Homework 7 Solutions

p539 # 3

The question is whether the distributions are different for men and women, i.e. if the marriage statistics are independent of sex. So we use test (iv), the χ^2-test for independence. First we need to reformulate our original table in terms of frequency.

The observed frequencies are:

	Men	Women
Never Married	45	30
Married	86	105
Widowed, divorced, separated	12	21

Now we have to compute the expected frequencies. There are 299 total people, and 75 of them are never married, 191 are married, and 33 are widowed, divorced or separated. Thus $\frac{75}{399} = 25\%$ of them are never married, $\frac{191}{299} = 64\%$ are married, and $\frac{33}{299} = 11\%$ are widowed, divorced or separated. If the distribution were independent of sex, then we would expect these same percentages to occur within each sex, so for instance we would expect $(0.25)143 = 34$ men who were never married.

This gives us the following table of expected frequencies:

	Men	Women
Never Married	36	39
Married	91	100
Widowed, divorced, separated	16	17

Now we can compute χ^2 as follows:

$$\begin{aligned}\chi^2 &= \frac{(45-36)^2}{36} + \frac{(86-91)^2}{91} + \frac{(12-16)^2}{16} + \frac{(30-39)^2}{39} + \frac{(105-100)^2}{100} + \frac{(21-17)^2}{17} \\ &= 6.8\end{aligned}$$

Here we have $(3-1)(2-1) = 2$ degrees of freedom, so our P value is just under 5%. So the distributions are different. We observed that more women are married than men, which begs the question of who the women are marrying. The sample is just both men and women aged 25-34, so the women are probably just marrying older (or younger) men.

p539 # 5

Here we are looking at the averages of two samples, and asking about their difference. Thus we should use the 2 sample z-test, which is option (ii). Here the SE for the difference of averages is $\sqrt{(\frac{25{,}000}{\sqrt{250}})^2 + (\frac{40{,}000}{\sqrt{250}})^2} = 2983.29$. Thus we get:

$$z = \frac{(30,000 - 50,000) - 0}{2983.29} = -6.7$$

This is a huge z value, so the P value is nearly 0, and the difference is real.

p540 # 6

Here the null hypothesis says that 55% of newborns are male. This could be done with the χ^2 test or the z-test, and they should give the same answer. To do it with the z-test we have an expected frequency of 550 and a SE of $\sqrt{1000}\sqrt{(0.55)(0.45)} = 15.7$. So we get:

$$z = \frac{568 - 550}{15.7} = 1.15$$

This gives a P-value of 12.5%, which means that the data are consistent with the theory.

If we used the χ^2-test we'd have:

$$\chi^2 = \frac{(568 - 550)^2}{550} + \frac{(432 - 450)^2}{450} = 2$$

Which gives a P-value between 10% and 30%.

p541 # 2

Here is the table in the question, with a column added for the expected number of jurors:

Educational Level	County	Number of jurors	Expected number of jurors
Elementary	28.4%	1	17.6
Secondary	48.5%	10	30
Some college	11.9%	16	7.4
College degree	11.2%	35	7

Now to answer the question we need to compute χ^2:

$$\chi^2 = \frac{(1-17.6)^2}{17.6} + \frac{(10-30)^2}{30} + \frac{(16-7.4)^2}{7.4} + \frac{(35-7)^2}{7} = 151$$

With 3 degrees of freedom this tells us that this distribition is extremely unlikely, which is option (ii).

p541 # 3

This is another χ^2-test for independence. First we need to compute the table with expected frequencies. This is done as in p539#3 and gives:

	Married	Widowed, divorced,..	Never Married
Employed	771.6	103.25	221.6
Unemployed	65.9	8.8	19
Not in Labor Force	28.6	3.8	8.2

Now we compute χ^2:

$$\begin{aligned}\chi^2 = &\frac{(790-771.6)^2}{771.6} + \frac{(98-103.25)^2}{103.25} + \frac{(209-221.6)^2}{221.6}\\ &+ \frac{(56-65.9)^2}{65.9} + \frac{(11-8.8)^2}{8.8} + \frac{(27-19)^2}{19}\\ &+ \frac{(21-28.6)^2}{28.6} + \frac{(7-3.8)^2}{3.8} + \frac{(13-8.2)^2}{8.2}\\ &= 14.35\end{aligned}$$

Here we have 3 rows and 3 columns, so we have $(3-1)(3-1) = 4$ degrees of freedom. Thus we get $P < 1\%$ and we can conclude that mariatal status and employment status are probably not independent.

p542 # 6

Again, we can make the table with a column for expected frequency:

Sum	Observed Frequency	Expected Frequency
2	11	10
3	18	20
4	33	30
5	41	40
6	47	50
7	61	60
8	52	50
9	43	40
10	29	30
11	17	20
12	8	10

Again, we compute χ^2 and get $\chi^2 = 2.01$ so $P = 99.6\%$. This means that in only 0.4% of cases would the results be this close to the expected results, so something is probably fishy.

p542 # 8

Here a χ^2 test is called for. The question is whether or not a die is fair. A fair die is defined to be one that comes up on each side with equal frequency. So to test whether a die is fair we need to test how close it is to the distribution where each side comes up equally.

Another way to think about it is that if the χ^2 test does not measure a statistically significant deviation from the expected distribution, then you can conclude that the die is almost definitely fair. In the case of the z-test for the average all you would know is that the average of the numbers rolled is almost definiely 3.5. This could happen with other distributions though, for instance the one below would also have an expected value of 3.5:

Outcome	Probability
1	$\frac{3}{12}$
2	$\frac{2}{12}$
3	$\frac{1}{12}$
4	$\frac{1}{12}$
5	$\frac{2}{12}$
6	$\frac{3}{12}$

p543 # 10

For this question I just wanted you to compute the χ^2-test. First here is the table of the expected values:

	Protestant	Catholic
Acquitted	6.6	28.4
Convicted	8.4	36.6

So we can compute the χ^2-test for this table versus the one in the book, which is what was observed. This gives $\chi^2 = 0.65$, and so we have $P > 20\%$, and we cannot say that the two are dependent.