Deep Dive Into Mann-Whitney and Spearman Rank

Deliverance Bougie

Sr. Statistician

August 2018

Deep Dive Into Mann-Whitney and Spearman Rank

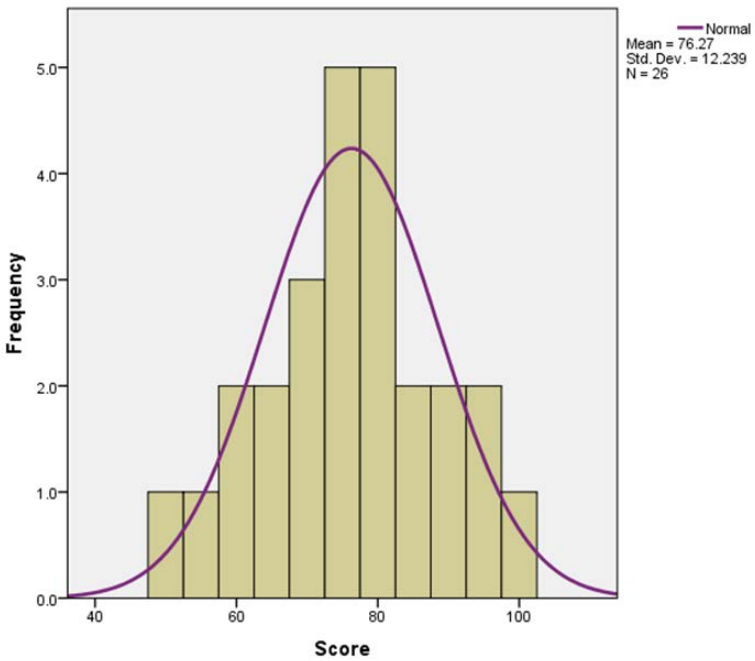
- Mann-Whitney Statistical Analysis
  - Why we use it.
  - Getting technical.
  - What do the results mean.
- Spearman Rank Statistical Analysis
  - Why we use it.
  - Getting technical.
  - What do the result mean.

Mann-Whitney Statistical Analysis

Why do we use it?

- Most statistical tests require certain "assumptions" to be made, such as having a normal distribution (Have you heard of the magical "Bell Curve"?).

- Mann-Whitney is a test that does not require all of these assumptions to be met.



Mann-Whitney Statistical Analysis

Why do we use it?

- Mann-Whitney tests the equality of two independent groups.

- Example: Is the average height of the men and women in this room statistically different?

Mann-Whitney Statistical Analysis

Hypothesis Testing

$$H_o = \mu_1 - \mu_2 = 0$$

$$H_a = \mu_1 - \mu_2 \neq 0$$

Level of significance: $\propto = .05$

Here’s your chance

Mann-Whitney Statistical Analysis

Men_Height	Women_Height	Height	Sex
60	70	60	1
60	70	60	1
70	62	70	1
70	62	70	1
70	62	70	1
71	62	71	1
72	62	72	1
69	62	69	1
68	62	68	1
	62	62	2
	62	62	2
	62	62	2
	62	62	2
	62	62	2
	62	62	2
	62	62	2
	62	62	2
	62	62	2
	62	62	2
	62	62	2
	62	62	2
	62	62	2
	62	62	2
	62	62	2
	62	62	2
	65	62	2
	65	62	2
	65	62	2
	65	62	2
	65	62	2
	68	62	2
	68	62	2
	69	65	2
	70	65	2
		65	2
		65	2
		65	2
		68	2
		68	2
		69	2
		70	2
		70	2
		70	2

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Men_Height	9	60	72	67.78	4.549
Women_Height	32	62	70	63.81	2.856
Valid N (listwise)	7				

Mann-Whitney Test

Ranks

	Sex	N	Mean Rank	Sum of Ranks
Height	1	9	28.44	256.00
	2	32	18.91	605.00
	Total	41		

Test Statistics[a]

	Height
Mann-Whitney U	77.000
Wilcoxon W	605.000
Z	-2.275
Asymp. Sig. (2-tailed)	.023
Exact Sig. [2*(1-tailed Sig.)]	.035[b]

a. Grouping Variable: Sex

b. Not corrected for ties.

Mann-Whitney Statistical Analysis

Equality of means

- If the groups are similar, each observation in the first group will have an equal probability of being greater than or less than each of the observations in the other group.

Mann-Whitney Statistical Analysis

Class Experiment

- Are those who had coffee as awake as those who did not have coffee?
- Are those who stay out late as awake as those who did not stay out late?

Mann-Whitney Statistical Analysis

- If the two conditions are similar, high and low ranks (how awake everyone is) will be distributed rather equally between the two conditions (caffeine/no caffeine or staying out late/not late). The smaller the test statistic, the less likely it is the results occurred by chance.

Mann-Whitney Statistical Analysis

$$U_1 = n_1 n_2 + \frac{n_2(n_1 + 1)}{2} - \sum_{i=n_1+1}^{n_2} R_i$$

Where:

U=Mann-Whitney U Test

n_1= Sample size one

n_2=Sample size two

R_i= Rank of the sample sizes

Mann-Whitney Statistical Analysis

Ratio Studies

- Is the percentage change in the group of sold parcels equal to the percentage change in the group of unsold parcels?

Mann-Whitney Statistical Analysis

Hypothesis test

H_0: Changes were made equally to sold and unsold properties.

H_A: Changes were not made equally to sold and unsold properties.

$$\alpha = .05$$

Mann-Whitney Statistical Analysis

PctChange	Sold
0.072412929	1
0.073389356	1
0.105960265	1
0.131406045	1
0.210801758	1

PctChange	Sold		PctChange	Sold
0.03536346	0		0.06882494	0
0.04735256	0		0.068970588	0
0.0493992	0		0.069044586	0
0.05157457	0		0.069165143	0
0.0520615	0		0.069586675	0
0.05371278	0		0.069751381	0
0.05728806	0		0.070158805	0
0.05851932	0		0.070167064	0
0.0610622	0		0.070290721	0
0.06138614	0		0.071036889	0
0.06158494	0		0.071361502	0
0.06164575	0		0.071915474	0
0.06181728	0		0.071953886	0
0.06189968	0		0.072463768	0
0.06214965	0		0.072566372	0
0.06294201	0		0.073643411	0
0.06350392	0		0.073662445	0
0.06397608	0		0.074056029	0
0.06422704	0		0.074142383	0
0.06496631	0		0.074737345	0
0.06501182	0		0.075288972	0
0.06504242	0		0.077089783	0
0.06630137	0		0.078503586	0
0.06654836	0		0.079029247	0
0.06722898	0		0.083404742	0
0.06820809	0		0.088034577	0
0.06841612	0			

When sample sizes are very different for each group, it can be difficult to determine if there is a (statistically) significant difference.

Mann-Whitney Statistical Analysis

Mann-Whitney Test

Ranks

	Sold	N	Mean Rank	Sum of Ranks
PctChange	0	53	27.45	1455.00
	1	5	51.20	256.00
	Total	58		

Test Statistics[a]

	PctChange
Mann-Whitney U	24.000
Wilcoxon W	1455.000
Z	-3.006
Asymp. Sig. (2-tailed)	.003
Exact Sig. [2*(1-tailed Sig.)]	.001[b]

a. Grouping Variable: Sold

b. Not corrected for ties.

Mann-Whitney Statistical Analysis

Some statistics on the Mann-Whitney test.

	2016	2017	2018	% Change
Total neighborhoods with 5+ sales	3100	4915	6245	101%
Total Mann-Whitney failed neighborhoods	683	1193	1700	149%
Total neighborhoods counties required to explain	61	290	435	613%

Spearman Rank Statistical Analysis

- **Why do we use it?**
  - Just as with the Mann-Whitney, certain assumptions are not required to be met.
  - Measures the strength of the relationship between two variables.

Spearman Rank Statistical Analysis

- Strength of the relationship between the AV/SP ratio and assessed value.
  - Are low- and high- valued properties being assessed at an equal percentage of the market?
  - ρ=1 Perfect positive relationship
  - ρ=-1 Perfect negative relationship
  - And then there is everything in between

Spearman Rank Statistical Analysis

Spearman Rank Formula

- $$\rho = \frac{n\sum_{i=1}^{n} R(x_i)R(y_i) - \left[\sum_{i=1}^{n} R(x_i)\right]\left[\sum_{i=1}^{n} R(y_i)\right]}{\sqrt{\left\{n\sum_{i=1}^{n}[R(x_i)]^2 - \left[\sum_{i=1}^{n} R(x_i)\right]^2\right\}\left\{n\sum_{i=1}^{n}[R(y_i)]^2 - \left[\sum_{i=1}^{n} R(y_i)\right]^2\right\}}}$$

Spearman Rank Statistical Analysis

Or…we could just use *this* formula:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

where:

d_i is the difference between the ranks

n is the number of observations

Spearman Rank Statistical Analysis

The Data

StudySectic	Grouping	StudySaleP	CurrentLan	CurrentImp	CurrentTota	Ratio
ResImp	This Group	177800	26500	140700	167200	0.940382
ResImp	This Group	185000	26900	137300	164200	0.887568
ResImp	This Group	144150	19500	143100	162600	1.127992
ResImp	This Group	174662	23100	139000	162100	0.928078
ResImp	This Group	163000	16400	133300	149700	0.918405
ResImp	This Group	156428	29300	137900	167200	1.068862
ResImp	This Group	171000	22200	122200	144400	0.844444
ResImp	This Group	158600	24800	131300	156100	0.984237
ResImp	This Group	169315	22400	124900	147300	0.869976
ResImp	This Group	166351	22400	127900	150300	0.903511
ResImp	This Group	177312	23500	139000	162500	0.916464
ResImp	This Group	180739	19300	145600	164900	0.912365
ResImp	This Group	165000	13100	131200	144300	0.874545

AV Rank	Ratio Rank	AVRankSq	RatioRankSq	AVRank*RatioRank	d	D-Squared		
12.5	10	156.25	100	125	-2.5	6.25		1-((6*s(d^2))/n(n^2-1))
10	4	100	16	40	-6	36		
9	13	81	169	117	4	16		
7	9	49	81	63	2	4		0.634615
4	8	16	64	32	4	16		
12.5	12	156.25	144	150	-0.5	0.25		
2	1	4	1	2	-1	1		
6	11	36	121	66	5	25		
3	2	9	4	6	-1	1		
5	5	25	25	25	0	0		
8	7	64	49	56	-1	1		
11	6	121	36	66	-5	25		
1	3	1	9	3	2	4		
91	91	818.5	819	751	0	135.5		

Spearman Rank Statistical Analysis

The Results

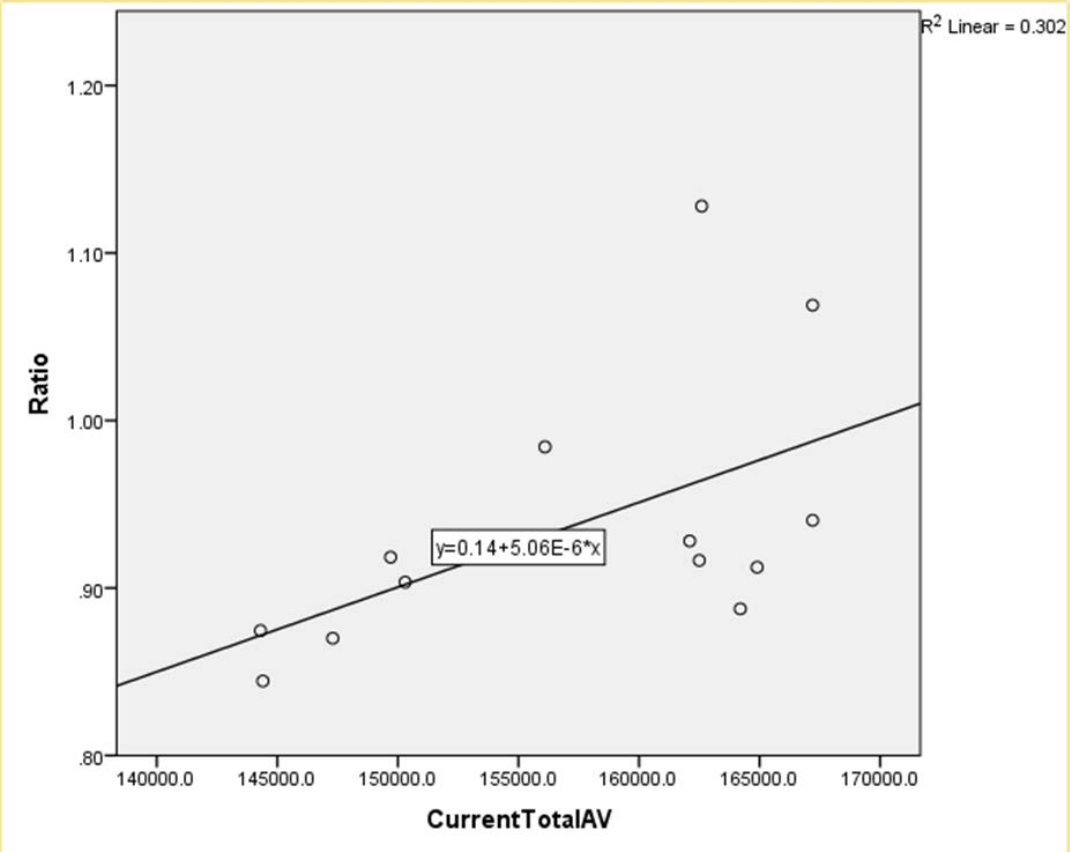
Correlations

			Ratio	CurrentTotalAV
Spearman's rho	Ratio	Correlation Coefficient	1.000	.627*
		Sig. (2-tailed)	.	.022
		N	13	13
	CurrentTotalAV	Correlation Coefficient	.627*	1.000
		Sig. (2-tailed)	.022	.
		N	13	13

*. Correlation is significant at the 0.05 level (2-tailed).

Spearman Rank Statistical Analysis

The Visual



Contact

Deliverance Bougie

- Senior Statistician
- 317.234.5861
- Dbougie@dlgf.in.gov
- www.in.gov/dlgf