War-Winning Capabilities … On Time, On Cost

RIDIT ANALYSIS FOR FLYING QUALITIES & Cooper-Harper Data

Arnon Hurwitz, PhD.

AFTC-PA-14539

POC: arnon.hurwitz@edwards.af.mil

U.S. AIR FORCE

Overview

- Introduction + BLUF
- Ridit Analysis Method
- Ridit Analysis DOE– Certification with CHR's
- Conclusion

Introduction +BLUF

- The Certification team wished to tell the difference between two different flight configurations
- Cooper-Harper Ratings (CHR’s) to be used as the ‘measurement device’
- CHR’s are not numbers – need a way to do the analysis in a statistically defensible way
- Ridit analysis was used to convert CHR’s to numeric items, analyze the situation, & make recommendations

Ridit Analysis

412TW

- Consider a simple example: 27 students are asked to answer 'Course was good?' from #1 (Strongly Disagree) to #5 (Strongly Agree)

This year's scores

Last year's scores

This year's proportions

Last year's proportions

Bad

Good

Preference	#	Comparison	Reference	ridits (r)	p	q	rp	rq
Strongly Disagree	*1*	5	3	0.056	0.185	0.111	0.010	0.006
Disagree	*2*	8	6	0.222	0.296	0.222	0.066	0.049
Neither A. nor D.	*3*	6	6	0.444	0.222	0.222	0.099	0.099
Agree	*4*	2	4	0.630	0.074	0.148	0.047	0.093
Strongly Agree	*5*	6	8	0.852	0.222	0.296	0.189	0.252
SUM		**27**	**27**				0.411	0.500

Ridit Analysis – (continued)

Preference	#	Comparison	Reference	ridits (r)	p	q	rp	rq
Strongly Disagree	*1*	5	3	0.056	0.185	0.111	0.010	0.006
Disagree	*2*	8	6	0.222	0.296	0.222	0.066	0.049
Neither A. nor D.	*3*	6	6	0.444	0.222	0.222	0.099	0.099
Agree	*4*	2	4	0.630	0.074	0.148	0.047	0.093
Strongly Agree	*5*	6	8	0.852	0.222	0.296	0.189	0.252
SUM		**27**	**27**				0.411	0.500

- Proportions **p** and **q** (i.e. estimated probabilities) are computed from the data. E.g. 0.185 = 5/27, etc.
- A population (Last year's) is set as the 'Reference'
- The *k-th ridit of the Ref. population* is defined as:

$$r_k = \begin{cases} \frac{q_1}{2} & for\ k = 1, \\ q_1 + \cdots + q_{k-1} + \frac{1}{2} q_k & for\ k > 1 \end{cases}$$

Ridit Analysis – (continued)

To the left ↑ / To the rt. ↓

Preference	#	Comparison	Reference	ridits (r)	p	q	rp	rq
Strongly Disagree	*1*	5	3	0.056	0.185	0.111	0.010	0.006
Disagree	*2*	8	6	0.222	0.296	0.222	0.066	0.049
Neither A. nor D.	*3*	6	6	0.444	0.222	0.222	0.099	0.099
Agree	*4*	2	4	0.630	0.074	0.148	0.047	0.093
Strongly Agree	*5*	6	8	0.852	0.222	0.296	0.189	0.252
SUM		**27**	**27**				0.411	0.500

- Form columns **rp** and **rq**, and sum (∑) each one
- ∑ **rp** = 0.411 is the probability that the **Reference** pop. will be '**to the left**' of the Comparison pop.
  - If the p's are 'bunched' to the right versus the q's, then ∑ **rp** > ∑ **rq**
  - that is, high ∑ **rp** ⟹ **Reference** pop. (the q's) is bunched '**to the left**'
- Our HYPOTHESIS is that ∑ **rp** ≥ 0.5 What does this mean?
  - If true, then last year's (**Reference**) scores are **worse** than this year's
  - However, it's obvious that ∑ **rp** = 0.411 ≤ 0.5 - So was last year better?
  - Can only say this *if experimental error = 0* → We need a statistical test!

Ridit Analysis – Hypothesis Test

- ‘Experimental error’ means that, if the underlying situation stays the same, but we draw a new sample, the numbers (p’s and q’s) we see will be somewhat different. So conclusions might change

- To test Ho: $\sum$ **rp** $\geq \sum$ **rq = 0.5** , form **t** = ($\sum$ **rp** – **0.5**) / $\sqrt{\left[\frac{1}{12m} + \frac{1}{12n} + \frac{1}{12mn}\right]}$

  m = n = 27. So t = (0.411- 0.5)/sqrt(0.0063)= -1.12, with d.f.= m+n-2 = 52

- Left-tail, critical t (at 95% confidence, d.f.=52) = -1.675, so <u>do not</u> reject Ho
  → We <u>cannot</u> say that this year’s scores are any better than last year’s

Ridit Analysis – FQ

- Consider a DOE proposed for a Flying Qualities test
- It has 2 factors and 8 rows
- 4 of the 8 rows are ‘Midpoint’ or ‘Control’ rows (i.e. nominal operation)
- After the test flights, we get CHR data – how do we compare the results at ‘High’ (+1) with ‘Low’ (-1) ?

#	Factor A	Factor B	
1	Mid	Mid	
2	+1	-1	
3	-1	+1	
4	Mid	Mid	
5	Mid	Mid	
6	+1	+1	
7	-1	-1	
8	Mid	Mid	

Ridit Analysis – FQ

Consider just Factor A:

- Stripping out the Midpoint runs from the 8-run design, and inserting CHR's from 2 pilots

Factor A	PILOT 1 -CH	PILOT 2 - CH
1	2	3
-1	4	5
1	4	3
-1	5	5

Rearranging so as to facilitate a ridit analysis

- for example, the low setting for A had one 4 and three 5's

CHR	A (-1)	A (+1)	SUM
1	0	0	0
2	0	1	1
3	0	2	2
4	1	1	2
5	3	0	3
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	0	0	0
SUMS	4	4	8

- These CHR's were chosen so that the 'median' of all entries was 4, and the difference (Δ) between the medians was 2

This places the medians near the CH boundaries: '*Satisfactory w/o improvement*,' & '*Deficiencies*'

Ridit Analysis – FQ

412TW

- Two pilots fly 4 runs each, for factor (*for example, let* A=speed)
  - **-1** → fly 'slow' configuration
  - **1** → fly 'fast' configuration
- If 'slow' is the Reference, is 'fast' better?
- Rearranging so as to facilitate a ridit analysis
  - For example, the slow setting for A had one 4 and three 5's
- Note: Now 'to the left' → 'better'

Factor A	PILOT 1 -CH	PILOT 2 - CH
1	2	3
-1	4	5
1	4	3
-1	5	5

CHR	A (-1)	A (+1)	SUM
1	0	0	0
2	0	1	1
3	0	2	2
4	1	1	2
5	3	0	3
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	0	0	0
SUMS	4	4	8

Ridit Analysis – with CHR's

412TW

Converting the above CHR data to ridit analysis, R(p|q) = 0.0313

	REFERENCE	COMPARISON
CHR	Slow	Fast
1	0.000	0.000
2	0.000	1.000
3	0.000	2.000
4	1.000	1.000
5	3.000	0.000
6	0.000	0.000
7	0.000	0.000
8	0.000	0.000
9	0.000	0.000
10	0.000	0.000

	p	q	r	rp	rq
	0.0000	0.0000	0.0000	0.0000	0.0000
	0.2500	0.0000	0.0000	0.0000	0.0000
	0.5000	0.0000	0.0000	0.0000	0.0000
	0.2500	0.2500	0.1250	0.0313	0.0313
	0.0000	0.7500	0.6250	0.0000	0.4688
	0.0000	0.0000	1.0000	0.0000	0.0000
	0.0000	0.0000	1.0000	0.0000	0.0000
	0.0000	0.0000	1.0000	0.0000	0.0000
	0.0000	0.0000	1.0000	0.0000	0.0000
	0.0000	0.0000	1.0000	0.0000	0.0000
SUM	1.0000	1.0000	5.7500	0.0313	0.5000

$$t = \frac{0.0313 - 0.5}{\sqrt{\frac{1}{12*4} + \frac{1}{12*4} + \frac{1}{12*4*4}}} = \frac{-0.469}{0.216} = -2.17$$

The 'critical' one-sided t-statistic at 95% confidence and 4+4-2=6 degrees of freedom is -1.94, so we reject the hypothesis: Ho: 'slow' is better (i.e. closer to '1) than 'fast'.

- Thus we conclude that 'fast' is, in fact, better than 'slow'

How many flights? – A Power Analysis

- When faced with resource constraints, some agents will want to do as few flights as possible
- How do we choose between, say, flight budgets of 8, 16 or 24 flights (including control runs)?

Ridit Analysis – Sample Size

Rearrangements of CHR's to give various delta's and medians

SLOW	FAST
0	0
0	0
0	1
1	2
3	1
0	0
0	0
0	0
0	0
0	0
Δ=1	**Median=4.5**
R(q|q) = 0.5	R(p|q) = 0.2188

SLOW	FAST
0	0
0	1
0	2
1	1
3	0
0	0
0	0
0	0
0	0
0	0
Δ=2	Median=4.0
R(q|q) = 0.5	R(p|q) = 0.0313

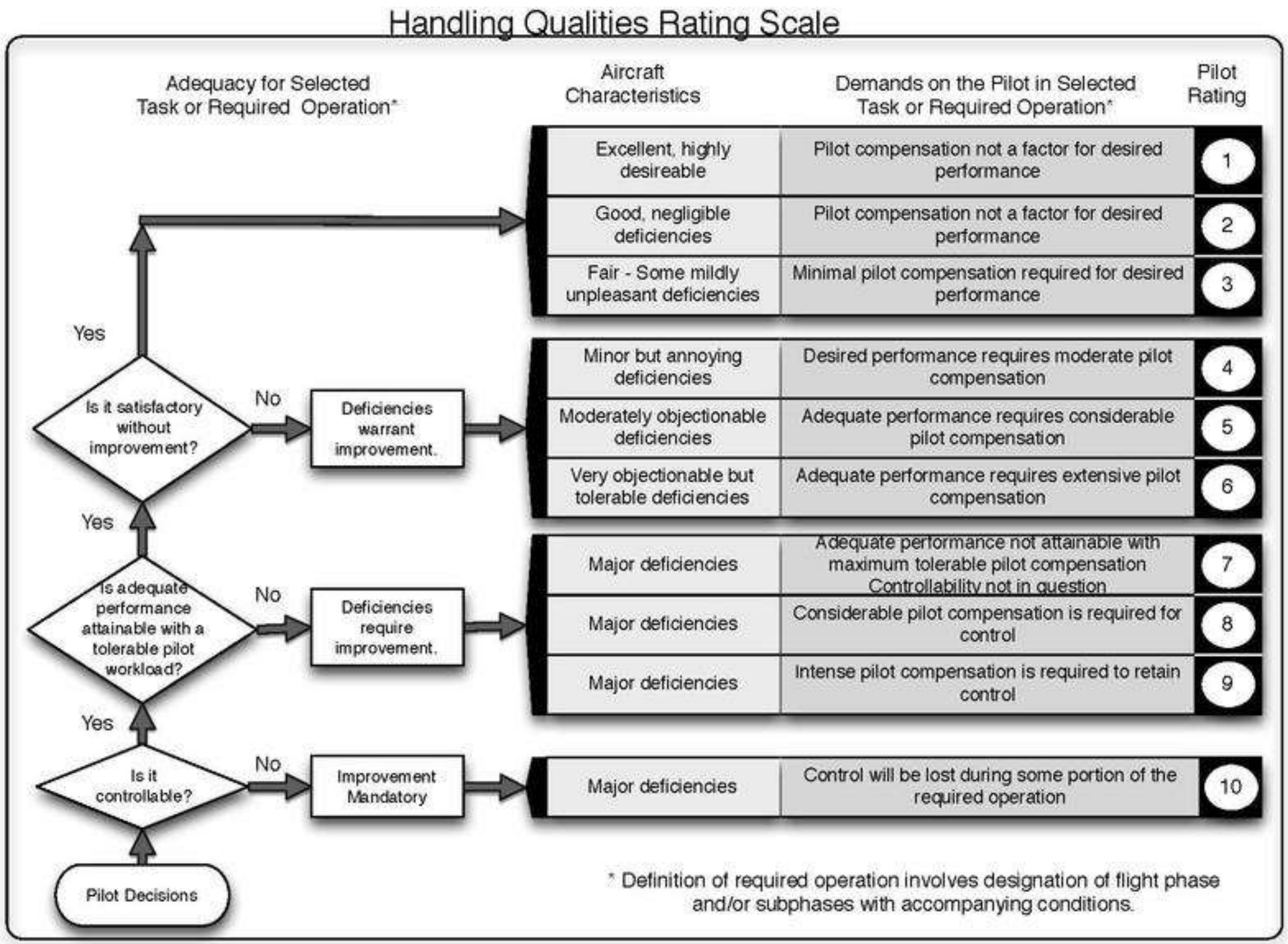
SLOW	FAST
0	0
0	0
0	1
0	2
0	1
1	0
3	0
0	0
0	0
0	0
Δ=3	**Median=5.5**
R(q|q) = 0.5	R(p|q) = 0.0

<u>No. of Pilots</u>	<u>Flight Budget</u>	Power to detect a CHR median diff. of Δ and at a 95% confidence level		
Each pilot flies 4 runs	Control runs = ½ of Budget	Δ=1 Median=4.5	Δ=2 Median=4.0	Δ=3 Median=5.5
1	8	15%	26%	28%
2	16	49%	85%	89%
3	24	83%	99%	99.9%

Conclusions

- **Ridit analysis is a simple and effective method for converting Cooper-Harper scores to values that can support a statistically defensible analysis**
  - **i.e. sample size, α-risks, and β-risks**
- **The fact that the mean ridit R(p|q) does not change if there is zero overlap between Reference and Comparison populations does not invalidate the analysis if we use variance formula** $\left(\frac{1}{12n}+\frac{1}{12m}+\right.$

Cooper-Harper flowchart



References

1. Agresti, A. (1984). Analysis of Ordinal Categorical Data. New York: Wiley & Sons
2. Beder and Heim (1990). On the use of Ridit Analysis. Psychometrika, vol. 55, No. 4. 603-616.
3. Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978). Statistics for Experimenters. Wiley, New York.
4. Bross, I.D.J. (1958). How to use ridit analysis. Biometrics, 14. 18-38.
5. Cooper, G. and Harper, R. (1969). The use of pilot rating in the evaluation of aircraft handling qualities. Technical Report TN D-5153, NASA
6. Croushore, D. & Schmidt, R.M. (2010). Ridit analysis of student score evaluations. Robins School of Business, Univ. of Richmond, VA 23173 (USA).
7. R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.
8. Selvin, S. (1977). A further note on the interpretation of ridit analysis. American Journal of Epidemiology, 105. 16-20.
9. Selvin, S. (2004) Statistical Analysis of Epidemiologic Data, 3rd ed. OUP. 176-177
10. Wikipedia, 2014. Cooper-Harper. http://en.wikipedia.org/wiki/Cooper-Harper
11. Wilson, D.J. & Riley, D.R. (1989). Cooper-Harper Pilot Rating Variability. American Institute of Aeronautics & Astronautics. (© McDonnell Douglas Aircraft Corp., St. Louis, MO 63166)
12. Wilson, D.J. & Riley, D.R. (1990). More on Cooper-Harper Pilot Rating Variability. American Institute of Aeronautics & Astronautics. (© McDonnell Douglas Aircraft Corp., St. Louis, MO)
13. Wu, C. 2007). On the application of Grey relational analysis and Ridit analysis to Likert Scale surveys. International Mathematical Forum, 2, No. 14. 675-687.

Summary

412TW

- **Ridit analysis is a simple and effective method for converting Cooper-Harper scores to values that can support a statistically defensible analysis**

- **Ridit analysis can be used even if the data is sparse, and is particularly useful if there is an overlap between Reference & Comparison populations where it is uncertain if one population is ‘to the left’**