Part II

Key

Practice Final covering Chapters 54-56 only

Formulas are on the last page.

Questions 1-7 pertains to the Wilcoxon Mann Whitney test

A randomized double-blind test was done to test the effectiveness of caffeine in improving short-term memory in patients suffering from dementia. 7 patients were given coffee and 7 were given decaf. All patients were given a pre and post memory test (a list of 25 words to memorize). The numbers below are their improvement scores (post test score – pre test score) $n_1 = 7$ $n_2 = 7$ $N = 14$

Caffeine Group: -5, -4, 5, 6, 8, 9, 10 — Decaf group: -10, -1, 0, 1, 2, 3, 4

2, 3, 10, 11, 12, 13, 14 — 1, 4, 5, 6, 7, 8, 9

1) What's the $R_{Caffeine}$, the rank sum for the Caffeine group?

a) 9 **b)** 29 **c)** 40 **d)** 51 **e)** 65

$2+3+10+11+12+13+14 = 65$

2) The sum of the 2 group R statistics must =________ for any 2 groups with a total of 14 members.

a) 38 **b)** 49 **c)** 91 **d)** 98 **e)** 105

$\frac{N(N+1)}{2} = \frac{14(15)}{2} =$

3) Now compute the $U_{Caffeine}$, the U statistic for the Caffeine group.

a) 9 **b)** 12 **c)** 29 **d)** 37 **e)** 40

1, 1, 7, 7, 7, 7, 7

(-5 and -4 outscore -10, 5, 6, 8, 9, 10 outscore all 7 #'s in Decaf group)

4) The sum of the 2 group U statistics must = ________ for any 2 groups with 7 members each.

a) 38 **b)** 49 **c)** 91 **d)** 98 **e)** 105

$n_1 \times n_2 = 7 \times 7$

To test **H_0 :** Caffeinated Coffee works the same as Decaf Coffee in this population.
H_A : Caffeinated Coffee *works **better*** than Decaf in this population.

5) I computed the Z stat for both the caffeine group and the Decaf group using the Rank Sum. How do the two Z stats compare?

- **a)** They're exactly the same.
- **b)** They'll be the same in absolute value, but opposite signs.
- **c)** They'll be close to each other but not exactly the same since the SE's are somewhat different.
- **d)** They'll be close to each other but not exactly the same because their expected values are different.
- **e)** They'll be the same in this case, because the sample sizes of the 2 groups are the same, but they won't be the same when the sample sizes differ.

6) I computed the Z stat for the caffeine group first using the Rank Sum and then using U statistic. How do the two Z stats compare?

- **a)** They're exactly the same.
- **b)** They'll be the same in absolute value, but opposite signs.
- **c)** They'll be close to each other but not exactly the same since the SE's are somewhat different.
- **d)** They'll be close to each other but not exactly the same because their expected values are different.
- **e)** They'll be the same in this case, because the sample sizes of the 2 groups are the same, but they won't be the same when the sample sizes differ.

7) Suppose the sample sizes were too small to use the normal approximation what test stat can we use?

- **a)** t statistic
- **b)** We can still use the z statistic since this is a non-parametric test.
- **c)** Tables that show the exact probability distribution of the U stat.

Never use t stat for rank tests because we know the SD once in ranks. Always use exact probability distributions for small samples and Z (or χ^2 when you have 3 or more groups) for large samples.

Practice Final covering Chapters 54-56 only

Questions 8-13 pertain to the Kruskal Wallis test
Suppose we wanted to test whether time of day affects performance on a stats exam, so we randomly divide 30 students into groups of 10 and assign them to take the same stats exam either in the morning, afternoon, or evening.

H_0: Time of day makes no difference in exam performance.
H_A: Time of day does make a difference for at least one of the groups.

8) The exams scores were ranked from 1 to 30 (lowest to highest) and ranks sums were computed for each group. The rank sum for the afternoon group is missing. What is it? $Sum = N(N+1)/2$

$R_{morning}$= 205 $R_{afternoon}$ =________________________ $R_{evening}$=105

a) 155 **b)** 465 **c)** 360 **d)** not enough info

Sum of all 30 $= \frac{30 \cdot 31}{2} = 465$

$R_{aft} = 465 - 205 - 105 =$ 155

9) The expected rank sum for each group = _________.

a) 150 **b)** 155 **c)** 160

$\frac{\underbrace{10}_{\text{sample size}}\underbrace{(31)}_{(N+1)}}{2}$

10) The H-stat=6.45 To find the p-value you would look at the _______ curve.

a) Z **b)** t **c)** χ^2 **d)** F

11) If 6.45 __$>$__the critical value at α=0.05 then you'd reject the null.

a) is greater than **b)** is less than

12) Rejecting the null at α =0.05 means that....

a) There's less than a 5% chance that time of day makes no difference in exam performance.
b) Morning students perform the best and evening students perform the worst 95% of the time.
c) If time of day made no difference in exam performance, the likelihood that we would see differences this extreme or more between the 3 groups is less than 5%.
d) There's at least a 95% chance that time of day makes a difference in exam performance.
e) All of the above

13) If we knew that the exams scores were normally distributed then instead of doing the Kruskall-Wallis test we could use ... **a)** the Z test **b)** the t-test **c)** ANOVA and the F-stat

Practice Final covering Chapters 54-56 only

Questions 14-18 pertain to the Spearman Rank-Order Correlation Coefficient

14) The four (x,y) pairs: (-6, 1) (-4, -10) (0, 3) (5, 20) have a
in ranks (1,2) (2,1) (3,3) (4,4)
Spearman Rank-Order Correlation Coefficient (r_s) =0.8 Which of the following (x,y) pairs also have a $r_s = 0.8$?

a) (1, 2) (2,1) (3,3) (4,4)

b) (-60,10) (-40,-100) (0,30) (50,200) → (1,2) (2,1) (3,3) (4,4)

c) (1, -6) (-10, -4) (3,0) (20,5) → (2,1) (1,2) (3,3) (4,4)

d) all of the above

e) none of the above

Questions 15-18 pertain to the following situation

Let's say we randomly sampled 4 points from a large population and after converting the points to ranks we got (1,1) (2,2) (3,3) (4,4). obs $r_s = 1$ exp $r_s = 0$ under Ho
We want to test: H_0: population correlation= 0 H_A: population correlation > 0 ← one-sided

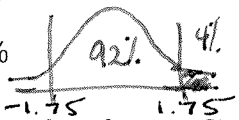
15) If we used the normal approximation our Z stat would be closest to.... $Z = \frac{obs - exp}{SE} = \frac{1-0}{\frac{1}{\sqrt{3}}} = 1.73$
(Answers are rounded to fit the lines on Normal table)

a) Z = 1.3 **b)** Z= 1.65 **c)** Z = 1.75 **d)** Z= 2 **e)** Z = 2.5

$SE_{r_s} = \frac{1}{\sqrt{n-1}} = \frac{1}{\sqrt{3}}$
↑ formula is given (see last page)

16) and the p-value would be closest to .... HA is one-sided

$p = \frac{100-92}{2} = 4$ **a)** <1% **b)** 2.5 % **c)** 4% **d)** 5% **e)** 10%



17) You could also figure the p-value using the exact probability distribution. Since there are 24 possible orders for 4 points the p-value figured that way would be closest to .... $4 \times 3 \times 2 \times 1 = 24$

a) <1% **b)** 2.5 % **c)** 4% **d)** 5% **e)** 10% $\frac{1}{24} \approx 4\%$

18) Suppose we lost the original values of the 4 points, is it possible to figure them out from the 4 ranked points?

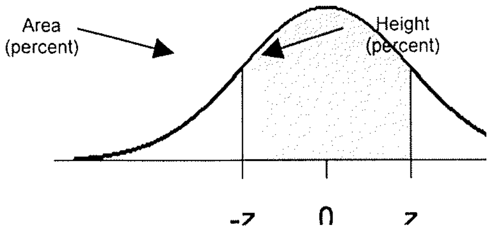
a) Yes, if we convert them to Z scores we can convert them back to their original values.

b) No, it is not possible to convert from rankings back to values.

c) Yes, it's possible by estimating the regression equation from the 4 points and using the equation to solve for their actual values.

There are an infinite # of possibilities for the actual values.

STANDARD NORMAL TABLE



z	Area		z	Area		z	Area
0.00	0.00		1.50	86.64		3.00	99.730
0.05	3.99		1.55	87.89		3.05	99.771
0.10	7.97		1.60	89.04		3.10	99.806
0.15	11.92		1.65	90.11		3.15	99.837
0.20	15.85		1.70	91.09		3.20	99.863
0.25	19.74		1.75	91.99		3.25	99.885
0.30	23.58		1.80	92.81		3.30	99.903
0.35	27.37		1.85	93.57		3.35	99.919
0.40	31.08		1.90	94.26		3.40	99.933
0.45	34.73		1.95	94.88		3.45	99.944
0.50	38.29		2.00	95.45		3.50	99.953
0.55	41.77		2.05	95.96		3.55	99.961
0.60	45.15		2.10	96.43		3.60	99.968
0.65	48.43		2.15	96.84		3.65	99.974
0.70	51.61		2.20	97.22		3.70	99.978
0.75	54.67		2.25	97.56		3.75	99.982
0.80	57.63		2.30	97.86		3.80	99.986
0.85	60.47		2.35	98.12		3.85	99.988
0.90	63.19		2.40	98.36		3.90	99.990
0.95	65.79		2.45	98.57		3.95	99.992
1.00	68.27		2.50	98.76		4.00	99.9937
1.05	70.63		2.55	98.92		4.05	99.9949
1.10	72.87		2.60	99.07		4.10	99.9959
1.15	74.99		2.65	99.20		4.15	99.9967
1.20	76.99		2.70	99.31		4.20	99.9973
1.25	78.87		2.75	99.40		4.25	99.9979
1.30	80.64		2.80	99.49		4.30	99.9983
1.35	82.30		2.85	99.56		4.35	99.9986
1.40	83.85		2.90	99.63		4.40	99.9989
1.45	85.29		2.95	99.68		4.45	99.9991

Formulas:

$$SE_{R_A} = SE_{R_B} = SE_U = \sqrt{\frac{n_A n_B (N+1)}{12}}$$

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{g} \frac{(obsR_i - expR_i)^2}{n_i}$$

$$SE_{r_s} = \frac{1}{\sqrt{n-1}}$$