STA 4502 - LESSON D1 - TWO INDEPENDENT MEDIANS TESTING

METHOD: WILCOXON RANK-SUM TEST

m = NUMBER IN SAMPLE 1

n = NUMBER IN SAMPLE 2

ASSIGN RANKS TO EACH DATA POINT IN THE JOINT SAMPLE, AND (BY CONVENTION) ADD THE RANKS FOR THE DATA IN SAMPLE 2.

EX POINTS SCORED IN NIT BASKETBALL GAMES IN THE PAST 10 YEARS:

SAMPLE 1 (GA): 63, 65, 69, 71, 93

SAMPLE 2 (FL): 74, 77, 97

JOINT SAMPLE: 63, 65, 69, 71, 74, 77, 93, 97

JOINT RANKS: 1, 2, 3, 4, 5, 6, 7, 8

$W = 5+6+8 = 19$ ← TEST STATISTIC

$W_{min} = 1+2+3 = 6$, $W_{max} = 6+7+8 = 21$

$\binom{8}{3} = \frac{8!}{3!5!} = 56$ PERMUTATIONS OF RANKS

USE dwilcox (W - W_{min} , m, n) FOR THE p.d.f.

	21	20	19	18	17	16	15	14
W	6	7	8	9	10	11	12	13
P(W)	$\frac{1}{56}$	$\frac{1}{56}$	$\frac{2}{56}$	$\frac{3}{56}$	$\frac{4}{56}$	$\frac{5}{56}$	$\frac{6}{56}$	$\frac{6}{56}$

TAIL

$H_0: \eta_{GA} - \eta_{FL} = 0 \qquad H_1: \eta_{GA} - \eta_{FL} < 0$

$$\text{P-VALUE} = \left(\frac{1}{56} + \frac{1}{56} + \frac{2}{56}\right) = \frac{4}{56} = \underline{\underline{0.0714}}$$

REJECT H_0 AT $\alpha = 0.10$, ENOUGH EVIDENCE TO CONCLUDE THAT THE MEDIAN POINTS SCORED BY UGA IN NIT GAMES IS LOWER THAN THE MEDIAN POINTS SCORED BY UF IN NIT GAMES.

R CODE:

```
wilcox.test ( sample1, sample 2, alternative =
   " less ", paired= FALSE, exact = TRUE, mu = 0)
```

NOTE: THE PROCEDURE ASSUMES A SHIFT IN THE DISTRIBUTIONS, SO THAT THEY HAVE AN UNKNOWN BUT EQUAL SHAPE.

STA4502 - LESSON D2 - TWO INDEPENDENT MEDIANS ESTIMATION

METHOD: MOSES INTERVAL

LIST ALL POSSIBLE VALUES OF $X_i - Y_j$ IN ORDER FROM LOWEST TO HIGHEST. THE MEDIAN IS THE POINT ESTIMATE FOR THE MEDIAN OF THE DIFFERENCE BETWEEN THE X AND Y OBSERVATIONS.

FLA \ GA	63	65	69	71	93
97	-34	-32	-28	-26	-4
77	-14	-12	-8	-6	16
74	-11	-9	-5	-3	19

$\hat{m} = -9$

$\hat{\hat{m}} = \hat{m}_1 - \hat{m}_2 = 69 - 77 = -8$

THESE ARE OFTEN NOT EQUAL TO EACH OTHER.

THE WILCOXON RANK-SUM TEST STATISTIC HAS A ONE-TO-ONE CORRESPONDENCE WITH THE MANN-WHITNEY U-STATISTIC:

$$U = \sum_{i,j} \phi_{ij} \qquad \phi_{ij} = \begin{cases} 1 & \text{IF } X_i < Y_j \\ 0 & \text{IF } X_i > Y_j \end{cases}$$

$$W = \left(\sum_{i,j} \#\text{ OF } X_i < Y_j\right) + \left(\sum_{j',j} \#\text{ OF } Y_{j'} < Y_j\right) + n$$

$$= U + \left(0 + 1 + 2 + \dots + (n-1)\right) + n = U + \frac{n(n+1)}{2}$$

TAKING ALL POSSIBLE DIFFERENCES OF X-Y OBSERVATIONS (SIMILAR TO WALSH AVERAGES) AND USING THE RANK-SUM DISTRIBUTION LEADS TO A CONFIDENCE INTERVAL FOR THE MEDIAN DIFFERENCE BETWEEN X AND Y OBSERVATIONS.

D 2-3

EX 80% CI FOR m_{X-Y}

	-34	-32	-28	-26	-14	-12	-11	-9	-8	-6	-5	-4	-3	16	19	
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
U →	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

AN 80% CI SHOULD INCLUDE ALL VALUES OF m FOR WHICH A ONE-SIDED RANK-SUM TEST WOULD NOT REJECT H_0 AT $\frac{\alpha}{2} = 0.10$.

WORK INWARD FROM EACH TAIL AND STOP WHEN CUMULATIVE PROBABILITY EXCEEDS 0.10.

$U = 3$, OR $W = 3 + \frac{3(3+1)}{2} = 9$, IS NOT RARE.

THEREFORE, THE CI ENDPOINTS ARE THE 3RD ORDERED DIFFERENCES FROM EACH END.

THE CI IS (-28, -3). I AM 80% CONFIDENT (ACTUALLY 85.71%) THAT THE TRUE MEDIAN POINTS SCORED IN NIT GAMES IS BETWEEN 3 AND 28 POINTS FEWER FOR UGA THAN FOR UF.

R CODE:
```
wilcox.test(sample 1, sample 2, alternative =
    "two.sided", paired = FALSE, exact = TRUE,
    conf.int = TRUE, conf.level = 0.80)
```

STA 4502 - LESSON D3 - TWO INDEP. MEDIANS

SYMMETRIC INFERENCE

WHAT IF THE TWO SAMPLES HAVE UNEQUAL VARIATION? WITH $m = 10$, $n = 3$, A WILCOXON RANK-SUM STATISTIC OF $W = 12$ MIGHT NOT GIVE A COMPLETE INDICATION OF THE MEDIAN DIFFERENCE, e.g. $1+4+7$ VS. $3+4+5$.

INSTEAD, USE A TEST STATISTIC THAT ACCOUNTS FOR VARIANCE BY STANDARDIZING (OR STUDENTIZING).

METHOD: FLIGNER-POLICELLO TEST AND INTERVAL.

$P_i = \#$ OF $Y < X_i$ $\qquad$ $Q_j = \#$ OF $X < Y_j$

$\bar{P} =$ AVG. OF P_i $\qquad$ $\bar{Q} =$ AVG OF Q_j

$V_1 = \sum_i (P_i - \bar{P})^2$ $\qquad$ $V_2 = \sum_j (Q_j - \bar{Q})^2$

$$\hat{U} = \frac{\sum Q_j - \sum P_i}{2\sqrt{V_1 + V_2 + \bar{P}\bar{Q}}}$$

EX

GA	63	65	69	71			93	
FL					74	77		97

$P_i = \{0, 0, 0, 0, 2\}$ $\qquad Q_j = \{4, 4, 5\}$

$\bar{P} = \frac{2}{5}$ $\qquad \bar{Q} = \frac{13}{3}$

$V_1 = 4\left(\frac{2}{5}\right)^2 + \left(\frac{8}{5}\right)^2 = 3.2$ $\qquad V_2 = 2\left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^2 = \frac{2}{3}$

$$\hat{U} = \frac{13-2}{2\sqrt{\frac{16}{5} + \frac{2}{3} + \left(\frac{2}{5}\right)\left(\frac{13}{3}\right)}} = \underline{\underline{2.3242}}$$

THIS IS STILL A NONPARAMETRIC METHOD, AND WE CAN CALCULATE $\hat{U}$ FOR ALL 56 PERMUTATIONS OF RANKS TO OBTAIN AN EXACT P-VALUE. USING EXCEL, TAIL VALUES OF $\hat{U}$ ARE: $\langle \infty, 4.1957, 2.3242, 2.3242 \rangle\!\!/\; 1.6323, 1.4524,$ $1.4524, 1.0853, 1.0853, 1.0853, 0.9223, 0.8216, 0.7217, \ldots$

THE P-VALUE FOR $H_0: \eta_{X-Y} = 0$ VS. $H_1: \eta_{X-Y} < 0$ IS STILL $\frac{4}{56} = \underline{\underline{0.0714}}$, BUT P-VALUES WILL OFTEN BE LOWER THAN WITH THE WILCOXON RANK-SUM TEST. WHY?

R CODE:

{package NSM3}

```
pFligPoli(samplelower, samplehigher, method="Exact")
```

NOTE: THE PROCEDURE ASSUMES SYMMETRIC DISTRIBUTIONS WITH POSSIBLY UNEQUAL VARIANCES.

ALTERNATIVE: MOOD'S MEDIAN TEST.

$$H_0: m_1 - m_2 = 0 \quad \text{VS} \quad H_1: m_1 - m_2 < 0$$

COUNT SUCCESSES AND FAILURES AGAINST A MEDIAN OF THE JOINT SAMPLE. SINCE THE TOTAL NUMBER OF SUCCESSES IS FIXED, THE DIST'N IS THE HYPERGEOMETRIC OF FISHER'S EXACT TEST.

EX JOINT MEDIAN IS 72.5.

	SUCC	FAIL	TOT
GA	1	4	5
FL	3	0	3
TOT	4	4	8

GA SUCC	P(GA SUCC)
1	1/14
2	6/14
3	6/14
4	1/14

MOOD'S MEDIAN TEST HAS RELATIVELY LOW POWER TO DETECT ACTUAL DIFFERENCES IN TWO MEDIANS

COMPARISON WITH 2-SAMPLE t-PROCEDURES:

EQUAL VARIANCES (CASE 1)

RANK-SUM: P-VALUE = 0.0714

$CI_{0.80} = (-28, -3)$

POOLED t: P-VALUE = 0.1424 ($t = -1.1745$, $df = 6$)

$CI_{0.80} = (-23.2974, 2.3641)$

UNEQUAL VARIANCES (CASE 2)

FLIGNER-POLICELLO: P-VALUE = 0.0714

UNPOOLED t: P-VALUE = 0.1535 ($t = -1.1619$, $df = 4.20$)

$CI_{0.80} = (-24.1533, 3.2200)$

NONE OF THE NONPARAMETRIC METHODS STUDIED ARE DESIGNED TO BE USED WITH SKEWED DATA, BUT THEY ARE ROBUST TO MODERATE AMOUNTS OF SKEW.

STAT 4502 LECTURE E1 NOTES

By Sayar Karmakar, Yuhan Tian

sayarkarmakar@ufl.edu, yuhan.tian@ufl.edu

In this lecture, we will discuss about a non-parametric method for testing equality of two variances.

Normal Theory.
When normality is assumed, F-test is the most basic procedure for testing the ratio of two variances. However, the test is highly sensitive to outliers and noticeable departures from normality. Other procedures exist (Bartlett, Breusch-Pagan and Brown-Forsythe) that are improvements, but still not distribution-free.

1. Ansari-Bradley Test.
Ansari-Bradley test is a rank-based test to test equality of two variances. To use this test, we assume equal population medians, form a joint sample, and assign ranks (1 = highest and lowest, 2 = 2nd highest and lowest, etc, up to $\frac{N+1}{2}$ for odd N or $\frac{N}{2}$ for even N). The test statistic is the sum of the ranks for one sample, and the distribution of the test statistic is found by permuting the ranks.

Example.
Triglyceride Levels (mg/dl):

Men: x_i	60	100	150	210		
Women: y_i	55	70	80	110	130	160

Assume population median for men is $\eta_1 = 125$ and population median for women is $\eta_2 = 90$, then we have $x_i - \eta_1 = \{-65, -25, 25, 85\}$ and $y_j - \eta_2 = \{-35, -20, -10, 20, 40, 70\}$.

Joint Sample:	-65*	-35	-25*	-20	-10	20	25*	40	70	85*
Joint Rank:	1*	2	3*	4	5	5	4*	3	2	1*

We are going to test $H_0 : \sigma_1^2 \leq \sigma_2^2$ vs. $H_1 : \sigma_1^2 > \sigma_2^2$, where σ_1^2 and σ_2^2 denote the population variances for men and women, respectively.
The test statistic is

$$C = \sum R_i = 9 \begin{cases} \text{min} & = 1 + 1 + 2 + 2 = 6 \\ \text{max} & = 4 + 4 + 5 + 5 = 18 \end{cases}$$

where lower values of C indicate more support for H_1. There are $\begin{pmatrix} 10 \\ 4 \end{pmatrix} = 210$ possible permutations of ranks and the P-value is Pr(C $\leq$ 9).

C	Ranks				# of Ways	P(C)
6	1	1	2	2	1	1/210
7	1	1	2	3	4	4/210
8	1	1	2	4	4	9/210
	1	1	3	3	1	
	1	2	2	3	4	
9	1	1	2	3	4	16/210
	1	1	3	4	4	
	1	2	2	4	4	
	1	2	3	3	4	

Thus, P-value is $\frac{1+4+9+16}{210} = 0.1429$. (When N is even, C is symmetric.)

2. Confidence Interval for $\frac{\sigma_1}{\sigma_2}$.
To find point estimate and interval, we will invert the Ansari-Bradley test and calculate the Bauer ratios for all (x_i, y_j) pairs with the same sign.

	-35	-20	-10	20	40	70
-65	1.8571	3.25	6.50			
-25	0.7143	1.25	2.50			
25				1.25	0.625	0.3571
85				4.25	2.125	1.2143

Scale ratio (point estimate): $\hat{\gamma} = \text{median}_{i,j}\,(x_i/y_j \mid x_i y_j > 0)$
$\hat{\gamma} = \frac{1.25+1.8571}{2} = 1.5536$
The 80% confidence interval for $\gamma = \frac{\sigma_1}{\sigma_2}$ must exclude the lowest 10% and highest 10% of the distribution of C. These are the 3 lowest and 3 highest values of C, which correspond to the 3 lowest/highest Bauer ratios. Thus, the 80% confidence interval in this example is $(0.7143, 3.2500)$.

3. R Code.
men ← c(-65,-25,25,85)
women ← c(-35,-20,-10,20,40,70)
ansari.test(men, women, alternative = “two.sided”, exact = TRUE, conf.int = TRUE, conf.level = 0.80)

4. Comparison with Normal Theory.
$H_0 : \sigma_1^2/\sigma_2^2 \leq 1$ vs. $H_1 : \sigma_1^2/\sigma_2^2 > 1$
$F = 2.6512$
P-Value $= 0.1602$
80% CI for $\sigma_1/\sigma_2 = (0.6707, 2.9399)$
$\hat{\sigma}_1/\hat{\sigma}_2 = 1.6283$

5. Large-Sample Approximation for C.

$C^* = \frac{9-12+\frac{1}{2}}{\sqrt{16/3}} = -1.0825$

P-Value = 0.1395

The "$\frac{1}{2}$" in the numerator is the continuity correction.

STAT 4502 LECTURE E2 NOTES

By Sayar Karmakar, Yuhan Tian

sayarkarmakar@ufl.edu, yuhan.tian@ufl.edu

In this lecture, we will discuss about a non-parametric method to simultaneously test for equal location and scale parameters for two populations. The core idea behind this test is that location and scale parameters can be constructed, so that they are independent.

1. Joint Location and Scale Testing (Lepage).

We will use the standardized Wilcoxon Rank-Sum statistic (W^*) and the standardized Ansari-Bradley statistic (C^*). These standardized statistics should be close 0 under the joint null hypothesis. Formally, our setup for the joint hypothesis test is

$$H_0 : \eta_1 - \eta_2 = 0 \text{ and } \frac{\gamma_1}{\gamma_2} = 1 \text{ vs. } H_1 : \eta_1 - \eta_2 \neq 0 \textbf{ or } \frac{\gamma_1}{\gamma_2} \neq 1.$$

The test statistic is $D = (W^*)^2 + (C^*)^2$. (D^2 is the Euclidean distance from H_0.) The exact p-value can be obtained from all permutations of ranks.

2. Example.

Nit Scores:

UGA = $\{63, 65, 69, 71, 93\}$

UF = $\{74, 77, 97\}$

E(W) = $\frac{27}{2}$ VAR(W) = $\frac{45}{4}$

E(C) = $\frac{15}{2}$ VAR(C) = $\frac{75}{28}$

There are $\binom{8}{3} = 56$ permutations of ranks.

PERM	J	W	C	D	PERM	J	W	C	D
ABC	15	6	6	5.8406	ADH	5	13	6	0.8622
ABD	12	7	7	3.8489	AEF	6	12	8	0.2933
ABE	10	8	7	2.7822	AEG	5	13	7	0.1156
ABF	10	9	6	2.6400	BCD	12	9	9	2.6400
ABG	10	10	5	3.4222	BCE	9	10	9	1.9289
ABH	10	11	4	5.1289	BCF	7	11	8	0.6489
*ACD	12	8	8	2.7822	BCG	7	12	7	0.2933
ACE	9	9	8	1.8933	BCH	7	13	6	0.8622
ACF	7	10	7	1.1822	BDE	9	11	10	2.8889
ACG	7	11	6	1.3956	BDF	6	12	9	1.04
ACH	7	12	5	2.5333	BDG	4	13	8	0.1156
ADE	9	10	9	1.9289	BEF	6	13	9	0.8622
ADF	6	11	8	0.6484	CDE	9	12	11	4.7733
ADG	5	12	7	0.2933	CDF	6	13	10	2.3556

For our sample, we have D = 2.7822 and P-value = $P(D \geq 2.7822) = \frac{8}{28} = 0.2857$

3. R Code.

Package: {NSM3}

```
pLepage(uga,uf)
```

STAT 4502 LECTURE E3 NOTES

By Sayar Karmakar, Yuhan Tian

sayarkarmakar@ufl.edu, yuhan.tian@ufl.edu

In this lecture, we will discuss about a non-parametric method to test **any** difference in the distributions of two populations. The idea behind is to estimate the cumulative distribution function (C.D.F) of each population based on sample data.

1. Distribution Difference Testing (Kolmogorov-Smirnov).

Hypothesis Test:. $H_0 : F(t) = G(t)$ for all t vs. $H_1 : F(t) \neq G(t)$ for some t

Test Statistic:. The highest difference between two CDFs across all possible values of the variable.

Theoretical Method:

$J = \frac{mn}{d} \max_t |F_m(t) - G_n(t)|$

Computational Method:

$J = \frac{mn}{d} \max_i \left|F_m\left(z_{(i)}\right) - G_n\left(z_{(i)}\right)\right|$

d is the greatest common factor of m and n.

$z_{(i)}$ is the ith ordered observation in the joint sample.

Example:.

Nit Scores:

UGA = {63,65,69,71,93}

UF = {74,77,97}

Joint = {63,65,69,71,74,77,93,97}

F_m	0	1/5	1/5	2/5	2/5	3/5	3/5	4/5	4/5	4/5	4/5	4/5	4/5	1	1	1	1
$Z_{(i)}$	$\rightarrow$	63	$\rightarrow$	65	$\rightarrow$	69	$\rightarrow$	71	$\rightarrow$	74	$\rightarrow$	77	$\rightarrow$	93	$\rightarrow$	97	$\rightarrow$
G_n	0	0	0	0	0	0	0	0	0	1/3	1/3	2/3	2/3	2/3	2/3	1	1

Based on the table above, the highest difference between two empirical CDFs is $|4/5 - 0|$ when $71 \leq Z_{(i)} < 74$. Thus, we have $J = \frac{5*3}{1} * |\frac{4}{5} - 0| = 12$.

PERM	J	PERM	J
ABC	15	ADH	5
ABD	12	AEF	6
ABE	10	AEG	5
ABF	10	BCD	12
ABG	10	BCE	9
ABH	10	BCF	7
*ACD	12	BCG	7
ACE	9	BCH	7
ACF	7	BDE	9
ACG	7	BDF	6
ACH	7	BDG	4
ADE	9	BEF	6
ADF	6	CDE	9
ADG	5	CDF	6

Based on the permutation table (part of the table from Lecture E2) above, we can summarize the distribution of J as below.

J	3	4	5	6	7	8	9	10	11	12	13	14	15
P(J)	0	1/28	3/28	5/28	6/28	0	5/28	4/28	0	3/28	0	0	1/28

Thus, we have P-value = $P(J \geq 12) = 4/28 = 0.1429$.

R Code:.
ks.test(uga, fl, alternative = “two.sided”, exact = TRUE)

STA 4502 - LESSON F1 - ANOVA COMPLETELY RANDOMIZED DESIGN

METHOD: KRUSKAL-WALLIS TEST

PURPOSE: COMPARE THREE OR MORE TREATMENTS

ASSUMPTIONS: INDEPENDENT SAMPLES, SAME SHAPE FOR ALL POPULATIONS, PERHAPS DIFFERENT LOCATION PARAMETERS.

NORMAL THEORY: F TEST STATISTIC IS THE RATIO OF THE SUM OF SQUARES BETWEEN TREATMENTS (SCALED) TO THE SUM OF SQUARES WITHIN TREATMENTS (ALSO SCALED).

NONPARAMETRIC VERSION USES JOINT RANKS:

TEST STATISTIC:

$$H = \frac{12}{N(N+1)} \sum_j n_j \left(\bar{R}_j - \frac{N+1}{2} \right)^2$$

EX HEAD IMPACT FORCE FROM FRONTAL COLLISION FOR THREE TYPES OF LACROSSE HELMETS:

CASCADE: 976, 990, 1033, 1071, 1099, 1240, 1348, 1388

AIRFIT: 866, 875, 1067, 1077, 1168, 1193

ULTRALITE: 835, 847, 852, 951, 1013, 1144

JOINT RANKS:

		n_j	R_j	$\bar{R}_j$
SAMPLE A:	7, 8, 10, 12, 14, 18, 19, 20	8	108	13.5
SAMPLE B:	4, 5, 11, 13, 16, 17	6	66	11.0
SAMPLE C:	1, 2, 3, 6, 9, 15	6	36	6.0

$$H = \frac{12}{(20)(21)}\left[8\,(13.5-10.5)^2 + 6\,(11.0-10.5)^2 + 6\,(6.0-10.5)^2\right]$$

$$= \frac{39}{7} = 5.5714$$

NOTE: $$H = \frac{12}{N(N+1)} \sum_j \left(n_j \bar{R}_j^{\,2} - 2\, n_j \bar{R}_j \frac{N+1}{2} + n_j \frac{(N+1)^2}{2} \right)$$

$$= \frac{12}{N(N+1)} \left(\sum_j \left(n_j \frac{R_j^2}{n_j^2} \right) - (N+1) \sum_j n_j \bar{R}_j + \frac{(N+1)^2}{2} \sum_j n_j \right)$$

$$= \frac{12}{N(N+1)} \sum_j \left(\frac{R_j^2}{n_j} \right) - \frac{12}{N} \cdot \frac{N(N+1)}{2} + \frac{3(N+1)}{N} \cdot N$$

$$= \left[\frac{12}{N(N+1)} \sum_j \frac{R_j^2}{n_j} \right] - 3(N+1)$$

$$H = \frac{12}{(20)(21)} \left[\frac{108^2}{8} + \frac{66^2}{6} + \frac{36^2}{6} \right] - 3(21) = \underline{\underline{5.5714}}$$

VALUES CLOSE TO ZERO SUPPORT $H_0: \eta_1 = \eta_2 = \eta_3$

VALUES FAR FROM ZERO SUPPORT H_1: AT LEAST ONE η_j IS DIFFERENT FROM OTHERS

EXACT DISTRIBUTION? THERE ARE $\dfrac{20!}{8!\,6!\,6!} = 116,396,280$

PERMUTATIONS OF RANKS. LET'S NOT DO THIS BY HAND.

LARGE-SAMPLE APPROXIMATION:

H IS APPROXIMATELY χ^2 WITH K-1 d.f. FOR k TREATMENTS. P-VALUE = 0.0617

{R package MultNonParam}

```
impact <- c(976, 990, ..., 1013, 1144)   <- ALL 20 VALUES
helmet <- c(rep(1,8), rep(2,6), rep(3,6))
aov.P(impact, helmet)
```

EXACT P-VALUE = 0.058594 (TAKES ~ 30 SECONDS!)

$= \frac{6820098}{116,396,280}$

REJECT H_0 AT $\alpha = 0.10$, ENOUGH EVIDENCE TO CONCLUDE THAT THE MEDIAN FRONTAL IMPACT FORCE IS NOT THE SAME FOR ALL THREE TYPES OF LACROSSE HELMETS.

NOTE THAT THE CONCLUSION IS SOMEWHAT AMBIGUOUS, SINCE IT DOES NOT IDENTIFY WHICH HELMET(S) HAVE A DIFFERENT MEDIAN, AND WHETHER THE DIFFERENT MEDIAN(S) IS/ARE HIGHER OR LOWER THAN THE OTHERS.

STA 4502 - LESSON F2 - ANOVA CRD

MULTIPLE COMPARISON INFERENCE

IDEA: DO A FOLLOW-UP ANALYSIS WHEN H_0 IS REJECTED BY THE KRUSKAL-WALLIS TEST.

METHOD: THE TREATMENTS ARE COMPARED PAIRWISE USING THE RANK-SUM TEST.

EX HELMET STUDY

A vs. B:

A vs. B:			3	4	5		7		9			12	13	14
	1	2				6		8		10	11			

A vs C:

A vs C:					5	6		8	9	10		12	13	14
	1	2	3	4			7				11			

B vs. C:

B vs. C:				4	5			8	9		11	12
	1	2	3			6	7			10		

$W_{AB} = 38 \qquad$ P-VALUE $= \frac{1242}{3003} = \underline{\underline{0.4136}}$

$W_{AC} = 28 \qquad$ P-VALUE $= \frac{88}{3003} = \underline{\underline{0.0293}}$

$W_{BC} = 29 \qquad$ P-VALUE $= \frac{122}{924} = \underline{\underline{0.1320}}$

NOTE: ONLY THE A-C DIFFERENCE IS SIGNIFICANT.

CONCLUSION: THE MEDIAN FRONTAL IMPACT IS HIGHER FOR HELMET A (CASCADE) THAN FOR HELMET C (ULTRALITE).

F2-2

WHEN PERFORMING MULTIPLE COMPARISONS, IT IS IMPORTANT TO CONTROL THE EXPERIMENT-WISE (FAMILY) ERROR RATE. SEVERAL OPTIONS ARE AVAILABLE (BONFERRONI, SCHEFFE, TUKEY). THE BONFERRONI METHOD IS EASIEST TO APPLY (BUT MOST CONSERVATIVE OPTION).

SET $\alpha' = \frac{\alpha}{k(k-1)/2}$ FOR EACH PAIRWISE COMPARISON TO ACHIEVE A FAMILY ERROR RATE NO HIGHER THAN α.

IF $\alpha = 0.12$ AND $k = 3$, THEN $\alpha' = 0.04$

96% CI FOR $\eta_A - \eta_B = (-101, 311)$ $\quad \hat{\eta}_{A-B} = 112.5$

96% CI FOR $\eta_A - \eta_C = (20, 405)$ $\quad \hat{\eta}_{A-C} = 192$

96% CI FOR $\eta_B - \eta_C = (-85, 321)$ $\quad \hat{\eta}_{B-C} = 90$

I AM 88% CONFIDENT THAT THE MEDIAN FRONTAL IMPACT FORCE ON HELMET A IS B/W 20 AND 405 FORCE UNITS HIGHER THAN THE MEDIAN IMPACT OF HELMET C, η_A IS B/W 101 LOWER AND 311 HIGHER THAN η_B, AND η_B IS B/W 85 LOWER AND 321 HIGHER THAN η_C.

COMPARISON WITH ANOVA F-TEST:

SOURCE	df	SS	MS	F	P-VAL
TRT	2	141970.34	70985.17	3.3912	0.0576
ERR	17	355846.21	20932.13	—	
TOT	19	497816.55	—		

$(S_p = 144.68)$

A : $\bar{X}_A = 1143.125$ $S_A = 161.23$ $P_{AB} = 0.2086$

B : $\bar{X}_B = 1041.000$ $S_B = 140.99$ $P_{AC} = 0.0189$

C : $\bar{X}_C = 940.333$ $S_C = 122.05$ $P_{BC} = 0.2447$

96% CI FOR $\mu_A - \mu_B = (-72, 276)$ $\hat{\mu}_A - \hat{\mu}_B = 102$

96% CI FOR $\mu_A - \mu_C = (29, 377)$ $\hat{\mu}_A - \hat{\mu}_C = 203$

96% CI FOR $\mu_B - \mu_C = (-85, 286)$ $\hat{\mu}_B - \hat{\mu}_C = 101$

NOTE: KRUSKAL-WALLIS TEST CAN ALSO BE PERFORMED IN R USING kruskal.test, AND THE P-VALUE CAN BE CALCULATED USING pKW.

STAT 4502 LECTURE G1 NOTES

By Sayar Karmakar, Steven Goodman

sayarkarmakar@ufl.edu, s.goodman@ufl.edu

1. Testing for Randomized Block Designs.

The Randomized Block Design is an experimental design that includes (usually) two factors. We call one of the factors the blocking factor, which often to divides the subjects into homogeneous subgroups in the hopes of reducing variance in the experiment. Within each block, we then apply different levels of the treatment factor to the subjects.

We are interested in analyzing the median responses for each treatment across the blocks to see if there exist significant treatment effects.

Suppose we have a randomized block design with n blocks and k treatments. Let c_{ij} denote the number of subjects in the i^{th} block and j^{th} treatment. To carry out the desired analyses, we must assume:

1. The $N = \sum_{i=1}^{n} \sum_{j=1}^{k} c_{ij}$ total observations are mutually independent.
2. For each block i and treatment j, the c_{ij} subjects are a random sample from a distribution with distribution function F_{ij}.
3. For each i and j, the distribution functions are related in the following way:

$$F_{ij}(t) = F(t - \beta_i - \tau_j)$$

   where F is a distribution function with unknown median Θ, β_i is the effect from the i^{th} block and τ_j is the effect from the j^{th} treatment.

2. Friedman Test for General Alternatives.

Now suppose we have a randomized block design with all $c_{ij} = 1$. In other words, there is 1 subject in every (block i, treatment j) combination, denoted by X_{ij}. Recall that we have n blocks and k treatments.

In this particular setting, the Friedman Test allows us to test for the existence of treatment effects across the k different treatments. Formally, the hypotheses are:

$$H_0 : \tau_1 = \tau_2 = ... = \tau_k \text{ vs. } H_1 : \tau_1, \tau_2, ..., \tau_k \text{ not all equal}$$

To compute the Friedman test statistic, we first order the k subjects from least to greatest within each block. Then for each block i, let r_{ij} denote the rank of X_{ij} in that ordering.

Then, set

$$R_j = \sum_{i=1}^{n} r_{ij} \text{ and } \bar{R}_j = \frac{R_j}{n}$$

In other words, R_j is sum of the ranks of treatment j subjects across all n blocks. $\bar{R}_j$ is the average rank of these observations across the blocks.

The Friedman statistic is:

$$S = \frac{12n}{k(k+1)} \sum_{j=1}^{k} \left(\bar{R}_j - \frac{k+1}{2} \right)^2$$

.

Example 2.1. *SAT Math scores by school and major.*

Note that in this example, the treatment factor is the university the students attend and the blocking factor is the students' major. Therefore we have $k = 3$ treatment levels and $n = 4$ blocks. The subscripts for each subject denote their within-block ranking, r_{ij}.

We are interested in testing

$$H_0 : \eta_{\text{UF}} = \eta_{\text{FSU}} = \eta_{\text{UWF}} \text{ vs. } H_1 : \text{Not all three medians equal.}$$

	Stat	Bio.	Bus.	Psych.	$\bar{R}_j$
UF	770_3	670_2	710_3	610_3	11/4
FSU	680_2	690_3	590_2	560_2	9/4
UWF	620_1	530_1	500_1	510_1	1

We calculate S by:

$$S = \frac{12(4)}{3(3+1)} \left[(\frac{11}{4} - 2)^2 + (\frac{9}{4} - 2)^2 + (1-2)^2 \right] = 6.5$$

Using R, the p-value is computed to be 0.038. At level $\alpha = 0.05$, we reject H_0 and conclude that at not all medians are equal.

Example 2.2. *Testing for block effects.*

Suppose we want to test

$$H_0 : \eta_{\text{Sta}} = \eta_{\text{Bio}} = \eta_{\text{Bus}} = \eta_{\text{Psy}} \text{ vs. } H_1 : \text{Not all four medians are equal.}$$

We can simply redefine the blocking factor as the school the students attend, and the treatment factor as the students' majors. We now have $k = 4$ treatments and $n = 3$ blocks:

	Stat	Bio.	Bus.	Psych.
UF	770_4	670_2	710_3	610_1
FSU	680_3	690_4	590_2	560_1
UWF	620_4	530_3	500_1	510_2
$\bar{R}_j$	11/3	3	2	4/3

Then S is calculated by:

$$S = \frac{12(3)}{4(4+1)}\left[(\frac{11}{3}-\frac{5}{2})^2+(3-\frac{5}{2})^2+(2-\frac{5}{2})^2+(\frac{4}{3}-\frac{5}{2})^2\right] = 5.8$$

3. Code. Using R for the Friedman Test as in Example 2.1:

```
SAT <- c(770,670,710,610,680,690,590,560,620,530,500,510)
school <- c(1,1,1,1,2,2,2,2,3,3,3,3)
major <- c(1,2,3,4,1,2,3,4,1,2,3,4)
friedman.test(SAT, school, major)
```

STAT 4502 LECTURE G2 NOTES

By Sayar Karmakar, Steven Goodman

sayarkarmakar@ufl.edu, s.goodman@ufl.edu

1. Inference for the Randomized Block Design.

When we use the Friedman Test to analyze treatment effects in a randomized block design experiment, we are testing for general differences in the medians of the k treatment groups. When we reject

$$H_0 : \tau_1 = \tau_2 = ... = \tau_k$$

and conclude

$$H_1 : \tau_1, \tau_2, ..., \tau_k \text{ not all equal,}$$

we are interested in conducting multiple comparisons to find exactly which treatment effects (medians) are significantly different.

We can use the two-sided Signed-Rank test to test for unequal medians between treatment groups

Example 1.1. *SAT Math Scores.*

Consider the data from the Lecture G1 of SAT scores across majors (the blocks) and universities (the treatments). Recall that the Friedman test rejected H_0.

	Stat	Bio.	Bus.	Psych.	$\bar{R}_j$
UF	770_3	670_2	710_3	610_3	11/4
FSU	680_2	690_3	590_2	560_2	9/4
UWF	620_1	530_1	500_1	510_1	1

First, we test

$$H_0 : \eta_{UF} = \eta_{FSU} \text{ vs. } H_1 : \eta_{UF} \neq \eta_{FSU}$$

with the Signed-Rank Test.

The set of differences Z_i of UF - FSU scores is $\{90, -20, 120, 50\}$. We take the sum of the ranks of the positive differences to calculate the Signed-Rank Test statistic:

$$T^+ = 2 + 3 + 4 = 9$$

We now derive the null distribution of T^+ for $n = 4$ (refer to Lecture B3):

x	0	1	2	3	4	5	6	7	8	9	10
$P(T^+ = x)$	1/16	1/16	1/16	2/16	2/16	2/16	2/16	2/16	1/16	1/16	1/16

and find the two-sided p-value of $p = \frac{1}{4}$. We fail to reject H_0, and conclude that there exists no significant difference between the in SAT score between UF and FSU students across the different majors.

To compute an 80% confidence interval for $\eta_1 = \eta_{UF} - \eta_{FSU}$, we order the Walsh Averages from least to greatest. Recall that Walsh Averages are of the form

$$\frac{Z_i + Z_j}{2}$$

for all $i \leq j$.

The ordered set of Walsh Averages is

$$\{-20, 15, 35, 50, 50, 70, 85, 90, 105, 120\}.$$

We now look for the lowest "reasonable" value of T^+ puts (approximately) $\frac{\alpha}{6} = 0.033$ in each tail, and then place these bounds on the ordered set of Walsh Averages. (Refer to lecture B4)

$$(\text{Lowest Walsh Average, Highest Walsh Average}) = (-20, 120).$$

We then test the next pair of treatments, FSU and UWF. Our hypotheses are:

$$H_0 : \eta_{FSU} = \eta_{UWF} \text{ vs. } H_1 : \eta_{FSU} \neq \eta_{UWF}$$

The set of differences Z_i of FSU - UWF scores is $\{60, 100, 90, 50\}$.

The ordered Walsh Averages are:

$$\{50, 55, 60, 70, 75, 90, 105, 110, 125, 160\}.$$

To complete the pairwise comparisons, we would need to repeat these procedures for $\eta_3 = \eta_{UF} - \eta_{UWF}$.

80% confidence interval for the pairs are

- UF-FSU : (-20,120)
- FSU-UWF: (50,160) and
- UF-UWF: (something positive, bigger positive) (Pleae complete)

Note that it often of interest to apply these same analyses to the blocks. However, is that specific instance, the Friedman Test didn't return a significant result when applied to the blocks, so the above follow-up procedures aren't required.

STAT 4502 LECTURE G3 NOTES

By Sayar Karmakar, Steven Goodman

sayarkarmakar@ufl.edu, s.goodman@ufl.edu

1. Mack-Skillings Test for Randomized Block Designs.

We are again concerned with analyzing the randomized block design with k treatments and n blocks. Our general assumptions the same as before:

1. The $N = \sum_{i=1}^{n} \sum_{j=1}^{k} c_{ij}$ total observations are mutually independent.
2. For each block i and treatment j, the c_{ij} subjects are a random sample from a distribution with distribution function F_{ij}.
3. For each i and j, the distribution functions are related in the following way:

$$F_{ij}(t) = F(t - \beta_i - \tau_j)$$

   where F is a distribution function with unknown median Θ, β_i is the effect from the i^{th} block and τ_j is the effect from the j^{th} treatment.

For the Mack-Skillings Test, we further require that all of the nk treatment-block combinations contain c observations. Therefore our total number of observations is simply

$$N = nkc.$$

As before, we are concerned with testing

$$H_0 : \tau_1 = \tau_2 = ... = \tau_k \text{ vs. } H_1 : \tau_1, \tau_2, ..., \tau_k \text{ not all equal}$$

where τ_i is the location shift of the i^{th} treatment.

To calculate the test statistic, we denote the q^{th} observation in the (block, treatment) combination (i, j) as X_{ijq}. Then, within each of the n blocks, we rank the kc observations from least to greatest. Let r_{ijq} denote the rank of X_{ijq} in this ordering. We then set, for $j = 1, 2, ...k$,

$$S_j = \sum_{i=1}^{n} \left[\sum_{q=1}^{c} \frac{r_{ijq}}{c} \right]$$

Note that for each fixed j,

$$\sum_{q=1}^{c} \frac{r_{ijq}}{c}$$

is the average rank of the c treatment j observations in the i^{th} block ordering. Therefore S_j is the sum, across the n blocks, of these averages.

We then calculate the test statistic MS by:

$$MS = \frac{12}{k(N+n)} \sum_{j=i}^{k} \left[S_j - \frac{N+n}{2} \right]^2 = \frac{12}{k(N+n)} \left[\sum_{j=i}^{k} S_j^2 \right] - 3(N+n)$$

Example 1.1. *SAT Math Scores*

Here, the blocking factor is whether the students' major is in STEM of not, and the treatment factor is the university they attend. Therefore $k = 3$, $n = 2$, and $c = 2$. The subscripts of each observation denote r_{ijq}.

	STEM	Non-STEM	Avg. Rank	S_j
UF	770_6, 670_3	$710_6, 610_5$	5	10
FSU	680_4, 690_5	$590_4, 560_3$	4	8
UWF	620_2, 530_1	$500_1, 510_2$	3/2	3

Our hypotheses in this case are:

$$H_0 : \eta_{\text{UF}} = \eta_{\text{FSU}} = \eta_{\text{UWF}} \text{ vs. } H_1 : \text{Not all three medians equal.}$$

The Mack-Skillings Test statistic is:

$$\frac{12}{3(12+2)} \left[(10-7)^2 + (8-7)^2 + (3-7)^2 \right] = 7.42$$

and the p-value can be shown to be approximately 0.0122.

Now, suppose we reverse how we defined the blocking and treatment factors. When we define the blocking factor as the universities and the treatment as STEM or non-STEM, we have $k = 2, n = 3$, and $c = 2$.

	STEM	Non-STEM
UF	$770_4, 670_2$	$710_3, 610_1$
FSU	$680_3, 690_4$	$590_2, 560_1$
UWF	$620_4, 530_3$	$500_1, 510_2$
Avg. Rank	10/3	5/3
S_j	10	5

Let η_S denote the median of scores for students with STEM majors and η_N denote those of non-STEM majors. We are testing:

$$H_0 : \eta_S = \eta_N \text{ vs. } H_1 : \eta_S \neq \eta_N.$$

We calculate

$$MS = \frac{12}{2(12+3} \left[(10 - \frac{15}{2})^2 + (5 - \frac{15}{2})^2 \right] = 5$$

and compute a p-value of 0.037. Therefore we reject H_0 and conclude that $\eta_S \neq \eta_N$.

STA 4502 - LESSON G4 - ANOVA 2-FACTOR INTERACTION TEST AND MULTIPLE COMPARISONS

METHOD: HETTMANSPERGER-ELMORE TEST

IDEA: DISTRIBUTION-FREE TEST FOR INTERACTION BETWEEN TWO FACTORS

THE TWO-FACTOR LAYOUT MODEL IS:

$$\underbrace{\mu_{ij}}_{\text{CELL MEAN}} = \underbrace{\mu}_{\text{GRAND MEAN}} + \underbrace{\alpha_i}_{\text{TREATMENT } i \text{ EFFECT}} + \underbrace{\beta_j}_{\text{BLOCK } j \text{ EFFECT}} + \underbrace{(\alpha\beta)_{ij}}_{\text{TRT/BLK } ij \text{ INTERACTION EFFECT}}$$

TO TEST FOR INTERACTION, RECENTER THE DATA USING THE GRAND MEAN, TREATMENT EFFECT, AND BLOCK EFFECT, THEN USE RANKS TO ASSESS SIGNIFICANT DIFFERENCES BETWEEN CELLS. P-VALUES CAN BE OBTAINED USING LARGE-SAMPLE APPROXIMATIONS, RANK PERMUTATIONS, OR MONTE-CARLO SIMULATION.

G 4-2

FACTOR A \ FACTOR B	STEM	NON-STEM	
UF	770, 670	710, 610	690
FSU	680, 690	590, 560	630
UWF	620, 530	500, 510	540
$\bar{X}$	660	580	620

$\hat{\mu} = 620$ $\hat{\alpha}_1 = 690-620 = 70$ $\hat{\alpha}_2 = 630-620 = 10$

$\hat{\alpha}_3 = 540-620 = -80$ $\hat{\beta}_1 = 660-620 = 40$ $\hat{\beta}_2 = 580-620$

SUBTRACT THE GRAND MEAN, THE iTH ROW FACTOR A EFFECT, AND THE jTH COLUMN FACTOR B EFFECT.

ADJUSTED DATA:

	STEM	NON-STEM	$\bar{R}$
UF	40 (10.5), -60 (1)	60 (12), -40 (3)	6.625
FSU	10 (7.5), 20 (9)	0 (5.5), -30 (4)	6.500
UWF	40 (10.5), -50 (2)	0 (5.5), 10 (7.5)	6.375
$\bar{R}$	6.750	6.250	6.500

JOINTLY RANK ALL 12 DATA POINTS.

NOTE: $N = 12$, $k = 3$, $n = 2$, $C = 2$

TEST STATISTIC:

$$Q = \frac{12c}{N(N+1)} \sum_i \sum_j \left(\overline{R}_{ij} - \overline{R}_{i\cdot} - \overline{R}_{\cdot j} + \frac{N+1}{2} \right)^2$$

$$= \frac{12(2)}{12(13)} \Big[(-1.125)^2 + (1.125)^2 + (1.5)^2 + (-1.5)^2 + (-0.375)^2 + (0.375)^2 \Big] = \underline{\underline{1.1250}}$$

UNDER H_0: NO INTERACTION BETWEEN SCHOOL & MAJOR, LARGE SAMPLE DIST'N IS $\chi^2_{(i-1)(j-1)} = \chi^2_2$

WITH APPROXIMATE P-VALUE = $\underline{\underline{0.5698}}$

FTR H_0, NOT EE TO CONCLUDE A NON-ADDITIVE INTERACTION EFFECT BETWEEN SCHOOL AND MAJOR. PROCEED WITH COMPARISONS OF LEVELS OF FACTOR A AND FACTOR B.

IF H_0 WAS REJECTED, EE TO CONCLUDE AN INTERACTION EFFECT. MUST ANALYZE EACH FACTOR SEPARATELY AT THE DIFFERENT LEVELS OF THE OTHER FACTOR.

A LARGE-SAMPLE METHOD FOR IDENTIFYING SIGNIFICANT TREATMENT DIFFERENCES IN TWO-WAY ANOVA WITH INTERACTION IS GIVEN BY:

$$LSD_{TUKEY(\Delta S)} = q_{\alpha,\infty}\sqrt{\frac{k(N+n)}{12}}$$

AN EXACT METHOD FOR IDENTIFYING SIGNIFICANT TREATMENT DIFFERENCES IN TWO-WAY ANOVA WITH INTERACTION IS GIVEN BY:

$$LSD_{MS(\Delta S)} = \sqrt{\frac{k}{6}(N+n)\,ms_{\alpha}}$$

EX $S_1 = 10$, $S_2 = 8$, $S_3 = 3$, $k = 3$, $n = 2$, $c = 2$, $N = 12$,

LET $\alpha = 0.20$, $q_{\alpha,3,\infty} = 2.425$

$$LSD_{TUKEY(\Delta S)} = 2.425\sqrt{\frac{3(12+2)}{12}} = 4.5368$$

$S_1 - S_3 = 7$, AND $S_2 - S_3 = 5$, SO EE AT $\alpha = 0.20$ (FAMILY) TO CONCLUDE THAT THE MEDIAN SAT-MATH SCORE IS LOWER AT UWF THAN AT EITHER UF OR FSU.

ALSO $MS_{0.20}$ = cMackSkil(0.20, 3, 2, 2, method = "Exact") = 3.5714

$$LSD_{MS(\Delta S)} = \sqrt{\frac{3}{6}(12+2)\left(\frac{25}{7}\right)} = 5$$

SO, SAME CONCLUSION WITH MACK-SKILLINGS EXACT PAIRWISE SIGNIFICANCE AS WITH TUKEY PAIRWISE SIGNIFICANCE.

R CODE FOR ANOVA PROCEDURES:

RBD: { package coin }

```
satm <- c(770, 670, ..., 500, 510)
schoolcat <- c(rep("UF",4), rep("FSU",4), rep("UWF",4))
majorcat <- c(rep("STA","BIO","BUS","PSY"), 3)
schoolfactor <- factor(schoolcat)
majorfactor <- factor(majorcat)
friedman_test(satm ~ schoolfactor | majorfactor
    {, distribution = approximate(nresample = 1000000)} )
```

2 WAY: { package NSM3 }

```
satm <- c(770, 670, ..., 500, 510)
school <- c(1,1,2,2,1,1,2,2,1,1,2,2)
major <- c(1,1,1,1,2,2,2,2,3,3,3,3)
pMackSkil(satm, school, major, method="Exact")
                               or "Asymptotic"
```

CRITICAL VALUES FOR MULTIPLE COMPARISONS:

```
TUKEY: cRangeNor(α, k)
NEMENYI: cWNMT(α, k, n)
MACK-SKILLINGS: cMackSkil(α, k, n, c, method="Exact")
```