Homework 3

Cody Frisby

10/26/2016

7.1

First, I display the ANOVA table from our model.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	760.4540	380.22698	6.780427	0.0079873
Residuals	15	841.1571	56.07714	NA	NA

And it can be seen that there is evidence to reject the null hypothesis that all group means are the same, $p = 0.0079873$.

This compares reasonably with the non parametric method used in the textbook for example 7.2, where $p = 0.0103279$ when the chi-squared approximation is used in R. Running `kruskal_test` from the `coin` package we can approximate the exact p value by using the argument `distribution = approximate(B = 100000)` or whatever value for B you would like for the number of simulations. When this is ran on the above data over and over we get p-values very close to 0.0047, the exact p-value mentioned in the text. **It's also worth mentioning that some the assumptions of the parametric model fit above are not valid**.

7.3

First, the test statistic is needed for the asymptotic Jonckheere-Terpstra test and we find that it is 32.5.

$$Z = \frac{U - E(U)}{\sqrt{Var(U)}} = \frac{32.5 - 17.5}{\sqrt{27.9166667}} = 2.8389613$$

The probability of Z under the standard normal distribution is 0.002263.

7.5

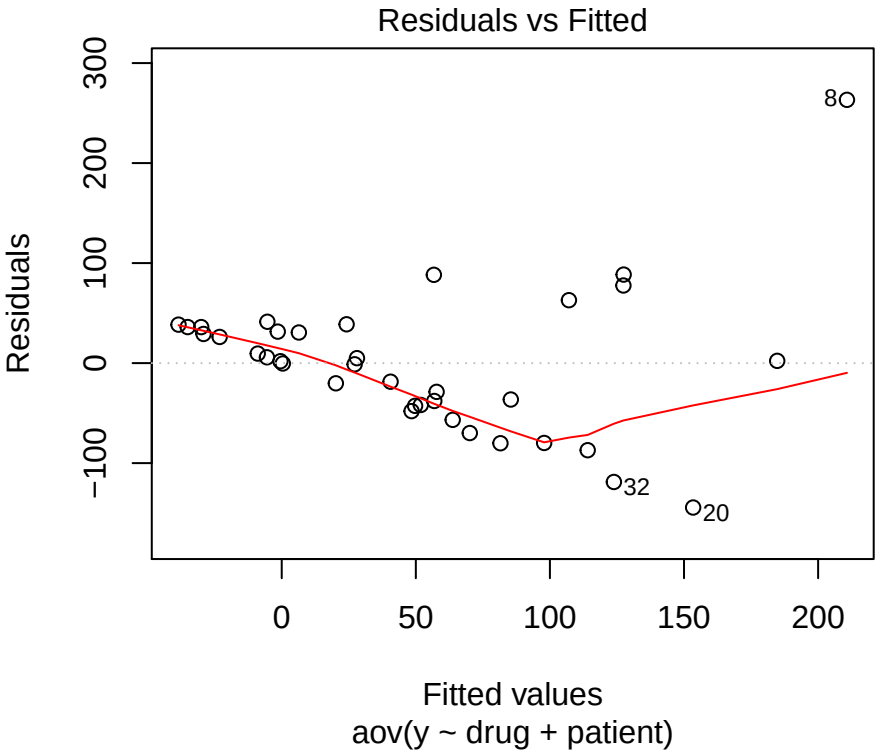
Running an Approximate (Monte Carlo) Friedman test we get $p = 0.00279$ and running a Friedman test using the chi-squared approximation we get $p = 0.0043202$. The ANOVA method displayed here

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
student	9	24.533333	2.7259259	5.183099	0.0014925
group	2	9.866667	4.9333333	9.380282	0.0016181
Residuals	18	9.466667	0.5259259	NA	NA

and the use of either test would lead to the same conclusion. We would conclude that there is evidence of a difference in recall rates between the three groups.

7.10

At $\alpha = 0.05$ we would reject the null hypothesis of no difference between groups with blocking by patient, $T = 8.0425532$ and $p = 0.0179301$ (the chi-squared approximation is used here). There appears to be a violation of some of the assumptions of the parametric model (see below).



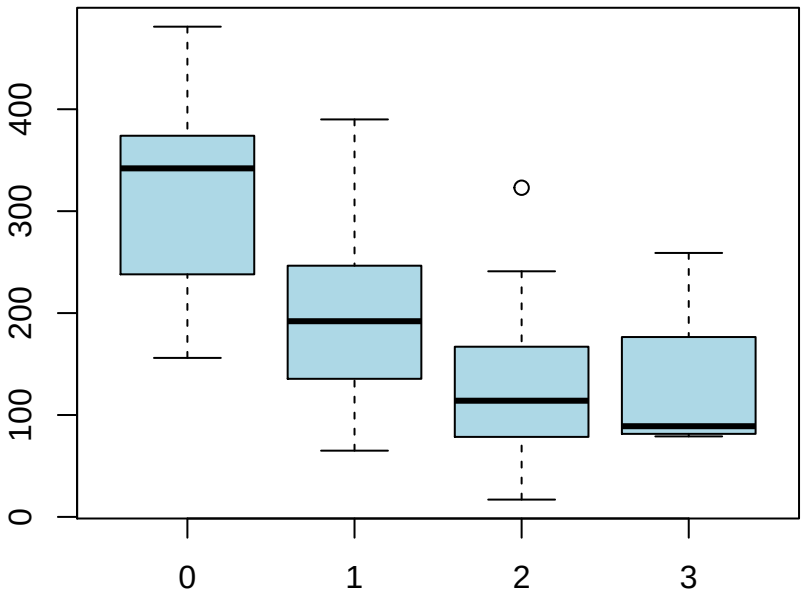
Here the assumption of equality of variance appears to be violated. The p-value for the parametric method is also below the same threshold for α.

7.12

The result from running The Friedman test is $T = 13$ and $p = 0.0015034$. The result from running a Page test is $L = 110$ and $p = 7.6803269 \times 10^{-5}$. The p-value is much smaller when using the Page method.

7.15

First, a look at a box plot of the observations.



And we can see that there appears to be a relationship between x and y from this plot.

Using the Jonckheere-Terpstra Test we conclude there is an association between spleen size and platelet counts, $JT = 115$ and $p = 2 \times 10^{-4}$.

10.2

Here are the probabilities for $n = 4$ for Spearman's Rank Correlation Coefficient.

	number	p-value
-1	1	0.0416667
-0.8	3	0.1250000
-0.6	1	0.0416667
-0.4	4	0.1666667
-0.2	2	0.0833333
0	2	0.0833333
0.2	2	0.0833333
0.4	4	0.1666667
0.6	1	0.0416667
0.8	3	0.1250000
1	1	0.0416667

10.6

Using

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

where $T = 18$ and $n = 7$ we get 0.6785714 for the value for Spearman's correlation coefficient. Testing the significance of this value using `cor.test` in R we get $p = 0.1095238$.

For Kendall's tau we can find the estimate using

$$t_k = \frac{n_c - n_d}{n(n-1)/2}$$

where $n_c = 15$ and $n_d = 6$, and the estimate on correlation is 0.4285714 and $p = 0.2388889$.

10.14

Using

$$\kappa = \frac{p_o - p_e}{1 - p_e} = \frac{0.76 - 0.6010667}{1 - 0.6010667} = 0.3983957$$

indicating moderate agreement between the two experts. Using the books asymptotic formula for the standard error of κ and then plugging in for $Z = \frac{\kappa}{se(\kappa)}$ we get $p = 2.5289311 \times 10^{-10}$, where $z = 6.2172988$.

12.8

	Introvert	Extrovert	total
Pass	14	34	48
Fail	31	41	72
total	45	75	120

The probability that $X \leq 14$ is 0.1767434 using fisher's exact test. There is not evidence to suggest a difference between introvert and extrovert and whether or not one is more or less inclined to pass or fail the test. Using the chi-squared test

$$p = 0.1779317$$

12.12

The test statistic is 5.3244343 with $df = 2$, $p = 0.0697933$. There is weak evidence of nationality having an influence on preference.

12.21

	True	False
True	523	230
False	345	554

We can see from this table that 345 is greater than 230 (the lower corner and the upper corner). Running the off-diagonal test described in the book we have

$$\chi^2 = \sum \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}} = 23$$

and the degrees of freedom are $n(n-1)/2 = 1$, where $n = 2$, and $p = 1.620014 \times 10^{-6}$. There is evidence that there is a change in attitude among the respondents.