Testing Significance Tests: A Simulation with Cliff's Delta, *t*-tests, and Mann-Whitney *U*

Tyler Barnes (Department of Veterans Affairs/Wright State University)
Scott C. Moore (Department of Veterans Affairs)
Katerine Osatuke (Department of Veterans Affairs)

*Any opinions, findings, conclusions, or recommendations expressed in this material are those of the presenters and do not necessarily represent the view of the Department of Veterans Affairs or U.S. Government.

Outline

- Introduction
- Why Cliff's Delta?
- Problem
- Method
- Results
- Discussion

Introduction

- All Employee Survey (AES)
- Action Planning
- We use Cliff's Delta (Cliff, 1993) to see if there are significant group differences

Why Cliff's Delta?

- There are many pairwise comparison statistics
- An effect size (for us this is practical).
- Made for ordinal data/non-parametric
- Power and error

Problem

- Much past research uses “small” sample sizes.
- Unequal sample sizes are not as drastic as they can be in AES
- In general, parameters of those studies don’t match the AES well
- This is a formal test of one aspect of the AES data analysis.

Method

Method

- 2 Simulations (Everything in R)
- First Simulation (equal sample sizes)
  - 5x2x6x3x3 fully-crossed design with 100 replications each.
  - sample size (20,50,100, 1000,10000).
  - Normal skewed or Standard Normal distribution.
  - mean differences (0-.02, .09-.11, .19-.21, .29-.31, .39-.41, .49-.51).
  - Standard Deviation of items (.5,.75,1)- fourth and fifth conditions
- Second Simulation (Unequal sample sizes)
  - Same conditions as above except sample size.
  - Fully crossed $N_1 = 17$ and $N_2 = 742$ and the reverse
- All data were generated from a random uniform distribution and then discretized from a specified marginal distribution to meet respective conditions (with checks to ensure data quality).
- 1-5 ordered discrete scale

Continued…

- Data Analysis
  - Student's *t*
  - Welch's *t*
  - Mann-Whitney U
  - Cliff's Delta- function written, tested, and used by VHA NCOD
  - Error- % of hits in the mean difference = 0 conditions
  - Power- % of hits in all other conditions

Results

Error

Error rates of pairwise comparison statistics for N=10,000.

		Normal				Skewed			
Diff	SD (G1, G2)	*t*	t_w	Δ	*U*	*t*	t_w	Δ	*U*
0 - .02	(.5, .5)	0.01	0.01	0.02	0.02	0.01	0.01	0	0
	(.5, .75)	0.01	0.01	0.03	0.03	0.03	0.03	0.02	0.02
	(.5, 1)	0	0	0.02	0.04	0	0	0.57	0.58
	(.75, .5)	0.02	0.02	0	0.02	0.05	0.05	0.03	0.03
	(.75, .75)	0	0	0	0	0	0	0	0
	(.75, 1)	0	0	0	0	0	0	0.51	0.51
	(1, .5)	0	0	0.02	0.05	0	0	0.56	0.6
	(1, .75)	0	0	0	0	0	0	0.48	0.51
	(1, 1)	0	0	0	0	0	0	0	0

Note. Diff= Mean difference; SD (G1,G2) = Standard deviations of group 1(G1) and group 2 (G2); *t* = Student's *t*-test; t_w = Welch's *t*-test; Δ = Cliff's Delta; *U* = Mann & Whitney *U*; Normal = Normal distribution; Skewed = Skewed distribution. All statistics had perfect power ratings for all other conditions.

Equal Sample Sizes(N = 20)

		Normal				Skewed			
Diff	SD (G1, G2)	*t*	t_w	Δ	*U*	*t*	t_w	Δ	*U*
0.2	(.5, .5)	0	0	0	0	0	0	0	0
	(.5, .75)	0	0	0	0	0	0	0.01	0.01
	(.5, 1)	0	0	0	0	0	0	0	0
	(.75, .5)	0	0	0	0	0	0	0	0
	(.75, .75)	0	0	0	0	0	0	0	0
	(.75, 1)	0	0	0	0	0	0	0	0
	(1, .5)	0	0	0	0	0	0	0	0
	(1, .75)	0	0	0	0	0	0	0	0
	(1, 1)	0	0	0	0	0	0	0	0
0.3	(.5, .5)	0.48	0.48	0.48	0	0	0	0.43	0.23
	(.5, .75)	0	0	0.02	0	0	0	0.11	0.09
	(.5, 1)	0	0	0	0	0	0	0.05	0.04
	(.75, .5)	0	0	0.01	0	0	0	0	0.03
	(.75, .75)	0	0	0	0	0	0	0	0
	(.75, 1)	0	0	0	0	0	0	0	0
	(1, .5)	0	0	0	0	0	0	0	0
	(1, .75)	0	0	0	0	0	0	0	0
	(1, 1)	0	0	0	0	0	0	0	0
0.4	(.5, .5)	1	1	1	1	1	1	1	1
	(.5, .75)	0.32	0	0.83	0.38	0.37	0.1	0.89	0.83
	(.5, 1)	0	0	0.02	0.03	0	0	0.4	0.34
	(.75, .5)	0.45	0.11	0.32	0.12	0.32	0	0.84	0.33
	(.75, .75)	0	0	0.01	0.01	0	0	0.27	0.21
	(.75, 1)	0	0	0	0	0	0	0.12	0.04
	(1, .5)	0	0	0.01	0.01	0	0	0	0
	(1, .75)	0	0	0	0	0	0	0	0
	(1, 1)	0	0	0	0	0	0	0	0
0.5	(.5, .5)	1	1	1	1	1	1	1	1
	(.5, .75)	1	1	1	1	1	1	1	1
	(.5, 1)	0.29	0.14	0.41	0.46	0.31	0.23	0.94	0.94
	(.75, .5)	1	1	1	1	1	1	1	1
	(.75, .75)	1	1	0.91	0.46	1	1	0.92	0.47
	(.75, 1)	0	0	0.14	0.06	0	0	0.73	0.42
	(1, .5)	0.41	0.18	0.56	0.39	0.29	0	0.06	0.09
	(1, .75)	0	0	0.09	0	0	0	0	0
	(1, 1)	0	0	0	0	0	0	0.04	0.02

Equal Sample Sizes(N = 100)

		Normal				Skewed			
Diff	SD (G1, G2)	*t*	t_w	Δ	*U*	*t*	t_w	Δ	*U*
0.2	(.5, .5)	1	1	1	1	1	1	1	1
	(.5, .75)	1	1	0.95	0.95	1	1	1	1
	(.5, 1)	0.02	0.01	0.17	0.31	0	0	1	1
	(.75, .5)	1	1	0.83	0.83	1	1	0.98	0.98
	(.75, .75)	0.2	0.2	0.3	0.28	0.22	0.22	0.96	0.96
	(.75, 1)	0	0	0.05	0.05	0	0	0.98	0.98
	(1, .5)	0	0	0.13	0.19	0	0	0.03	0.09
	(1, .75)	0	0	0.01	0.01	0	0	0.04	0.07
	(1, 1)	0	0	0	0	0	0	0.11	0.12
0.3	(.5, .5)	1	1	1	1	1	1	1	1
	(.5, .75)	1	1	1	1	1	1	1	1
	(.5, 1)	1	1	1	1	1	1	1	1
	(.75, .5)	1	1	1	1	1	1	1	1
	(.75, .75)	1	1	1	1	1	1	1	1
	(.75, 1)	1	1	1	1	1	1	1	1
	(1, .5)	1	1	0.96	0.98	1	1	1	1
	(1, .75)	1	1	0.95	0.95	1	1	0.99	0.99
	(1, 1)	0.98	0.98	0.82	0.74	0.99	0.99	1	0.99

Unequal Sample Sizes(Skewed)

		N (G1=17, G2=742)				N (G1=742, G2=17)			
Diff	SD (G1, G2)	t	t_w	Δ	U	t	t_w	Δ	U
0.2	(.5, .5)	0	0	0	0.03	0	0	0	0
	(.5, .75)	0	0	0.24	0	0	0	0.12	0.12
	(.5, 1)	0	0	1	0	0	0	0	0.15
	(.75, .5)	0	0	0	0.09	0	0	0	0
	(.75, .75)	0	0	0	0	0	0	0.12	0
	(.75, 1)	0	0	0	0	0	0	0	0
	(1, .5)	0	0	0	0	0	0	0	0
	(1, .75)	0	0	0	0	0	0	0	0
	(1, 1)	0	0	0	0	0	0	0	0
0.3	(.5, .5)	1	1	1	1	1	1	1	1
	(.5, .75)	0	1	1	1	1	0	0.57	0.84
	(.5, 1)	0	1	1	0.29	1	0	0.17	0.61
	(.75, .5)	1	0	0.05	1	0	1	1	0.19
	(.75, .75)	0	0	0.33	0.61	0	0	0.14	0.03
	(.75, 1)	0	0	0.67	0.03	0	0	0.09	0.19
	(1, .5)	1	0	0	0.23	0	1	0.9	0
	(1, .75)	0	0	0	0	0	0	0.02	0
	(1, 1)	0	0	0	0	0	0	0	0
0.4	(.5, .5)	1	1	1	1	1	1	1	1
	(.5, .75)	1	1	1	1	1	0.72	0.99	1
	(.5, 1)	0	1	1	1	1	0	0.42	0.98
	(.75, .5)	1	0.66	0.98	1	1	1	1	1
	(.75, .75)	1	0.59	1	1	1	0.76	0.97	0.97
	(.75, 1)	0	0.51	1	0.98	1	0	0.31	0.82
	(1, .5)	1	0	0	0.94	0	1	1	0
	(1, .75)	1	0	0	0.83	0	0.66	0.96	0.03
	(1, 1)	0	0	0.09	0.01	0	0	0.19	0.07
0.5	(.5, .5)	1	1	1	1	1	1	1	1
	(.5, .75)	1	1	1	1	1	1	1	1
	(.5, 1)	0.91	1	1	1	1	0.03	0.89	1
	(.75, .5)	1	1	1	1	1	1	1	1
	(.75, .75)	1	1	1	1	1	1	1	1
	(.75, 1)	0.84	1	1	1	1	0.07	0.9	1
	(1, .5)	1	0	0.2	1	0.99	1	1	1
	(1, .75)	1	0	0.28	0.99	0.96	1	1	0.95
	(1, 1)	0.82	0	1	1	0.96	0.14	0.88	0.94

Discussion

- Recommendations:
  - Low error rates
  - Use *t*-test if SD of large group is .5 and the mean differences are small. If SD of large group is 1, use Cliff's Delta or Mann-Whitney *U.*
  - In general, for equal sample sizes less than 1,000 use Cliff's Delta. Larger than that use *t* – test.
- Limitations
  - Results may not be generalizable
  - Error conceptualization and possible alternatives to error
- Conclusions
  - Choice should be based on the research question, properties of the data, and the situation.

Equal Sample Sizes(N = 50)

		Normal				Skewed			
Diff	SD (G1, G2)	*t*	t_w	Δ	*U*	*t*	t_w	Δ	*U*
0.2	(.5, .5)	0.71	0.71	0.79	0.5	0.52	0.52	0.75	0.65
	(.5, .75)	0	0	0.01	0.02	0	0	0.62	0.56
	(.5, 1)	0	0	0	0	0	0	0.46	0.45
	(.75, .5)	0	0	0	0	0	0	0.02	0.05
	(.75, .75)	0	0	0	0	0	0	0.08	0.09
	(.75, 1)	0	0	0	0	0	0	0.11	0.11
	(1, .5)	0	0	0	0	0	0	0	0
	(1, .75)	0	0	0	0	0	0	0	0
	(1, 1)	0	0	0	0	0	0	0	0
0.3	(.5, .5)	1	1	1	1	1	1	1	1
	(.5, .75)	1	1	1	1	1	1	1	1
	(.5, 1)	0.09	0.07	0.31	0.39	0	0	0.97	0.98
	(.75, .5)	1	1	0.95	0.95	1	1	1	1
	(.75, .75)	0.63	0.63	0.49	0.39	0.69	0.69	0.97	0.93
	(.75, 1)	0	0	0.07	0.04	0	0	0.82	0.75
	(1, .5)	0.04	0.02	0.15	0.17	0.07	0.04	0.11	0.19
	(1, .75)	0	0	0.03	0.02	0	0	0.07	0.09
	(1, 1)	0	0	0	0	0	0	0.01	0
0.4	(.5, .5)	1	1	1	1	1	1	1	1
	(.5, .75)	1	1	1	1	1	1	1	1
	(.5, 1)	1	1	0.97	1	1	1	1	1
	(.75, .5)	1	1	1	1	1	1	1	1
	(.75, .75)	1	1	1	1	1	1	1	1
	(.75, 1)	1	1	0.91	0.85	1	1	1	1
	(1, .5)	1	1	0.95	0.96	1	1	1	1
	(1, .75)	1	1	0.93	0.84	1	1	1	1
	(1, 1)	0.6	0.6	0.51	0.38	0.68	0.68	1	0.98

Equal Sample Sizes(N = 1,000)

		Normal				Skewed			
Diff	SD (G1, G2)	*t*	t_w	Δ	*U*	*t*	t_w	Δ	*U*
0.1	(.5, .5)	1	1	1	1	1	1	1	1
	(.5, .75)	1	1	1	1	1	1	1	1
	(.5, 1)	1	1	1	1	1	1	1	1
	(.75, .5)	1	1	1	1	1	1	1	1
	(.75, .75)	1	1	1	1	1	1	1	1
	(.75, 1)	1	1	1	1	1	1	1	1
	(1, .5)	1	1	0.96	0.97	1	1	0.8	0.87
	(1, .75)	1	1	0.91	0.91	1	1	0.6	0.63
	(1, 1)	1	1	0.9	0.9	1	1	0.92	0.92

Unequal Sample Sizes(Normal)

		N (G1=17, G2=742)				N (G1=742, G2=17)			
Diff	SD (G1, G2)	*t*	t_w	Δ	*U*	*t*	t_w	Δ	*U*
0.2	(.5, .5)	0	0	0	0	0	0	0	0
	(.5, .75)	0	0	0	0	0	0	0	0.05
	(.5, 1)	0	0	0	0	0	0	0	0.03
	(.75, .5)	0	0	0	0	0	0	0	0
	(.75, .75)	0	0	0	0	0	0	0	0
	(.75, 1)	0	0	0	0	0	0	0	0
	(1, .5)	0	0	0	0.01	0	0	0	0
	(1, .75)	0	0	0	0	0	0	0	0
	(1, 1)	0	0	0	0	0	0	0	0
0.3	(.5, .5)	1	1	1	1	1	1	1	1
	(.5, .75)	0	1	1	0	1	0	0.06	0.82
	(.5, 1)	0	1	1	0	1	0	0.01	0.44
	(.75, .5)	1	0	0.05	0.34	0	1	1	0
	(.75, .75)	0	0	0.04	0.01	0	0	0.04	0.02
	(.75, 1)	0	0	0.07	0	0	0	0	0.02
	(1, .5)	1	0	0	0.09	0	1	1	0
	(1, .75)	0	0	0	0	0	0	0.01	0
	(1, 1)	0	0	0	0	0	0	0	0
0.4	(.5, .5)	1	1	1	1	1	1	1	1
	(.5, .75)	1	1	1	1	1	0.56	1	1
	(.5, 1)	0	1	1	0.47	1	0	0.21	0.94
	(.75, .5)	1	0.64	0.76	1	1	1	1	1
	(.75, .75)	1	0.62	0.74	0.93	1	0.49	0.98	0.85
	(.75, 1)	0	0.72	0.98	0.05	1	0	0.03	0.27
	(1, .5)	1	0	0.03	0.76	0	1	1	0.06
	(1, .75)	1	0	0.04	0.14	0	0.52	0.93	0.02
	(1, 1)	0	0	0	0	0	0	0	0
0.5	(.5, .5)	1	1	1	1	1	1	1	1
	(.5, .75)	1	1	1	1	1	1	1	1
	(.5, 1)	0.98	1	1	1	1	0	0.16	1
	(.75, .5)	1	1	1	1	1	1	1	1
	(.75, .75)	1	1	1	1	1	1	1	1
	(.75, 1)	0.99	1	1	1	1	0	0.22	0.96
	(1, .5)	1	0	0.66	1	0.93	1	1	1
	(1, .75)	1	0.11	0.62	0.85	0.92	1	1	0.97
	(1, 1)	0.95	0.11	0.91	0.67	0.89	0	0.29	0.35