Major Assignment 3

Topic 6 - Probability · Descriptive Stats · Empirical Rule · Histogram

STUDENT WALKTHROUGH · SPRING 2026

Two components. One probability story.

Major Assignment 3 (MA3) is the descriptive-statistics and probability capstone of MAT-144. Two worksheet components - Analysis of a 50-student tutoring dataset, and a Visualization (histogram) of the score changes. 100 points total, due at the end of Topic 6. No separate Word write-up - the one paragraph answer lives inside the workbook. Every dataset is personalized off your name.

POINTS	COMPONENTS	PERSONALIZED
100	**2**	**1 of 23**
54 Analysis · 46 Visualization	tabs of calculations	name hash picks your dataset

The Excel template lives next to this PDF on the course site. Open the .xlsx in desktop Excel for Microsoft 365. Work the Analysis tab first - the Visualization tab pulls its 50 data points from Analysis!D14:D63 by reference. Submit the completed .xlsx to Halo. No other files.

WHAT'S IN THIS WALKTHROUGH

THE TEMPLATE'S COLOR LEGEND

Blue = enter text (your name, your written paragraph). **Green** = enter a number (MA3 has zero green cells - the personalization comes from the name hash, not typed numbers). **Gold** = enter an Excel formula with cell references. **Any other color** = leave alone. Get the colors right and you've eliminated most of the easy point losses.

Major Assignment 3

Topic 6 - Probability · Descriptive Stats · Empirical Rule · Histogram

01 - COMPONENT 1 · ANALYSIS · 54 PTS

The Analysis tab simulates a tutoring study. Fifty students each took a math test before tutoring (column B) and again after (column C). You compute the per-student change (column D), then four summary statistics for all three columns, two personalized percentiles on the change column, and the 68% empirical-rule interval. Finish with a short written interpretation in cell F43.

1. **Type your full name in B10 (blue cell).**

   At least 5 characters - if your name is shorter, pad with X's. The hash in cell F9 uses your name to pick one of 23 personalized datasets. Until B10 is filled, the score table at B14:C63 stays blank.

2. **Compute the 50 score changes in D14:D63.**

   Click cell D14 and type **`=C14-B14`** (after minus before). Grab the fill handle (bottom-right corner) and drag down through D63. All 50 gold cells now show formulas (not typed numbers), and column D becomes your "Change in Score" column.

3. **Fill the 4x3 stats grid (G18:I21).**

   Twelve gold cells. Rows are Mean / Median / SD / Range. Columns are Before / After / Change. Use **`=AVERAGE(B14:B63)`**, **`=MEDIAN(B14:B63)`**, **`=STDEV(B14:B63)`**, and **`=MAX(B14:B63)-MIN(B14:B63)`**. Repeat for columns C and D. Format every cell as Number with 2 decimals.

4. **Compute the two personalized percentiles (G27, G28).**

   Cells F27 and F28 already display labels like "12th" and "47th" (auto-generated from your name hash). Convert the label to a decimal by hand: "12th" becomes 0.12, "47th" becomes 0.47. Then **`=PERCENTILE(D14:D63, 0.12)`** in G27 and the same shape in G28. Format Number, 2 decimals.

■ COMMON SLIPS - ANALYSIS

(1) Typed the stats by hand. The grader checks that each stats cell shows an =FUNCTION in the formula bar, not just a number. **(2) Passed "12th" to =PERCENTILE().** The label is text. The function wants 0.12. **(3) Range left as a single MAX or MIN.** Range is MAX – MIN, both in one formula. **(4) Number formatting forgotten.** The rubric checks for 2 decimals on every stat. **(5) Filled D14 without dragging.** All 50 cells need the formula, not just the first.

Major Assignment 3

Topic 6 - Probability · Descriptive Stats · Empirical Rule · Histogram

02 - THE EMPIRICAL RULE · THE PROBABILITY PIECE

MA3 lives under Topic 6 (Probability) because the Empirical Rule turns standard deviation into a probability statement: about 68% of values in a roughly normal distribution fall within one SD of the mean. You apply it to the Change column and write a short paragraph interpreting the bounds.

1. **Compute the lower bound in G36 (Mean - 1 SD).**
   The mean of the Change column lives in H18; the SD lives in H20. Type **`=H18 - H20`**. Cell references only - don't retype the decimal numbers.
2. **Compute the upper bound in G37 (Mean + 1 SD).**
   Same shape, plus instead of minus: **`=H18 + H20`**. Format both bounds as Number with 2 decimals.
3. **Write the 2-3 sentence interpretation in F43:L49.**
   It's a merged blue cell. Answer two things: did tutoring help (yes / no based on the sign of the mean change), and what does the 68% interval tell you? Cite the actual lower and upper bounds you just computed - the grader is looking for your specific numbers, not a generic empirical-rule definition.

■ TEMPLATE PARAGRAPH (FILL IN THE BOUNDS)

About 68% of the 50 students (roughly 34) had a change in score between **[lower bound from G36]** and **[upper bound from G37]**. Because the mean change is **[positive / negative]**, the data suggests tutoring **[did / did not]** improve scores on average. The 68% interval comes from the Empirical Rule applied to a roughly normal distribution of change scores. Use this as a skeleton, but swap in your own numbers and your own observation.

■ COMMON SLIPS - EMPIRICAL RULE

(1) Used 95% (2 SD) or 99.7% (3 SD) instead of 68%. The prompt asks for the 68% interval - one standard deviation, not two or three. **(2) Confused empirical rule with Chebyshev's theorem.** Chebyshev gives a lower-bound percentage for any distribution; empirical rule gives the specific 68/95/99.7 percentages for normal distributions. MA3 wants empirical. **(3) Wrote a generic definition instead of citing the actual bounds.** The grader wants your specific lower and upper numbers in the paragraph. **(4) Forgot to mention the ~34 students.** 68% of 50 is about 34 - say it explicitly.

Major Assignment 3

Topic 6 - Probability · Descriptive Stats · Empirical Rule · Histogram

03 - COMPONENT 2 · VISUALIZATION · 46 PTS

The Visualization tab pulls your 50 change values from Analysis!D14:D63 into Visualization!B12:B61 automatically. Your job is to set up an 11-bin frequency distribution, then build a histogram. The bin width formula is mechanical - the trap is the 0.1 buffer on Bin Min and Bin Max.

1. **Compute Bin Min, Bin Max, and Bin Width (E22, E23, E24).**

   Bin Min: **`=MIN(B12:B61) - 0.1`**. Bin Max: **`=MAX(B12:B61) + 0.1`**. Bin Width: **`=(E23 - E22) / 11`**. Eleven bins is the template's fixed choice.

2. **Fill the Lower Limit column (D28:D38).**

   Row 1: **`=E22`** (just the Bin Min). Rows 2-11: each cell equals the PREVIOUS row's Upper Limit. So D29 = E28, D30 = E29, etc. This creates contiguous bins with no gaps.

3. **Fill the Upper Limit column (E28:E38) and Title column (F28:F38).**

   Upper Limit: **`=D28 + $E$24`** (the lower limit plus the bin width - lock E24 with $ so it doesn't shift when you drag). Title of Bin: **`=(D28 + E28) / 2`** (midpoint of the bin). Drag both down 11 rows.

4. **Count frequencies (G28:G38) and relative frequencies (H28:H38).**

   Frequency (preferred): **`=COUNTIFS($B$12:$B$61, ">"&D28, $B$12:$B$61, "<="&E28)`**. Note the dollar signs and the > / <= operators (right-open bins). Relative frequency: **`=G28 / COUNT($B$12:$B$61)`**, formatted as a percentage.

■ THE 0.1 BUFFER - WHY IT EXISTS

Without subtracting 0.1 from MIN and adding 0.1 to MAX, the smallest data value lands exactly on the boundary of bin 1, and **`COUNTIFS(">x", "<=y")`** excludes it (because it uses strict greater than on the lower side). Your frequencies sum to 49 instead of 50. Same thing on the top - without +0.1, the largest value falls outside bin 11. The buffer fixes both edges in one move and is the #1 reason frequency totals don't match.

04 - THE HISTOGRAM · CHART CHROME COUNTS

Once the bin table is filled, build the histogram from scratch. There's no placeholder chart in the template - you insert it. The grader checks four things: custom title, both axis labels, bars touching (gap width zero), and Title of Bin used as the x-axis labels.

1. **Select Title of Bin (F28:F38) and Frequency (G28:G38).**
   Click F28, drag down to F38, then Ctrl-click G28 and drag to G38. Two columns selected - the labels and the counts. Insert > Chart > 2-D Clustered Column. A bar chart appears below the table.
2. **Rename the chart title.**
   Click the title text "Chart Title" once to select it, click again to edit, and type something specific - for example "Distribution of Change in Math Scores". The grader is looking for any custom title; the literal default text is an automatic deduction.
3. **Add both axis labels.**
   Click the chart. Chart Design > Add Chart Element > Axis Titles > Primary Horizontal (label it "Change in Score" or "Title of Bin"). Then again for Primary Vertical (label it "Frequency"). Both labels are required; missing either costs points.
4. **Set gap width to zero (bars touching).**
   Right-click any bar > Format Data Series > Series Options > Gap Width slider all the way to the left (0%). The bars now touch - this is what makes it a histogram and not a bar chart. The rubric explicitly checks this.

■ TWO ACCEPTABLE FREQUENCY FORMULAS

The instructor deck explicitly says either `=COUNTIFS()` or `=FREQUENCY()` earns full credit. COUNTIFS is the simpler option (one row at a time, no array formula). FREQUENCY is faster (one array formula for all 11 bins at once, using all upper limits except the last as the bins_array). Pick whichever feels comfortable - the video links in the explanation deck cover both.

■ COMMON SLIPS - HISTOGRAM

(1) Used Excel's built-in Histogram chart type. It auto-bins differently than the template's 11 fixed bins, so the chart doesn't match your bin table. Use Clustered Column with Title of Bin as the x-axis instead. **(2) Gap width left at 150%.** Bars must touch - drag the slider to 0%. **(3) Y-axis label missing.** Frequency on the vertical, Change in Score on the horizontal. Both required. **(4) Chart inserted on the wrong tab.** The chart belongs on the Visualization tab, near row 48 - not as a separate chart sheet.

Major Assignment 3

Topic 6 - Probability · Descriptive Stats · Empirical Rule · Histogram

05 - BEFORE YOU SUBMIT

Most point losses on MA3 aren't math errors - they're typed numbers where formulas were expected, the 0.1 buffer left off, or the empirical-rule paragraph that defines the rule instead of citing the student's specific bounds. Catch them here.

- Analysis: Name in B10, all 50 changes in D14:D63 formulas ✓
- Analysis: 12 stats cells use =AVERAGE / =MEDIAN / =STDEV / =MAX-MIN ✓
- Analysis: Percentiles use 0.12, not "12th"; format 2 decimals ✓
- Analysis: G36/G37 are Mean - SD and Mean + SD on Change; F43 cites them ✓
- Viz: Bin Min has -0.1, Bin Max has +0.1; bin width = (max-min)/11 ✓
- Viz: Frequencies sum to 50; histogram has title + labels + 0% gap ✓

If all six check, you're at 100%.

06 - GRADING RUBRIC BREAKDOWN

The 100 points are split 54 / 46 between Analysis and Visualization. Inside each component, formatting (decimals, percentages) and audit-trail (functions, cell references) checks live alongside the math.

Component 1 - Analysis (name + diff table + stats + percentiles + bounds + paragraph)	54 pts
Component 2 - Visualization (bin table + frequency distribution + histogram)	46 pts
Total	100 pts

STUCK? OPEN THE LIVE PAGE ON MAT144.COM

Every component on this PDF has a corresponding page at **MAT144.com/topics/6/ma** with the same hints plus deep links into the relevant lessons and DQs. Topic 6 DQ 1 and DQ 2 cover the probability concepts (empirical rule, percentiles) that show up on MA3.