Green Belt Content using Copilot vs Minitab

Agenda

- Summary Statistics
- Boxplots
- Histogram
- Normality Test
- Individual Value Plot
- SPC
- Capability Analysis
- ANOVA
- T Test
- Regression
- DOE
- Q&A

Summary Statistics

Weight

87.79
78.8
82.9
81.5
80.26
76.74
83.02
79.55
84.76
80.57
86.5
82.44
81.44
79.64
86.14
84.14
79.69
…

“Provide summary statistics for this data set”

MINITAB

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Weight	100	0	82.8685	0.307163	3.07163	70.58	81.3725	83.205	84.6975	90.05

COPILOT

Central tendency

- **Mean:** 82.8685
- **Median:** 83.205
- **Mode:** none (no repeated dominant value)

Spread

- **Minimum:** 70.58
- **Maximum:** 90.05
- **Range:** 19.47
- **Sample standard deviation:** 3.071
- **Sample variance:** 9.431
- **Coefficient of variation:** 3.71%

Quartiles

- **Q1:** 81.37
- **Q3:** 84.69
- **Interquartile range (IQR):** 3.32

Five-number summary

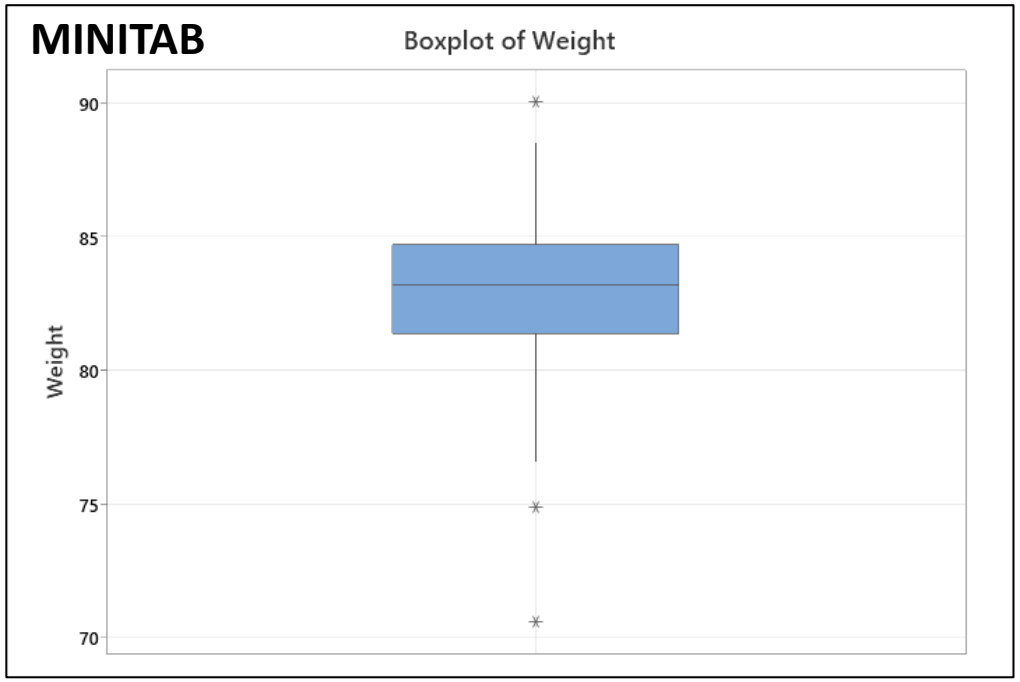
- **Min:** 70.58
- **Q1:** 81.37
- **Median:** 83.205
- **Q3:** 84.69
- **Max:** 90.05

Boxplots

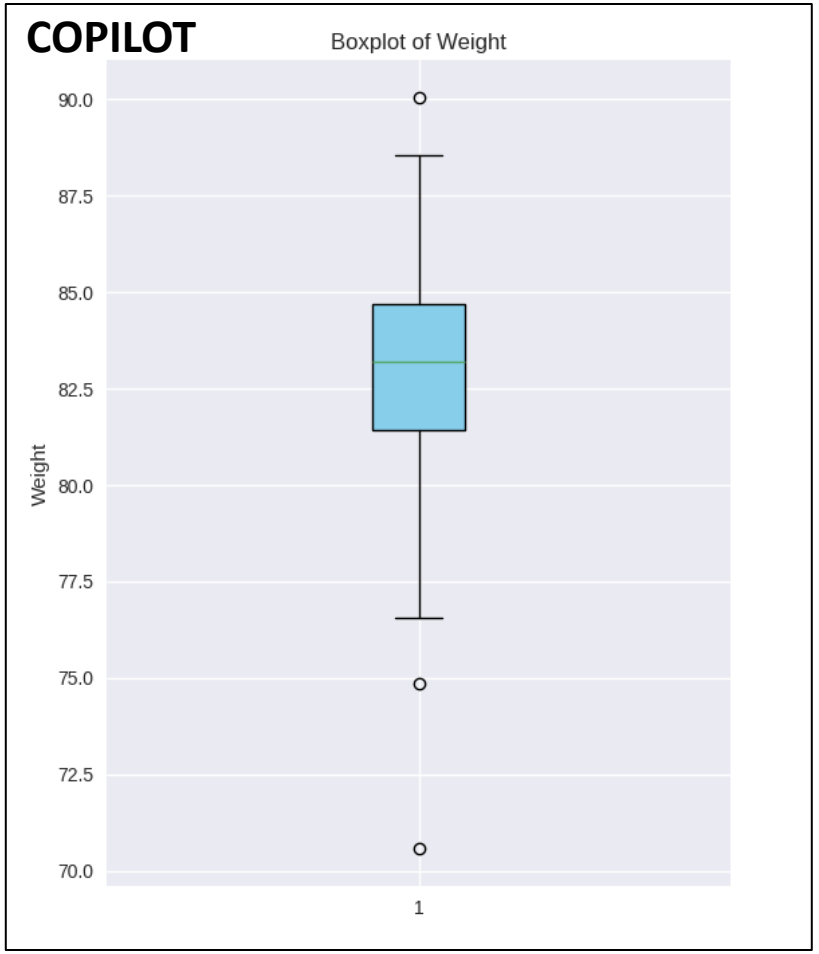
“Create a boxplot for the weight data”

Note: Can hover over chart to see values

MINITAB



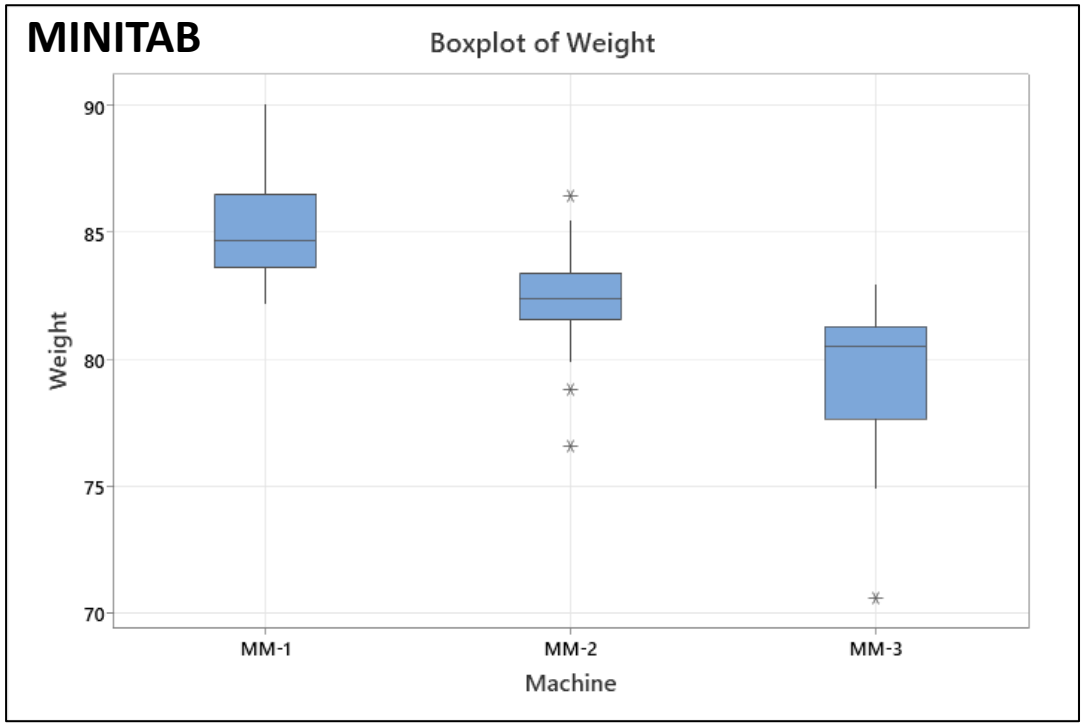
COPILOT



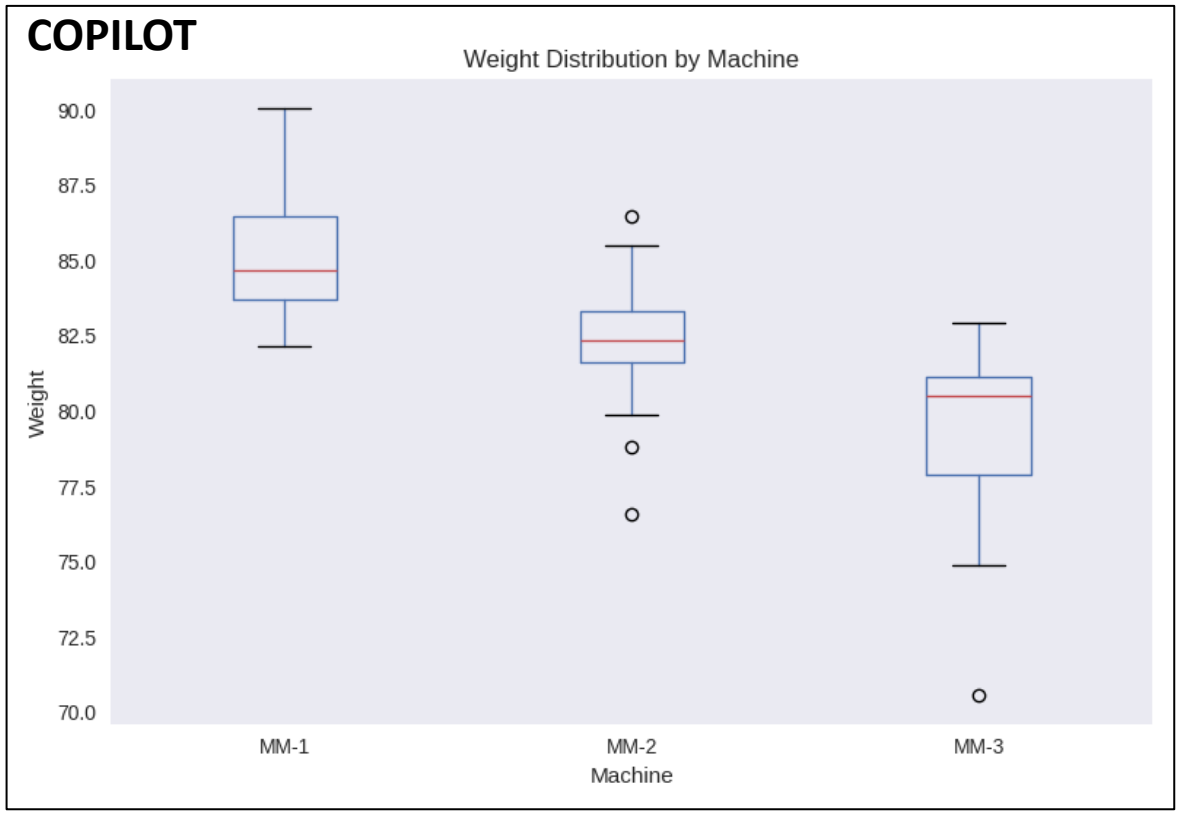
Multiple Boxplots

Create boxplots for the weight broken out by the machine column, all in the same chart?

MINITAB



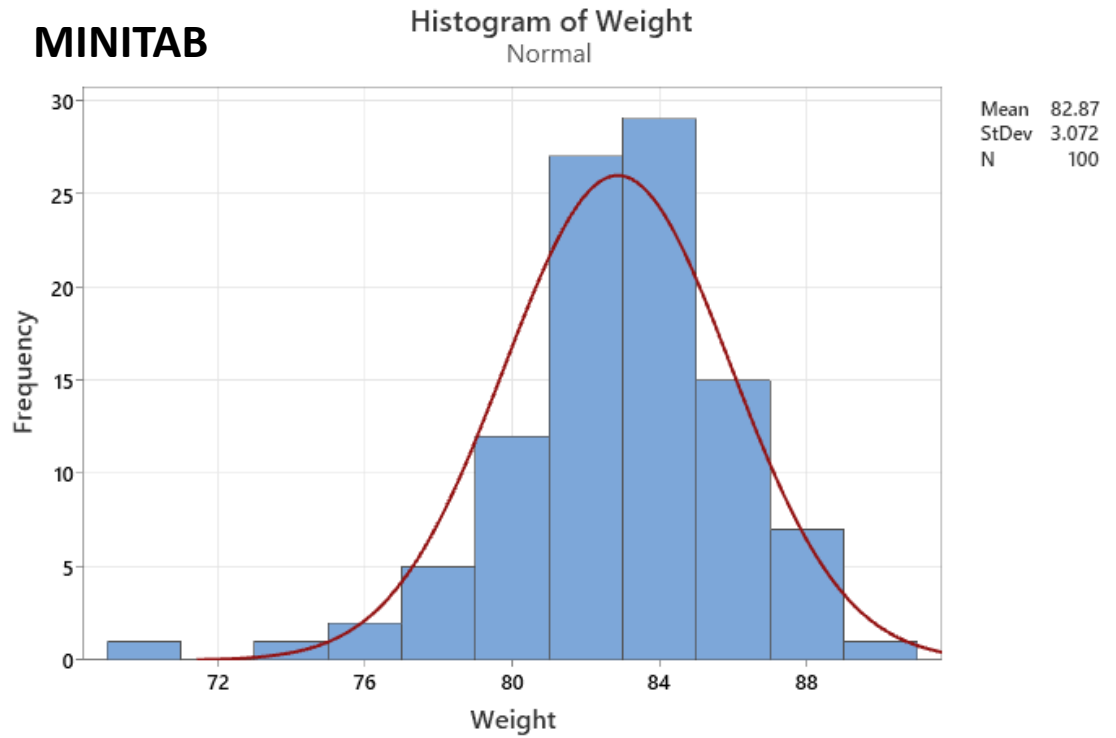
COPILOT



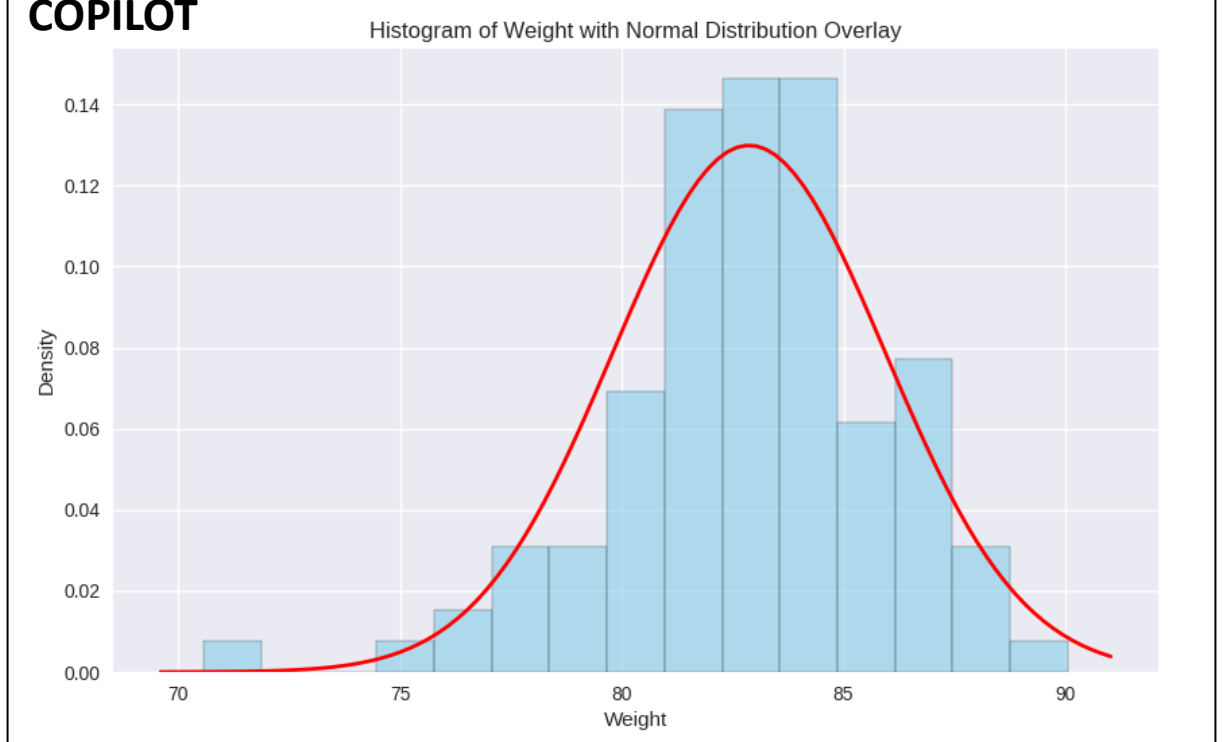
Histograms

“Create a histogram with a normal curve overlay for the weight data”

MINITAB



COPILOT



Note: Minitab has proprietary way of creating histogram bins

Normality Test

“Run an Anderson-Darling normality test against the weight data”

Anderson–Darling results interpreted

The test statistic is:

- **$A^2 = 0.6671$**

And the critical values are:

- **15% → 0.5550**
- **10% → 0.6320**
- **5% → 0.7590**
- **2.5% → 0.8850**
- **1% → 1.0530**

To interpret this, compare the test statistic to the critical values:

- 0.6671 is **greater than** the 10% critical value (0.6320)
- 0.6671 is **less than** the 5% critical value (0.7590)

This places the result in the range:

p-value ≈ 0.05–0.10

COPILOT

What this means for normality

- At the **5% significance level**, you **do not reject** normality.
- At the **10% level**, you **would reject** normality.

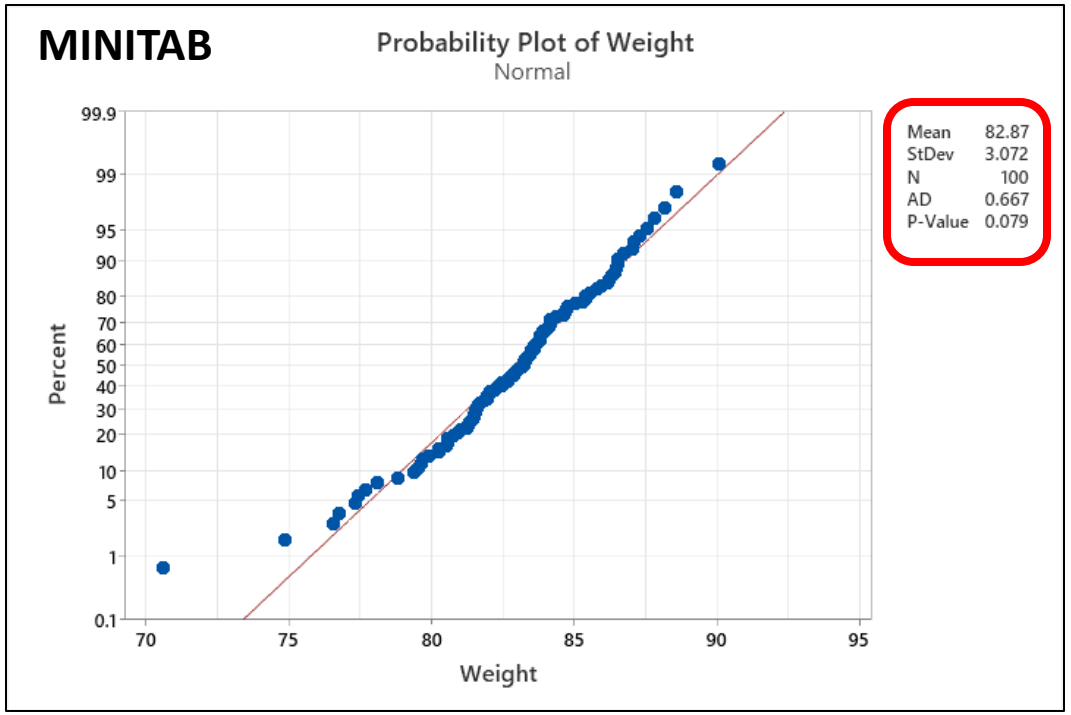
In practical terms:

- The data is **close to normal**, but with mild deviations—consistent with the histogram and boxplot you generated earlier.
- The deviations are driven mainly by:
  - A handful of low values (70–78)
  - A few high values (88–90)
  - Slight skewness visible in the histogram

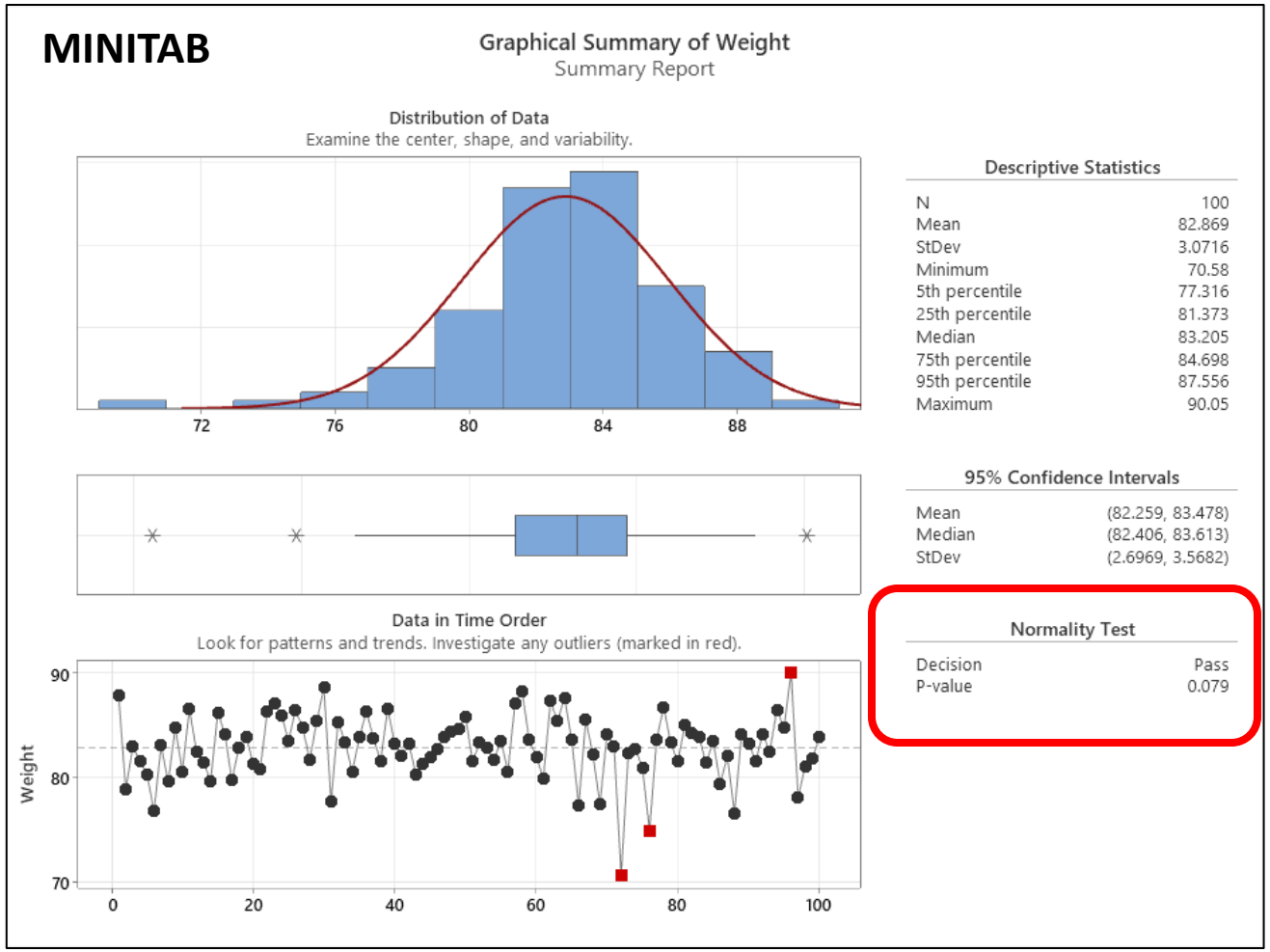
For most quality and capability analyses, this level of deviation is **not severe**, and normal-based methods (Cp/Cpk, X-bar/R charts) remain reasonable unless your process requires strict normality.

Note: Very helpful explanation and suggestions, but not an exact p-value provided

Normality Test



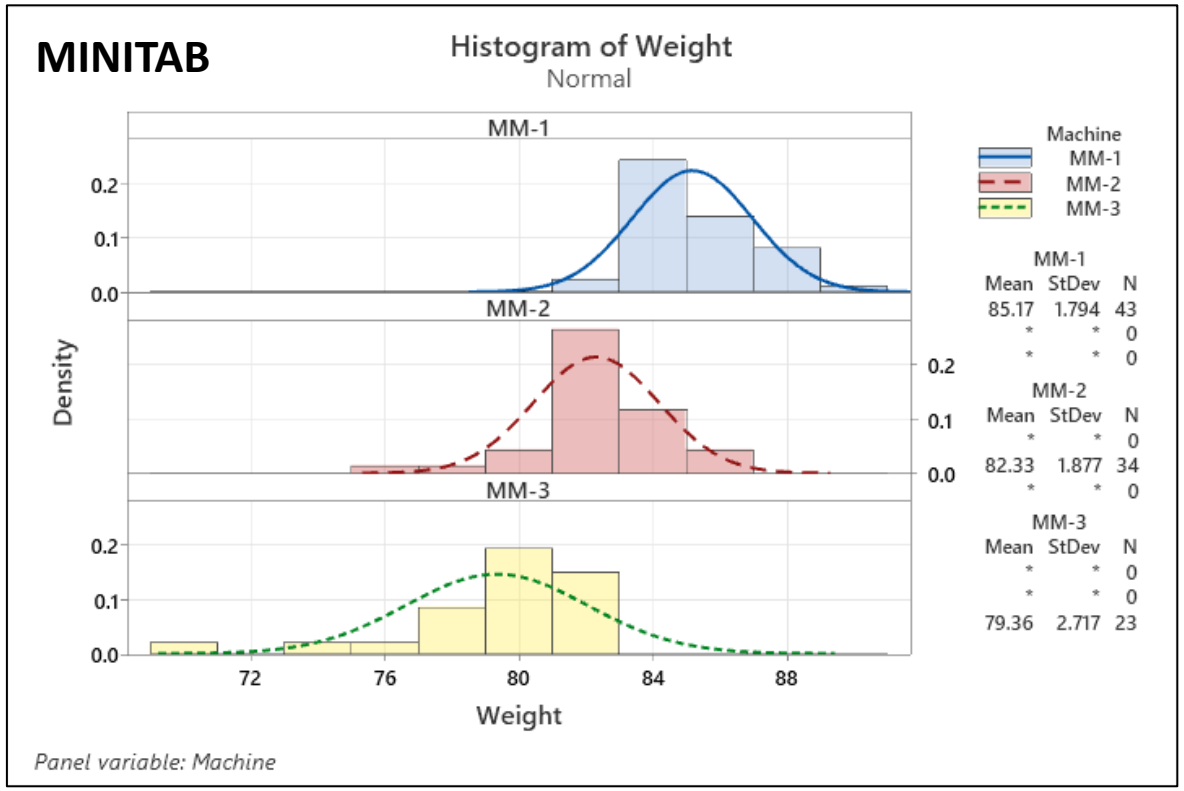
Note: AD values match Copilot, but gives exact p-value (not range of p-values)



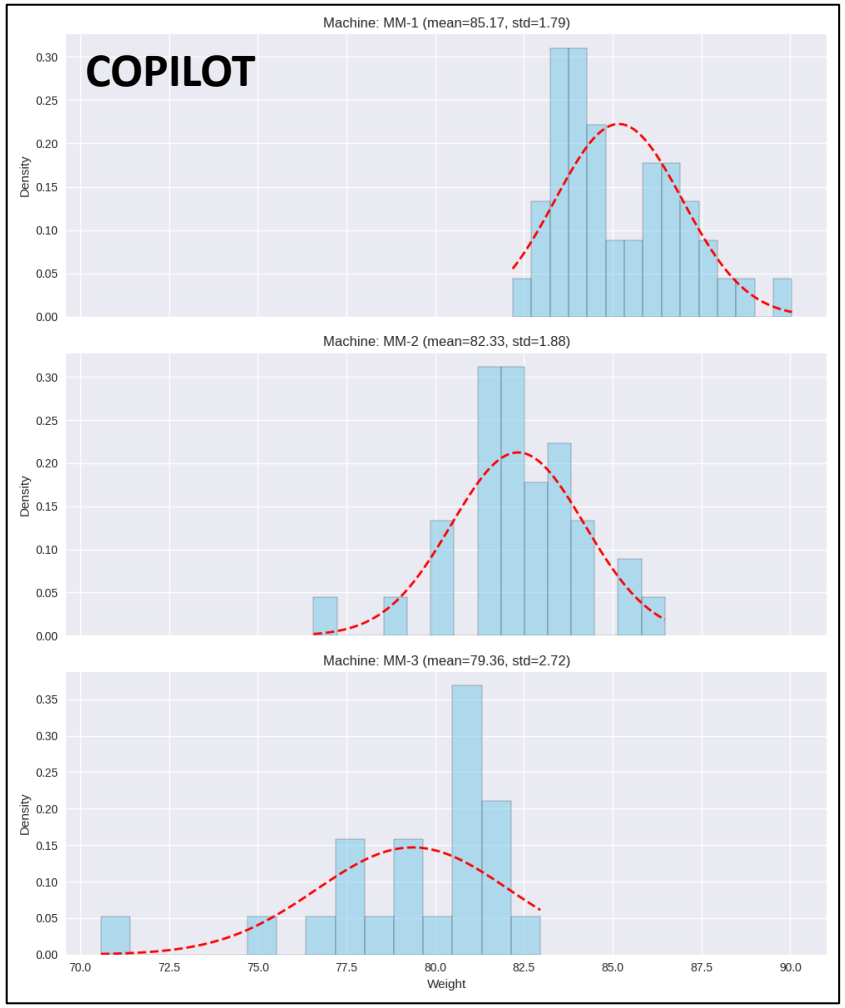
Multiple Histograms

“Create histograms of the weight data with a normal curve overlay, separated by the machine column and stacked on top of each other (using the same x-axis) on one chart”

MINITAB



COPILOT

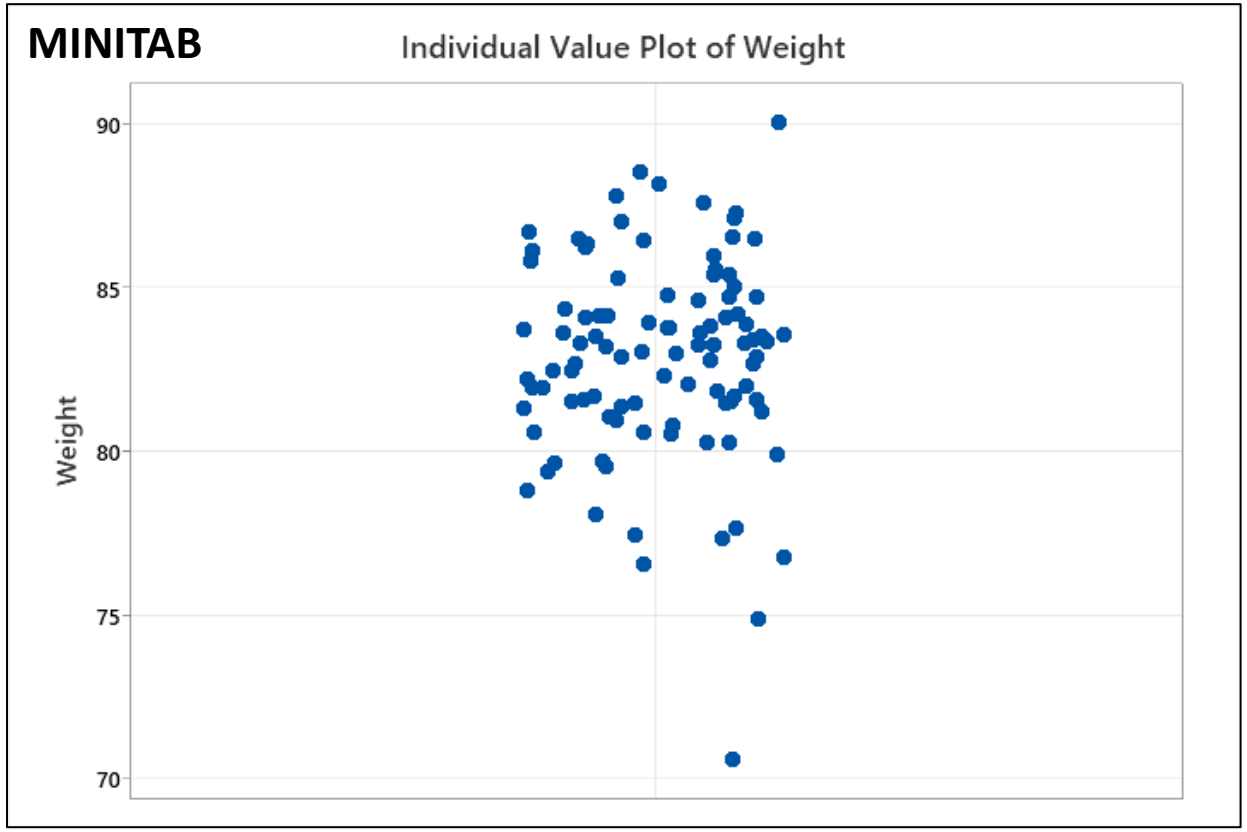


Note: Requires some editing in options to stack vertically

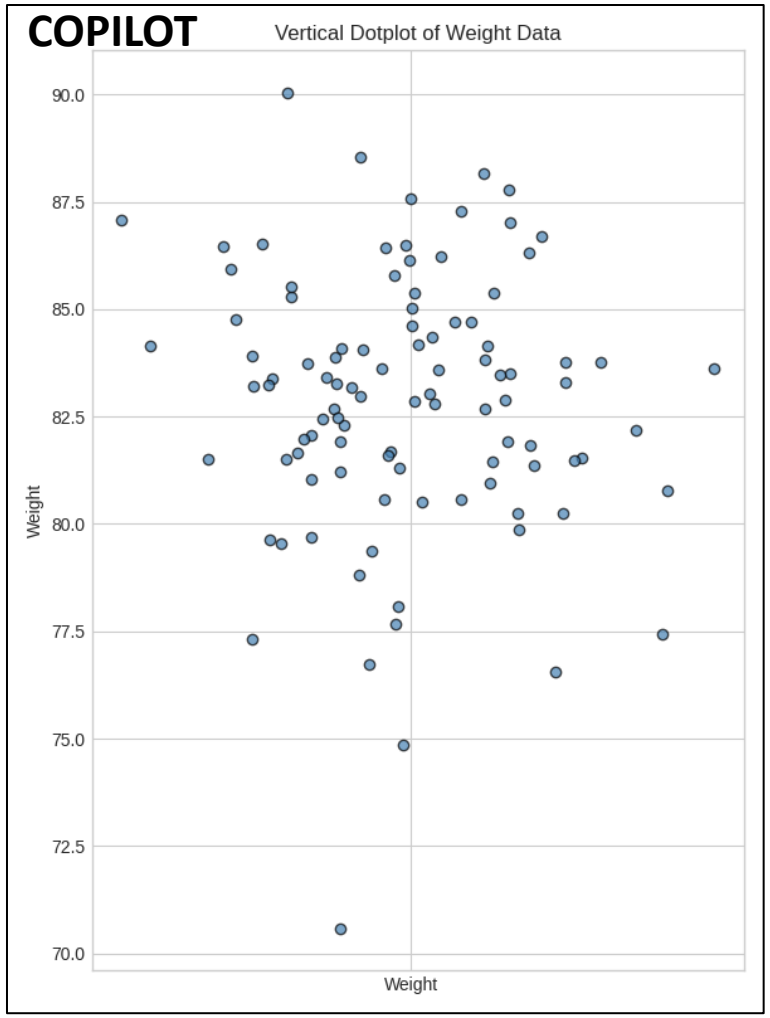
Dot Plot / Individual Value Plot

“Create a vertical dotplot for the weight data”

MINITAB



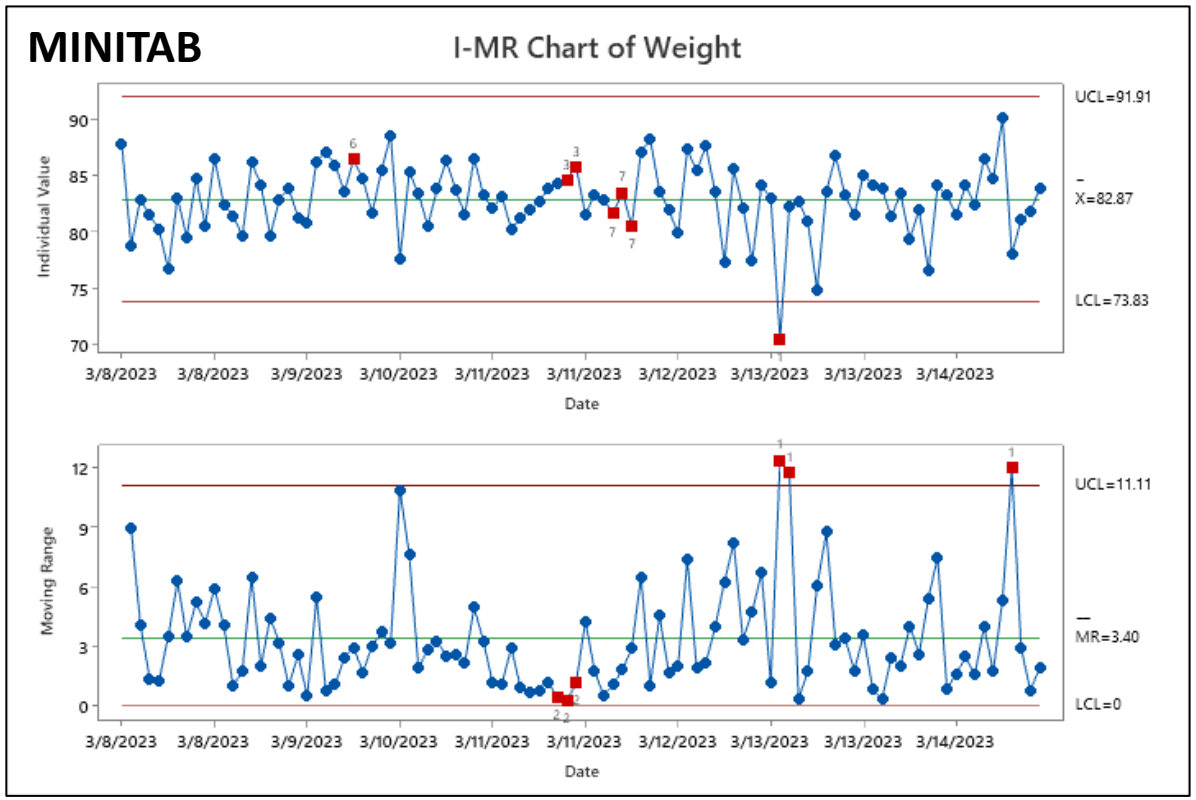
COPILOT



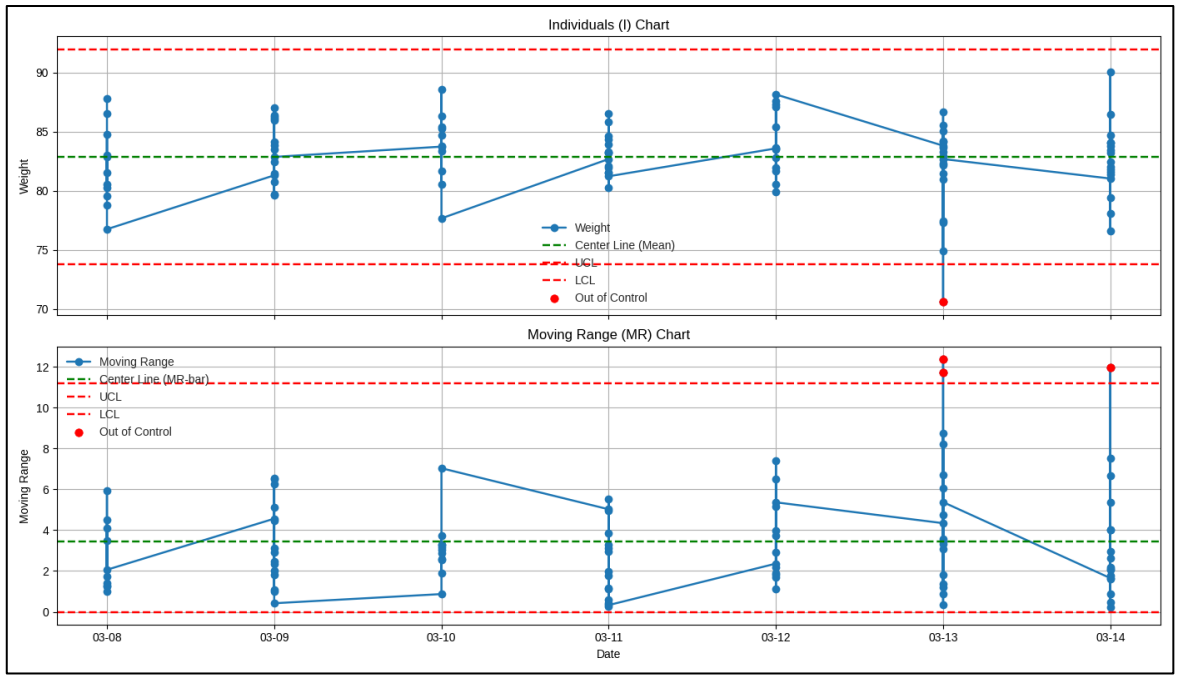
SPC (I-MR chart)

“Create an Individuals and Moving Range control chart for weight data with the date on the x-axis and check for any out of control conditions”

MINITAB



COPILOT

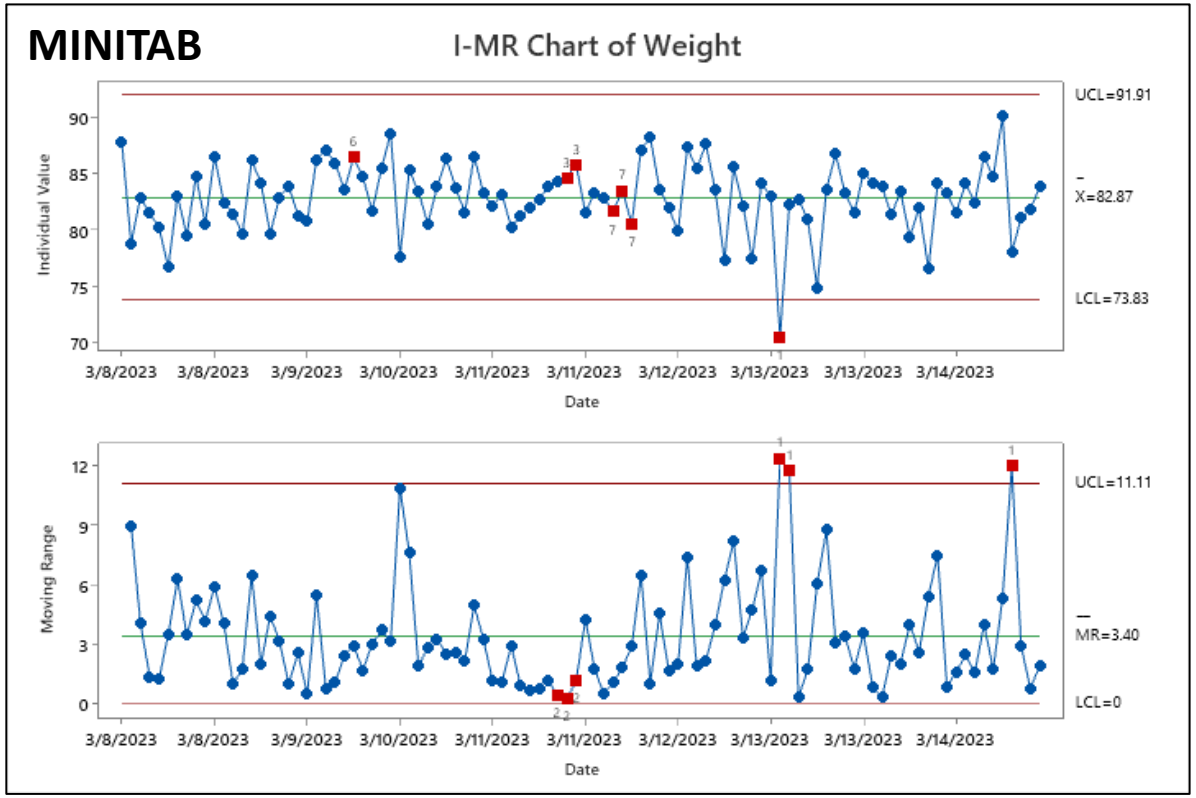


Note: Handled it how Excel treats Date fields

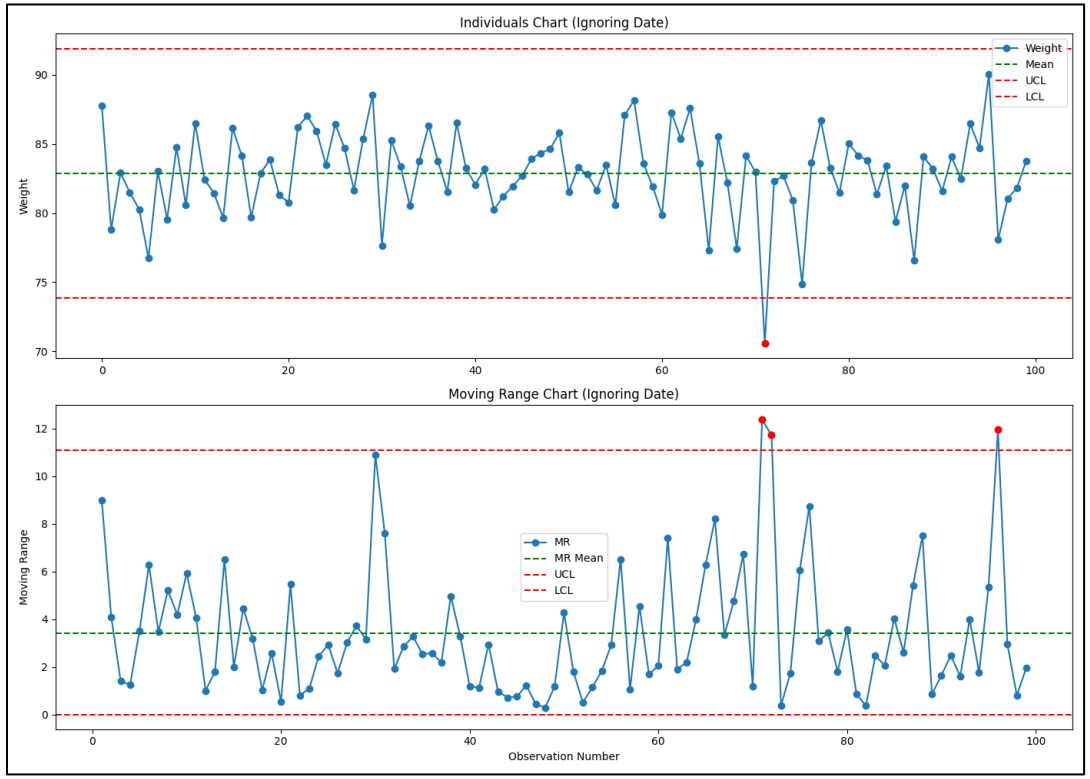
SPC (I-MR chart)

“Re-create the control chart, except ignore the date column”

MINITAB



COPILOT

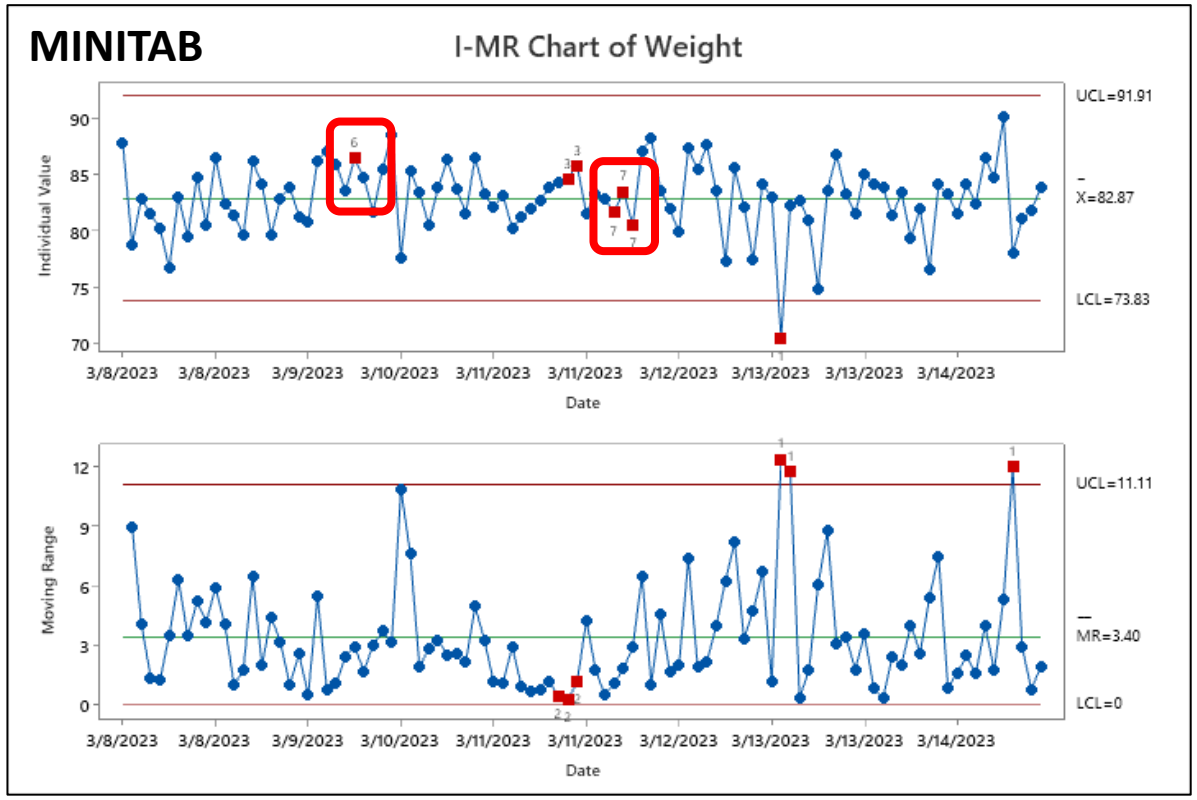


Note: Only used rule 1, any data point outside control limits

SPC (I-MR chart)

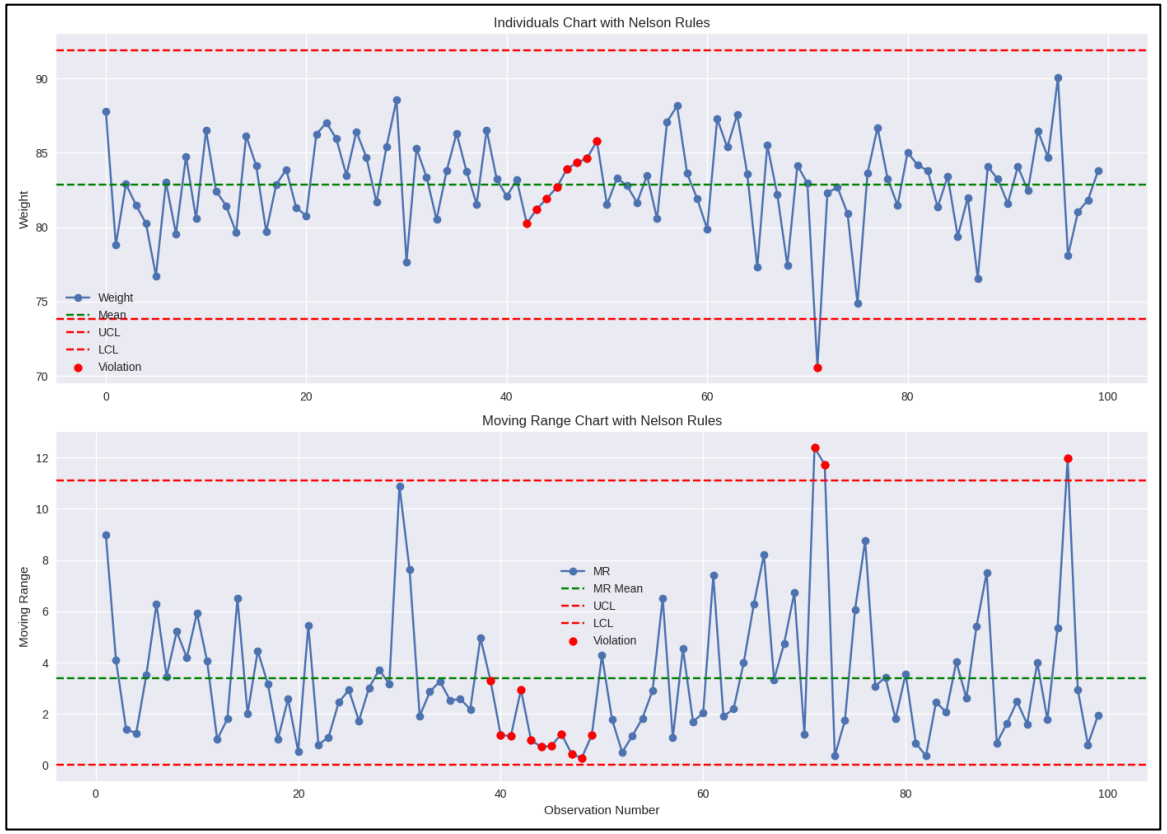
“Regenerate the control chart with Nelson rules”

MINITAB



Note: Detected Rule 6 and Rule 7 correctly

COPILOT



Note: Flags all related data points, not just violation points, but did not detect all rule violations

Capability Analysis (Cpk/Ppk)

“Perform a capability analysis of the weight data, with spec limits set from 76-88, including a histogram showing the spec limits on the chart with a normal curve overlay”

📌 Capability results (spec limits 76–88)

COPILOT

Using the dataset’s established summary statistics:

- **Mean** ≈ 82.87
- **Standard deviation (within, sample)** ≈ 3.07
- **Standard deviation (overall, population)** ≈ 3.05

Capability indices

$$Cp = \frac{USL - LSL}{6\sigma}$$

$$Cpk = \min\left(\frac{USL - \bar{X}}{3\sigma}, \frac{\bar{X} - LSL}{3\sigma}\right)$$

Index	Value	Interpretation
Cp	**0.65**	Process spread is too wide for the spec window
Cpk	**0.58**	Centering is off; closer to LSL than USL
Pp	**0.66**	Overall variation also too large
Ppk	**0.59**	Long-term centering issues persist

📊 What these numbers mean

COPILOT

1. The process is not capable

All indices are **well below 1.00**, meaning the natural variation of the process exceeds the specification width.

- Cp < 1 → The process spread is too wide.
- Cpk < Cp → The process is **not centered** within the spec limits.

2. The process is slightly shifted toward the lower spec

Because:

- Mean = 82.87
- Midpoint of specs = 82

The mean is slightly above center, but the **lower tail (values in the 70s)** is causing the limiting Cpk term.

3. Short-term and long-term capability are similar

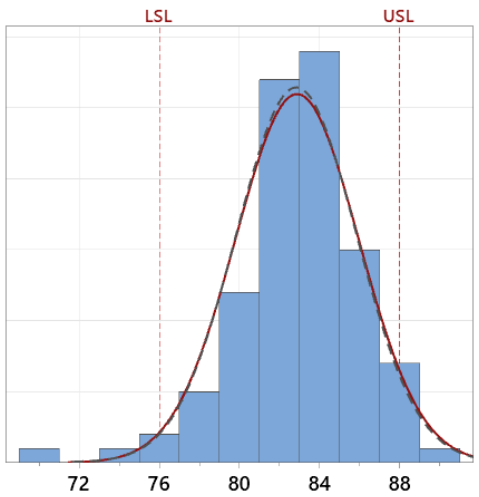
Cp ≈ Pp and Cpk ≈ Ppk, meaning:

- No major long-term drift.
- Variation is consistent over time.
- The issue is **inherent spread**, not instability.

Capability Analysis (Cpk/Ppk)

“Perform a capability analysis of the weight data, with spec limits set from 76-88, including a histogram showing the spec limits on the chart with a normal curve overlay”

MINITAB Process Capability Report for Weight



Process Data

LSL	76
Target	*
USL	88
Sample Mean	82.8685
Sample N	100
StDev(Overall)	3.07163
StDev(Within)	3.01338

Overall
Within

Overall Capability

Pp	0.65
PPL	0.75
PPU	0.56
Ppk	0.56
Cpm	*

Potential (Within) Capability

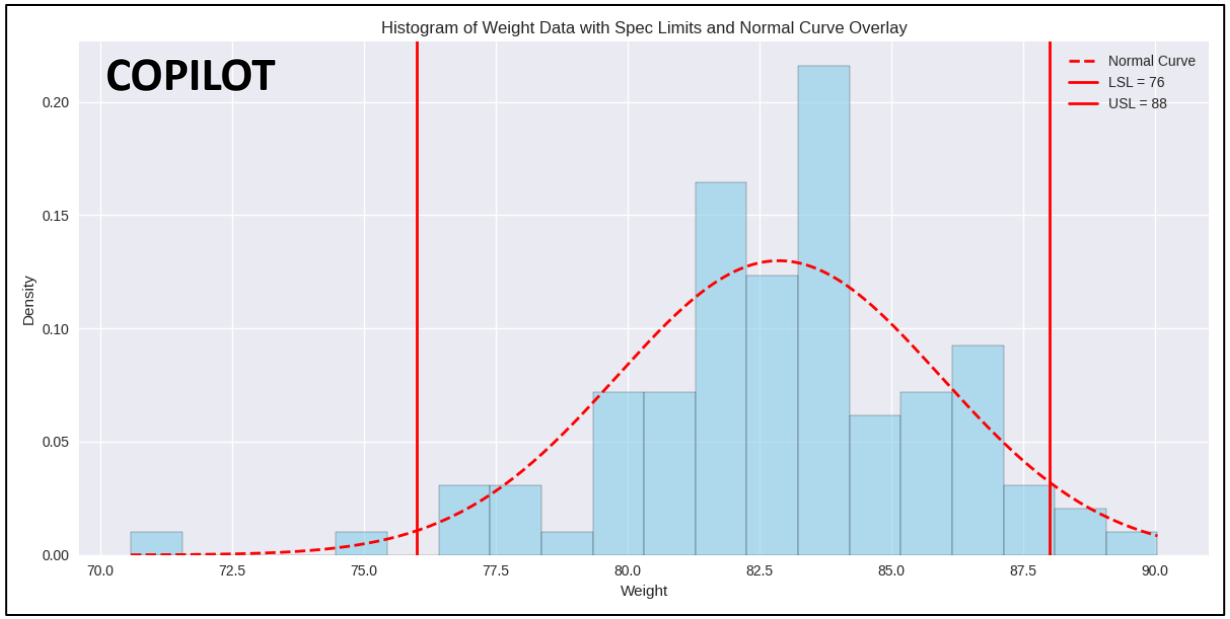
Cp	0.66
CPL	0.76
CPU	0.57
Cpk	0.57

Performance

	Observed	Expected Overall	Expected Within
% < LSL	2.00	1.27	1.13
% > USL	3.00	4.74	4.43
% Total	5.00	6.01	5.56

The actual process spread is represented by 6 sigma.

COPILOT



Capability Indices Summary

ORIGINAL

Statistic	Minitab	Copilot
Mean	82.87	82.87
Standard Deviation (Overall)	3.07	3.04
Standard Deviation (Within)	3.01	3.07
Cp	0.66	0.65
Cpk	0.57	0.58
Pp	0.65	0.66
Ppk	0.56	0.59

AFTER CORRECTION

Statistic	Minitab	Copilot
Mean	82.87	82.87
Standard Deviation (Overall)	3.07	3.04
Standard Deviation (Within)	3.01	3.01
Cp	0.66	0.66
Cpk	0.57	0.57
Pp	0.65	0.65
Ppk	0.56	0.56

Note: Assumed population (not sample) using n (not n-1), but still arrived at the correct answer

ANOVA

“Can you perform a hypothesis test for the weight broken out by the machine column, comparing each machine to each other, and include the summary ANOVA table and R-squared values?”

COPILOT

ANOVA Summary Table

Source	Sum of Squares	df	F-statistic	p-value
Machine	520.29	2	**60.986**	7.07×10^{-18}
Residual	413.77	97	—	—

Interpretation

- The p-value is far below 0.001.
- **Reject the null hypothesis.**
- At least **one machine’s mean weight is significantly d**

Model Fit Statistics

Metric	Value
R-squared	0.5570
Adjusted R-squared	0.5479

Interpretation

- About **55.7% of all weight variation** is explained solely by which machine produced the part.
- This is a **strong effect**, especially for manufacturing data.

MINITAB

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Machine	2	520.3	260.145	60.99	0.000
Error	97	413.8	4.266		
Total	99	934.1			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.06534	55.70%	54.79%	52.62%

Note: **Identical results.** Minitab assumes equal variance but can be changed. Copilot chose ANOVA and provided Tukey comparison without prompt, also assumed equal variance, but not sure if it checked beforehand. Copilot didn’t initially include ANOVA table or R^2 values.

Tukey HSD Pairwise Comparisons Test

📌 Tukey HSD Pairwise Comparisons — COPILOT

Comparison	Mean Difference	p-value	Significant?
MM-1 vs MM-2	−2.840		Yes
MM-1 vs MM-3	−5.807		Yes
MM-2 vs MM-3	−2.967		Yes

Interpretation

- **All three machines differ significantly** from each oth
- **MM-1 produces the heaviest weights.**
- **MM-3 produces the lightest weights.**
- MM-3's lower mean is consistent with the left-tail fail

📌 What this means for your process

Machine differences are:

- **Real**
- **Statistically significant**
- **Operationally meaningful**

This strongly suggests:

- Machine-level calibration differences
- Potential mechanical wear or setup variation
- A need for **machine-specific capability analysis (Cp/Cpk)**
- A need for **machine-specific control charts**

Given the size of the differences, MM-3 is almost certainly the primary contributor to:

- Low Cpk/Ppk
- LSL failures
- Left-tail skew in the histogram

Tukey Pairwise Comparisons — MINITAB

Grouping Information Using the Tukey Method and 95% Confidence

Machine	N	Mean	Grouping
MM-1	43	85.170	A
MM-2	34	82.330	B
MM-3	23	79.363	C

Means that do not share a letter are significantly different.

Means

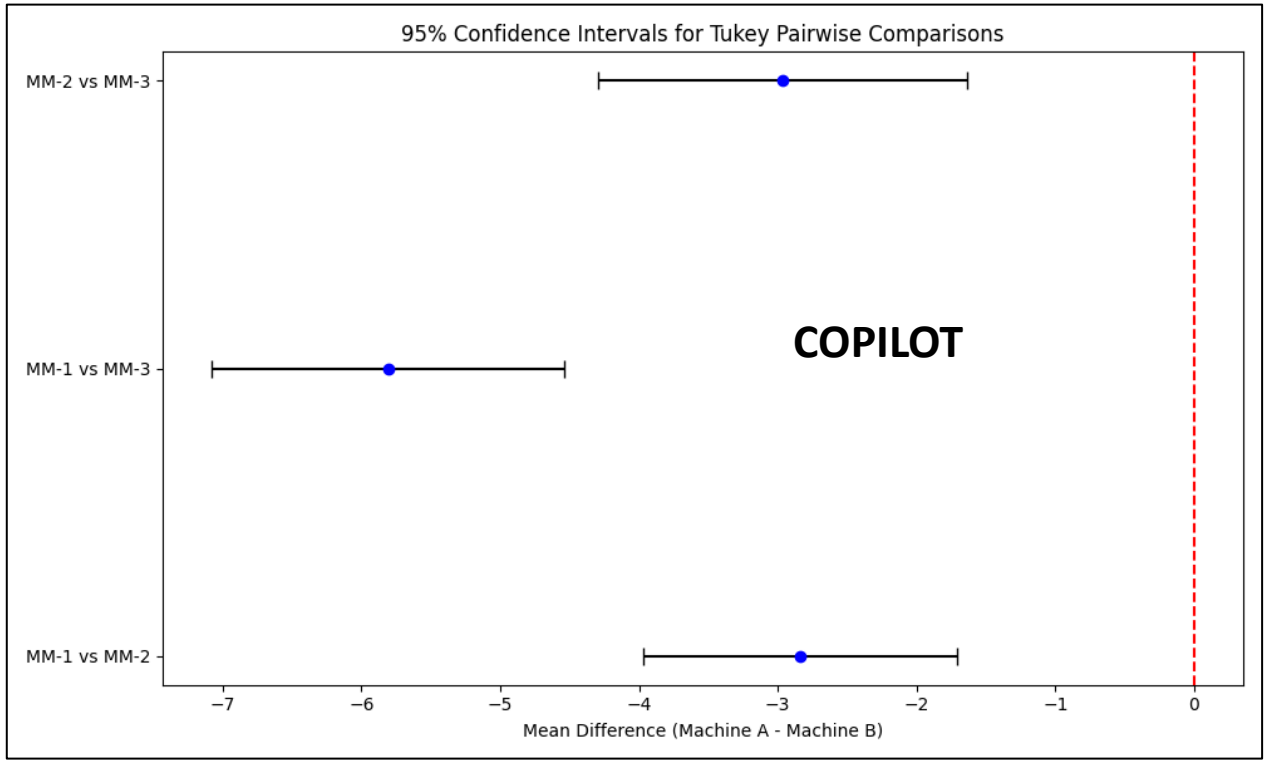
Machine	N	Mean	StDev	95% CI
MM-1	43	85.170	1.794	(84.545, 85.795)
MM-2	34	82.330	1.877	(81.627, 83.033)
MM-3	23	79.363	2.717	(78.508, 80.217)

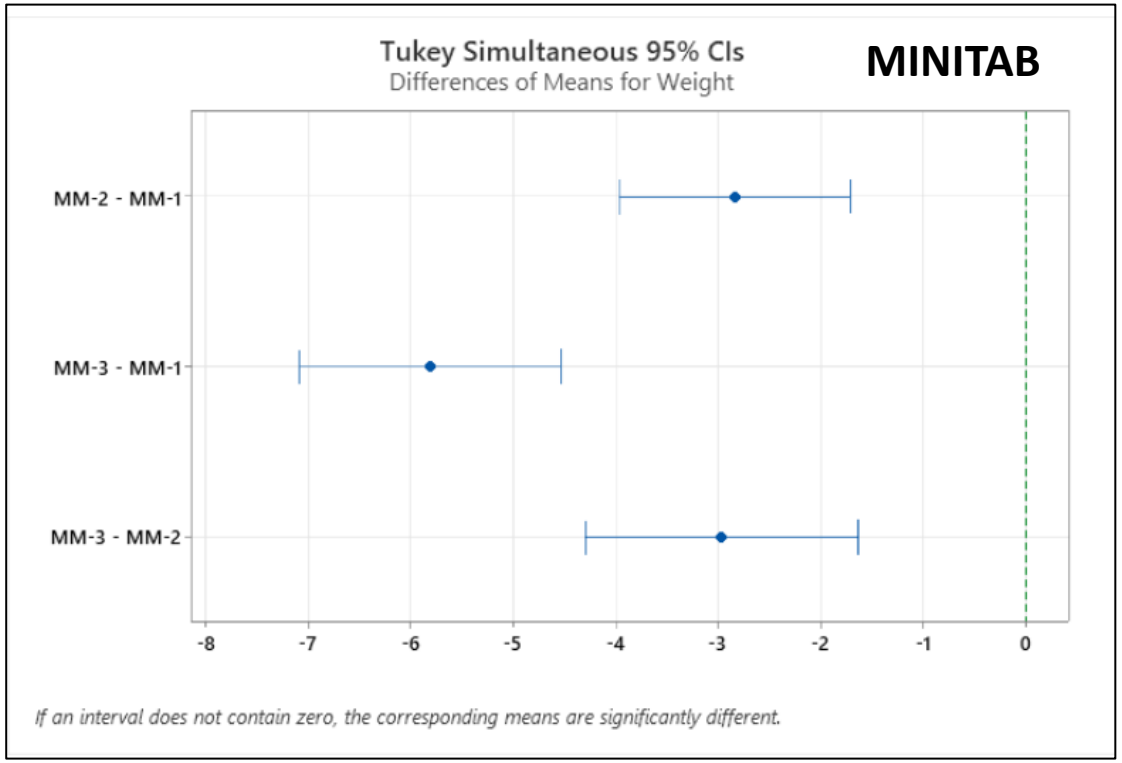
Pooled StDev = 2.06534

Note: **Results match** but displayed differently. Good summary of results and recommendations from Copilot.

Tukey HSD Pairwise Interval Plot

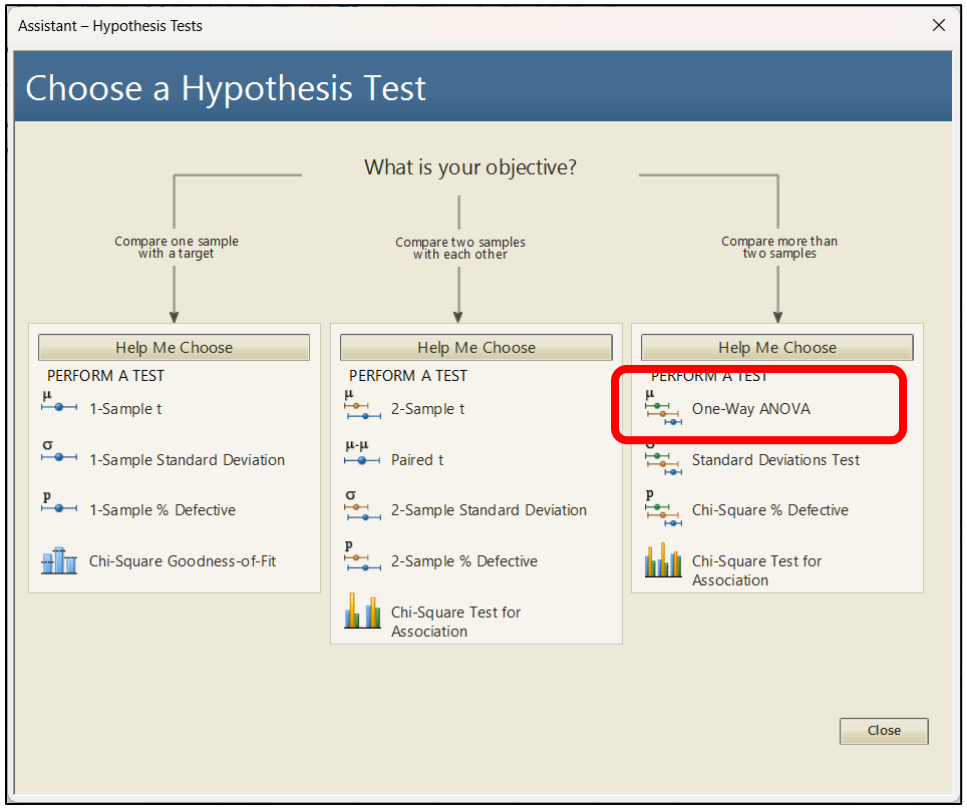
Can you create one chart showing the 95% confidence intervals of each comparison performed in the Tukey test?

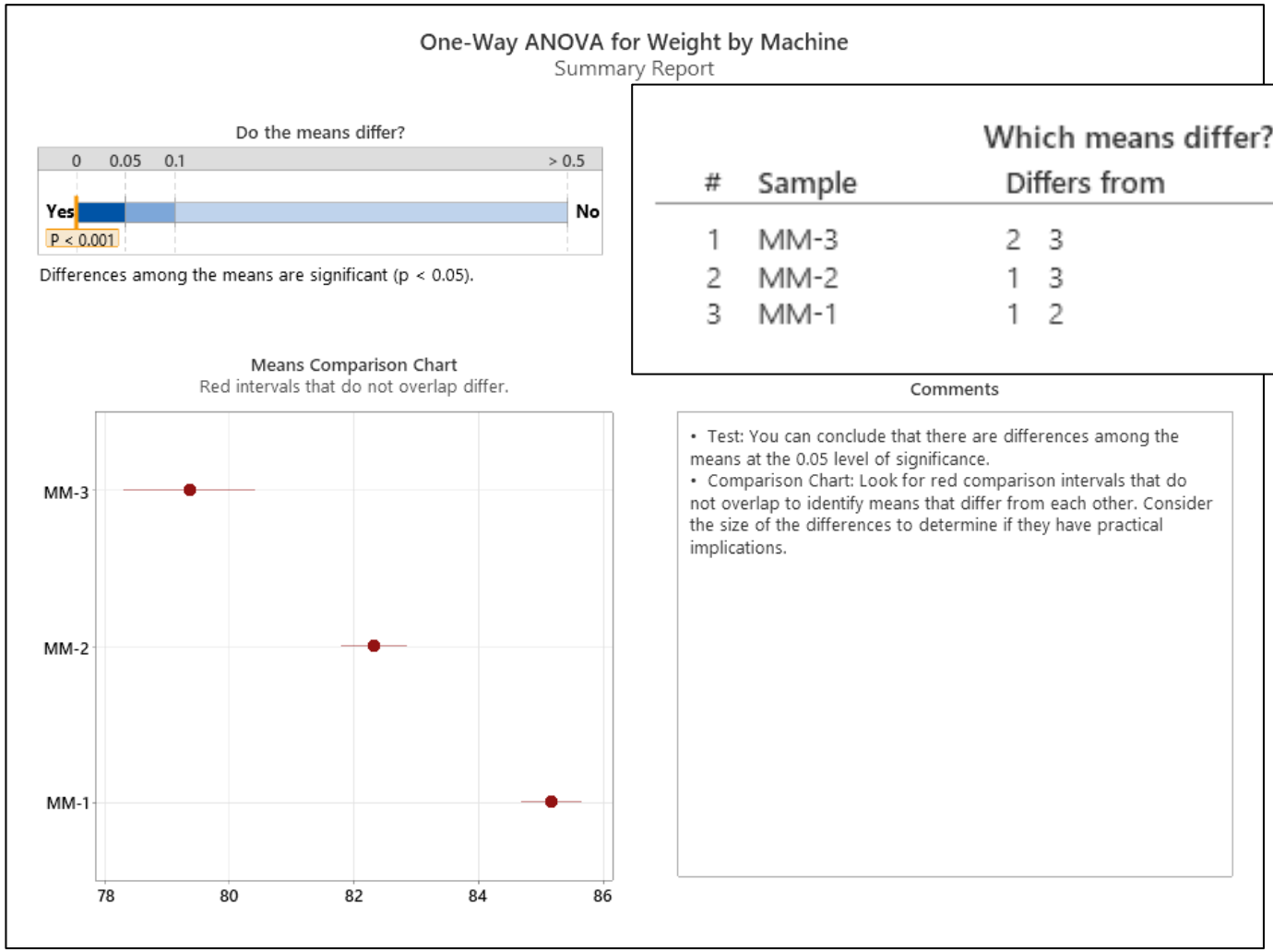




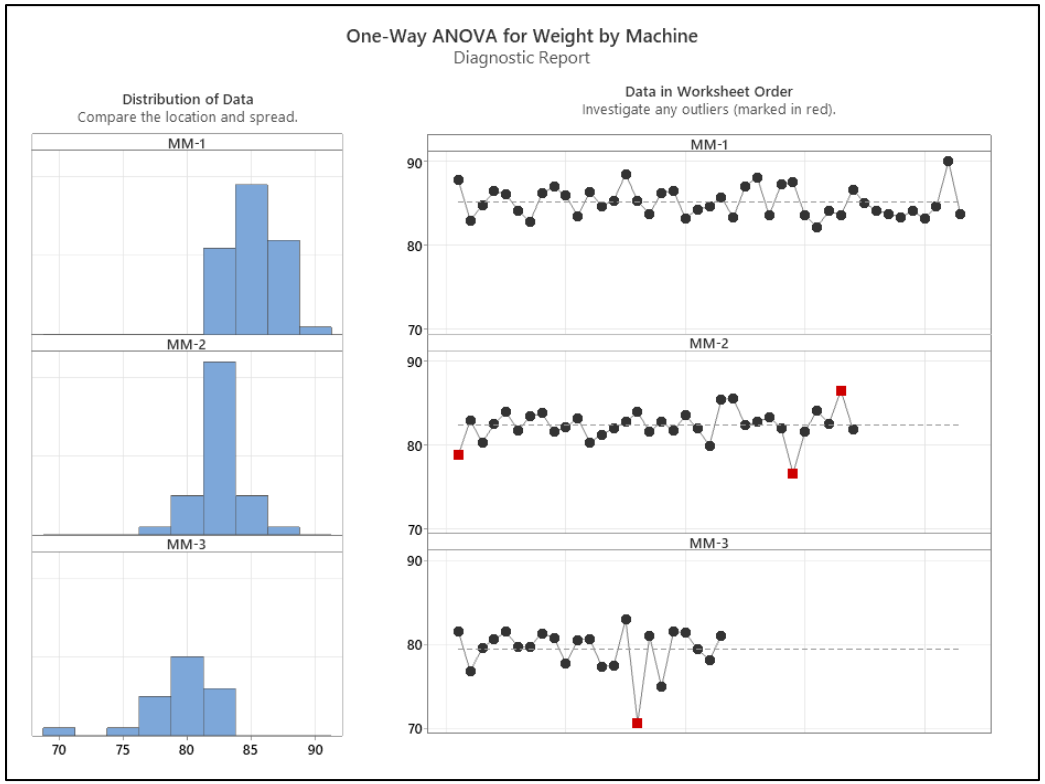
Note: **Similar results.** Minitab automatically creates the chart when Comparison is selected.

Minitab Assistant





Minitab Assistant (cont'd)



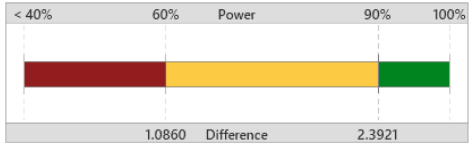
Note: Assistant checks stability, normality, outliers, sample sizes. Ran a multiple comparison test and switched to Welch's test which does not assume equal variance but doesn't seem to check for equal variance.

One-Way ANOVA for Weight by Machine
Report Card

Check	Status	Description
Unusual Data	!	Some of the data points are unusual compared to the others in the same sample. Because unusual data can have a strong influence on the results, you should try to identify the cause of their unusual nature. These points are marked in red on the Diagnostic Report. You can hover over a point or use Minitab's brushing feature to identify the worksheet row. Correct any data entry or measurement errors. Consider removing data that are associated with special causes and repeating the analysis.
Sample Size	✓	The sample is sufficient to detect differences among the means.
Normality	✓	Because all your sample sizes are at least 15, normality is not an issue. The test is accurate with nonnormal data when the sample sizes are large enough.
Equal Variance	i	Minitab's Assistant uses Welch's method, which does not assume or require that the samples have equal variances. Research shows that the test performs well with unequal variances, even when the sample sizes are not equal.

One-Way ANOVA for Weight by Machine
Power Report

What is the chance of detecting a difference?



Based on your samples and α level (0.05), you have at least a 90% chance of detecting a difference of 2.3921, and at most a 60% chance of detecting a difference of 1.0860.

What difference can you detect with your sample sizes?

Difference	Power
1.0860	28.8 - 60.0%
1.6718	60.0 - 94.2%
1.8647	70.0 - 97.9%
2.0884	80.0 - 99.5%
2.3921	90.0 - 100.0%

Power is a function of the sample sizes and the standard deviations. To detect differences smaller than 2.0884, consider increasing the sample sizes.

Statistics

Machine	Sample Size	Mean	Standard Deviation	Individual 95% CI for Mean
MM-1	43	85.170	1.7936	(84.618, 85.722)
MM-2	34	82.330	1.8772	(81.675, 82.985)
MM-3	23	79.363	2.7166	(78.188, 80.537)

T Test with Copilot

“Can you perform 2-sample t tests between each pair of the different machines to see if there is a difference in weights?”

📌 Pairwise 2-Sample t-Tests (Welch’s)

MM-1 vs MM-2

- **t-statistic:** 6.723
- **p-value:** 4.11×10^{-9}
- **Conclusion: Significant difference** in mean weight

MM-1 vs MM-3

- **t-statistic:** 9.232
- **p-value:** 1.32×10^{-10}
- **Conclusion: Highly significant difference** in mean weight

MM-2 vs MM-3

- **t-statistic:** 4.554
- **p-value:** 5.82×10^{-5}
- **Conclusion: Significant difference** in mean weight

🎯 Interpretation

All three pairwise comparisons show **statistically significant differences**

- **MM-1 produces the heaviest parts**
- **MM-3 produces the lightest parts**
- **MM-2 sits in the middle**

These results perfectly align with:

- Your **ANOVA**
- Your **Tukey HSD**
- Your **boxplots and histograms**
- Your **capability analysis**, where MM-3 drives the left-tail failures

🧠 What this means for your process

Machine differences are:

- **Real**
- **Statistically significant**
- **Operationally meaningful**

This strongly suggests:

- Machine-specific calibration differences
- Potential mechanical wear or setup variation
- A need for **machine-level capability analysis (Cp/Cpk)**
- A need for **machine-level control charts**

If you want, I can now generate:

- **Machine-specific Cp/Cpk and Pp/Ppk**
- **Machine-specific control charts**
- **A Tukey mean plot**
- **A regression model including date or product**

Note: P-value more detailed (not rounded to zero). Provides interpretation of p-value.

T Test with Minitab

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Weight_MM-1	43	85.17	1.79	0.27
Weight_MM-2	34	82.33	1.88	0.32

Estimation for Difference

Difference	95% CI for Difference
2.840	(1.997, 3.683)

Test

Null hypothesis	$H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis	$H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
6.72	69	0.000

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Weight_MM-1	43	85.17	1.79	0.27
Weight_MM-3	23	79.36	2.72	0.57

Estimation for Difference

Difference	95% CI for Difference
5.807	(4.526, 7.088)

Test

Null hypothesis	$H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis	$H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
9.23	32	0.000

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Weight_MM-2	34	82.33	1.88	0.32
Weight_MM-3	23	79.36	2.72	0.57

Estimation for Difference

Difference	95% CI for Difference
2.967	(1.646, 4.288)

Test

Null hypothesis	$H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis	$H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
4.55	36	0.000

Note: Data had to be unstacked to perform t test in Minitab, and I chose unequal variance assumption. **Results matched exactly with Copilot.** More details provided and could have added Individual Value Plots and Boxplots.

Regression Data Set

CATEGORICAL INPUTS | CONTINUOUS INPUTS | CONTINUOUS OUTPUT

Date	Machine	Day of Week	Labor Hours	Backlog Inventory	Volume Output
3/6/2023	1	Mon	20.7	602	580
3/6/2023	2	Mon	23.6	623	681
3/7/2023	1	Tue	20.7	475	536
3/7/2023	2	Tue	20.0	655	598
3/8/2023	1	Wed	18.9	606	628
3/8/2023	2	Wed	19.8	653	720
3/9/2023	1	Thu	20.9	635	647
3/9/2023	2	Thu	24.5	642	664
3/10/2023	1	Fri	21.3	654	627
3/10/2023	2	Fri	26.6	572	612
3/13/2023	1	Mon	23.2	519	597
3/13/2023	2	Mon	20.4	649	682
3/14/2023	1	Tue	20.4	...	...
3/14/2023	2	Tue	19.3		

Note: Data imported, but Copilot said there were 38 rows instead of 31. When questioned, it agreed there were 31 rows.

Regression Analysis

“Perform a multiple linear regression analysis using Volume Output as the response variable, and Machine, Day of Week, Labor Hours and Backlog Inventory as the dependent variables with 2-way interactions in the model. Include residual plot analysis, ANOVA table with R-squared values, and regression equations?”

regression_output.j...
JSON

You picked a rich model here—nice choice. I'll lay out the model structure, the ANOVA/R^2 pieces, the regression equations, and how to read the residual diagnostics so you can line it up with your own software output.

> Example rows from your Regression sheet:
> “03/06/2023 — 1 — Mon — 20.749547120370444 — 602 — 580”
> “03/24/2023 12:00:00 — 2 — Fri — 21.382911561902503 — 554 — 654”


```
{"anova_table": [{"term": "Machine", "sum_sq":
3165.9254096375385, "df": 1.0, "F":
6.793917677278346, "PR(>F)": 0.035090859748155255},
{"term": "Day_of_Week", "sum_sq":
19825.127936747973, "df": 4.0, "F":
10.635933393579013, "PR(>F)":
0.0042465667077034586}, {"term":
"Machine:Day_of_Week", "sum_sq":
1020.0888549919515, "df": 4.0, "F":
0.547264923174382, "PR(>F)": 0.7075270167361702},
{"term": "Labor_Hours", "sum_sq":
257.0539556574693, "df": 1.0, "F":
0.5516249397535703, "PR(>F)": 0.481829613422802},
{"term": "Machine:Labor_Hours", "sum_sq":
366.09637494061025, "df": 1.0, "F":
0.7856245209457705, "PR(>F)": 0.4048465614139666},
{"term": "Day_of_Week:Labor_Hours", "sum_sq":
1290.9405278994907, "df": 4.0, "F":
0.6925734609944201, "PR(>F)": 0.620215670766775},
{"term": "Backlog_Inventory", "sum_sq":
6442.791049674094, "df": 1.0, "F":
13.825907543539662, "PR(>F)":
0.0074725626374264695}, {"term":
"Machine:Backlog_Inventory", "sum_sq":
343.01175749135047, "df": 1.0, "F":
0.7360860858062988, "PR(>F)": 0.4193208937393577},
{"term": "Day_of_Week:Backlog_Inventory", "sum_sq":
```


Note: Copilot created a JSON output file that contained the results, but didn’t display on the screen

Regression Analysis with Copilot

“Perform a multiple linear regression analysis using Volume Output as the response variable, and Machine, Day of Week, Labor Hours and Backlog Inventory as the dependent variables with no interactions”

regression_output.j...
JSON

You picked a rich model here—nice choice. I'll lay out the model structure, the ANOVA/R^2 pieces, the regression equations, and how to read the residual diagnostics so you can line it up with your own software output.

> Example rows from your Regression sheet:
> “03/06/2023 — 1 — Mon — 20.749547120370444 — 602 — 580”
> “03/24/2023 12:00:00 — 2 — Fri — 21.382911561902503 — 554 — 654”


```
{"anova_table": [{"term": "Machine", "sum_sq":
3165.9254096375385, "df": 1.0, "F":
6.793917677278346, "PR(>F)": 0.035090859748155255},
{"term": "Day_of_Week", "sum_sq":
19825.127936747973, "df": 4.0, "F":
10.635933393579013, "PR(>F)":
0.0042465667077034586}, {"term":
"Machine:Day_of_Week", "sum_sq":
1020.0888549919515, "df": 4.0, "F":
0.547264923174382, "PR(>F)": 0.7075270167361702},
{"term": "Labor_Hours", "sum_sq":
257.0539556574693, "df": 1.0, "F":
0.5516249397535703, "PR(>F)": 0.481829613422802},
{"term": "Machine:Labor_Hours", "sum_sq":
366.09637494061025, "df": 1.0, "F":
0.7856245209457705, "PR(>F)": 0.4048465614139666},
{"term": "Day_of_Week:Labor_Hours", "sum_sq":
1290.9405278994907, "df": 4.0, "F":
0.6925734609944201, "PR(>F)": 0.620215670766775},
{"term": "Backlog_Inventory", "sum_sq":
6442.791049674094, "df": 1.0, "F":
13.825907543539662, "PR(>F)":
0.0074725626374264695}, {"term":
"Machine:Backlog_Inventory", "sum_sq":
343.01175749135047, "df": 1.0, "F":
0.7360860858062988, "PR(>F)": 0.4193208937393577},
{"term": "Day_of_Week:Backlog_Inventory", "sum_sq":
```


Note: Simplified the request. It confirmed what to do, then asked me if I wanted to “run the model”
It gave more summary information but not detailed analysis and created another JSON file. I asked how to view a JSON file and it asked if I’d like to view a readable form.

Regression Analysis with Copilot (cont'd)

"Show the regression results in readable form"

COPILOT

⭐ Model Fit Statistics

Metric	Value
R^2	0.8500
Adjusted R^2	0.8023

Interpretation:
The additive model explains **85%** of the variation in Volume Output.
After adjusting for the number of predictors, it still explains **80%**, which is strong for only 30 observations.

📈 ANOVA Table (Type II)

This shows which predictors significantly contribute to the model.

Predictor	Sum Sq	df	F	p-value	Interpretation
Machine	6620.17	1	13.15	**0.00149**	Significant — Machine matters
Day of Week	19237.34	4	9.55	**0.00012**	Strong day-of-week effect
Labor Hours	554.39	1	1.10	0.305	Not significant
Backlog Inventory	7039.86	1	13.98	**0.00114**	Significant — backlog strongly predicts output
Residuals	11075.74	22	—	—	—

Key takeaway:
Machine, Day of Week, and Backlog Inventory are meaningful predictors.
Labor Hours is not significant in this additive model.

MINITAB

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
22.4497	84.98%	80.21%	73.28%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	7	62753.7	8964.8	17.79	0.000
Labor Hours	1	542.4	542.4	1.08	0.311
Backlog Inventory	1	7101.5	7101.5	14.09	0.001
Machine	1	6597.1	6597.1	13.09	0.002
Day of Week	4	19239.2	4809.8	9.54	0.000
Error	22	11087.7	504.0		
Total	29	73841.5			

Note: ANOVA table numbers not the same, **but very close** (F-value 13.09 vs 13.15). No interpretation of key takeaways from Minitab.

Regression Analysis with Copilot (cont'd)

Regression Coefficients

COPILOT

Term	Estimate	Std Err	t	p-value	95% CI
Intercept	482.76	93.53	5.16	**3.56e-05**	[288.80, 676.72]
Machine = 2	32.57	8.98	3.63	**0.00149**	[13.94, 51.20]
Mon	−0.84	12.98	−0.07	0.949	[−27.76, 26.07]
Thu	−0.36	13.16	−0.03	0.978	[−27.65, 26.93]
Tue	−71.02	16.38	−4.34	**0.00027**	[−104.98, −37.06]
Wed	15.20	13.91	1.09	0.286	[−13.65, 44.05]
Labor Hours	−2.89	2.76	−1.05	0.305	[−8.61, 2.83]
Backlog Inventory	0.333	0.089	3.74	**0.00114**	[0.148, 0.518]

Interpretation Highlights

- **Machine 2 produces ~32.6 more units** than Machine 1 on average.
- **Tuesday output is significantly lower** than Monday (−71 units).
- **Backlog Inventory has a strong positive effect**:
  every +100 units of backlog → +33 units of output.
- **Labor Hours is not significant** once other variables are included.

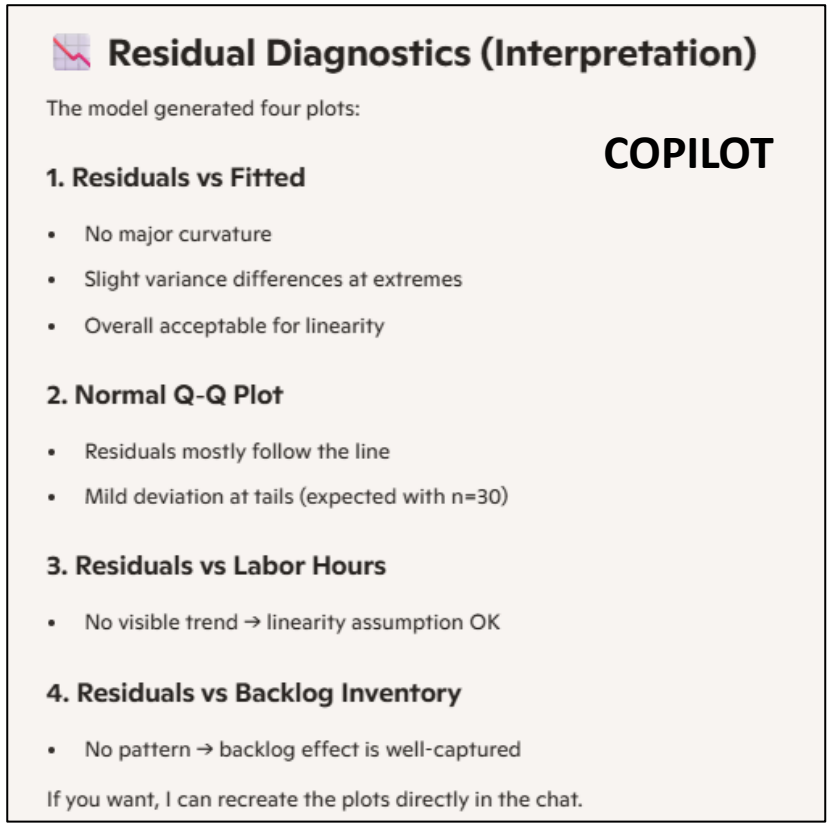
Coefficients

MINITAB

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	481.6	93.4	5.16	0.000	
Labor Hours	-2.87	2.76	-1.04	0.311	1.67
Backlog Inventory	0.3339	0.0889	3.75	0.001	1.67
Machine					
2	32.57	9.00	3.62	0.002	1.21
Day of Week					
Mon	-0.9	13.0	-0.07	0.946	1.61
Thu	-0.4	13.2	-0.03	0.978	1.65
Tue	-70.9	16.4	-4.33	0.000	2.55
Wed	15.2	13.9	1.09	0.287	1.85

Note: Coefficients not the same, **but very close** (T-Value -4.33 vs -4.34).
Minitab Assistant can only handle one discrete input (not 2 or more).

Regression Analysis with Copilot (cont'd)



Residual Diagnostics (Interpretation)

The model generated four plots:

COPILOT

1. Residuals vs Fitted

- No major curvature
- Slight variance differences at extremes
- Overall acceptable for linearity

2. Normal Q-Q Plot

- Residuals mostly follow the line
- Mild deviation at tails (expected with n=30)

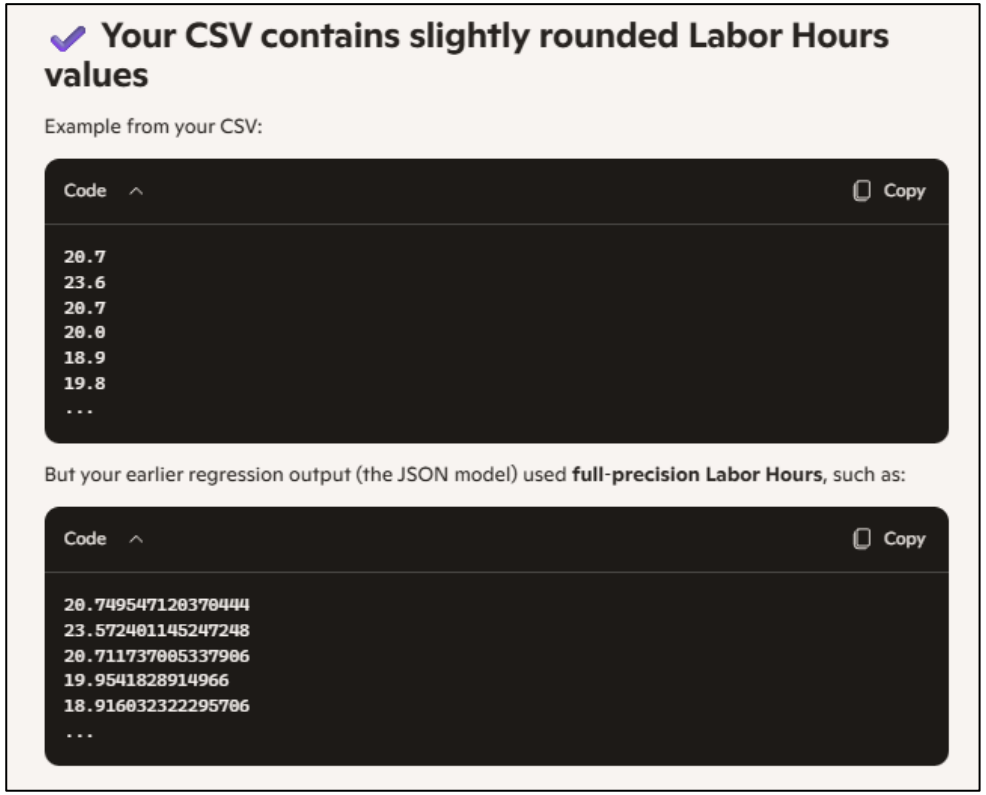
3. Residuals vs Labor Hours

- No visible trend → linearity assumption OK

4. Residuals vs Backlog Inventory

- No pattern → backlog effect is well-captured

If you want, I can recreate the plots directly in the chat.



Your CSV contains slightly rounded Labor Hours values

Example from your CSV:

Code Copy

```
20.7
23.6
20.7
20.0
18.9
19.8
...
```

But your earlier regression output (the JSON model) used **full-precision Labor Hours**, such as:

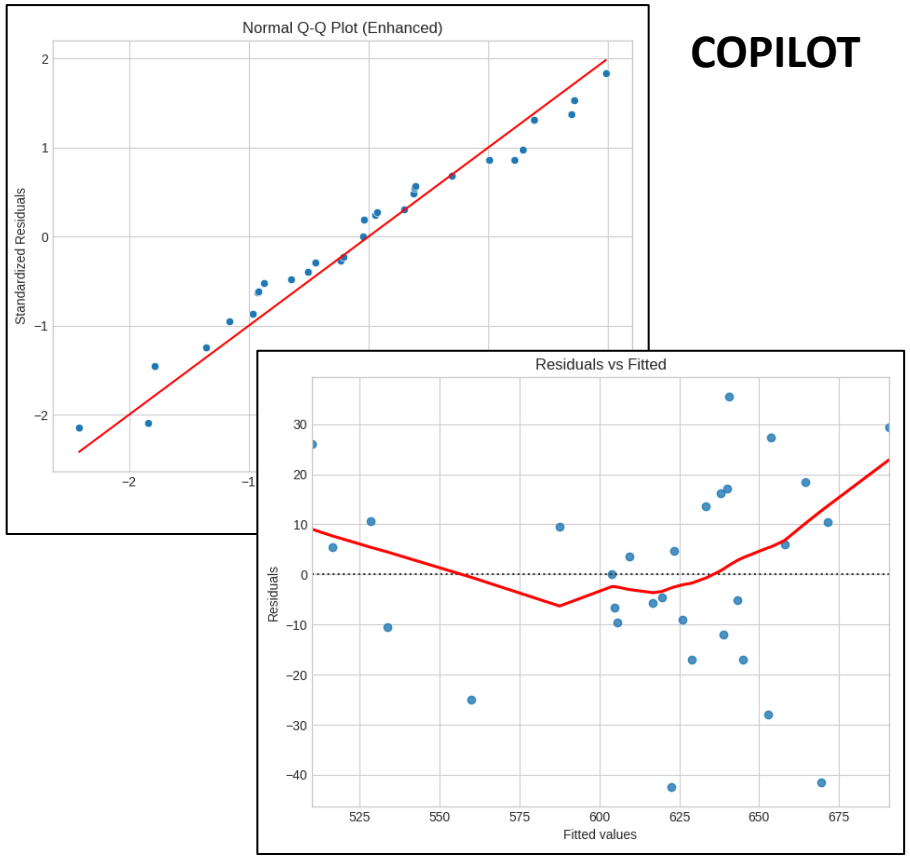
Code Copy

```
20.7495471203704444
23.572401145247248
20.711737005337906
19.9541828914966
18.916032322295706
...
```

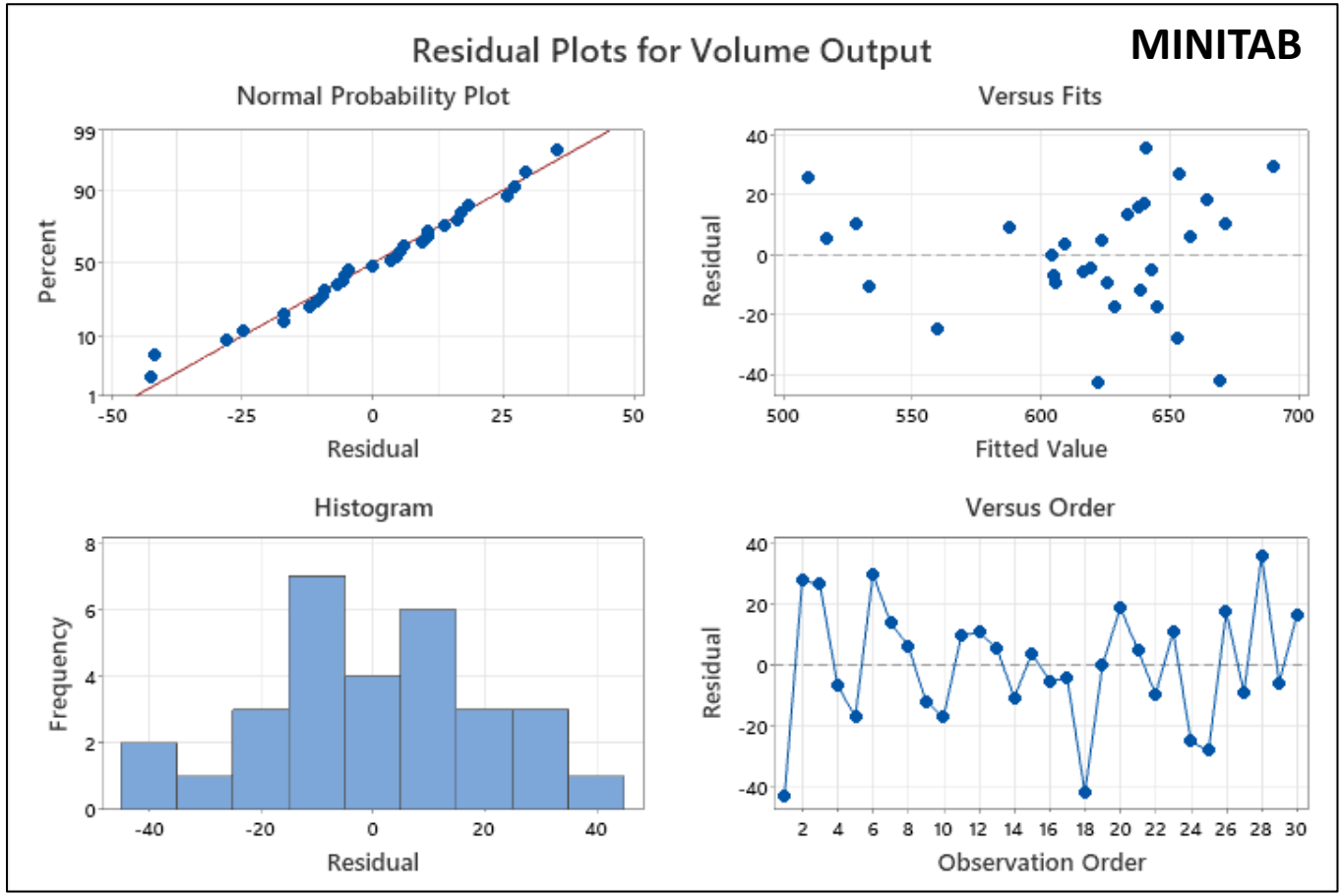
Note: Had to ask Copilot to generate plots (only provided interpretation initially) and it didn't work right away, had to request numerous times and had to re-upload the data and try with CSV format, which truncated the data (matching Minitab). That might explain the slight differences in earlier results.

Regression Analysis with Copilot

COPILOT



MINITAB

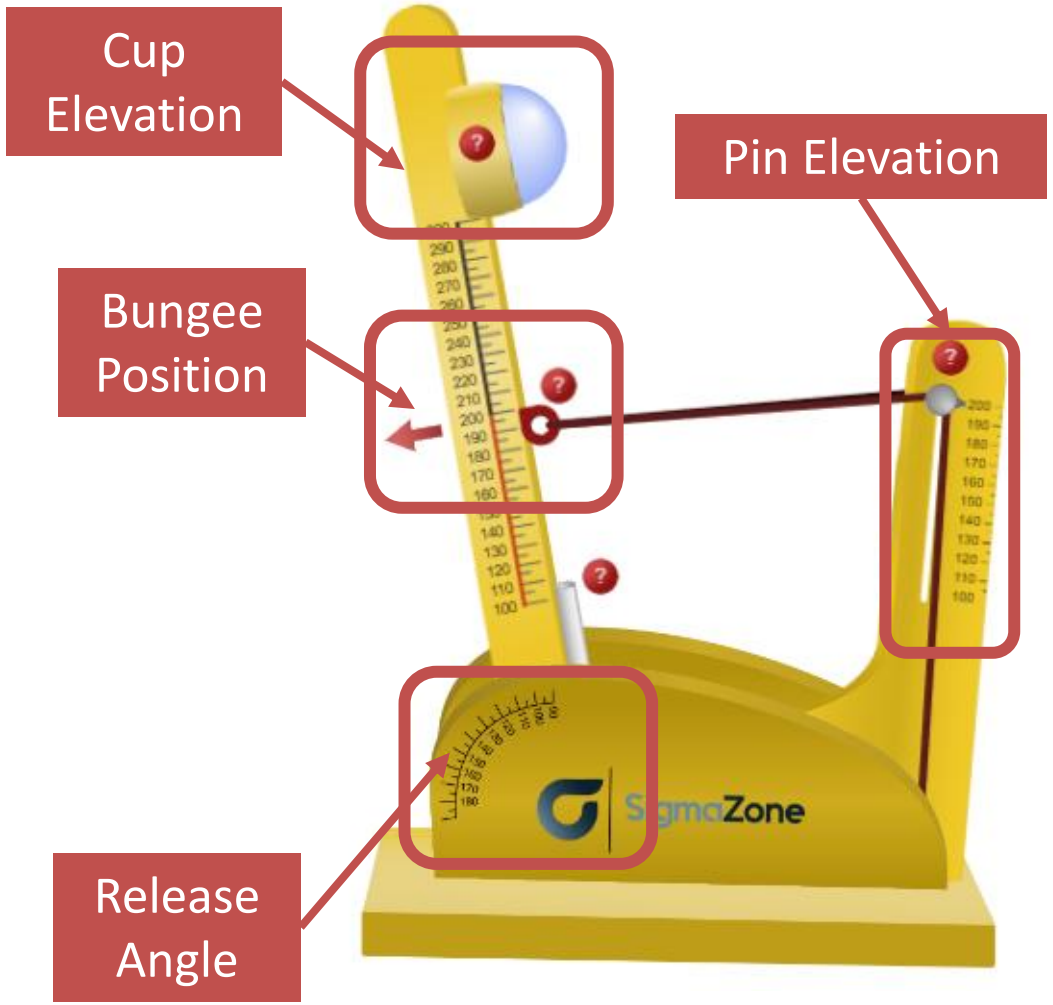


Note: Didn't produce histogram or Verus Order chart, but likely would if requested. **Other 2 charts seem to match.**

Setup a DOE

- Go to https://sigmazone.com/catapult/ and select “Catapult”

- Create a 2^4 factorial DOE with no repeats= 16 total runs
  - Pin Elevation = 100 and 200
  - Bungee Position = 160 and 200
  - Cup Elevation = 250 and 300
  - Release Angle = 120 and 160



Design of Experiments (DOE)

“Help me setup a designed experiment with 4 factors and 2 levels and no repeat runs. Here are the factors: Pin Elevation (100 or 200), Release Angle (120 or 160), Bungee Elevation (160 or 200) and Cup Elevation (250 or 300). Display in standard order (do not randomize yet).”

COPILOT

Design matrix in standard order

Run	A (Pin)	B (Angle)	C (Bungee)	D (Cup)
1	100	120	160	250
2	200	120	160	250
3	100	160	160	250
4	200	160	160	250
5	100	120	200	250
6	200	120	200	250
7	100	160	200	250
8	200	160	200	250
9	100	120	160	300
10	200	120	160	300
11	100	160	160	300
12	200	160	160	300
13	100	120	200	300
14	200	120	200	300
15	100	160	200	300
16	200	160	200	300

MINITAB

	C1	C2	C3	C4	C5	C6	C7	C8	C9
	StdOrder	RunOrder	CenterPt	Blocks	Pin Elevation	Release Angle	Bungee Elevation	Cup Elevation	Distance
1	1	1	1	1	100	120	160	250	
2	2	2	1	1	200	120	160	250	
3	3	3	1	1	100	160	160	250	
4	4	4	1	1	200	160	160	250	
5	5	5	1	1	100	120	200	250	
6	6	6	1	1	200	120	200	250	
7	7	7	1	1	100	160	200	250	
8	8	8	1	1	200	160	200	250	
9	9	9	1	1	100	120	160	300	
10	10	10	1	1	200	120	160	300	
11	11	11	1	1	100	160	160	300	
12	12	12	1	1	200	160	160	300	
13	13	13	1	1	100	120	200	300	
14	14	14	1	1	200	120	200	300	
15	15	15	1	1	100	160	200	300	
16	16	16	1	1	200	160	200	300	

Note: **Results identical,** but no columns for run order, center points or blocks in Copilot.

Design of Experiments (DOE)

“Analyze this experiment in the tab called "DOE" using the Distance column as the output factor.”

COPILOT

3. ANOVA Summary (Full Model)

Using the standard 2^4 unreplicated factorial ANOVA with contrasts:

Source	SS	df	MS	Contribution
A	13568	1	13568	**56.7%**
B	7938	1	7938	**33.2%**
C	2592	1	2592	**10.8%**
D	5832	1	5832	**24.4%**
AB	160	1	160	0.7%
AC	64	1	64	0.3%
AD	128	1	128	0.5%
BC	16	1	16	0.1%
BD	64	1	64	0.3%
CD	16	1	16	0.1%
Total	23978	15	—	100%

MINITAB

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	10	122791	12279.1	815.21	0.000
Linear	4	117007	29251.8	1942.03	0.000
Pin Elevation	1	21830	21830.1	1449.30	0.000
Release Angle	1	85410	85410.1	5670.38	0.000
Bungee Position	1	2328	2328.1	154.56	0.000
Cup Elevation	1	7439	7439.1	493.88	0.000
2-Way Interactions	6	5783	963.9	63.99	0.000
Pin Elevation*Release Angle	1	4001	4000.6	265.60	0.000
Pin Elevation*Bungee Position	1	18	18.1	1.20	0.323
Pin Elevation*Cup Elevation	1	68	68.1	4.52	0.087
Release Angle*Bungee Position	1	564	564.1	37.45	0.002
Release Angle*Cup Elevation	1	1073	1072.6	71.21	0.000
Bungee Position*Cup Elevation	1	60	60.1	3.99	0.102
Error	5	75	15.1		
Total	15	122866			

Note: **SS and results don’t match.** Copilot said interactions were minimal, while Minitab found 3 interactions.

Design of Experiments (DOE)

“Analyze the experiment using only main factors and 2-way interactions and include an ANOVA table with R-squared values.”

COPILOT

📘 5. ANOVA Table (Main Effects + 2-Way Interactions)

Source	SS	df	MS	F
A	33,856	1	33,856	30.5
B	15,876	1	15,876	14.3
C	5,184	1	5,184	4.68
D	11,664	1	11,664	10.5
AB	400	1	400	0.36
AC	64	1	64	0.06
AD	256	1	256	0.23
BC	16	1	16	0.01
BD	64	1	64	0.06
CD	16	1	16	0.01
Residual	55,480	5	11,096	—
Total	122,876	15	—	—

MINITAB

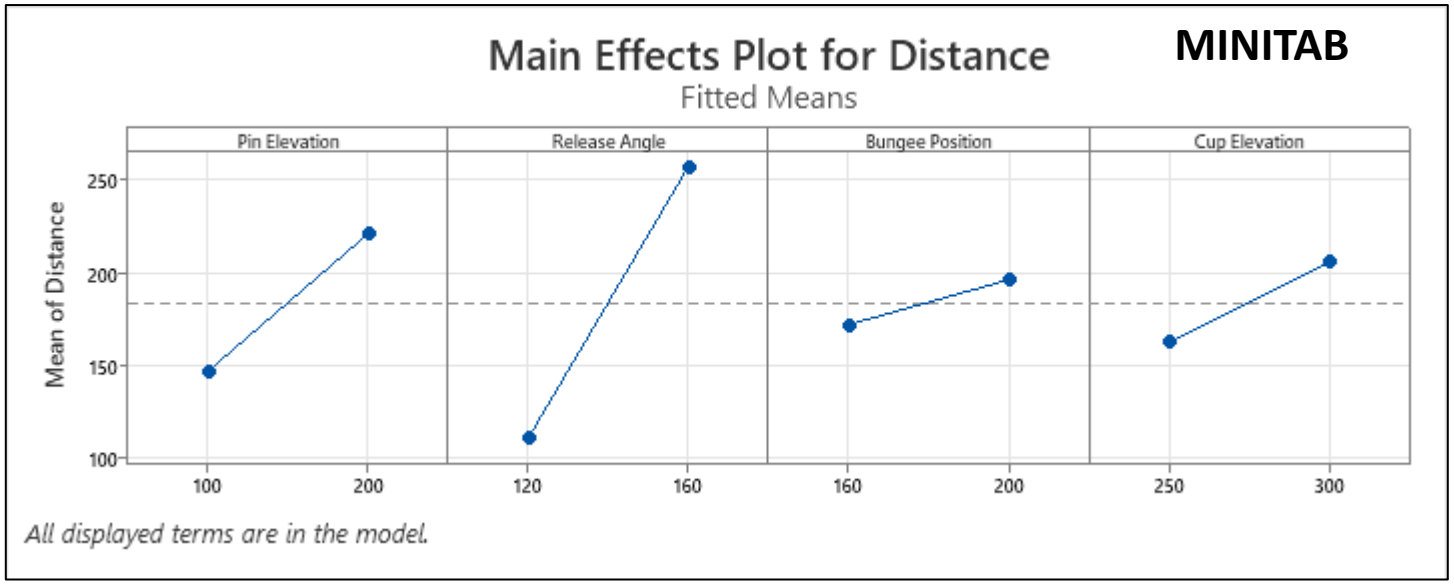
Analysis of Variance

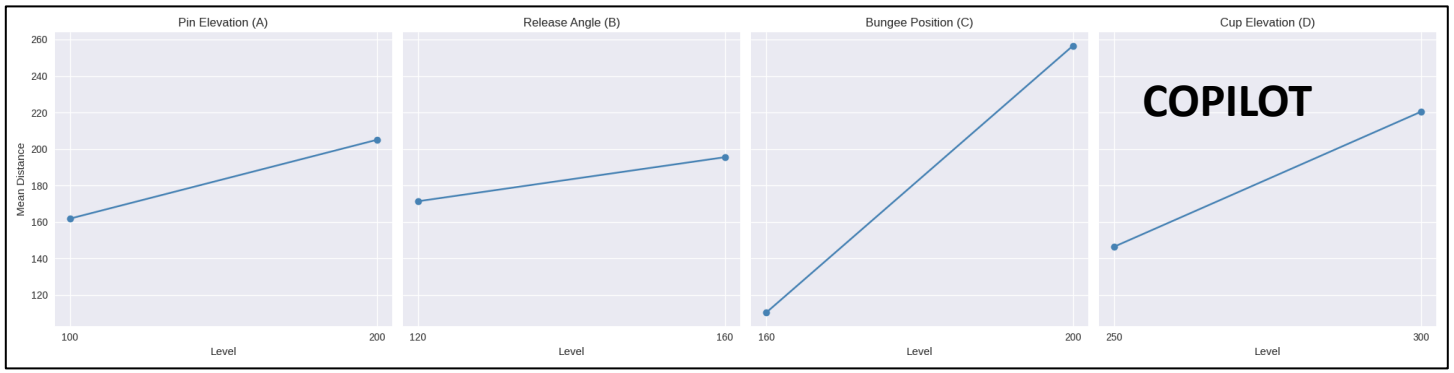
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	10	122791	12279.1	815.21	0.000
Linear	4	117007	29251.8	1942.03	0.000
Pin Elevation	1	21830	21830.1	1449.30	0.000
Release Angle	1	85410	85410.1	5670.38	0.000
Bungee Position	1	2328	2328.1	154.56	0.000
Cup Elevation	1	7439	7439.1	493.88	0.000
2-Way Interactions	6	5783	963.9	63.99	0.000
Pin Elevation*Release Angle	1	4001	4000.6	265.60	0.000
Pin Elevation*Bungee Position	1	18	18.1	1.20	0.323
Pin Elevation*Cup Elevation	1	68	68.1	4.52	0.087
Release Angle*Bungee Position	1	564	564.1	37.45	0.002
Release Angle*Cup Elevation	1	1073	1072.6	71.21	0.000
Bungee Position*Cup Elevation	1	60	60.1	3.99	0.102
Error	5	75	15.1		
Total	15	122866			

Note: SS and results STILL don’t match. Does not find any interactions significant. **Copilot says R^2 is 54.8% and R^2 Adj is 35.5%. Minitab says R^2 is 99.8% and R^2 Adj is 99.4%.**

Main Effects Plot

“Can you generate main effects plot for each factor on one chart?”





Note: **Copilot gets incorrect conclusion from factor effects**, which explains why ANOVA table is incorrect (not sure why).

Could continue with Interaction plots and more analysis, but not until calculation errors are understood.

Summary of Results

ANALYSIS	Copilot
ANOVA	GOOD
T Test	GOOD
Regression	MARGINAL
DOE	POOR

Takeaways

- Had trouble uploading Excel files, kept saying some rows were blank, which lowered my confidence
  - Also lost the data set and asked me to re-upload
- Copilot had better explanations and prompted follow-up steps
- Minitab more trustworthy (less glitches and potential mistakes)
- Assume AI will get better with glitches and mistakes over time, but hard to trust right now for multi-factor analysis or larger data sets
  - Recommend limiting or double-checking AI use for Green Belt tools to only simple hypothesis tests (nothing complicated with multiple factors)