Solution to home assignment III

Some parts are more detailed than expected to obtain a 100% mark. Because Fisher's exact test is not a core part of the course, its use is not strictly required but at the very least an assessment should be given of the adequacy of the Pearson chi-square test and the validity of the conclusions obtained from it. Generally speaking, if more than one test is reported for the same hypothesis, the rationale for doing that and the choice between them should be addressed in the text.

Data and Notation

The (castrated) cats are cross-classified according to 3 variables, but the analyses requested correspond to different two-way tables. In order to have a single, consistent notation we introduce 3 indices (alternatively, one may use different notations for the different analyses):

$$\begin{aligned} N_{ufe} &= \text{ number of cats with urinary disease level } u, \text{ dry feed level } f \text{ and exercise level } e, \text{ where} \\ u &= 0, 1 \sim \text{ (non-diseased, diseased)}, \\ f &= 0, 1 \sim \text{ (little, much)}, \\ e &= 0, 1 \sim \text{ (low, high)}. \end{aligned}$$

Among the 3 variables, the two questionnaire responses are clearly response (outcome) variables, but the grouping of cats into diseased or non-diseased in the urinary system may seem less clear. It should, however, almost certainly be considered as an explanatory variable, corresponding to targeting the questionnaires to cat-owners when it was established that their cat would fall into one of the disease categories. It may be helpful to ask the question whether the number of diseased and non-diseased cats are (partly) controlled by the experimenter or represent the observed proportions of diseased and non-diseased cats in a sample from some population. The answer should be obvious: it is not realistic to assume that these are true proportions of diseased and non-diseased cats. The description of the study spells out the fact that the data are two separate samples pretty clearly. In epidemiologic terms, the study is a *case-control design*.

1) Separate analyses for diseased and non-diseased cats

Looking at the diseased cats first, the data constitute a 2×2 table with 2 response variables (feed and exercise), so that our model is a multinomial distribution on the 4 cells of the table (a model II, or in IPS terminology: a model for examining independence):

$$(N_{100}, N_{101}, N_{110}, N_{111}) \sim \text{ multinomial } (50; p_{100}, p_{101}, p_{110}, p_{111}).$$

The parameters p_{1fe} are estimated by the observed proportions of cats in the respective categories, e.g. $\hat{p}_{101} = 3/50 = 0.06$. The hypothesis of interest is H_0: independence between feed and exercise responses, and 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 $(2-1)\cdot(2-1) = 1$ degrees of freedom, so the critical value is $\chi^2_{.95}(1) = 3.84$. For the non-diseased cats, the statistical model and analysis are entirely similar, and results are shown in the same table below.

	Urinary disease		Non-diseased	
Count (exp. value)	Exercise: low	Exercise: high	Exercise: low	Exercise: high
Dry feed: little	5 (3.04)	3 (4.96)	12 (12.04)	5 (4.96)
Dry feed: much	14 (15.96)	28 (26.04)	5 (4.96)	2 (2.04)
X^2 test (P)	2.43 (0.12)		0.002 (0.97)	
Fisher's exact P	0.23		1.00	

For the diseased cats, $X^2 = 2.43$, and with two cells with expected values below 5, there should be concern about the accuracy of the P-value. However, Fisher's exact test (with its two-sided P-value obtained from Minitab) shows that the data are quite far from significant. For the non-diseased cats, the X^2-value is very small, and despite the low expected values in 3 out of 4 table cells, it seems clear that the test is non-significant (and Fisher's P-value confirms that). We conclude that in none of the groups of cats there is any evidence of an association between the feed and exercise groupings.

The two responses can therefore be summarized separately by estimates and associated 95% confidence intervals (CIs). For both responses, the marginal distributions are used, and both of these correspond to a single binomial distribution $B(50, p)$ (urinary disease group) or $\mathrm{B}(24, p)$ (non-diseased group), where p is the probability for one of the categories in question (there is little gain in summarizing the information obtained for both p and $1-p$). Because in three out of the four situations, the conditions for the classical CI are not met, all CIs below are by the "plus four" method (exact CIs would also be perfectly ok, see the footnotes of the table). The CIs below are given with 3 decimals although normally two decimals would be adequate for the precision we have here.

Parameter: Estimate (CI)	Urinary disease	Non-diseased
prob. of high exercise level	$31/50 = 0.62\ (0.481, 0.741)^a$	$7/24 = 0.29\ (0.148, 0.494)^c$
prob. of much dry feed	$42/50 = 0.84\ (0.711, 0.918)^b$	$7/24 = 0.29\ (0.148, 0.494)^c$

[a] classical: (0.485,0.755); Adjusted-Blaker: (0.480,0.746); Clopper-Pearson: (0.472,0.573)

[b] Adjusted Blaker: (0.714,0.927); Clopper-Pearson: (0.709,0.928)

[c] Adjusted Blaker: (0.133,0.500); Clopper-Pearson: (0.126,0.511)

The two groups of cats seem to differ clearly with respect to both the amount of exercise and dry feed consumption: both are larger in the diseased group.

2) Combined questionnaire answers

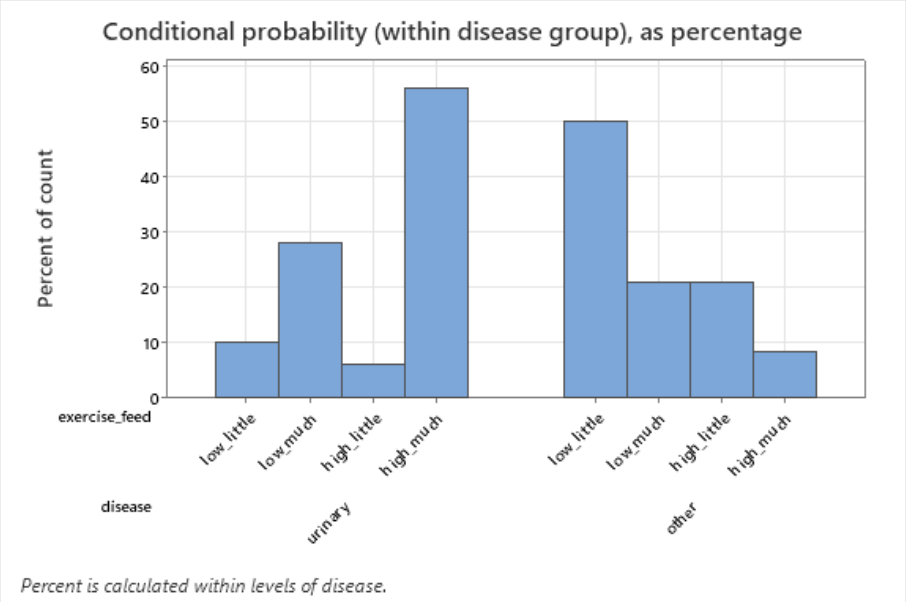
When combining the two questions, each cat is grouped into one of 4 possible categories by its exercise level and dry feed level. The combined variable is a response variable as well (because both of its components are). For the comparison between disease groups we have a 2×4 or 4×2 table (depending on whether we put the disease groups as rows or columns) with one explanatory and one response variable. The statistical model therefore assumes two independent multinomial distributions on 4 cells each, i.e. the two multinomial distributions from **1)**, which is a model I situation (or in IPS terminology: a model for comparing two populations). The null hypothesis of interest is that the two multinomial distributions have the same probabilities:

$$\mathrm{H}_0: \ p_{100} = p_{000},\ p_{101} = p_{001},\ p_{110} = p_{010},\ p_{111} = p_{011}.$$

We test this hypothesis by a X^2 test for the two-way table, and summarize the results again in a table of counts and expected values:

Count (exp. value)	Exercise: low		Exercise: high	
	Dry feed: little	Dry feed: much	Dry feed: little	Dry feed: much
urinary diseased cats	5 (11.49)	14 (12.84)	3 (5.41)	28 (20.27)
non-diseased cats	12 (5.51)	5 (6.16)	5 (2.59)	2 (9.73)

The X^2 value is 24.0 with $(2-1)\cdot(4-1) = 3$ degrees of freedom, and the P-value may be computed as $\mathrm{P}(\chi^2(3) \geq 24.0) = 0.000025$ (using Minitab). There is one cell with an expected value below 5, but that does not violate the guidelines for use of the $\chi^2(3)$-distribution as our reference distribution. In view of the clear significance, there is no doubt that the null hypothesis should be rejected. thus, there *are* some differences in responses between the two groups of cats. When comparing the observed and expected values, the largest differences (and contributions to the X^2-statistic, not shown) are seen in the (low,little) and (high,much) groups, indicating once more that diseased cats are higher with respect to both exercise and dry feed consumption. The relevant probabilities to estimate following the test would be the conditional distributions for the four exercise × feed categories within each disease group. We show these estimates in the figure below, but abstain from further analysis at this point (e.g. confidence intervals), as the data analysis progresses in the next question.



3) Separate analyses of exercise and feed

As suggested by the wording for the third question, it is not automatically allowed (meaningful) to analyze the response variables separately with respect to their distribution among diseased and non-diseased cats. The reason is that the two variables may show association/interaction, and in such case one variable (e.g. exercise) becomes a lurking variable for the analysis of the other (feed) vs. the disease groups. In a worst case, misinterpretations like Simpson's paradox may arise. For these data, however, we found in **1)** that the two responses showed no significant interaction when examined separately for the two disease groups. According to the theory of log-linear models, this is exactly the condition needed to allow separate analysis of the two responses. In particular, one may add the X^2-values for the disease groups to obtain a combined test for no interaction (adding also the degrees of freedom). In our case, $X^2 = 2.43 + 0.00 = 2.43$, which is far from significant in a $\chi^2(2)$-distribution.

We consider first the dry feed consumption. By aggregating (summing up) the data across the exercise levels, we obtain a 2×2 cross-classification of cats with respect to dry feed and disease groups. The statistical model with one response (feed) and one explanatory variable (cat groups) is two independent binomial distributions, or model I:

$$N_{11\cdot} \sim \mathrm{B}(50, p_{11\cdot}) \quad \text{and} \quad N_{01\cdot} \sim \mathrm{B}(24, p_{01\cdot}),$$

where $N_{11\cdot}$ is the total number of cats in the diseased group fed with much dry feed, and $N_{01\cdot}$ is the same number for the non-diseased group. The null hypothesis is H_0: $p_{11\cdot} = p_{01\cdot}$, i.e. same probability of high dry feed consumption in the two disease groups, against a two-sided alternative H_0: $p_{11\cdot} \neq p_{01\cdot}$ (in absence of knowledge to suggest otherwise). We analyze the 2×2-table by a X^2-statistic with $(2-1)\cdot(2-1) = 1$ degrees of freedom, and give also the observed and expected values in the table below. Also shown are the corresponding values for the exercise responses, where the binomial counts are $N_{1\cdot1}$ and $N_{0\cdot1}$ for the two disease groups. (Note: One may also use the methods for comparison of two independent proportions; the z- and X^2-statistics are equivalent.)

	Dry feed consumption		Exercise		
Count (exp. value)	little	much	low	high	Total
Urinary disease	8 (16.89)	42 (33.11)	19 (24.32)	31 (25.68)	50
No urinary disease	17 (8.11)	7 (15.89)	17 (11.68)	7 (12.32)	24
X^2 test (P)	21.8 (0.000003)		7.00 (0.008)		

Both X^2-tests are clearly significant (with all expected counts > 5), and the one for dry feed mostly so with a very small P-value. We conclude that there *are* differences between the diseased and non-diseased cats with respect to both questions. Estimates and confidence intervals were already given in the table under **1)**. Cats with urinary disease eat more dry feed (a possible causal relationship?) and get more exercise (which may seem rather strange).

Finally, a remark on a *conceptually wrong* but perhaps tempting model and analysis for **3)**, namely to compare the proportions of diseased cats between feed groups (or exercise groups). Despite giving the same X^2-statistic, it is a wrong analysis because whether the cats were diseased or not is not a response variable, therefore the binomial p's would not correspond to population parameters. To illustrate, if the experimenter (researcher) had chosen to take roughly the same number of cats in the two groups instead of a ratio of 2:1, the proportions of diseased and non-diseased cats in the data would have been completely different. In a case-control design like this one and without additional information about the population prevalences (of either exposure or disease), it is *impossible* to estimate prevalences/incidences of disease or differences of risk or relative risk. But are we then not able to say anything about the impact of exposure on disease from such a design? Yes, surprisingly! — it turns out that under some further assumptions the null hypothesis of equal exposure distribution among the cases and controls is equivalent to an equal distribution of cases and controls in the two exposure groups (that is, our H_0 is indeed the interesting one!), and we can estimate the impact of exposure on disease from the case-control data by another statistic called the odds-ratio. The discussion hereof is beyond the biostats VHM 801 course (and should be in the epidemiology course).

Specific biological interpretations of the findings for these cats are beyond what would be expected of students in VHM 801.