S352 Midterm Exam 2: Brief answers

Fall 2024

Part I.

1. (a) $\hat{\lambda} = 596/380 \approx 1.568$

   (b) $\text{Var}(\hat{\lambda}) = \lambda/n = \lambda/380$. The estimate of $\text{Var}(\hat{\lambda})$ is $1.568/380 \approx 0.00413$.

   (c) i. $H_0 : \lambda = 2$ vs $H_1 : \lambda \neq 2$

   ii. $W = \dfrac{(\hat{\lambda} - 2)^2}{\text{Var}(\hat{\lambda})}$. The Wald statistic is 35.46 (use $\lambda_0 = 2$ in the variance; preferred) or 45.23 (use $\hat{\lambda} = 1.568$ in the variance; acceptable).

   iii. `1 - pchisq(w, df = 1)`

   (d) $1.568 \pm 1.645\sqrt{0.00413} = 1.568 \pm 0.1057$ or about 1.462 to 1.674 goals per game.

2. (a) $f(4) = 0.1$

   (b)

x (Second digit)	$f(x)$	Number of regions	Expected number of regions
0	0.120	573	550.3
1	0.114	505	522.8
2	0.109	492	499.9
3	0.104	478	476.9
4	0.100	459	458.6
5	0.097	455	444.8
6	0.093	440	426.5
7	0.090	393	412.7
8	0.088	407	403.6
9	0.085	384	389.8

   (c) Note: log in the following formula is natural logarithm.

$$G = 2\sum_{i=1}^{m} o_i \cdot \log\left(\frac{o_i}{e_i}\right)$$

   o_i is the observed count in category i and e_i is the expected count in category i under the null. $m = 10$ since there are 10 categories. $o_i \cdot \log\left(\frac{o_i}{e_i}\right)$ equals the following for $i = 1, \ldots, 10$: (23.14, -17.50, -7.81, 1.06, 0.40, 10.27, 13.71, -19.26, 3.45, -5.77).

   (d) `1 - pchisq(3.392, df = 9)`

(e) The P-values is large, close to 1. We fail to reject the null hypothesis. The data seems to be consistent with the second-digit Benford's law. (Whether this has anything to do with voter fraud is arguable.)

3. (a) The predicted life expectancy in Gabon is $20.138 + 4.631 \times \log(13000) = 64$. (Note: log() in R computes natural logarithms by default.)

   (b) For the countries in Asia, the regression line's intercept is $20.138 + 10.114 = 30.252$ and its slope is 4.631, i.e.

$$\text{Predicted life expectancy} = 30.252 + 4.631 \times \log(\text{gdpPercap})$$

   (c) For the countries in Americas, the regression line's intercept is $20.138 + 11.694 = 31.832$ and its slope is 4.631, i.e.

$$\text{Predicted life expectancy} = 31.832 + 4.631 \times \log(\text{gdpPercap})$$

   (d) Making predictions at the extremes, the Asia countries prediction goes from 62.669 to 76.562 as log(gdpPercap) goes from 7 to 10, while the Americas countries prediction goes from 64.249 to 78.142. We can plot this by joining the dots.

   These two lines are parallel, meaning they have the same slope. We can see that predicted life expectancy is always 1.58 higher in the Americas compared to the Asia, for the same GDP per capita.

Part II.

1. (a) Exponential model is better as it has a lower AIC, 836.1503 vs 837.5869.

   (b) $W = 0.5893562$. p-value: 0.4426682. Fail to reject H_0.

   (c) LR test stat: 0.5633861. p-value: 0.4528991. Fail to reject H_0.

2. First, you need to find the MLE of λ, which is the sampe mean of the number of children. $\hat{\lambda} = 1.57297$.

   (a) Remember to group some of the categories as instructed. Eventually there are only 7 categories. Calculate the observed and expected counts for each category and compute the G test stat. $G = 71.46548$. p-value: 5×10^{-14}. (Note: df = 7-1-1=5.) Strong evidence to reject H_0.

   (b) Compared to a Poisson distribution with rate parameter $\hat{\lambda} = 1.57297$, there are a lot fewer women with one child and a lot more with two kids in the sample data. Also, in the sample, having 4 children is surprising unpopular.

3. (a) Only showing the histogram or density plot is not sufficient. A QQ plot should be presented when checking normality is the goal.

   (b) $\hat{\mu} = 0.1741212, \hat{\sigma}^2 = 198.4192866$. Note: this question asks for MLEs. The sample variance is not accepted here.

(c) All three tests you have practiced are accepted: Wald test (need the normality assumption), t test (no assumption needed as the sample size is large), LR test (need the normality assumptioin). The conclusion is the same: fail to reject the null hypothesis and cannot conclude that μ is not 0.

(d) Sunday.

(e) Model 1: Predicted result $= -0.91 + 0.959 \times$Speed

(f) Model 2:
For Sunday: Predicted result $= -0.46 + 0.948 \times$Speed
For Monday: Predicted result $= -1.67 + 0.948 \times$Speed

(g) Model 1 has lower AIC, 2134.978 vs 2135.878.

(h) Model 1 has lower MSE, 184.9453 vs 188.5711.