(TA) 9.34: Ch 8 P55

- Find MLE and probabilities:

$$lik(\theta) = \binom{3839}{1997, 906, 904, 32} .25 (2+\theta)^{1997} (1-\theta)^{906} (1-\theta)^{904} \theta^{32}$$

$$l(\theta) = \log\left(\binom{3839}{1997,906,904,32} .25\right) + 1997 \log(2+\theta) + 906 \log(1-\theta) + 904 \log(1-\theta) + 32 \log\theta$$

$$l'(\theta) = \frac{1997}{2+\theta} + \frac{906+904}{1-\theta} + \frac{32}{\theta} = 0$$

$$\frac{-3839\theta^2 - 1655\theta + 64}{(2+\theta)(1-\theta)\theta} = 0$$

$$-3839\theta^2 - 1655\theta + 64 = 0$$

$$\Rightarrow \hat{\theta}_{ML} = 0.0357 \quad \Rightarrow \quad \begin{aligned} P_1(\hat{\theta}) &= 0.509 \\ P_2(\hat{\theta}) &= 0.241 \\ P_3(\hat{\theta}) &= 0.241 \\ P_4(\hat{\theta}) &= 0.009 \end{aligned}$$

- Expected vs Observed:

i	O_i	$E_i = 3839 P_i(\hat{\theta})$
1	1997	1953.67
2	906	925.58
3	904	925.58
4	32	34.17

- Test Stat:

$$TS = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} = \frac{1877.49}{1953.67} + \frac{383.49}{925.58} + \frac{465.70}{925.58} + \frac{4.696}{34.17} = 2.015$$

$$TS \sim \chi^2_{4-1-1=2}$$

p value $= 1 - \text{pchisq}(2.015, 2) = 0.365 > \alpha$ (for any typical α)

So the model is a good fit for the data.

7B) 9.38

$H_0: p_{Jan} = p_{Feb} = \ldots = p_{dec} = \frac{1}{12}$

H_1: not all month's probabilities are equal

n = number of men = 43229 $\Rightarrow E_i = 43229 \cdot \frac{1}{12} = 3602.42 \quad \forall i = 1, \ldots, 12$

m = number of women = 16379 $\Rightarrow E_i = 16379 \cdot \frac{1}{12} = 1364.92 \quad \forall i = 1, \ldots, 12$

$$TS_{men} = \sum_{i=1}^{12} \frac{(O_i - 3602.42)^2}{3602.42} = 74.56 \quad \text{(in R)}$$

$$TS_{wom} = \sum_{i=1}^{12} \frac{(O_i - 1364.92)^2}{1364.92} = 53.79 \quad \text{(in R)}$$

$$TS \sim \chi^2_{12-1=11}$$

$\text{pvalue}_{men} = 1 - \text{pchisq}(74.56, 11) = 1.65 \times 10^{-11}$

$\text{pvalue}_{wom} = 1 - \text{pchisq}(53.79, 11) = 1.29 \times 10^{-7}$

$\Big\}$ less than any typical α

So we reject H_0 and conclude that the suicide rates are seasonal.

7C)

a) H_0: these are 2 independent RVs of size 3,3, ie. $\pi_{ij} = \pi_i \pi_j$

"marital status and employment status are independent"

H_1: these are 2 dependent RVs...

"marital status and employment status are not independent"

Degrees of freedom $= \dim \Omega - \dim \omega_0$

$= (3 \cdot 3 - 1) - (2 + 2) = 8 - 4 = 4$

They are dependent. (low p-value)

b) R gives a warning message because their are very low values in the expected table.

7D

$$\sum \frac{(O_i - E_i)^2}{E_i} = 13.369 \quad \text{(in R)}$$

$$2 \sum O_i \log\left(\frac{O_i}{E_i}\right) = 12.389$$

So the first technique was used by R to do the test.

7E

$n = 1231$

$\hat{p}_1 = 0.076$

$z = 1.96$

$$95\% \text{ CI}: \hat{p}_1 \pm z\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}} = 0.076 \pm 1.96(.0076)$$

$$= [0.062, 0.091]$$

7F

$\hat{p}_2 = 0.891$

$\hat{p}_1 - \hat{p}_2 = 0.815$

$$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{s_{\hat{p}_1}^2 + s_{\hat{p}_2}^2}{n}} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1) + \hat{p}_2(1-\hat{p}_2)}{n}} = 0.011$$

$$95\% \text{ CI} = [0.794, 0.836]$$

7G

$$\sum_{i=1}^{2} \frac{(x_i - np_i)^2}{np_i} = \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2} = \frac{(X_1 - np_1)^2}{np_1} + \frac{(\overbrace{n - X_1}^{=X_2} - n\overbrace{(1-p_1)}^{=p_2})^2}{n(1-p_1)}$$

$$= \frac{(1-p_1)(X_1 - np_1)^2 + p_1(n - X_1 - n(1-p_1))^2}{np_1(1-p_1)}$$

$$= \frac{(1-p_1)(X_1 - np_1)^2 + p_1(-X + np_1)^2}{np_1(1-p_1)}$$

$$= \frac{(1-p_1)(X_1 - np_1)^2 + p_1(X_1 - np_1)^2}{np_1(1-p_1)} = \frac{(X_1 - np_1)^2}{np_1(1-p_1)}$$

7H

We have six categories : 0, 1, 2, 3, 4, 5 so the $df = 6 - 1 = 5$ which what R used for the test.

The p-value is 0.9641, which is too large to reject the null hypothesis at any standard significance level. Thus, the data appears to follow a binomial distribution. (which we know to be true!)

7I

Yes, these histograms agree with the conclusions on 341-343 of the text. They show that the two test statistics are approximately equal. We can see that they are very similar density histograms compared to the χ^2_5 distribution.