

ECON 0150 | MiniExam 2 | Skillset

This document lists the skills covered in Part 2 (2.1, 2.2, 2.3) that students should be able to demonstrate on MiniExam 2.

Part 2.1: Bivariate Relationships

Students should be able to:

1. **Identify the direction of a relationship** — Given two numerical variables, determine whether the relationship is positive, negative, or unclear.
 2. **Judge the strength of a relationship** — Distinguish between strong relationships (points cluster tightly around a trend) and weak ones (points are widely scattered).
 3. **Understand why visualization matters** — Recognize that datasets can share identical summary statistics (mean, standard deviation, correlation) yet show very different patterns when plotted (Anscombe's quartet).
 4. **Apply log transformations** — Calculate log-transformed values (log base 2 or log base 10) and understand what they do: compress large values, spread small values, and make skewed data easier to compare.
 5. **Interpret log-transformed axes** — On a log2 scale, recognize that each unit represents a doubling. On a log10 scale, each unit represents a tenfold increase.
-

Part 2.2: Numerical Variables by Category

Students should be able to:

6. **Count observations by category** — Given a categorical column, determine how many rows fall into each category.
7. **Filter data on a single condition** — Given a condition (e.g., Salary > 50,000), identify which rows remain in the filtered dataset.
8. **Filter data on multiple values** — Keep only rows where a categorical variable matches one of several values (e.g., Department is "Sales" or "HR"). Note: this semester covers single-column filters and multi-value matching, not compound AND/OR logic across columns.
9. **Investigate and clean data** — When a visualization looks unexpected (e.g., a boxplot compressed at zero), use counting to diagnose the issue, identify which rows are causing the problem, and filter to clean the data before

re-visualizing.

10. **Read and compare boxplots** — Identify the median, spread (IQR), and range from a boxplot. Compare distributions across categories: which group has a higher center, more variation, or outliers.
 11. **Interpret a stripplot overlay** — Understand that a stripplot shows individual data points on top of a boxplot, revealing clusters, gaps, or patterns that the box alone might hide.
 12. **Calculate grouped statistics** — Given a dataset grouped by category, compute the mean, standard deviation, and count for each group.
 13. **Perform multi-step operations** — Chain together a filter, a group-by, and an aggregation to answer a specific question about a subset of the data.
-

Part 2.3: Bivariate Relationships by Category

Students should be able to:

14. **Merge two datasets on a shared key** — Given two tables with a common column, predict which rows match and what the merged table looks like. Understand that unmatched rows are dropped (in a default merge).
15. **Reshape data from wide to long format** — Given a table where each column represents a time period or category, convert it to long format where each row represents one observation. Understand why long format is needed for certain visualizations.
16. **Read a scatter plot colored by category** — Identify separate trends for each group in a scatter plot with color-coded categories. Determine whether the relationship differs across groups.
17. **Describe how a relationship differs by group** — Articulate specific differences: one group shows a stronger/weaker relationship, or the direction reverses, or one group is shifted up/down.
18. **Choose the right visualization** — Given a research question and data description, select the appropriate tool: scatter plot (two numerical variables), boxplot by category (one numerical + one categorical), or scatter by category (two numerical + one categorical).