Attribute Agreement Analysis

Scenario

Let's imagine that we work in the purchasing department and our team's ability to identify errors on Purchase Orders is critical. As such, we decide to conduct a Measurement Systems Analysis.

For this study, we collect 20 Purchase Order forms. 10 of the forms have mistakes and are considered bad and 10 had no mistakes and are considered good. For the initial study, we randomly select three associates to take part in the study.

The facilitator will bring each associate into a quiet room and randomly hand them a Purchase Order form.

The associate will have 30 seconds to review the form and decide if it's good or bad. Their answer will be documented accordingly by the facilitator.

Once the associate is done with all 20 forms, the next person will come into the room and the process will repeat itself until all three associates have done the study twice. Since each associate is simply stating whether they believe each form is good or bad, which is binary data, a form of attributes data, we'll be conducting an Attributes Agreement Analysis instead of a Gage R&R study.

Attribute Agreement Analysis Study

Setting up an Attribute Agreement Analysis study is similar to setting up a Gage R&R study. First, you should randomly select at least three operators. Next, the number of parts used should be enough so that when we multiply P, the number of parts, by O, the number of operators, the result is greater than 15.

We also need to determine which parts are "good" and "bad." It's also important to have an even mix of good and bad parts in the study. Finally, we always want to randomize things during the study which means we'll randomly hand each part to the associates.

Of course, the facilitator will know which part is which so they can record the results accordingly which most definitely takes good organization and planning. It's very important to ensure anyone asked to participate in the study is at ease and not scared or worried about being in trouble if they do poorly.

I can't help but remember a situation where a young lady was asked to take part in a study just like this. She was so nervous that she literally made herself sick with worry. Never take for granted how important this step of the process is since in the end, punching buttons in Minitab is easy, anyone can learn to do this. But the ability to work with people in a respectful and kind manner is something we all need to constantly work at.

Analyzing the Results in Minitab

Once the study is complete, it's time analyze the results in Minitab.

Within Appraisers

Assessment Agreement

Appraiser	# Inspected	# Matched	Percent	95% CI
Cindy	20	20	100.00	(86.09, 100.00)
George	20	17	85.00	(62.11, 96.79)
Steve	20	16	80.00	(56.34, 94.27)

Matched: Appraiser agrees with him/herself across trials.

Fleiss' Kappa Statistics

Appraiser	Response	Kappa	SE Kappa	Z	P(vs > 0)
Cindy	Bad	1.00000	0.223607	4.47214	0.0000
	Good	1.00000	0.223607	4.47214	0.0000
George	Bad	0.69925	0.223607	3.12713	0.0009
	Good	0.69925	0.223607	3.12713	0.0009
Steve	Bad	0.60000	0.223607	2.68328	0.0036
	Good	0.60000	0.223607	2.68328	0.0036

The first section we see deals with how well each operator agreed with themselves. For example, Cindy inspected 20 forms across two separate trials. As it turned out, she agreed with herself every time. In other words, if she thought form 1 was good during the first trial, she agreed with herself on the second trial.

The thing to remember here is that this doesn't mean Cindy matched the standard all 20 times. We'll look at this in a bit. Instead, this section simply tells us how well each associate agreed with themselves.

George agreed with himself 17 times and Steve agreed with himself 16 times. In the section below, we see what are called Kappa statistics, as well as, P-values.

The P-values help us determine whether or not these agreements are due to chance or are statistically valid.

The null hypothesis, or Ho, for these tests is that the agreement within each appraiser is due to chance and the alternate hypothesis, or Ha, is that the agreement within each appraiser is not due to chance.

For example, when we look at Steve's results, we see the P-values are .0036 which is less than the typical Alpha value of .05, so we'd reject Ho and conclude that the agreements within his study are not due to chance.

In this section, we see Kappa statistics which help us determine the level of agreement within each appraiser. A Kappa value of 1 means there was perfect agreement and a Kappa value of 0 means the agreements were the same as random chance.

We typically prefer to see these Kappa values be greater than .75, but as with any scoring system, each organization will need to come up with what works best for them. If we use the .75 Kappa value as a cutoff value, we see that both George and Steve have opportunity for improvement with Kappa values of .69 and .60 respectively.

The next section focuses on how well each appraiser did versus the standard. In other words, if Purchase Order form 1 was indeed "good", we'll now check to see how many times each appraiser got it right and said this form was good. As we can see, no appraiser was right every time.

Cindy and George matched 17 times while Steve matched the known standard 16 times. This section focuses on assessment disagreement. The legend towards the bottom of this section does a good job explaining what we're looking at.

Good/Bad

For example, Good/Bad means the parts were noted as good both times during the assessments even though the parts were really bad. Bad/Good means the parts were called bad both times but the parts were really good and Mixed means

the assessments weren't the same meaning the appraiser said the parts were good one time and bad the other.

As we can see here, Cindy called one part good twice even though it was bad. She then called two parts bad when they were really good. Now this sort of information is extremely helpful as it can be used to help us better train associates.

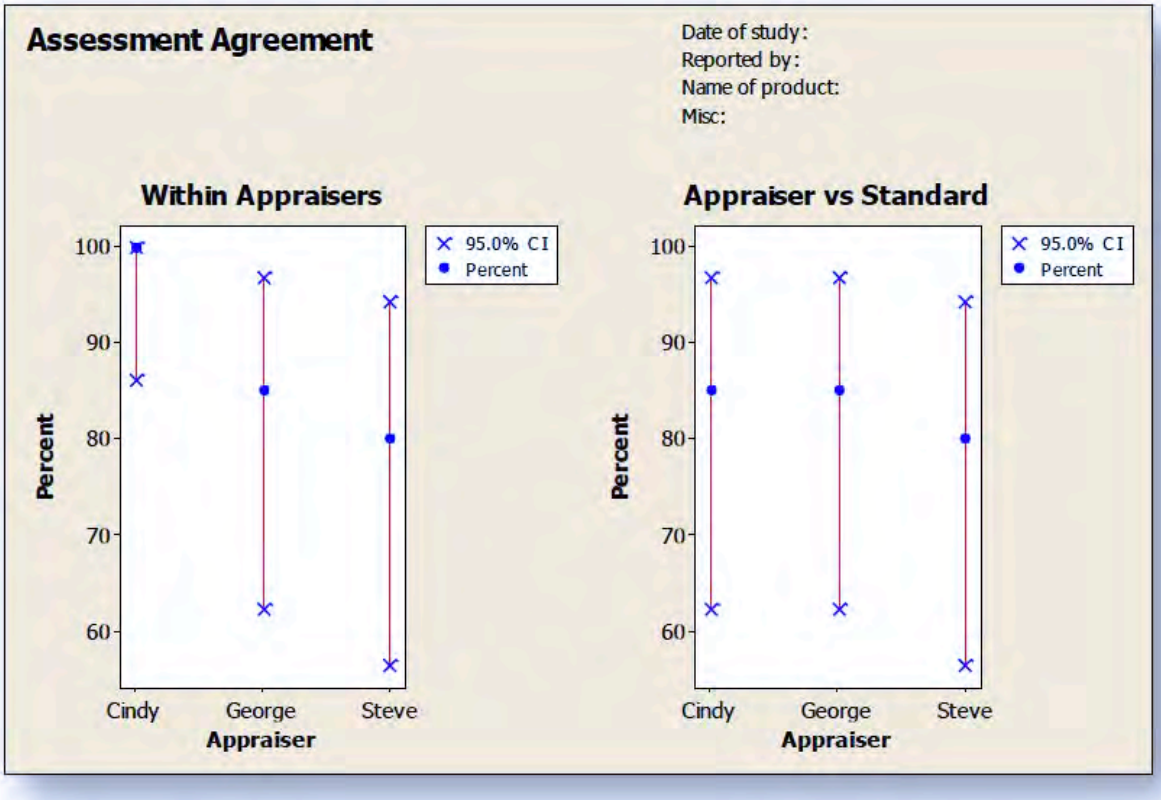
Once again, we see our Kappa statistics and P-Values in this section. All the P-values are zero meaning these results are not due to chance. We also see that Cindy has the most opportunity for improvement in this section since her Kappa values are only .69.

Her score was low like this since she struggled the most in the Assessment Disagreement section. The next section focuses on how well each appraiser's assessments agree with each other.

As we can see, all appraisers agreed with each other 12 times or 60%. This returned a Kappa value of .646 which shows we definitely have opportunities for improvement in this area.

The last, and arguably most important section of all, focuses on how well all appraisers' assessments agreed with the known standard. As we see here, this occurred 12 times for a Kappa value of .78.

Minitab also provides some graphs showing how well each appraiser does within themselves, as well as, how well they do against the standard.



95% confidence intervals are also calculated giving you an idea of what kind of measurement performance you could expect. For example, we could expect Steve's performance against the standard to range from around 95% to less than 60%.

Conclusion

In summary, after looking at the results of this MSA, we'd conclude that this measurement system is capable to marginally capable depending on the Kappa values you decide to use. Some next steps may include working with each associate on the forms they missed trying to understand what they saw and didn't see.

Additionally, this would also be an excellent time to improve standards and training plans on how to inspect forms consistently. Finally, we'll want to redo the assessment as needed including after these associates have been trained.