Biostats VHM 801 Course
Fall 2014, Atlantic Veterinary College, PEI
Henrik Stryhn

Solution to Final Exam, December 2014

Question 1

The data originate from the article, Rendle, D. I., Cottle, H. J., Love, S. & Hughes, K. J. (2007): Comparative study of doramectin and fipronil in the treatment of equine chorioptic mange, *Veterinary Record* **161**, 335–338.

Subquestion a)

The study is an experiment because the treatments D and F were allocated to the cases (horses) using randomization. The two treatment groups constitute independent samples unless there are some relations between the cases; we do not have any information to indicate such relations. Two measurements (e.g. lesion scores at days 0 and day 28) on the same horse constitute paired samples. So the design contains both paired and independent samples. The main use of blinding would be if the veterinarians scoring the lesions did not know the horse's treatment group. That may be impossible to achieve in practice.

Subquestion b)

If we denote the lesion scores at day 14 in one of the groups by $X_1, \ldots, X_n$, the mean is computed as $\bar{X} = (X_1 + \ldots + X_n)/n$. The formula for the median is different for odd and even group sizes, but in both cases one would start by ordering the data from lowest to highest, say as $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$. For n odd, the median is the observation ranked $(n+1)/2$. For n even, the median is the average of observations ranked $(n/2)$ and $(n/2)+1$. For the two groups we get:

$$\begin{aligned} \text{Group D} &: \; n = 8;\; \bar{X} = 11/8 = 1.375;\; \text{median}(X) = (0+2)/2 = 1, \\ \text{Group F} &: \; n = 9;\; \bar{X} = 4/9 = 0.44;\; \text{median}(X) = 0. \end{aligned}$$

All of these values are seen to agree with those listed in the table (article).

Subquestion c)

The proportions of cases with no lesions in the two groups are:

$$\begin{aligned} \text{Group D} &: \; n = 8;\; \hat{p}_D = 2/8 = 0.25, \\ \text{Group F} &: \; n = 9;\; \hat{p}_F = 4/9 = 0.44. \end{aligned}$$

The two proportions appear to be quite different. Statistical inference would be based on assuming two independent samples of data with binomial distributions; we estimated the proportions above. A confidence interval for the difference between the proportions would have to be based on the "plus four" method, in which one "positive" and one "negative" is added to each sample, before a CI based on the normal approximation is computed. The adjusted proportions are: $\tilde{p}_D = 3/10 = 0.3$ and $\tilde{p}_F = 5/11 = 0.455$. Using $z^* = 1.96$, we calculate:

$$\begin{aligned} \text{95\% CI for } p_D - p_F &: \; \tilde{p}_D - \tilde{p}_F \pm z^* \sqrt{\tilde{p}_D(1-\tilde{p}_D)/(n_D+2) + \tilde{p}_F(1-\tilde{p}_F)/(n_F+2)} \\ &\quad 0.3 - 0.455 \pm 1.96 \sqrt{(0.3)(0.7)/10 + (0.455)(0.545)/11} = -0.155 \pm 0.409. \end{aligned}$$

It is seen (maybe not too surprisingly) that the confidence interval easily includes 0. An explanation of the "plus four" method and/or a reference to menu for two proportions with the expanded data was considered a sufficient answer for this part.

Subquestion d)

Both mean and median were higher in group D than F at day 0. Also, 3 scores in group D were higher than all scores in group F, hence it could be said that some of the lesions were higher in group D. The summary would seem to quite reasonable, but could be criticized by not taking into account whether the difference was statistically significant (it is not).

The lesion scores are semi-quantitative, as ordinal scores added up across the four limbs. For such an outcome a nonparametric method is valid because the ranks only use the ordering between the values, not the actual scale. In favour of a nonparametric method also counts that the distributions seem too discrete for normal distribution inference. Analysis by normal distribution methods must therefore be considered as a poor alternative. One could also think about using methods for count data, whereby one would count the number of cases with the different scores. Due to the small data set, these counts would however be very small which would cause problems for methods for counts. Also, looking only at the counts would ignore the ordering in the outcome. Therefore methods for counts (at least those covered in VHM 801) are not very attractive either. In summary, nonparametric methods is a reasonable choice.

Subquestion e)

The analysis in the Minitab listing corresponds to a Mann-Whitney-Wilcoxon test for comparison of two independent samples. The test is for H_0 : equal distributions in the two treatment groups, against H_a : some difference in distributions, and the test is formally non-significant ($P > 0.05$). With this test there would be no evidence against H_0. The lesion scores at days 0 and 28 are however not independent because they are for the same animals. Therefore the analysis is totally inappropriate. The reporting of the test outcome as non-significant would have been formally correct, but it can also be criticized for omitting the important information that the test was very close to being significant. Finally, if the focus is on a reduction in lesion scores from day 0 to day 28, a one-sided alternative hypothesis could have been used, and that would have produced statistical significance, with $P \approx 0.025$.

Subquestion f)

For two paired samples we should form the differences (say, day 0 minus day 28), and use one-sample methods to assess whether the center in the distribution of differences equals zero. As discussed above, nonparametric methods are preferable, and there is a choice between the sign test and Wilcoxon's signed rank test. The latter test assumes the distribution (of differences) to be symmetrical which is difficult to assess from the small and quite discrete dataset. Therefore the sign test would be the most obvious choice, and indeed the only nonparametric test that can be computed by hand. We assume that the differences $D_1, \ldots, D_8$ are independent and from the same (continuous) distribution.

Among the 8 differences, 2 are zero, 6 are positive (day 0 $>$ day 28) and none are negative. Testing equal medians in the two groups corresponds to testing $H_0 : p = 0.5$ in $B(6, p)$. We choose a one-sided alternative $H_a : p > 0.5$, because of the focus on a reduction in lesion scores. The P-value of our test is $P(X \geq 6) = P(X = 6) = 0.5^6 = 0.0156$ (where $X \sim B(10,0.5)$), or 0.016 from Table 1 of Stevens. The P-value is clearly below the significance level, so the data constitute evidence of a reduction in lesion scores in group D.