Biostats VHM 801 Course
Fall 2016, Atlantic Veterinary College, PEI
Henrik Stryhn

Brief solution to home assignment III

The solution is more detailed than required for a 100% mark, by including (very abbreviated) answers for multiple of the requested analyses for the last two questions.

1. Estimates and confidence intervals for pain scores

The pain score is a categorical response variable. We assume the $n = 174$ women to form a simple random sample from a suitable population. In other words, the 174 pain scores are independent and follow the same distribution (i.i.d.). As there are four possible outcomes (categories), these assumptions correspond to a multinomial setting (as discussed in Session 9). If we denote by N_0, N_1, N_2 and N_3 the counts of responses in each of the four categories, the statistical model for (N_0, N_1, N_2, N_3) is therefore a multinomial distribution $(n; p_0, p_1, p_2, p_3)$. The four probability parameters are estimated by the corresponding observed proportions in the dataset: $\hat{p}_i = N_i/n$. Confidence intervals are computed by the formulae for a proportion in a binomial distribution; effectively, when focusing on one of the categories, the other categories are all considered "negative" responses (not equal to the category under consideration).

$$\begin{aligned}
\hat{p}_0 &= 13/174 = 0.075, \quad && 95\%\ \text{CI}: \ (0.044, 0.125),\\
\hat{p}_1 &= 77/174 = 0.443, \quad && 95\%\ \text{CI}: \ (0.369, 0.516),\\
\hat{p}_2 &= 59/174 = 0.339, \quad && 95\%\ \text{CI}: \ (0.269, 0.409),\\
\hat{p}_3 &= 25/174 = 0.144, \quad && 95\%\ \text{CI}: \ (0.092, 0.196).
\end{aligned}$$

All confidence intervals except the first one were computed by the normal approximation formula, because the requirements of at least 15 positives and 15 negatives were met. The first interval was computed by the plus four method, because the sample size met the requirement of being at least 10.

2. Categorical predictor

For the solution, three variables will be considered: *month*, *nkids* and *prevsev*. Combined with the categorical response, the most natural analysis is by methods for two-way tables. It is also possible to use a non-parametric analysis for the ordered response categories, but due to the few categories and hence a very large number of ties the non-parametric procedure may have little power only. It does, however, directly utilize the ordering of the severity scores, contrary to the two-way table analysis.

Questions to be addressed in the two-way table analysis are: removal of observations found inconsistent in home assignment I, choice of model, pooling of categories to achieve expected counts large enough to meet the assumptions for the chi-square test, and interpretation of a significant test result by the conditional distributions. The table below describe one way these considerations could be implemented for the 3 analyses. Generally, for predictors determined prior to the pregnancy the most natural model seems the one for multiple populations, whereas for predictors determined during the pregnancy the model for a single population seems most natural. To keep the solution at a reasonable length, no estimates are presented for the conditional distributions considered after a significant chi-square test. These should be presented to quantify any description of the pattern(s) of the observed effects. Note also that without back pain, *month* and the questions *a1-a5* and *r1-r5* are pretty meaningless, so observations with $sev = 0$ should be excluded from such analyses.

Issue	Predictor Number of kids	Previous severity score	Month where pain started
data modifications	none	*prevsev* $= 0, 1$ dropped	*sev* $= 0$ dropped
statistical model	model I (multiple populations)	model I (multiple populations)	model II (single population)
pooling of categories	*nkids* $\geq 1 \rightarrow 1$ category	*prevsev* $> 2 \rightarrow 1$ category sev $\leq 1 \rightarrow 1$ category	*month* $\rightarrow 3$ categories (0-2, 3-5, 6-9) months
chi-square analysis	$X^2 = 12.98$, df $= 3$ $P = 0.005$	$X^2 = 0.93$, df $= 2$ $P = 0.63$	$X^2 = 30.96$, df $= 4$ $P < 0.0005$
conclusion	strong significance: higher *sev* scores for *nkids* ≥ 1	no significance: same *sev* distributions regardless of previous pain scores	strong significance: low pain scores more likely to occur in first trimester

3. Continuous predictor

For the solution, six variables will be considered: *age*, *height*, and the four variables for weight, including weight gain (*wgain*). As indicated in the question, the recommended analysis is a comparison of multiple (four) populations defined by the severity scores with respect to each of these variables. Thus, for all analyses the statistical model is that of four independent samples from respective distributions. If normal distributions with equal variances can be assumed (i.e., $X_{ij} \sim \mathrm{N}(\mu_i, \sigma)$) a one-way ANOVA results; otherwise, a non-parametric Kruskal-Wallis analysis is the obvious choice. In addition, the removal of observations found inconsistent or outlying in the first home assignment should be addressed. The table below summarizes these considerations and gives results for the six variables. The P-value for the normality test is based on the A-D test for residuals; alternatively, each group could be tested separately. The s-ratio is the ratio between the largest and smallest standard deviation among the four groups. Finally, NS means non-significant at $P < 0.05$.

Issue	Predictor *age*	*height*	*weight0*	*weight1*	*weightb*	*wgain*
data modifications	none	none	none	none	2 outliers	1 outlier
s-ratio	1.83	1.16	1.39	1.21	2.03	1.41
P normality	< 0.005	0.009	< 0.005	0.40	0.017	0.19
ANOVA F-test P-value	2.02 0.11	0.99 0.40	0.13 0.94	0.87 0.46	1.07 0.36	2.78 0.043
Kruskal-W P-value	0.071	0.48	0.93	(0.52)	0.12	(0.13)
conclusion	NS	NS	NS	NS	NS	weak sign.

The results reported are for the data with the two large birth weights and the one very large weight gain excluded because all of these were thought to be likely errors or unrepresentative of the population. Note that for a rank-based procedure one could have kept the outliers in if they were not thought to be errors. Only one of the analyses (for *wgain*) shows a significant comparison between severity groups, whereas some of the other results could be termed as "close to significant" and interpreted as a possible indication of an effect. For weight gain, the significance is weak and occurs in the ANOVA only. The group means (medians) show that the weight gains are largest in the group with *sev* $= 3$, with the other groups being quite similar.