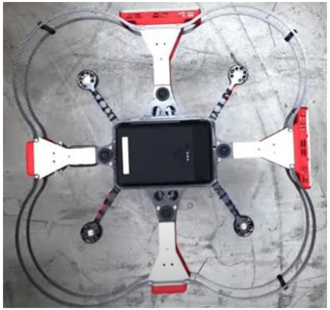
Drones Flying in Warehouses: An Application of Attribute Gauge Analysis



David Trindade

STAT-TECH

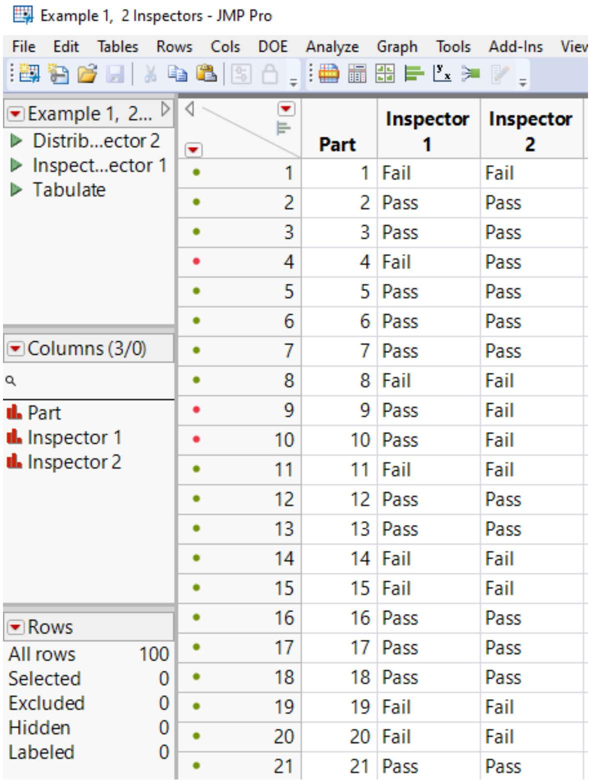
July 2022

Attribute Gauge Analysis

- Attribute gauge analysis is typically applied to compare agreement or lack thereof between two or more rating approaches to a problem.
- For example, two inspectors may have differences of opinion as to whether a part is conforming (Pass) or non-conforming (Fail) based on consideration of specific quality indicators for individual parts.
- How do we quantitively measure the degree of agreement?

Example 1: Two Inspectors

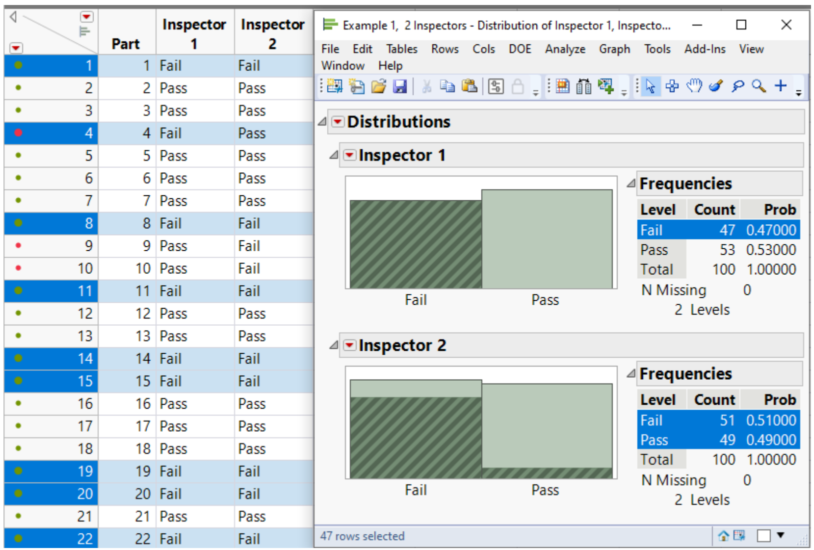
- Assume two inspectors (Inspector 1, Inspector 2) are presented with a list of critical characteristics on 100 parts and asked to determine whether each part should be classified as a “Pass” or a “Fail”.
- The results (partial) are shown in the data table.
- Note variables are all nominal.



	Part	Inspector 1	Inspector 2
1	1	Fail	Fail
2	2	Pass	Pass
3	3	Pass	Pass
4	4	Fail	Pass
5	5	Pass	Pass
6	6	Pass	Pass
7	7	Pass	Pass
8	8	Fail	Fail
9	9	Pass	Fail
10	10	Pass	Fail
11	11	Fail	Fail
12	12	Pass	Pass
13	13	Pass	Pass
14	14	Fail	Fail
15	15	Fail	Fail
16	16	Pass	Pass
17	17	Pass	Pass
18	18	Pass	Pass
19	19	Fail	Fail
20	20	Fail	Fail
21	21	Pass	Pass

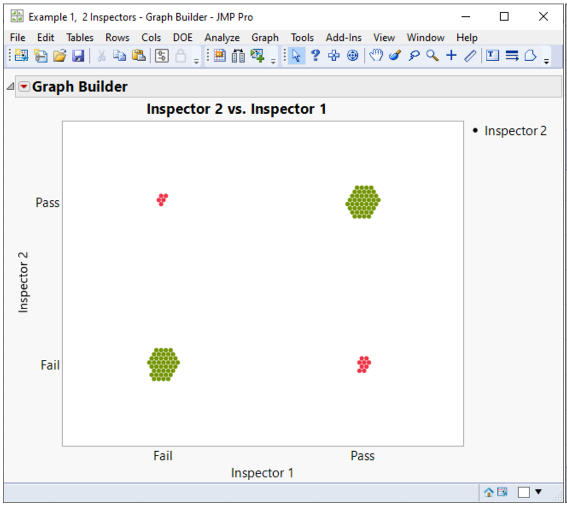
Analysis of Inspector Comparison Data

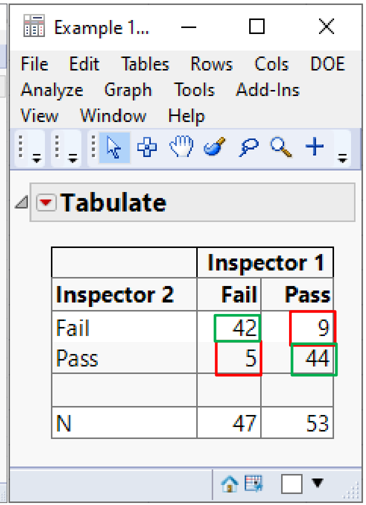
- A first step could be to look at the two classification distributions and use dynamic linking to compare.
- For example, if we click on Fail histogram bar for Inspector 1, we see mostly matches for Inspector 2 (Fail, Fail rows) but note five instances of disagreement (Fail, Pass rows) in the data table.



Visualization of Inspector Comparison Data

We can use **Graph Builder** with **Tabulate** to view agree and disagree counts between the two Inspectors.

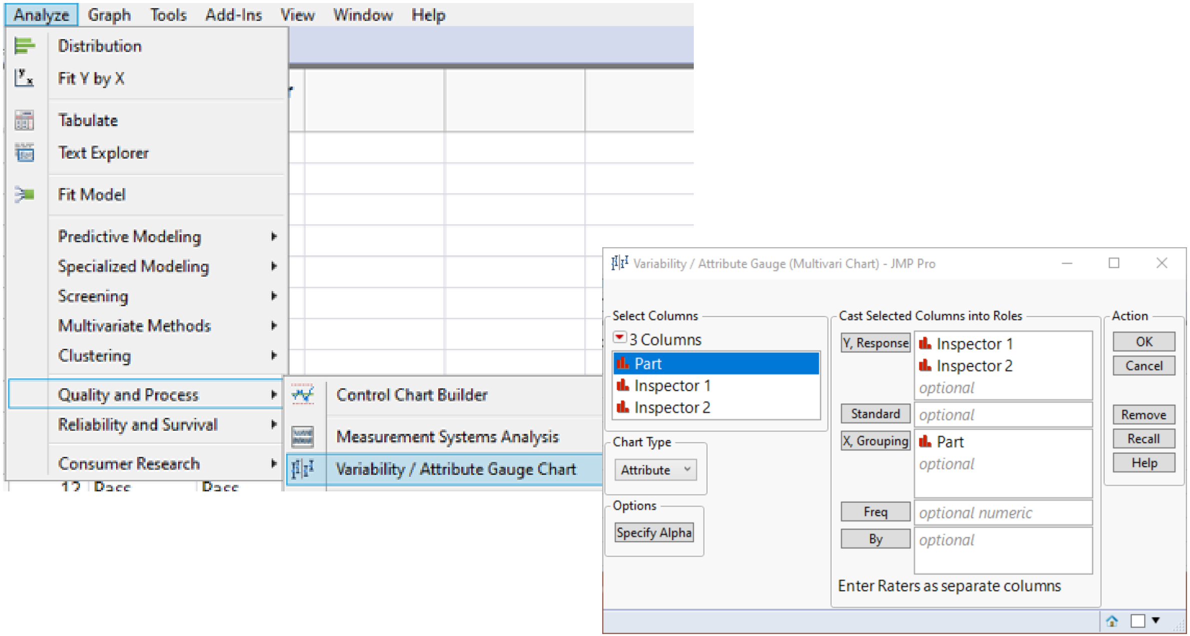




	Inspector 1	
Inspector 2	**Fail**	**Pass**
Fail	42	9
Pass	5	44
N	47	53

The inspectors agree on a classification for (42+44)/100 = 86% of the parts and disagree on (9+5)/100 =14%.

Attribute Gauge Analysis in JMP

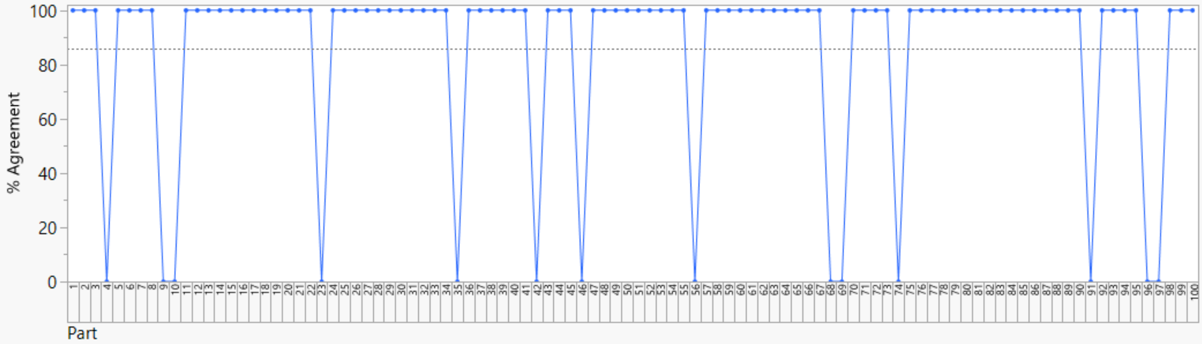


Select **Analyze > Quality and Process > Variability/ Attribute Gauge Chart**, and cast roles as shown.

Attribute Gauge Analysis Report

Attribute Gauge

Gauge Attribute Chart



- The **Gauge Attribute Chart** shows the % agreement (100% for agree, 0% for disagree) for each part.
- The left chart shows the overall **% Agreement** by Inspector. Since the comparison is between only the two Inspectors, both Inspectors have the same 86% agreement value.

Agreement Report

The **Agreement Report** table is a numerical summary of the overall 86**% Agreement** with 95% confidence intervals.

Agreement Report

Rater	% Agreement	95% Lower CI	95% Upper CI
Inspector 1	86.0000	83.3769	88.2675
Inspector 2	86.0000	83.3769	88.2675

Number Inspected	Number Matched	% Agreement	95% Lower CI	95% Upper CI
100	86	86.000	77.863	91.474

Agreement Comparisons Report

The **Agreement Comparisons** report includes the Cohen **Kappa** index (0.7203) which is designed **to correct for agreement by chance alone.**

Agreement Comparisons

Rater	Compared with Rater	Kappa	.2 .4 .6 .8	Standard Error
Inspector 1	Inspector 2	0.7203		0.0691

Agreement within Raters

Rater	Number Inspected	Number Matched	Rater Score	95% Lower CI	95% Upper CI
Inspector 1	100	100	100.000	96.3007	100.000
Inspector 2	100	100	100.000	96.3007	100.000

Agreement across Categories

Category	Kappa	.2 .4 .6 .8	Standard Error
Fail	0.7199		0.1000
Pass	0.7199		0.1000
Overall	0.7199		0.1000

Agreement by Chance

- What is “agreement by chance” and how can we estimate it?
- Consider two raters, R1 and R2. We’ll assume totally random choices for each rater for each sample, e.g., each part.
- We further assume that the probability a rater selects either choice (Pass or Fail) over the other is 50%.
- 100 samples or trials are therefore randomly categorized by Pass/Fail for each rater, similar to flipping a coin for each choice.
- What’s the expect fraction of agreements by chance?

Agreement by Chance for Two Raters

- Similar to tossing two coins, there are only four possible and equally likely chance outcomes between the two Inspectors for each part:

	Rater 1 Fail	**Rater 1 Pass**
Rater 2 Fail	Agree	Disagree
Rater 2 Pass	Disagree	Agree

- Therefore, the probability of agreement by chance alone is 2/4 = 50%.

The Cohen[1] Kappa Statistic

- The Kappa Statistic is meant to **correct for the expected probability of agreement by chance**.
- The simple formula for the Kappa statistic κ is

$$\frac{\left(\%\ Agreement - Expected\ by\ Chance\ from\ Data\right)}{\left(1 - Expected\ by\ Chance\ from\ Data\right)}$$

- How do we estimate the *Expected Agreement by Chance from Data*?

Estimation[2] of Cohen Kappa Statistic for Two Inspector Example

Here is the tabulated data. **Agreement by chance** is estimated as the **sum of the products** of the marginal fractions for each Pass/Fail type.

	A	B	C	D	E	F
1						
2				Inspector 1		
3				Fail	Pass	Sum
4		Inspector 2	Fail	42	9	51
5			Pass	5	44	49
6				47	53	100
7						
8			Agree	86%		
9			Disagree	14%		
10						
11			Agree by Chance	49.94%		
12						
13			Kappa	72.03%		

Agree: =(D4+E5)/100 = (42+44)/100

Disagree: =(D5+E4)/F6 = (5+9)/100

Agree by Chance: =(F4/F6)*(D6/F6)+(F5/F6)*(E6/F6) = (51/100)*(47/100) + (49/100)*(53/100)

Kappa: =(D8-D11)/(1-D11) = (86% - 49.94%)/(1 - 49.94%)

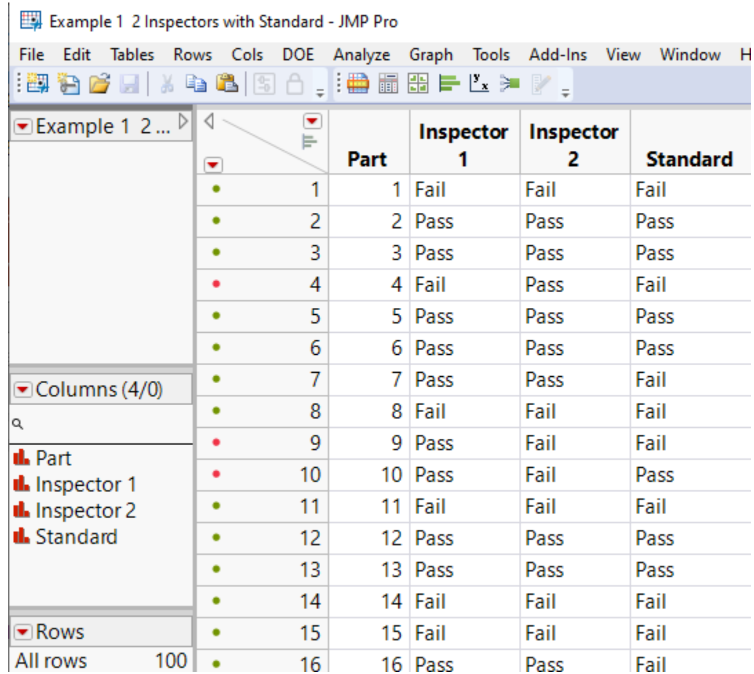
Interpreting Kappa κ

Here are some guidelines[3] for interpreting Kappa κ.

Kappa	Agreement
$\kappa > .75$	Excellent
$.40 < \kappa < .75$	Good
$0 < \kappa < .40$	Marginal/Poor

Incorporating a Standard (“Effectiveness”)

- Returning to the two Inspectors example, assume the correct part classification was either known or subsequently confirmed.
- How **accurate** are the Inspectors’ choices?
- We enter the true determination in a separate “**Standard**” column as shown in the partial table.

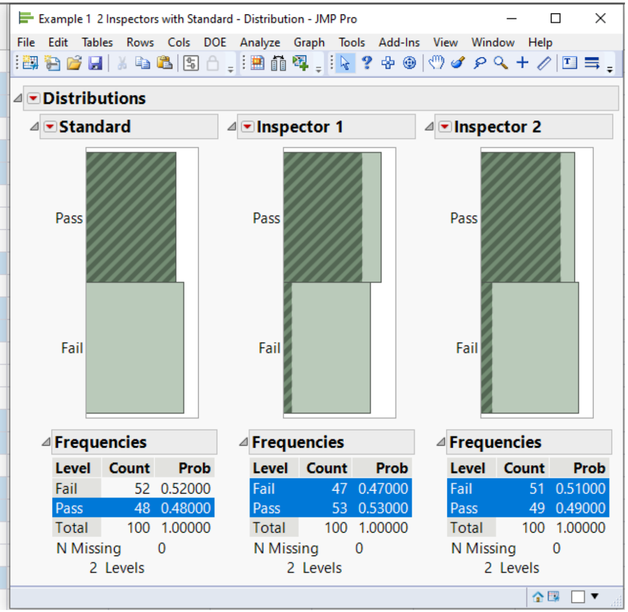


	Part	Inspector 1	Inspector 2	Standard
1	1	Fail	Fail	Fail
2	2	Pass	Pass	Pass
3	3	Pass	Pass	Pass
4	4	Fail	Pass	Fail
5	5	Pass	Pass	Pass
6	6	Pass	Pass	Pass
7	7	Pass	Pass	Fail
8	8	Fail	Fail	Fail
9	9	Pass	Fail	Fail
10	10	Pass	Fail	Pass
11	11	Fail	Fail	Fail
12	12	Pass	Pass	Pass
13	13	Pass	Pass	Pass
14	14	Fail	Fail	Fail
15	15	Fail	Fail	Fail
16	16	Pass	Pass	Fail

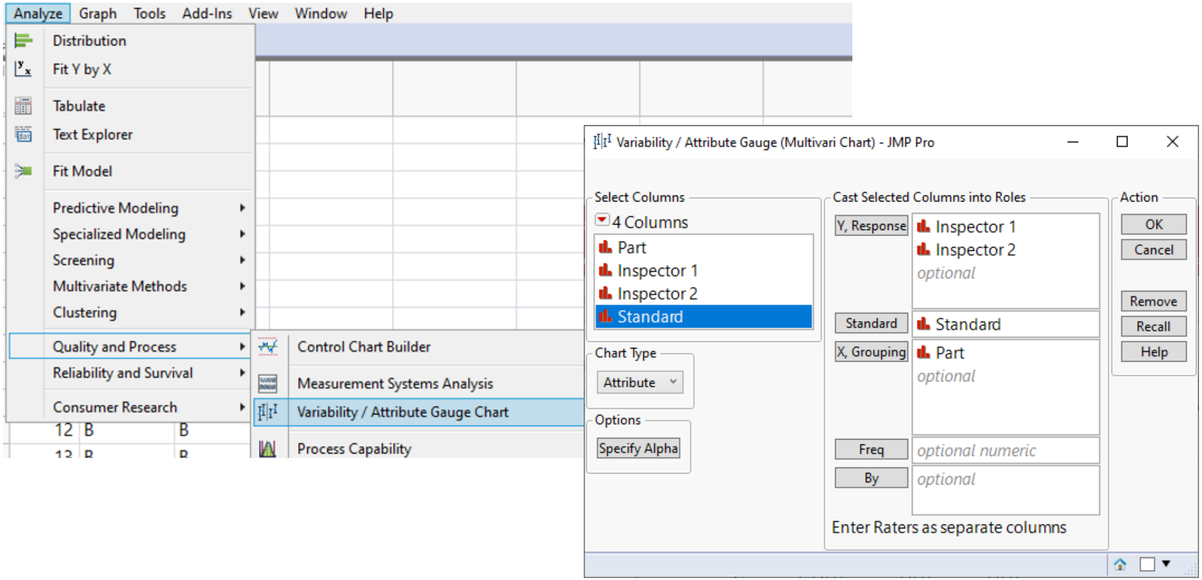
Distributions and Dynamic Linking

By selecting, for example, “Pass” on the Standard histogram bar, we can see several incorrect “Fails” (false alarms) by each inspector.

	Part	Inspector 1	Inspector 2	Standard
1	1	Fail	Fail	Fail
2	2	Pass	Pass	Pass
3	3	Pass	Pass	Pass
4	4	Fail	Pass	Fail
5	5	Pass	Pass	Pass
6	6	Pass	Pass	Pass
7	7	Pass	Pass	Fail
8	8	Fail	Fail	Fail
9	9	Pass	Fail	Fail
10	10	Pass	Fail	Pass
11	11	Fail	Fail	Fail
12	12	Pass	Pass	Pass
13	13	Pass	Pass	Pass
14	14	Fail	Fail	Fail
15	15	Fail	Fail	Fail
16	16	Pass	Pass	Fail
17	17	Pass	Pass	Pass
18	18	Pass	Pass	Pass
19	19	Fail	Fail	Fail
20	20	Fail	Fail	Pass
21	21	Pass	Pass	Pass
22	22	Fail	Fail	Fail
23	23	Pass	Fail	Fail
24	24	Pass	Pass	Pass
25	25	Fail	Fail	Fail



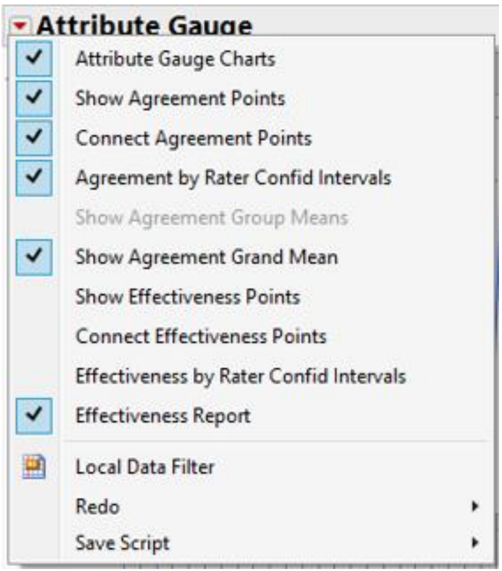
Attribute Gauge Analysis in JMP

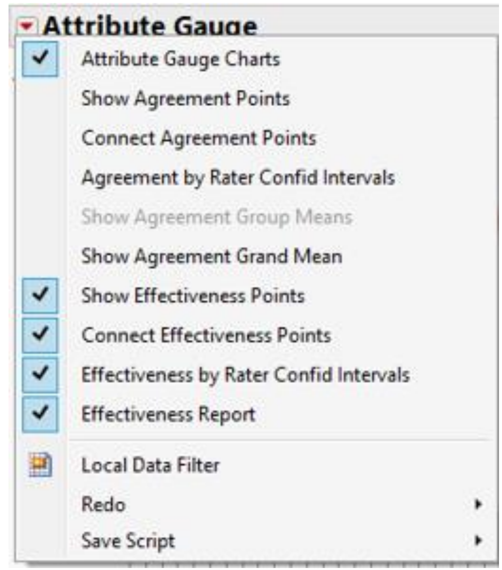


Select **Analyze > Quality and Process > Variability/ Attribute Gauge Chart**, and cast roles, including Standard column, as shown.

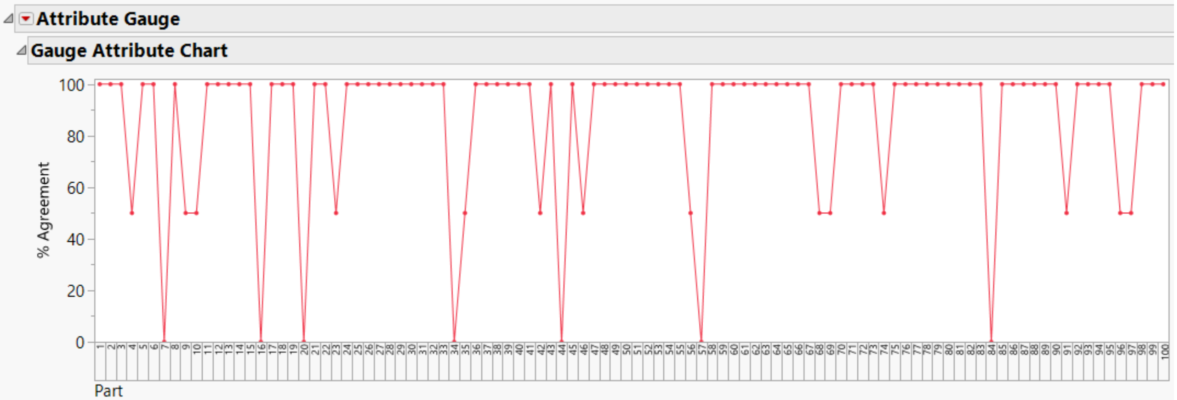
Effectiveness Report

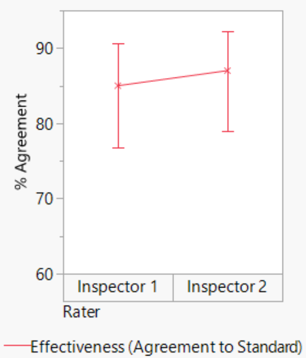
Under the **Attribute Gauge** red hotspot, unselect **Agreement** checkboxes (default settings) and select **Effectiveness** boxes as shown below.





Effectiveness: Agreement to Standard





- The **Gauge Attribute Chart now** shows the % Agreement (0%, 50%, or 100%) of the two Inspectors **to the Standard** for each part. A “0%” implies both Inspectors misdiagnosed a part (7 events). A “50%” signifies one inspector with correct classification.
- The left chart shows the overall **%** Agreement **to the Standard** for each Inspector.

Effectiveness Report

The **Effectiveness Report** incorporates Pass/Fail comparisons to the Standard for each Inspector.

Effectiveness Report

Agreement Counts

Rater	Correct(Fail)	Correct(Pass)	Total Correct	Incorrect(Fail)	Incorrect(Pass)	Grand Total
Inspector 1	42	43	85	10	5	100
Inspector 2	45	42	87	7	6	100

Effectiveness

Rater	Effectiveness	95% Lower CI	95% Upper CI	Error rate
Inspector 1	85.0000	76.7164	90.6940	0.1500
Inspector 2	87.0000	79.0196	92.2428	0.1300
Overall	86.0000	80.5101	90.1330	0.1400

Incorrect(Fail) means a Fail was incorrectly classified as a Pass.
Incorrect(Pass) means a Pass was incorrectly classified as a Fail.

Effectiveness Report: Misclassifications

The **Misclassifications** Summary shows 17 actual Fail parts misclassed as Pass and 11 Pass parts misclassed as Fail.

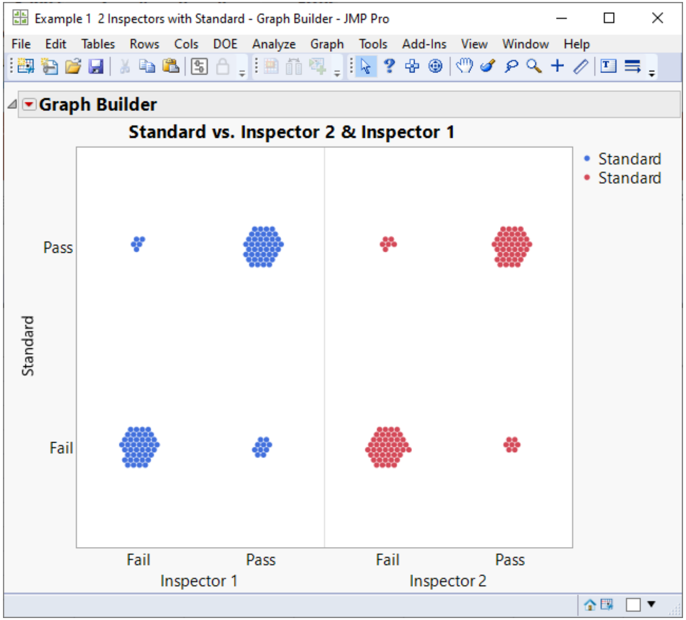
Misclassifications

Standard Level	Fail	Pass
Fail	.	11
Pass	17	.
Other	0	0

Classifications		
	Inspector 1	Inspector 2
Standard: Pass		
Classified as Pass	43	42
Misclassified as Fail	5	6
Standard: Fail		
Classified as Fail	42	45
Misclassified as Pass	10	7

Misclassifications Visualization

Using **Graph Builder**, we can view the classifications and misclassifications by each inspector.



Effectiveness Report: Conformance Report

Defining the “conformance” can be useful when classifying parts as pass-fail or as defective or not. Here, NonConform is defined as Fail, and Conform, as Pass. JMP provides probability estimates of **False Alarms** and **Misses**.

Conformance Report

Rater	P(False Alarms)	P(Misses)
Inspector 1	0.1042	0.1923
Inspector 2	0.1250	0.1346

Assumptions	
NonConform =	Fail
Conform =	Pass

Conformance Report: False Alarm

- **False Alarm:** Occurs the part is incorrectly classified as a Fail when it is correctly a Pass. (False positive)
- **P(False Alarms)** The number of parts that have been incorrectly judged to be Fails divided by the total number of parts that are judged to be Passes.
- For Inspector 1, for example, 5/(43+5) = 0.1042.

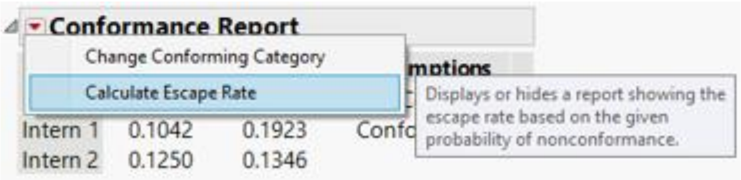
Conformance Report: Misses

- **Miss** The part is incorrectly classified as a Pass, when it actually is a Fail. (False negative)
- **P(Miss)** The number of parts that have been incorrectly judged to be Passes divided by the total number of parts that are judged to be Fails.
- For Inspector 1, for example, 10/(42+10) = 0.1923.

Conformance Report: Options

The **Conformance Report** red triangle menu contains the following options:

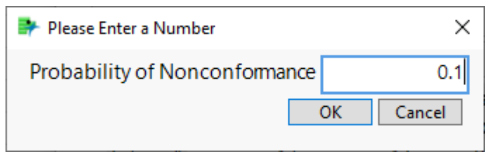
- **Change Conforming Category** Reverses the response category that is considered conforming.
- **Calculate Escape Rate** Calculates the Escape Rate, which is the probability that a **non-conforming part is produced and not detected**.



Conformance Report: Escape Rate

The **Escape Rate** is calculated as the probability that the process will produce a Fail part times the probability of a miss.

We specify a probability estimate that the process will produce a Fail part, also called the **Probability of Nonconformance**



Escape Rate

Rater	Escape Rate
Inspector 1	0.01923
Inspector 2	0.01346

Probability of Nonconformance = 0.1

Attribute Gauge Analysis in Practice

- Now that we have a feeling for the concepts of **agreement**, **effectiveness**, and **Kappa** index, let us see how we can apply the approach to a more complex problem in gauge analysis: **inventory tracking**.
- As part of a consulting project with a robotics company*, I was first introduced to the problem of drones flying in warehouse using OCR to read inventory labels on boxes in shelves.

Note: *Any data presented in this presentation is fictitious and not the actual results of studies by the company.*

*https://vimaan.ai/

Measurement System Analysis

- In measurement system analysis (MSA) the purpose is to determine if the variability in the measurement system is low enough to accurately detect differences in product-to-product variability.
- A further objective is to verify that the measurement system is **accurate**, **precise**, and **stable**.

Inventory Tracking

- In this study, the product to be measured via OCR on drones is the label on containers stored on racks in a warehouse. The measurement system must read the labels accurately.
- Furthermore, the measurements system will also validate the ability to detect “empty bins”, damaged items, counts, dimensions, etc.

Measurement System Analysis Features

- In gauge R&R studies, one concern addresses pure error, that is the **repeatability** of repeat measurements of the same label. Repeatability is a measure of **precision**.
- In addition, in Gauge R&R studies, a second concern is the bias associated with differences in tools, that is, differences among drones reading the same labels. This aspect is called **reproducibility**, which is a measure of **accuracy**.

Design for Measurement System Analysis

The design proposed will be a **crossed** study in which the same locations are measured multiple times (**repeatability**) across different bias factors (the drones for **reproducibility**).

The proposal will define several standards for the drones to measure. Thus, the comparisons will involve:

- ✓ **within- drone repeatability**
- ✓ **drone-to-drone agreement consistency**
- ✓ **drone-to-standard accuracy**.

Proposal for Drone Attribute Gauge Analysis

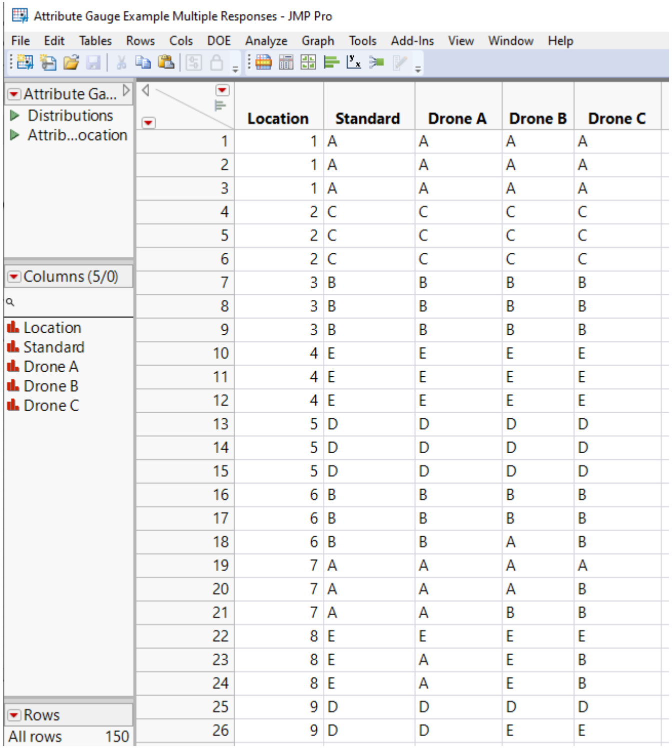
- The plan is to measure 50 locations (1 through 50). Three drones will be used to measure reproducibility, that is, drone-to-drone comparisons. There will be three passes for each location by each drone to measure repeatability.
- Multiple responses can be measured against each specific standard. The reading can be binary, that is, classified as either correct or incorrect. The reading also can provide status reporting for a location.

Possible Responses for Drone Attribute Gauge Analysis

Examples of different responses

1. How accurately can a drones read a standard label?
2. Are there missing or inverted labels?
3. Are inventory items in the correct location?
4. Is the quantity of boxes in a location correct?
5. Are any boxes damaged?

Proposal for Drone Attribute Gauge Inventory Analysis



	Location	Standard	Drone A	Drone B	Drone C
1	1	A	A	A	A
2	1	A	A	A	A
3	1	A	A	A	A
4	2	C	C	C	C
5	2	C	C	C	C
6	2	C	C	C	C
7	3	B	B	B	B
8	3	B	B	B	B
9	3	B	B	B	B
10	4	E	E	E	E
11	4	E	E	E	E
12	4	E	E	E	E
13	5	D	D	D	D
14	5	D	D	D	D
15	5	D	D	D	D
16	6	B	B	B	B
17	6	B	B	B	B
18	6	B	B	A	B
19	7	A	A	A	A
20	7	A	A	A	B
21	7	A	A	B	B
22	8	E	E	E	E
23	8	E	A	E	B
24	8	E	A	E	B
25	9	D	D	D	D
26	9	D	D	E	E

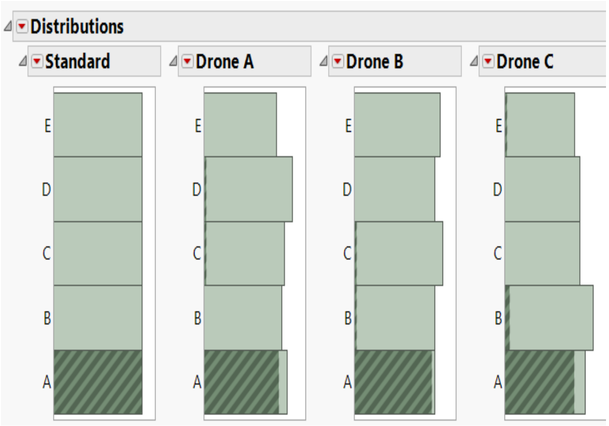
- **Multiresponses: Five characteristics (A,B,C,D,E) to check**
- **One characteristic is randomly specified for each of 50 locations** (1 through 50)
- 3 Drones (**Reproducibility**)
- 3 Passes for each location by each drone (**Repeatability**)
- Standards are specified for each location

Note: Data is made-up for illustration and not actual experimental results.

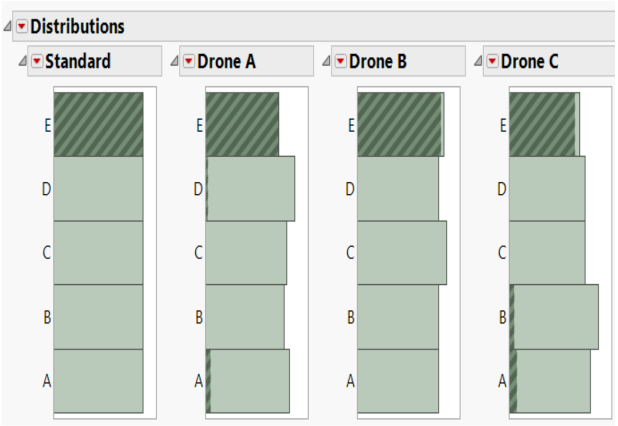
Distributions and Dynamic Linking

By selecting different standards on histogram bar, we can see misclassifications by drone.

Shows A's Misclassified

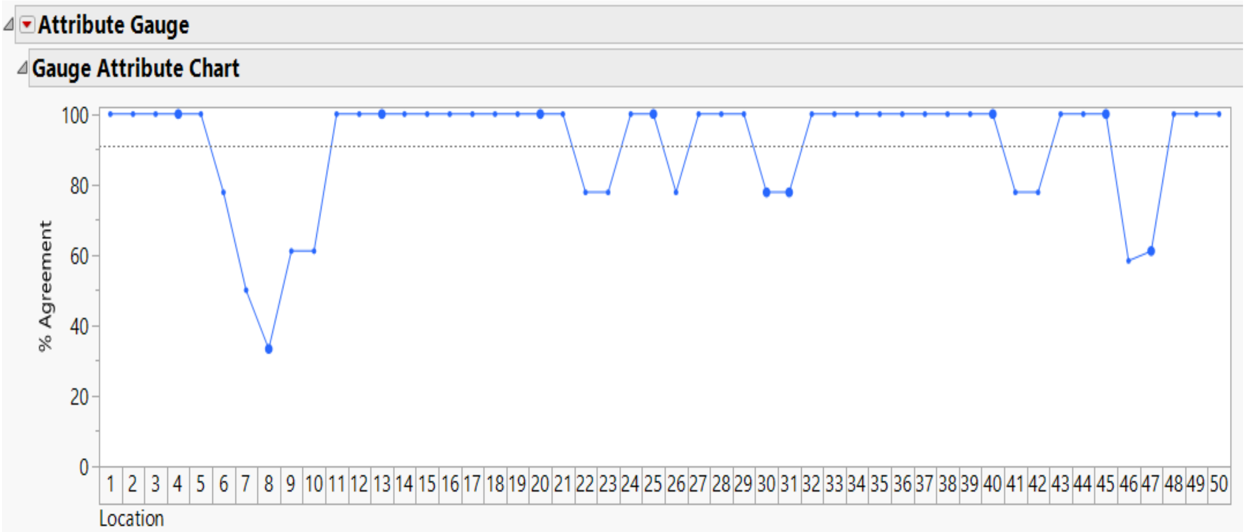


Shows E's Misclassified

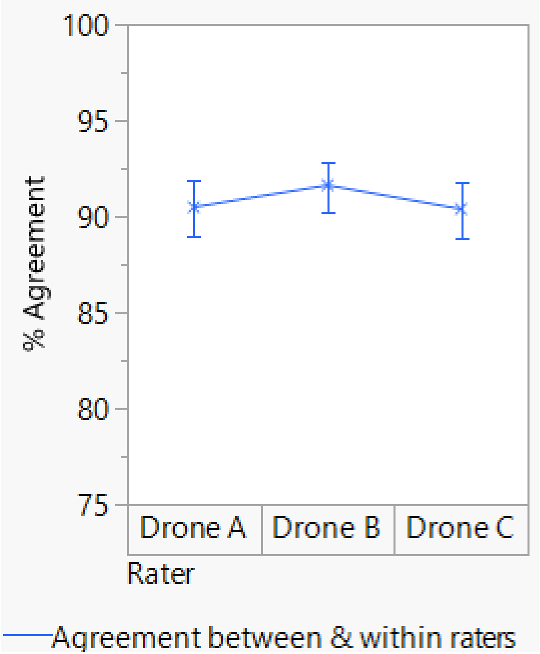


Analysis: Locations and Percent Agreement

Chart shows how well the **drones agreed** with each other for each location. Percent agreement dropped for locations 5 through 10, indicating locations were more difficult to categorize, prompting further investigation.



Analysis: % Agreement Between & Within Drones



Agreement Report

Rater	% Agreement	95% Lower CI	95% Upper CI
Drone A	90.4762	88.9082	91.8429
Drone B	91.6190	90.2190	92.8346
Drone C	90.3810	88.8043	91.7562

Number Inspected	Number Matched	% Agreement	95% Lower CI	95% Upper CI
50	36	72.000	58.335	82.526

Report shows agreement values and 95% confidence intervals of each drone with other drones or themselves.

Analysis: Agreement Comparisons

Agreement Comparisons

Rater	Compared with Rater	Kappa	Standard Error
Drone A	Drone B	0.8917	0.0287
Drone A	Drone C	0.8666	0.0315
Drone B	Drone C	0.9084	0.0266

Rater	Compared with Standard	Kappa	Standard Error
Drone A	Standard	0.9333	0.0229
Drone B	Standard	0.9583	0.0183
Drone C	Standard	0.9167	0.0254

Agreement within Raters

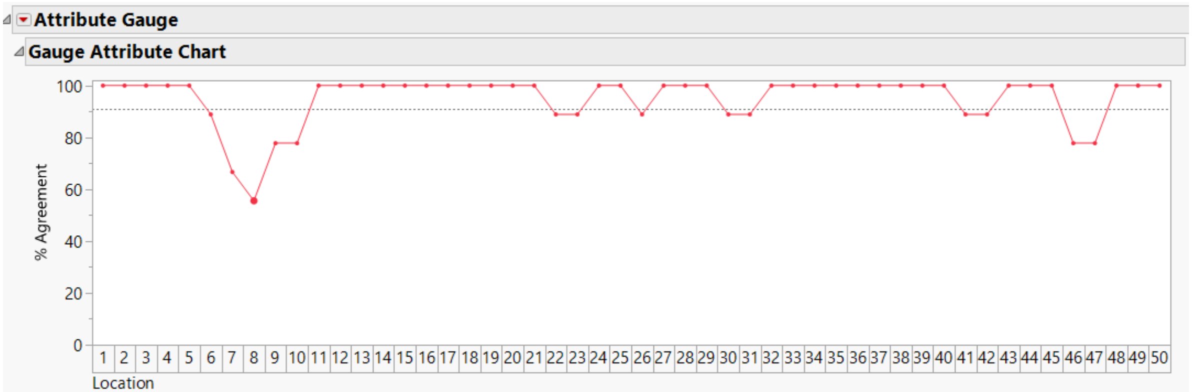
Rater	Number Inspected	Number Matched	Rater Score	95% Lower CI	95% Upper CI
Drone A	50	44	88.0000	76.1952	94.3824
Drone B	50	45	90.0000	78.6398	95.6524
Drone C	50	43	86.0000	73.8138	93.0492

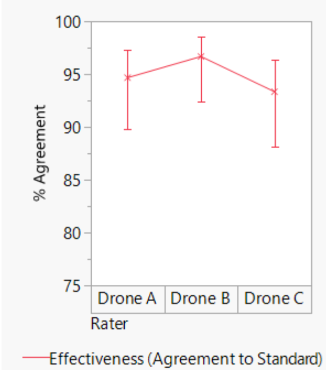
Agreement across Categories

Category	Kappa	Standard Error
A	0.8175	0.0236
B	0.9044	0.0236
C	0.9070	0.0236
D	0.9346	0.0236
E	0.8695	0.0236
Overall	0.8868	0.0118

- Tables shows agreement values comparing pairs of drones and drones to the standard.
- Kappa Indices are showing excellent agreement.
- Repeatability (within drones) and reproducibility (between drones) are very good.
- Agreement across categories is also excellent.

Effectiveness: Agreement to Standard





- The **Gauge Attribute Chart now** shows the average % Agreement of the three drones **to the Standard** for each location. Locations 7 and 8 had the lowest level of agreement to the **Standard**.
- The left chart shows the overall **%** Agreement **to the Standard** by drone.

Effectiveness Report: Agreement Counts

The **Effectiveness Report** summarizes the comparisons of the drones to the Standards. There are agreement differences among the five characteristics, and the counts are shown.

Effectiveness Report

Agreement Counts

Rater	Correct(A)	Correct(B)	Correct(C)	Correct(D)	Correct(E)	Total Correct	Incorrect(A)	Incorrect(B)	Incorrect(C)	Incorrect(D)	Incorrect(E)	Grand Total
Drone A	28	29	28	30	27	142	2	1	2	0	3	150
Drone B	28	28	30	29	30	145	2	2	0	1	0	150
Drone C	27	30	29	29	25	140	3	0	1	1	5	150

Analysis: Effectiveness Report

Effectiveness

Rater	Effectiveness	95% Lower CI	95% Upper CI	Error rate
Drone A	94.6667	89.8296	97.2730	0.0533
Drone B	96.6667	92.4348	98.5680	0.0333
Drone C	93.3333	88.1638	96.3388	0.0667
Overall	94.8889	92.4475	96.5704	0.0511

- Effectiveness = # correct decisions/total opportunities for a decision
- Table shows comparisons of drones to the standard
- All drones appear highly effective.

Effectiveness Report

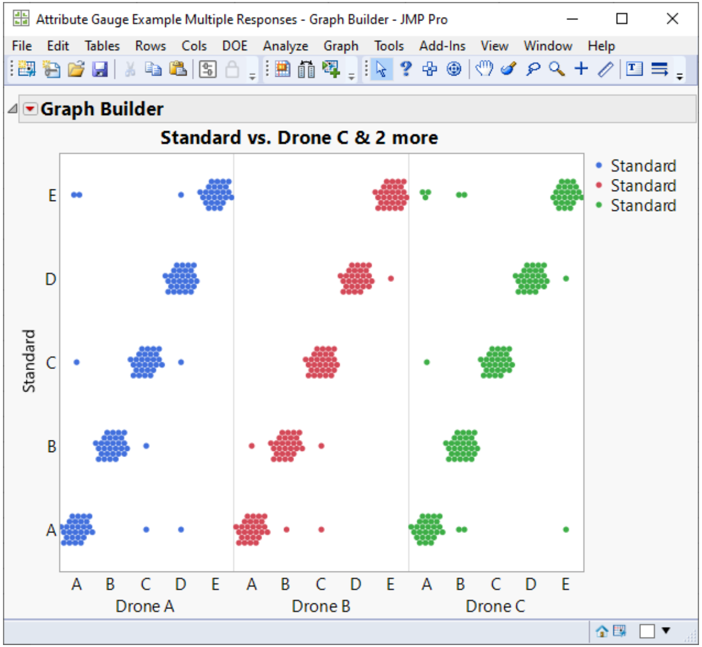
There is a detailed analysis by level, provided in a **Misclassifications** summary. We see that characteristics A and E had higher misclassification rates than the other three options.

Misclassifications

Standard Level	A	B	C	D	E
A	.	1	2	0	5
B	3	.	0	0	2
C	2	2	.	0	0
D	1	0	1	.	1
E	1	0	0	2	.
Other	0	0	0	0	0

Misclassifications Visualization

Using **Graph Builder**, we can view the classifications and misclassifications by each drone.



Summary

- The use of **attribute gauge analy**sis allowed the company to provide solid data on the agreement and effectiveness of drones for inventory management.
- Subsequent results reported on the company's website show **inventory counts to be 35% faster, inventory costs reduced by 40%, and reduced missed-shipment and damage claims by 50%** compared to previous methods.
- In addition, the system generates more **actionable data** for accurate, effective, safer, more cost-effective, and faster inventory control.

References

1. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, **20**, 37-46.

2. Fleiss, J.L., Levin, B., Paik, M.C. (2003). *Statistical Methods for Rates and Proportions*, 3rd ed., New York, John Wiley & Sons

3. Le, C.T. (1998). *Applied Categorical Data Analysis*, New York, John Wiley & Sons

Thank You