Biostats VHM 801 Course
Fall 2020, Atlantic Veterinary College, PEI
Henrik Stryhn

Solution to home assignment 3

Some parts are more detailed than expected to obtain a 100% mark.

Data and Notation

The dogs (offspring tested) are cross-classified according to 3 variables, but the analyses requested correspond to different one- and two-way tables. Rather than introducing a consistent notation throughout involving 3 indices, we use different notations for the different analyses. The HD-status of the dogs is clearly a response variable, and the year is clearly an explanatory variable. The sire is most naturally considered an explanatory variable. If we would take sire as a response variable we would implicitly assume that the number of offspring for the sires are representative of the population (or population totals). This may or may not be the case, we have no information about that. Nor does it seem to be of any particular interest to compare the *numbers* of offspring. We therefore consider those numbers as fixed and consequently sire as an explanatory variable.

1. Estimates and confidence intervals for proportions

We are asked to compute proportions and 99% confidence intervals for the population proportions of offspring predisposed to HD based on the data for year 1. For each of the sires we assume the number of HD-positive offspring dogs to follow a binomial distribution: $X \sim \text{Bin}(n, p)$, where n is the total number of dogs tested (in year 1) and p is the probability of being diagnosed as predisposed for HD. In this model, we ignore any dependence/heterogeneity between the offspring dogs due to their maternal genes (because we don't have that information). Estimates and confidence intervals are listed in the table below. For reference purposes all three types of confidence intervals are included, although only interval is needed. The choice between intervals is discussed below.

			99% confidence interval		
Sire	total n	estimate $\hat{p}$	"classical"	"plus four"	"exact binomial"
Caesar	193	0.435	(0.343, 0.527)	(0.346, 0.528)	(0.343, 0.530)
Quatro	36	0.472	(0.258, 0.687)	(0.272, 0.678)	(0.260, 0.692)
Obelix	32	0.250	(0.053, 0.447)	(0.085, 0.470)	(0.087, 0.489)

The values in the table have been computed using software but the classical (i.e., normal approximation) and plus four intervals could also be obtained from the formulae, using $z^* = 2.576$:

$$\text{classical}: \quad \hat{p} \pm 2.576\sqrt{\hat{p}(1-\hat{p})/n}, \qquad \hat{p} = X/n,$$
$$\text{plus four}: \quad \tilde{p} \pm 2.576\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}, \qquad \tilde{p} = (X+2)/(n+4).$$

The data for Obelix do not meet the requirement that both the number of cases and non-cases exceed 15, and the data for Quatro just barely do. Therefore it seems most natural to use the plus four interval for these two sires. For Caesar, the classical interval is perfectly fine (and indeed all three intervals are very close). The "exact binomial" interval is largest in all cases, as expected from its generally too large coverage (i.e., above 99%). The interval for Caesar is much narrower than the two other intervals, but still quite wide. A fairly narrow 99% interval for a proportion requires a very large sample size.

2. Comparison with population proportion

The task here is to compare the proportion of HD-positive offspring for each sire with the given, recent reference value of 0.37. As the second year's data is stated to be better comparable with the recent value, we concentrate on data for year 2. For each of the sires we assume a similar binomial model as in the first question: $X \sim \text{Bin}(n, p)$, where n is the total number of dogs tested (in year 2) and p is the probability of testing positive. Based on the model we test the hypothesis $\text{H}_0\colon p = 0.37$ against the two-sided alternative $\text{H}_a : p \neq 0.37$. Estimates, test statistics and P-values are given in the table below. Alternatively, confidence intervals could be computed and checked for whether 0.37 was inside the intervals.

Sire	total n	estimate $\hat{p}$	SE($\hat{p}$)	z-test	P for z	exact P
Caesar	94	0.553	0.051	3.68	<0.001	<0.001
Quatro	60	0.400	0.063	0.48	0.63	0.69
Obelix	72	0.278	0.053	-1.62	0.105	0.113

The values in the table have been computed using software but could also be obtained from the formulas,

$$\hat{p} = X/n, \quad \text{SE}(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/n}, \quad z = (\hat{p} - 0.37)/\text{SE}(\hat{p}).$$

The guidelines for use of the normal approximation (i.e., that $(n \cdot 0.37)$ and $(n \cdot 0.63)$ exceed 10) are satisfied for all datasets, but with access to software we generally prefer the exact test based on the binomial distribution ("why use an approximation, when we can get the exact P-value?").

In conclusion, the proportion of HD-positive offspring for Quatro is quite close to 0.37 and there is no evidence at all of a difference. For Obelix, the observed proportion is somewhat lower but not enough to claim evidence (at the 5% significance level). For Caesar, the proportion is considerably higher (0.55) and clearly significant so that we can say with great confidence the proportion of HD-positive offspring for Caesar exceeds 0.37. The biological significance of this depends on the reference population of the reference value (0.37); if we assume that it refers to a similar population of offspring dogs in a recent breeding program, we can certainly say that Caesar seems to be an above-average sire (with respect to HD predisposition, hence not a good breeding option) measured by recent standards.

3. Comparison of sires

Looking at year 1 first, the data constitute a 3×2 table with rows $\sim$ sires as an explanatory variable and HD-status $\sim$ columns as a response variable. Therefore, the model is three independent binomials (one for each row: $X_i \sim \text{Bin}(n_i, p_i)$, $i = 1, 2, 3$), or in IPS terminology a model for comparing independent populations (model I, although the data are displayed with rows and columns reversed compared to the examples in Lecture 8). This model is essentially the same as in 1), except that we consider all three sires together. We estimate the p_i's by the respective sample proportions X_i/n_i. Our interest is in the null hypothesis $\text{H}_0 : p_1 = p_2 = p_3$, and the (two-sided) alternative hypothesis H_a is that some differences between p_i's exist. We test the null hypothesis by a X^2-test; the table below gives expected values under the null hypothesis, as well as the X^2-statistic and P-value. The X^2-statistic has $(3-1) \cdot (2-1) = 2$ degrees of freedom, so the critical value is $\chi^2_{.95}(2) = 5.99$. For year 2, the statistical model and analysis are entirely similar, and results are shown in the same table below. Note that the guidelines for use of the X^2-statistic are easily met, with all expected counts above 10.

Count (Expected count)	Year 1 HD-negative	Year 1 HD-positive	Year 2 HD-negative	Year 2 HD-positive
Caesar	109 (112.4)	84 (80.6)	42 (54.1)	52 (39.9)
Quatro	19 (21.0)	17 (15.0)	36 (34.5)	24 (25.5)
Obelix	24 (18.6)	8 (13.4)	52 (41.4)	20 (30.6)
X^2 test (P)	4.38 (0.112)		12.9 (0.002)	

For year 1, there is no evidence (at the 5% significance level) to say that the three sires have different offspring HD predisposition rates. A P-value of 0.11 is however close enough (to 0.05) to be of some interest, and the table shows the largest deviations between observed and expected values for Obelix. Obelix is also seen (in the table of 1)) to have a lower rate than the other two sires. So we have some indication in that direction.

For year 2, there is strong evidence against same rates for the three sires, and the deviations are considerable for both Caesar and Obelix. The rates given in 1) are quite different between the three sires, and it is tempting to conclude that Obelix does better than Quatro who again does better than Caesar. However, the only pairwise comparison significant at the (unadjusted) 5% level is the one between Caesar and Obelix. (Pairwise comparisons can be done by restricting the chi-square analysis to the two rows under consideration.)

In conclusion, the sire Obelix had the lowest HD-rate among offspring in both years but only in year 2 there was a significant difference between the sires, and only the sires Obelix and Caesar could be declared different.

4. Comparison of years

The potential risk in combining the sires for the comparison of years is that the pattern seen in the combined (marginal) data does not reflect the patterns for the separate data. In an extreme form, this could lead to Simpson's paradox. In less extreme cases "opposite" patterns for two sires could neutralize each other when combining the data to indicate no difference between years. The general rule is to be very careful when combining values across one or several variables, and in particular always check that the separate and marginal patterns agree reasonably well. Statistical methods to assess whether combining across the levels of a variable is allowed exist but are beyond VHM 801 (log-linear models or logistic regression). If the patterns are similar for marginal and separate data, the combined analysis has the stronger power.

Considering our results in previous questions, it seems most natural to look at the sires individually, because we established differences between them in year 2 but not in year 1. This suggests that the years might not compare equally for the three sires.

For completeness, we show results both for the separate and marginal analyses. In both cases the data are of the form of a two-way table with (say) year as the row variable and HD-status as the column variable. As previously discussed, year is an explanatory and HD-status a response variable. The statistical model is therefore model I: two independent binomial distributions, one in each row — the same setup as for 2) except that we have only two groups (rows). We test the null hypothesis $H_0: p_1 = p_2$ against the alternative $H_a: p_1 \neq p_2$ (in absence of any information about how the panel changes between the years could affect the HD categorization). Again, we use a X^2-statistic to test the null hypothesis (equivalently, the classical z-statistic could be used, where $z^2 = X^2$). The table also gives the estimate and 95% confidence interval (plus four) for $p_1 - p_2$ corresponding to the analysis as two independent proportions. Again, the guidelines for the use of the X^2-statistic are easily met.

Data	$\hat{p}_1 - \hat{p}_2$ (95% CI)	z-test	X^2-test	P-value
Caesar	-0.118 (-0.238,0.005)	-1.88	3.53	0.060
Quatro	0.072 (-0.130,0.271)	0.69	0.48	0.49
Obelix	-0.028 (-0.199,0.161)	-0.29	0.09	0.77
combined	-0.007 (-0.097,0.082)	-0.16	0.03	0.87

The table shows that all analyses (with a slight reservation for Caesar) show no evidence against the hypothesis of equal proportions of HD-positive offspring in the two years. For Caesar, the proportion is higher in the second year, and the test is close to (borderline) significant. For the other two sires there is no indication of a difference whatsoever, and the same for the combined analysis. It seems fair to report this as a non-significant difference between years but also to mention that the data for Caesar shows some difference. It is then left up to the investigator to decide whether the result for Caesar is worth reporting or studying in further detail.

Appendix: Minitab commands and listing

Without further comments we give Minitab commands for the last two questions (the calculations could have been done in many other ways, so this is just for demonstration purposes). Before running these commands, the counts should be typed into the first row of the worksheet in the row-wise order shown in the table, one row at a time.

```
Name c1 'n'
Name c2 'sire'
TSet 'sire'
 1(  "Caesar" "Quatro" "Satan" )4
 End.
Name c3 'year'
Set 'year'
 3( 1 : 2 / 1 )2
 End.
Name c4 'HD'
Set 'HD'
 6( 0 : 1 / 1 )1
 End.
Print 'n' 'sire' 'year' 'HD'.

XTabs 'sire' 'HD' 'year';
 Layout 1 1;
 Frequencies 'n';
 Counts;
 ChiSquare;
 Expected;
 DMissing 'sire' 'HD' 'year'.
```

Data

Row	n	sire	year	HD
1	109	Caesar	1	0
2	84	Caesar	1	1
3	42	Caesar	2	0
4	52	Caesar	2	1
5	19	Quatro	1	0
6	17	Quatro	1	1
7	36	Quatro	2	0
8	24	Quatro	2	1
9	24	Satan	1	0
10	8	Satan	1	1
11	52	Satan	2	0
12	20	Satan	2	1

Results for year = 1

Rows: sire   Columns: HD

	0	1	All
Caesar	109	84	193
	112.40	80.60	
Quatro	19	17	36
	20.97	15.03	
Satan	24	8	32
	18.64	13.36	
All	152	109	261

Cell Contents
Count
Expected count

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	4.384	2	0.112
Likelihood Ratio	4.614	2	0.100

Results for year = 2

Rows: sire   Columns: HD

	0	1	All
Caesar	42	52	94
	54.07	39.93	
Quatro	36	24	60
	34.51	25.49	
Satan	52	20	72
	41.42	30.58	
All	130	96	226

Cell Contents
Count
Expected count

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	12.862	2	0.002
Likelihood Ratio	13.080	2	0.001

```
XTabs 'year' 'HD' 'sire';
 Layout 1 1;
 Frequencies 'n';
 Counts;
 ChiSquare;
 Expected;
 DMissing 'year' 'HD' 'sire'.
```

Results for sire = Caesar

Rows: year   Columns: HD

	0	1	All
1	109	84	193
	101.54	91.46	
2	42	52	94
	49.46	44.54	
All	151	136	287

Cell Contents
Count
Expected count

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	3.528	1	0.060
Likelihood Ratio	3.529	1	0.060

Results for sire = Quatro

Rows: year   Columns: HD

	0	1	All
1	19	17	36
	20.63	15.38	
2	36	24	60
	34.38	25.63	
All	55	41	96

Cell Contents
Count
Expected count

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	0.480	1	0.489
Likelihood Ratio	0.478	1	0.489

Results for sire = Satan

Rows: year   Columns: HD

	0	1	All
1	24	8	32
	23.38	8.62	
2	52	20	72
	52.62	19.38	
All	76	28	104

Cell Contents
Count
Expected count

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	0.087	1	0.768
Likelihood Ratio	0.088	1	0.767

```
XTabs 'year' 'HD';
 Layout 1 1;
 Frequencies 'n';
 Counts;
 ChiSquare;
 Expected;
 DMissing 'year' 'HD'.
```

Rows: year   Columns: HD

	0	1	All
1	152	109	261
	151.1	109.9	
2	130	96	226
	130.9	95.1	
All	282	205	487

Cell Contents
Count
Expected count

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	0.025	1	0.873
Likelihood Ratio	0.025	1	0.873