; comparative dot plots; simple linear regression; confidence intervals; bar chart
Formula Editor:	Creating a new variable
Multivariate:	Scatterplot matrix; correlation
Fit Y by X:	Simple linear regression, one factor ANOVA
Fit Model:	Multivariable regression

Problem statement

Semiconductor manufacturing requires very clean environments to help prevent small particles in the air from contaminating the wafers. Though all the workers use clean suits and follow certain procedures before entering a clean room, the presence and activity of people in those rooms are known to be a major source of these unwanted particles.

To better understand how human activity in a clean room affects the number of particles present, data has been collected for the past 45 days measuring the number of particles detected in a clean room under a variety of conditions. The study was conducted during typical operations so it was not possible to deliberately choose the number of technicians that would be entering/exiting the lab and how frequently, as well as how much activity (amount of movement technicians engage in) when in the lab. Two experimental factors were chosen for the study (type of clean suit and a pre-entrance cleaning procedure). These two factors were able to be deliberately changed. Each day the technicians were instructed to wear one of the three different types of clean suits and to perform or not perform a particular pre-entrance cleaning, which included face washing and hair brushing, among other steps to shed potential particles.

A set of measurement instrumentation was installed throughout the room positioned at two heights (ground and head level). The instruments take air samples in multiple locations throughout the room across various points in the day. The total amount of particles counted in aggregate throughout the day will be used as the response data.

Data

IDS_33.jmp

Day	Day of data collection (1, 2, … , 45)
Technicians	Number of technicians that were in the clean room that day
Person Hours	Total hours technicians were in the clean room
Entrance/Exits	Number of times technicians entered or exited the clean room
Suit	One of three different clean suit types all the technicians wore on that given day (A, B, C)
Pre-entrance Clean	Whether or not all technicians followed a specific pre-entrance cleaning process that day (Yes, No)
Activity	One of three levels indicating the amount of activity (i.e., movement of the technicians, materials, equipment, etc.) that occurred that day as reported by the lab manager (Low, Med, High)
Particles_Ground	Aggregate number of airborne particles measured at ground level
Particles_Head	Aggregate number of airborne particles measured at head level

Exercises

The exercises consist of creating a variety of different visualizations and conducting formal statistical analyses (Regression and ANOVA) to assess the potential effects that various conditions have on the number of particles present.

1. Graphically and numerically summarize the two response variables ('Particles_Ground' and 'Particles_Head') to get some initial understanding of the data.

> ***Instructions:***
> *Analyze > Distribution. Place the variables 'Particles_Ground' and 'Particles_Head' in the Y, Columns role. Click OK.*

2. Examine how the number of particles measured at the ground level and head level compare by creating two graphs: a scatterplot of ‘Particles_Ground’ vs ‘Particles_Head’and a bar chart of the difference between ‘Particles_Ground’ and ‘Particles_Head’ over time. Describe what these graphs convey about the data.

> ***Instructions:***
> *For the scatterplot. Graph > Graph Builder. Place the variable ‘Particles_Ground’ on the Y axis, and ‘Particles_Head’ on the X. Remove the smoother by unselecting it in the graph palette.*
>
> *To visualize the amount of agreement between the two, add a diagonal reference line at y=x. Do this by right-clicking on the graph and selecting ‘Customize’. Select* **+** *at the top. Under ‘Templates’, select ‘Y Function’. Edit the script to:* `Y Function(y=x, x);`
>
> *Click OK.*
>
> *For the bar chart of the difference. First create a new column that is the difference between ‘Particles_Ground’ and ‘Particles_Head’. Do this by highlighting both columns. Right-click and choose New Formula Column > Combine > Difference. Graph > Graph Builder, place this difference column on the Y axis, ‘Day’ on the X. Choose the Bar Chart symbol from the palette.*

3. Create a graph that can display either of the particle count variables on the Y axis vs. any of the three numeric candidate explanatory variables on the X axis. Describe some of the features this graph reveals about the possible effect that these numeric explanatory variables have on the number of particles.

> *Instructions: Graph > Graph Builder. Place the variable ‘Particles_Ground’ on the Y axis, and ‘Technicians’ on the X. Choose to add a linear regression line to the plot by selecting Fit Line from the graph palette.*
>
> *Choose to add the Equation to the graph by selecting those options in the controls on the left-hand side.*
> *To setup the Column Switcher for the outcome variables, select Redo > Column Switcher under the red triangle. Select ‘Particles_Ground’ for the Initial Column to Switch and then select both ‘Particles_Ground’ and ‘Particles_Head’ as the Replacement Columns. Now either variable can be chosen for the Y axis.*
> *To setup the Column Switcher for the explanatory variable, follow the same steps as above but use ‘Technicians’ for the Initial Column to Switch and ‘Technicians’, ‘Person Hours’, and ‘Entrance/Exits’ as the Replacement Columns. Now any of these three variables can be chosen for the X axis.*

4. Create a graph that can display either of the particle count variables on the Y axis vs. any of the three categorical candidate explanatory factors on the X axis. Describe some of the features this graph reveals about the possible effect that these categorical explanatory factors have on the number of particles.

> ***Instructions:***
> *Graph > Graph Builder. Place the variable 'Particles_Ground' on the Y axis, and 'Suit' on the X. Choose to add the mean and confidence intervals for the means on the graph by selecting Fit Line from the graph palette.* 
>
> *Choose to add the Means and Standard Deviations to the graph by selecting those options in the controls on the left-hand side.*
> *Follow the same steps as in Exercise 3 to set up the column switchers.*

5. Build separate simple linear regression models for each of the response variables vs. each of the three numeric candidate explanatory variables. Interpret the results of these analyses. Is there statistically significant evidence indicating these variables could influence the number of particles? Describe those effects. Are your conclusions any different from you concluded in observing just the graphs in Exercise 3?

> ***Instructions:***
> *Analyze > Fit Y by X. Choose the variables 'Particles_Ground' and 'Particles_Head' for the Y, Response role. Choose 'Technicians', 'Person Hours', and 'Entrance/Exits' as the X, Factors. The initial output shows scatterplots for each combination. To display more complete details for a simple linear regression analysis, select Fit Line under the red triangle for each.*

6. Build two separate multivariable linear regression models, one for each outcome variable vs. the three numerical explanatory variables. Examine the p-values testing the significance of each explanatory variable on the responses. How are these results different than what you observed in the three separate simple linear regression models from Exercise 5?
Note: You will be asked to explain why this is happening in Exercise 7.

> ***Instructions:***
> *Analyze > Fit Model. Choose the variable 'Particles_Ground' for the Y role. Choose 'Technicians', 'Person Hours', and 'Entrance/Exits' for the Model Effects. Choose Minimal Report in the Emphasis drop down.*
> *Repeat the same steps using 'Particles_Head' for the Y.*

7. Create a Scatterplot Matrix and Correlation Table of the outcome variables and the three numerical explanatory variables. Examine the scatterplots and the correlations of the explanatory variables. What does this reveal about those variables? Based upon this, explain the phenomenon you observed when comparing the results of the simple linear regression models in Exercise 5 to the multivariable regression models in Exercise 6.

   Instructions:
   Analyze > Multivariate Methods > Multivariate. Choose 'Particles_Ground' , 'Technicians', 'Person Hours', and 'Entrance/Exits' for the Y, Columns. Select Fit Line from the Red Triangle next to the name Scatterplot Matrix.
   Do the same analysis using the other response variable 'Particles_Head'.

8. Build separate One Factor ANOVA models for each of the response variables vs. each of the three categorical candidate explanatory factors. Interpret the results of these analyses. Is there statistically significant evidence indicating these factors influence the number of particles? Describe those effects. Are your conclusions different from what you concluded in observing just the graphs in Exercise 4?

   Instructions:
   Analyze > Fit Y by X. Choose the variables 'Particles_Ground' and 'Particles_Head' for the Y, Response role. Choose 'Suit', 'Pre-Entrance Clean, and 'Activity' as the X, Factors. The initial output shows comparative dot plots for each combination. To display more details for the One factor ANOVA model, select Means/ANOVA under the red triangle for each.

9. Build a final multivariable linear model for each outcome variable containing only the variables and factors that are deemed to have a statistically significant impact on the outcomes. Conduct model diagnostics to see if any model assumptions are being violated.

Instructions:
Analyze > Fit Y by X. Choose the variable ‘Particles_Ground’ for the Y. Highlight all the candidate numeric and categorial explanatory variables/factors in the column list on the left. Click on the button Macro and then Factorial to Degree. Note: Make sure the Degree is specified as “2”. The Model Effects should have 6 main effects and 15 2-way interactions. Choose Minimal Report for the Emphasis. Click Run.
Examine the Effects Summary Table. Begin removing the non-significant 2-way interactions (one at a time) in order of largest p-value to smallest until only those that have a p-values < 0.05 remain. Repeat this process with the main effects. Note: if a factor is contained in a significant 2-way interaction, you must retain the main effect, even if has a p-value > 0.05.
Perform model diagnostics. (Row Diagnostics > Residuals vs Predicted and Row Diagnostics > Residuals by Normal Quantiles under the red triangle).
Produce visualizations describing the final model. (Row Diagnostics > Plot Actual by Predicted ; Row Diagnostics > Plot Regression ; Factor Profiling > Profiler under the red triangle.
Repeat this process for the ‘Particles_Head’ outcome variable.

10. Use these final models to describe the effects that the significant variables have on the amount of particles. For example, how much do the particles change for each unit change in the numerical explanatory variables? How much do the particles change at different levels of the categorial variables. Illustrate this by choosing a few different conditions and showing the predicted value and a 95% confidence interval for the average amount of particles that would be produced under these conditions. Do you know for certain that the variables you didn’t include in your final model have no effect on causing more/less particles? Why or why not?

11. What are some ideas you have on further data that could be collected to increase the understanding about causes to the number of particles present in the lab?