Biostats VHM 801 Course
Fall 2023, Atlantic Veterinary College, PEI
Henrik Stryhn

Solution to assignment III

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

Data

The persons responding to the survey may be considered as a random sample from the population of community-dwelling seniors in the Atlantic region. A possible response bias might be a concern with the real data but we don't have information to discuss that. All 3 variables (*support*, *gender*, *age*) were obtained from the survey, so they may all be considered as response variables. The interest is however clearly on the support responses, and when we compare response distributions between gender or age groups we can also consider these variables as explanatory. Analysis of the real data has been published in McInnis-Perry, Weeks and Stryhn (2013), *Canadian Journal of Nursing Research* **45**, 50−68.

a) Support responses cross-classified by gender

The support responses cross-classified by gender form a two-way table of counts. Denote by N_{ij} the number of responses for gender i in support category j, where $i = 1, 2 \sim$ Gender (0 = male, 1 = female) and $j = 1, 2, 3, 4, 5 \sim$ Support (categories 0−4). The counts are tabulated below together with conditional distributions within each gender and the marginal distribution across gender categories. Note that we for convenience display the table with 2 rows and 5 columns, thus contrasting the standard layout in the textbooks where the (primary) response variable corresponds to the rows.

Count N_{ij} (Proportion $\hat{p}_{ij}$)		Support category					Total
		0 (None)	1 (Little)	2 (Some)	3 (Most)	4 (All)	
Gender	0 (Male)	58 (.143)	71 (.175)	75 (.185)	113 (.279)	88 (.217)	405
	1 (Female)	34 (.061)	121 (.218)	107 (.193)	166 (.299)	127 (.229)	555
Marginal		92 (.096)	192 (.200)	182 (.190)	279 (.291)	215 (.224)	960

The support distributions (the estimates $\hat{p}_{ij}$) mainly differ between men and women in their proportions for the no support category: 14.3% for men and 6.1% for women. The proportions in the other categories are all larger for women than men, mostly in the little support category (17.5% vs. 21.8%) and only little (at most 2.0%) in the other categories. The marginal distribution across gender is intermediate between the distributions for men and women. In a statistical model assuming two independent multinomial distributions (for one response and one explanatory variable, as discussed above), we can use test the hypothesis of equal distributions in the two rows ($H_0 : p_{1j} = p_{2j}$ for all j) by a Pearson chi-square test (X^2). The next table gives all the intermediate calculations (expected counts and chi-square contributions for each cell).

Expected count e_{ij} (X^2-contribution)		Support category				
		0 (None)	1 (Little)	2 (Some)	3 (Most)	4 (All)
Gender	0 (Male)	38.8 (9.49)	81.0 (1.23)	76.8 (.04)	117.7 (.18)	90.7 (.08)
	1 (Female)	53.2 (6.92)	111.0 (.90)	105.2 (.03)	161.3 (.14)	124.3 (.06)
		$X^2 = 19.08$, df $= (2-1) \cdot (5-1) = 4$, P-value $= 0.001$				

The test statistic is highly significant, thus giving us strong evidence to reject the hypothesis of equal support distributions in men and women. All expected counts are far above 5, so there are no concerns with the $\chi^2(4)$ reference distribution. By far the largest contributions to X^2 come from the no support category, where we already noted the largest difference between men and women. Our analysis therefore supports our previous description, and we may conclude the main difference between men and women to be that a larger proportion of men have no support (specifically, nobody "to do things with to help you get your mind off things").

b) Nonparametric comparison support distributions across gender

The previous analysis did not take the ordering of support response categories ($0 < 1 < \ldots < 4$) into account. A simple approach to involve the ordering is by determining ranks from considering the response categories considered as quantitative. When using ranks, the actual values (or codes, $0-4$) for the categories are irrelevant, only their ordering matter. One might also consider to use normal distribution models, but normal distributions are clearly inappropriate for the discrete values. More seriously perhaps, the results would depend on the actual values (codes), which are more or less arbitrary $(0, \ldots, 4)$. In many applications, such a dependence on scale is not desirable.

None of these concerns exist with a rank-based nonparametric approach; another example hereof is in IPS Exercise 15.39. The appropriate nonparametric approach for two independent samples (men and women) is the Wilcoxon-Mann-Whitney rank sum test. The W-M-W test can be carried out under the additional assumption that the two distributions have the same shape. Because the actual values have no intrinsic meaning, such an assumption is difficult to interpret, and our table of counts in **a)** indicated that the distributions were not moved up or down across the scale relative to each other. We therefore test the hypothesis that the two distributions are equal ($\mathrm{H}_0 : \mathrm{Dist}P_1 = \mathrm{Dist}P_2$) versus that they are different ($\mathrm{H}_a : \mathrm{Dist}P_1 \neq \mathrm{Dist}P_2$). As discussed in PSLS Chapter 27 (page $10-11$) and in more detail in IPS Chapter 15 (page 10), the alternative hypothesis may also be phrased that "one distribution is systematically larger than the other". This means that the right tail probabilities of one distribution are always larger than those of the other distribution, thus one distribution is essentially "further to the right" than the other distribution, without any assumptions about the shape.

In order to compute the W-M-W test in Minitab, we need the ungrouped data for men and women in two separate columns. The ungrouped data were provided in the long version of the dataset, but the values for men and women need to be unstacked into separate columns in order to the Mann-Whitney menu; alternatively, the Kruskal-Wallis test menu can be used without unstacking the values. Minitab computes the (Mann-Whitney) test as

$$W = 187\,785, \quad P = 0.099 \quad \text{(adjusted for ties)}$$

so there is no significant difference between the two distributions, and we cannot claim that one distribution is systematically larger than the other. (Note: A one-sided alternative hypothesis would give $P = 0.05$, but there is no justification for a one-sided alternative.)

The reason for the different outcomes in the two tests is primarily that they are targeted towards detecting different alternative hypotheses. The X^2-test will detect any differences in the probabilities across the 5 response categories, whereas the W-M-W test is most powerful against alternatives involving systematic differences between the distributions in one direction. The two distributions differ mostly in the no support category, and although this means that men's distribution is somewhat lower on the scale than the women's distribution, the difference between the distributions from category 3 onward is small. Therefore, overall the difference between the two distributions on an ordinal scale is not as strong as is the specific difference in the no support group. As a combined conclusion, we can say that there is a significant difference between the two distributions, and this difference is manifested in the no support category and is not a general tendency for the responses of the men to be lower on the support scale than those of the women.

c) Support responses cross-classified by age

For this question, we lay out the cross-classified data in a similar way as for **a)**, whereby the rows now correspond to the 3 age groups. The calculations and statistical model are also analogous to those of **a)**. First we construct the table of counts and estimated distributions per age group. The marginal distribution was already shown in **a)** but is repeated here to facilitate the comparisons.

Count N_{ij} (Proportion $\hat{p}_{ij}$)		Support category 0 (None)	1 (Little)	2 (Some)	3 (Most)	4 (All)	Total
Age group	0 (65-69)	27 (.083)	53 (.163)	60 (.185)	104 (.320)	81 (.249)	325
	1 (70-74)	26 (.099)	53 (.202)	37 (.141)	82 (.312)	65 (.247)	263
	2 (75+)	39 (.105)	86 (.231)	85 (.229)	93 (.250)	69 (.186)	372
Marginal		92 (.096)	192 (.200)	182 (.190)	279 (.291)	215 (.224)	960

The main difference between the 3 distributions is perhaps that the oldest age group has substantially lower (6-7%) proportions in the two highest response categories (most and all); these lower proportions are then balanced by higher proportions in all the first 3 categories. The distributions for the first two age groups are fairly similar except for response categories 1 and 2 which show a 4% difference in either direction.

Expected count e_{ij} (X^2-contribution)		Support category 0 (None)	1 (Little)	2 (Some)	3 (Most)	4 (All)
Age group	0 (65–69)	31.1 (.55)	65.0 (2.22)	61.6 (.04)	94.5 (0.96)	72.8 (.93)
	1 (70–74)	25.2 (.03)	52.6 (.00)	49.9 (3.32)	76.4 (.41)	58.9 (.63)
	2 (75+)	35.7 (.31)	74.4 (1.81)	70.5 (2.97)	108.1 (2.11)	83.3 (2.46)
		$X^2 = 18.75$, $\text{df} = (3-1)\cdot(5-1) = 8$, $P\text{-value} = 0.016$				

The test statistic is also significant here, but only moderately so ($P = 0.02$). The X^2-contributions are scattered across the table, and no single value is very large. About half of the test statistic originates from the oldest age group, and in the other two age groups the largest contributions are in categories 1 and 2. We conclude that statistical significant differences between the distributions exist, and that these are mostly related to the oldest age group (with responses indicating lower support) and the particular difference between the two younger age groups already discussed.

d) Separate assessments of age and gender effects

Often when describing the marginal distribution across a particular factor, it is implicitly assumed that the conditional distributions within the factor ignored are similar. For example, when discussing the gender distributions and ignoring the factor age, we might have assumed that our findings were similar across all age groups. This is not automatically the case, and we will address this question in **e)**. Non-similarities across age groups would really be an indication of an interaction between age and gender in their relations with support.

Another danger of considering only the marginal distributions is that one may run into a situation akin to Simpson's paradox, where all conditional distributions show one pattern but the marginal distribution shows a different one. Simpson's paradox may occur when strong lurking factors are ignored; we could e.g. hypothesize that gender is a strong lurking varible for the association between age and support. Whether such a situation occurs, is best assessed by simply looking at the conditional distributions and comparing them to the marginal distribution; this we will also do in **e)**.

e) Separate comparisons for age and gender groups

When considering age and gender groups together instead of one at a time, there are in total $3 \cdot 2 = 6$ combined groups whose response distributions needs to be explored. Our first table gives within-group probabilities for the response categories, similar to the distributions given previously. The marginal distributions of the table, across age and gender groups, correspond to the distributions already displayed for gender and age groups in **a)** and **c)**, respectively.

Age group	Count N_{ij} (Proportion $\hat{p}_{ij}$)		Support category					Total
			0 (None)	1 (Little)	2 (Some)	3 (Most)	4 (All)	
0	Gender	0 (Male)	19 (.138)	15 (.109)	25 (.181)	41 (.297)	38 (.275)	138
(65-69)		1 (Female)	8 (.043)	38 (.203)	35 (.187)	63 (.337)	43 (.230)	187
1	Gender	0 (Male)	15 (.136)	16 (.145)	16 (.145)	33 (.300)	30 (.273)	110
(70-74)		1 (Female)	11 (.072)	37 (.242)	21 (.137)	49 (.320)	35 (.229)	153
2	Gender	0 (Male)	24 (.153)	40 (.255)	34 (.217)	39 (.248)	20 (.127)	157
(75+)		1 (Female)	15 (.070)	46 (.214)	51 (.237)	54 (.251)	49 (.228)	215

We will comment on these distributions as we interpret the statistical tests for the requested comparisons. Our models and methods will be similar to those of **a)-c)**. We present the results of these multiple analyses in a more condensed, tabular format: the expected cell counts are not shown; these all exceed 5, so there are no problems with using the reference χ^2-distributions. The easiest way to perform the (X^2) calculations in Minitab is to split the worksheet by the (age or gender) groups requested in **i.)** and **ii.)**.

i) Gender comparisons in different age groups

X^2-contribution		Support category					Tests for Gender	
Age	Gender	0 (None)	1 (Little)	2 (Some)	3 (Most)	4 (All)	X^2 (P)	W (P)
0	0 (Male)	4.95	2.50	0.01	0.23	0.38	14.02	22561
(65-69)	1 (Female)	3.66	1.85	0.01	0.17	0.28	(.007)	(.94)
1	0 (Male)	1.57	1.72	0.02	0.05	0.29	6.26	14746
(70-74)	1 (Female)	1.13	1.23	0.01	0.04	0.21	(.175)	(.70)
2	0 (Male)	3.45	0.38	0.10	0.00	2.86	11.75	26275
(75+)	1 (Female)	2.52	0.28	0.07	0.00	2.09	(.019)	(.003)

Some differences across the age groups are seen. First, only the first and third age groups showed significant gender differences with the X^2-test, contrary to the overall significant gender difference. Large X^2-contributions were seen for the no support category, in all age groups, all reflective of a larger proportion of men than women in this category. The difference ranges from 6% (age group 1) to nearly 10% (age group 0). In the two youngest age groups, this gender difference is counterbalanced mostly by support category 1, similarly to the analysis for all age groups in **a)**, whereas in the oldest group the gender difference is counterbalanced by support category 5 (with a 10% gender difference).

Second, the third age group showed strong significance with the W-M-W test, whereas the test was nonsignificant in the other two groups (just as overall). The reason for the significance in the oldest group is that the two distributions disagree at opposite ends of the scale, so that the distribution for women would generally appear to "higher" than the one for men. The overall W-M-W test became nonsignificant as a result of combining age groups with different patterns.

ii) Age comparisons in different gender groups

X^2-contribution		Support category					Tests for Age
Gender	Age	0 (None)	1 (Little)	2 (Some)	3 (Most)	4 (All)	X^2 (P)
0	0 (65-69)	0.03	3.49	0.01	0.16	2.14	22.06
(Male)	1 (70-74)	0.04	0.56	0.94	0.17	1.56	(.005)
	2 (75+)	0.10	5.66	0.83	0.53	5.84	
1	0 (65-69)	1.04	0.19	0.03	0.89	0.00	9.64
(Female)	1 (70-74)	0.28	0.40	2.45	0.23	0.00	(.29)
	2 (75+)	0.25	0.02	2.20	1.65	0.00	

Here we see a clear gender difference in the comparison between age groups: fairly similar distributions in the 3 age groups for the women, and strong differences between the age groups for the men. The previously displayed distributions for each age and gender group as well as the above X^2-contributions show that the main source of the age difference for men is a much lower (15% drop) proportion of men with support all of the time in the oldest group, counterbalanced by larger counts in the group showing little support. The data for the two other age groups showed larger proportions of men than women with support all of the time.

iii) Conclusion

The analysis showed significant differences in the support available across both age and gender groups. These differences were not totally consistent across age or gender groups, indicating the presence of a (mild) interaction between age and gender. One fairly consistent difference between men and women was that among the former a larger proportion had no support. In the oldest age group, a lower proportion of men than women had support all of the time, and in this age group it thus appeared that the support level for the women was generally larger. The low proportion of men with support all of the time in the oldest age group, combined with a larger proportion of men with little support, was also the strongest contributor to age differences among the men, while no major (statistically significant) age differences were found for the women. Finally, we previously referred to the possibility of confounding effects of age and gender for each other, but it turns out that there is absolutely no association between these two variables ($X^2 = 0.02$, df $= 2$, $P = 0.99$). Hence no confounding effects can exist.