

What happens after data collection? Qualtrics data cleaning and analysis

Katie Stone

ORS Research Design & Data Analysis Lab Consultant

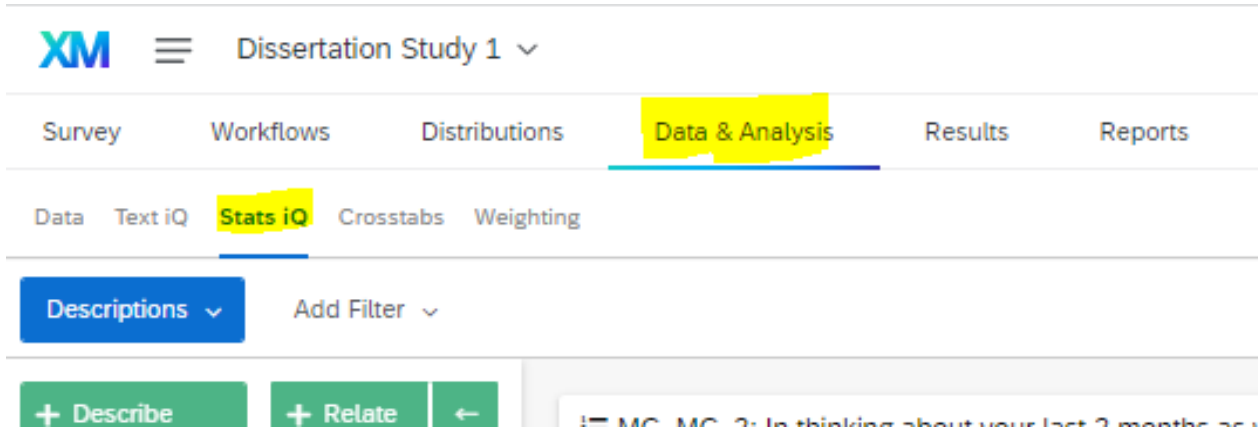


UTTyler[™]

THE UNIVERSITY OF TEXAS AT TYLER



What is Stats IQ?



A protected analytical tool within the platform.

Allows users to safely explore relationships and experiment with new variables in the data set without impacting other areas of the platform

[Video](#)

Stats iQ has three main parts, numbered in the screenshot below:

The screenshot shows the Stats iQ interface with three numbered callouts:

- 1**: Points to the **variable pane** on the left, which includes a search bar, a list of variables (e.g., "Overall, how satisfied ...", "CES: How would you r...", "ContactFreq: How ofte...", "ResSat: How satisfied ...", "CSAT: Overall, how sat...", "Product: Overall, how ..."), and a "+ Create or Clean Variable" button.
- 2**: Points to the **analyses** area in the center, which displays "No analysis results" and a "Run your first analysis" button.
- 3**: Points to the **workspace** area on the right, which is currently empty.

1. The [variable pane](#), where you'll see all your variables listed, and be able to either create new ones or select what variables you want to perform analyses with.
2. The [analyses](#), where you choose the type of statistical test you'd like to perform.
3. The [workspace](#), where completed analyses appear. You can add more workspaces as needed, select from the ones you've made, and share them with colleagues.

To summarize a variable, select it and hit **Describe**. If it's a numeric variable, we're given info on median, mean, and so on. Different [types of variables](#) are [described in different ways](#).

Typical Analysis Workflow

Stats iQ is a flexible tool that can be used however you'd like. However, you can find below a very common analysis workflow that you might follow to get started:

1. **Clean Variables:** Before running any statistical analyses, it's important to verify your variables are scored and labeled correctly.
2. **Examine the Distribution of the Data (Describe):** It is a good idea to first understand the data you are working with. Select some variables that you're interested in (or just select them all), then use the Describe analysis to visualize and summarize your data. Look out for interesting or surprising results, get a feel for distributions and outliers, and think about interesting questions to ask.
3. **Find Relationships (Relate):** Once you have a good understanding of how your data is distributed, explore the relationships between variables with the Relate analysis option. Put the "key" by the variable you're most interested in, select other variables, then Relate. Stats iQ will run the appropriate statistical test for each relationship, translate the results into plain English, and bring the strongest relationships to the top. Examine the results, noting which variables were and weren't statistically related to your key variable.
4. **Run Regression:** The above may be enough to surface some very interesting findings, but if you'd like to go further, you can use Regression. Keep the key by an output variable, such as "Satisfaction," and then select possible drivers of that output, such as "Income." Run a regression to understand which of the drivers are most related to the outcome and how they are related.

Cleaning Data & Variables

The screenshot shows the XM software interface for a 'Dissertation Study 1'. The 'Data & Analysis' tab is active, with sub-tabs for 'Data', 'Text IQ', 'Stats IQ', 'Crosstabs', and 'Weighting'. The 'Stats IQ' sub-tab is selected, showing a list of variables on the left and a detailed view for 'MC_MC_2' on the right. The 'MC_MC_2' variable is described as 'In thinking about your last 2 months at work, use the scale below to indicate your level of agreement with each here.' The interface includes buttons for '+ Describe', '+ Relate', '+ Regression', and '+ Create or Clean Variable' (highlighted in yellow). The 'Ranked Correlation' section shows a P-Value of <math>< 0.00001</math>, an Effect Size (Spearman's rho) of 0.393, a Confidence Interval of Effect Size from 0.316 to 0.465, and a Sample Size of 500. The 'Simple Linear Regression' section shows an R-squared value and a line of best fit equation.

Items to Check and Verify:

- Variable Name
- Variable Type (i.e., numbers, categories, times, dates)
- Recode Values

The screenshot shows the 'Responses' section of the XM software interface. It displays 'Responses: 373' and a 'Settings' button. The interface is partially obscured by a large grey rectangle, suggesting a transition or a specific view.

← Two places to verify →

Key Concepts

- **Numerical Data.** Any data that is only numbers. Examples include height, weight, age, and number of children.
- **Categorical Data.** Non-numeric data with no specific order. Examples include favorite color, country of residence, and brand of jeans.
- **P-value.** In a statistical test, the probability of getting the same or more extreme value if our data actually had no trend (i.e, was random). Generally, P-values less than .05 (5%) are considered significant.
- **Population.** The range of all possible values that could be sampled from.
- **T-Test.** A significance test that compares a numeric value between two group.
- **ANOVA.** A significance test that compares a numeric value for 3 or more groups.
- **Chi-Squared.** In a statistical test, the probability of getting the same or more extreme value if our data actually had no trend (i.e, was random).
- **Regression.** A formula that calculates a predicted value for one numeric variable based on its relationships with other variables.
- **Times Series.** Similar to a regression, time series predicts the future value of a variable based on its past values and trends
- **Logistic Regression.** A special type of regression that calculates the odds of someone belonging to two or more groups

Resources

- Qualtrics XM Basecamp
 - Provides free on-demand training
- Qualtrics Communities
 - Commonly asked and answered questions from the regular to the obscure.
 - Answers provide full, complete, and reproducible information
- Research Design & Data Analysis Lab:
 - <https://www.uttyler.edu/research/ors-research-design-data-analysis-lab/>
- Schedule a consultant appointment with me:
 - <https://www.uttyler.edu/research/ors-research-design-data-analysis-lab/ors-research-design-data-analysis-lab-consultants/>
- Check out the Lab Resources (including recording of this webinar):
 - <https://www.uttyler.edu/research/ors-research-design-data-analysis-lab/resources/>