Matt Isaac (A01515095)

Biostatistics

Homework 2

Due: 2/9/18

1. Create a visual summary that meaningfully compares the distribution of the scores in each group

\* Note: Histograms were not created for groups 1 and 3 because both groups only contained one level of severity (level 1).





Distribution for Group 2



Distribution for Group 4

1. Test whether the Group IV scores are much different than 3 using:
   1. A parametric method.

Method: One-sample t-test

Hypotheses: H0: µ = 3 vs HA: µ ≠ 3

Test statistic: 0.32

P-value: 0.7560

Conclusion: We fail to reject H0. These data show no evidence that the severity scores in group 4 are much different from 3 (“Mild – evident but not widespread”).

* 1. A nonparametric method.

Method: Wilcoxon Signed Rank test

Hypotheses: H0: µ = 3 vs HA: µ­­ ≠ 3

Test statistic: 1.5

P-value: 1.00

Conclusion: Fail to reject H0. We have no evidence in these data that the severity scores in group 4 are much different than 3.

1. Test whether the Group IV scores are much different than the Group III scores using:
   1. A parametric method.

Method: 2-sample t-test

Hypotheses: H0: µ3 = µ4 vs HA: µ­3­ ≠ µ4

\* Note: because the equality of variances test was significant (p < 0.0001), the Satterthwaite adjustment was used.

Test statistic: -7.35

P-value: < 0.0001

Conclusion: We reject H0. There is very strong evidence that the severity scores in group 3 are very different from the scores in group 4.

* 1. A nonparametric method.

Method: Wilcoxon Rank Sum (aka Mann-Whitney test)

Hypotheses: H0: µ3 = µ4 vs HA: µ­3­ ≠ µ4

Test statistic: 55.0

P-value: < 0.0001 (Two-Sized Pr > |Z|)

Conclusion: We would again reject H0. We have very strong evidence that the severity scores in group 3 are very different from the scores in group 4.

1. Test whether the scores are much different among groups using:
   1. A parametric method.

Method: One-way ANOVA

Hypotheses: H0: µi = µfor i = 1,2,3,4 vs HA: µ­i­ ≠ µfor some i

\* Note: The test for homogeneity of variance was significant (p = 0.0021), so Welch’s ANOVA was used.

Test statistic: 0.27

P-value: 0.6123

Conclusion: We fail to reject H0. We did not find evidence that the scores are significantly different across the four groups.

* 1. A nonparametric method.

Method: Kruskal-Wallis test

Hypotheses: H0: µ1 = µ2 = µ3 = µ4 vs HA: µi ≠ µj for some i,j

Test statistic: 34.0314

P-value: < 0.0001

Conclusion: Reject H0. We have found strong evidence that there is a significant difference between the severity scores in at least two of the groups.

1. Discuss which method (parametric or nonparametric) is most appropriate in exercises 2-4 and why.

Exercise 2: The non-parametric method (Wilcoxon Signed-Rank test) is more appropriate. Parametric assumptions not met.

Exercise 3: The non-parametric method (Wilcoxon Rank-Sum test) is more appropriate. Parametric assumptions not met.

Exercise 4: The non-parametric method (Kruskal-Wallis test) is more appropriate. Parametric assumptions not met.

The reason for the non-parametric methods being more appropriate is the same in all three cases. The data that we were examining cannot be assumed to follow a normal distribution. One could maybe argue that the data from group 4 follows a normal distribution. However, group 2 clearly does not follow a normal distribution as it shows signs of bi-modality. Likewise, groups 1 and 3 do not follow a normal distribution because all observations in these groups have the same value. Another reason to be skeptical of normality in these data is that the observations are on a 5-point ordinal scale.

Because these data do not meet the assumption of approximate normality, non-parametric methods are safer and more appropriate to use.

Appendix – SAS Code

/\* Biostatistics - Homework 2 \*/

data poultry;

input group $ severity @@;

cards;

g1 1 g1 1 g1 1 g1 1 g1 1 g1 1 g1 1 g1 1 g1 1 g1 1

g2 5 g2 3 g2 5 g2 3 g2 2 g2 3 g2 2 g2 4 g2 3

g3 1 g3 1 g3 1 g3 1 g3 1 g3 1 g3 1 g3 1 g3 1 g3 1

g4 3 g4 2 g4 3 g4 2 g4 3 g4 4 g4 3 g4 3 g4 2 g4 5 g4 4

;

run;

proc print data = poultry;

run;

/\* 1. Create a visual summary that meaningfully compares

the distribution of the scores in each group \*/

proc univariate data = poultry plots;

by group;

var severity;

run;

/\* 2(a) One-sample t-test \*/

proc ttest data = poultry HO = 3;

where group = 'g4';

var severity;

title1 'One-sample t-test';

run;

/\* 2(b) One-sample non-parametric test (Wilcoxon Signed Rank) \*/

proc univariate data = poultry mu0 = 3;

where group = 'g4';

var severity;

title1 'Wilcoxon Rank-Sum';

run;

/\* 3(a) Two-sample t-test \*/

proc ttest data = poultry;

where group = 'g3' | group = 'g4';

class group;

var severity;

title1 'Two-sample t-test';

run;

/\* 3(b) Wilcoxon Rank Sum test \*/

proc npar1way wilcoxon data = poultry;

where group = 'g3' | group = 'g4';

class group;

var severity;

title1 'Wilcoxon Rank-Sum Text';

run;

/\* 4(a) One-way ANOVA \*/

proc glm data = poultry;

class group;

model severity = group;

meas group / hovtest = BF welch;

title1 'One-way ANOVA';

run;

/\* 4(b) Kruskal-Wallis test \*/

proc npar1way wilcoxon data = poultry;

class group;

var severity;

title1 'Kruskal-Wallis test';

run;