SCIENCE

STATS 210: Statistical Theory

Assignment Tracking Sheet

Student Information

University ID:	107664618	Username:	mhu138
Family Name:	Hu	Given Names:	Joy

Assignment Information

Assignment Name:	Assignment 2	Due:	11:00 p.m. - 03 Sep, 2020 (NZ Time)
Department:	statistics		
Lab / Tutorial Day:		Time:	
Lab / Tutorial Group:		Tutor:	
Notes:		Word Count:	

Declaration: (please read and sign)

By submitting this assignment, I confirm that I am aware of The University expectation that all students complete coursework with integrity and honesty as stated in the Student Academic Conduct Statute.
http://www.auckland.ac.nz/uoa/home/about/teaching-learning/honesty/tl-uni-regs-statutes-guidelines

- I understand that the University of Auckland will not tolerate cheating or assisting other to cheat, and views cheating in coursework as a serious academic offence.
- I declare that where work from other sources (including sources on the world-wide web) has been used, it has been properly acknowledged and referenced.
- I confirm that this work represents my individual/ our team's effort and does not contain plagiarised material.
- I have checked the above details and verify them to be correct for the assignment I am submitting.
- I understand that the University of Auckland takes no responsibility for lost assignments and that I agree to provide a duplicate copy if requested.
- I understand that uncollected assignments will be retained in secure storage until the end of the examination period and thereafter destroyed.
- I agree that I will provide or submit an electronic version of my work for computerised review if requested.

Signed: ____________________ Date: ____________________

Note:

1. Assignments are not accessible after they have been handed in. No additions/removals will be permitted.
2. Marks may be withheld for students who have not submitted their work to Turnitin.com if required in the course outline.
3. The University of Auckland views cheating in coursework as a serious academic offence. Accordingly it may require submitted work to be reviewed against electronic source material using computerised detection mechanisms.

ID: 107664618
Joy HU
UPI: mhu138

1. (a) X is the number of races that Team A wins out of a fixed number of 15 races. We are counting the success out of a fixed number of trial, it's a Binomial distribution. Team A, B are equality likely to win, therefore, $p = 0.5$

$$X \sim \text{Binomial}(15, 0.5)$$

(b) We wants to check whether the probability of winning of both teams are equally likely. X is the number of winning in 15 races. The distribution is

$$X \sim \text{Binomial}(15, p_A)$$
$$H_0 : p_A = 0.5$$
$$H_1 : p_A \neq 0.5 \quad \text{(two-sided test)}$$

(c) The graph peaks at $x = 15 \times 0.5 = 7.5$

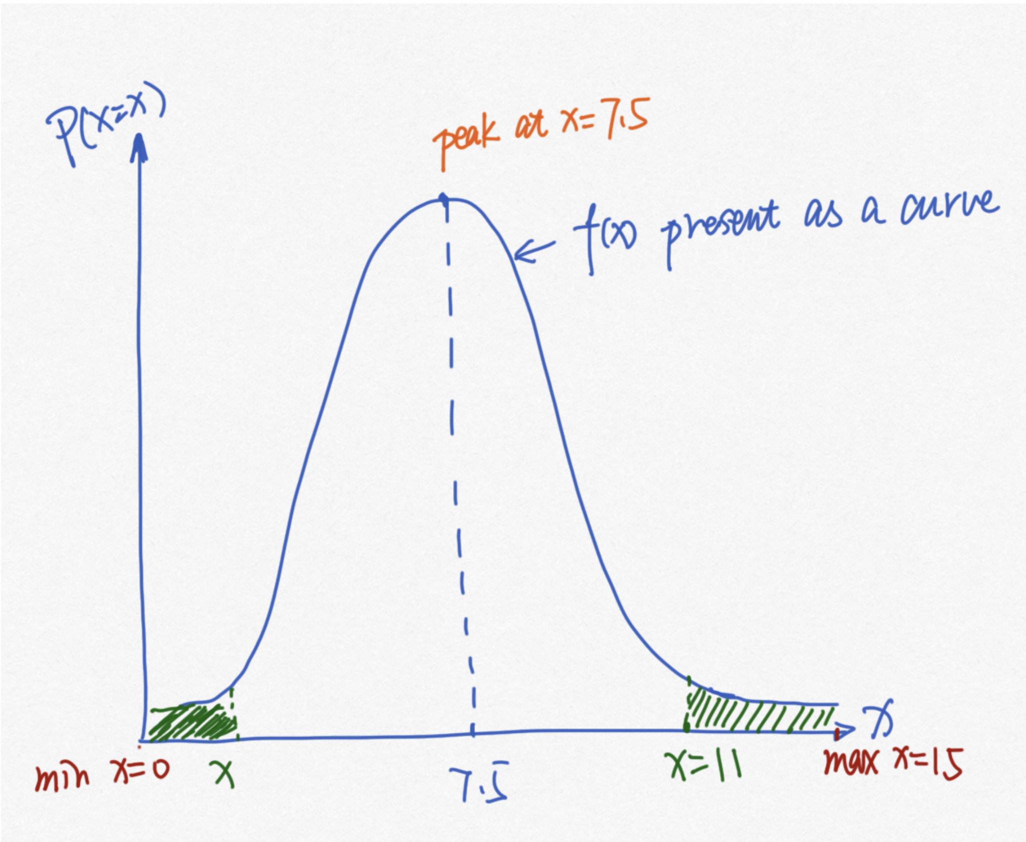

Figure 1: shape of probability function of X under null hypothesis

(d) For the p-value:

$$\begin{aligned}\mathbb{P}(X \geq 11) &= 1 - \mathbb{P}(X < 11) \\ &= 1 - \mathbb{P}(X \leq 10) \\ &= 1 - F_X(10)\end{aligned}$$

The R code for the p-value is: `2 * (1 - pbinom(10,15,0.5))`
p-value is calculated as 0.12. If the true probability is 0.5, we have close to 12% of the

chance to see that out of 15 races, team A will win 11 times. The p-value is big, therefore we don't have evidence against of null hypothesis. The null hypothesis just to test whether the winning probability is 0.5 and we don't have evidence that Team A is quicker than Team B.

(e) If team A won 110 out of 150 times, and conducting the same test, the p-value will become smaller than we will have more evidence against H_0. It's because the more samples we have, the less uncertainty about the population, and the probability will get more close to the true parameter, and less probability that our observed outcome just by chance. If our true probability is 0.5, when the sample is big as 150, we are expected to see that team A and team B will win roughly the same amount of number. Since out of 150 races, team A won 110 times, it's more likely that team A has higher winning probability.

2. (a) X is the number of customer who do not switch to a chocolate fish(failure) before 16 chocolate finish has been used up(success) with probability p. The distribution of X is negative binomial.

$$X \sim \text{Negative Binomial}(16, p)$$

(b) X is # of failures before 16 chocolate fishes have been used up, total number of customer that has been ask is 25, so the number of failure is $x = 25 - 16 = 9$

$$L(p; 9) = P(X = 9) \text{ when } X \sim \text{NegBin}(16, p)$$
$$L(p; 9) = \binom{16+9-1}{9} p^{16}(1-p)^9 = \binom{24}{9} p^{16}(1-p)^9 \qquad \text{for } 0 < p < 1$$

When $\dfrac{dL}{dp} = 0$, the probability is maximum.

$$\begin{aligned}\frac{dL}{dp} &= \binom{24}{9}\left(16p^{15}(1-p)^9 + p^{16} \times 9 \times (1-p)^8 \times (-1)\right) \quad \text{product rule}\\ &= \binom{24}{9} p^{15}(1-p)^8\left(16(1-p) - 9p\right)\\ &= \binom{24}{9} p^{15}(1-p)^8(16 - 25p) \quad \text{for } 0 < p < 1\end{aligned}$$

when $\left.\dfrac{dL}{dp}\right|_{p=\hat{p}} = 0$

$$\implies \binom{24}{9}\hat{p}^{15}(1-\hat{p})^8(16 - 25\hat{p}) = 0$$
$$\implies \hat{p} = 0 \text{ or } \hat{p} = 1 \text{ or } 16 - 25\hat{p} = 0$$

According the likelihood function of graph does dot peak at $\hat{p} = 0$ or $\hat{p} = 1$, the peak is close to 0.6. Maximizing L with respect to p to find MLE when $x = 9$, $16 - 25\hat{p} = 0 \implies \hat{p} = \dfrac{16}{25} = 0.64$

(c) when observed $X = x$, The maximum likelihood function for $0 < p < 1$ is given below

$$
\begin{aligned}
L(p; x) &= P(X = x) \text{ when } X \sim \text{NegBin}(16, p) \\
L(p; x) &= \binom{16 + x - 1}{x} p^{16}(1-p)^{x} \\
&= \binom{15 + x}{x} p^{16}(1-p)^{x} \qquad \text{for } 0 < p < 1
\end{aligned}
$$

When $\frac{dL}{dp} = 0$, the value is at its maximum.We first calculate $\frac{dL}{dp}$

$$
\begin{aligned}
\frac{dL}{dp} &= \binom{15 + x}{x} \left(16p^{15}(1-p)^{x} + p^{16} \times x(1-p)^{x-1} \times (-1)\right) \quad \text{product rule} \\
&= \binom{15 + x}{9} p^{15}(1-p)^{x-1} \left(16(1-p) - xp\right) \\
&= \binom{15 + x}{9} p^{15}(1-p)^{x-1}(16 - (16 + x)p) \quad \text{for } 0 < p < 1
\end{aligned}
$$

Then, we calculate $\left.\frac{dL}{dp}\right|_{p=\hat{p}} = 0$

$$
\begin{aligned}
\left.\frac{dL}{dp}\right|_{p=\hat{p}} &= \binom{15 + x}{9} \hat{p}^{15}(1-\hat{p})^{x-1}(16 - (16 + x)\hat{p}) = 0 \\
&\implies 16 - (16 + x)\hat{p} = 0 \\
\implies \hat{p} &= \frac{16}{16 + x}
\end{aligned}
$$

We know that the MLE of L with respect to p is $\hat{p} = \frac{16}{16+x}$. Replace x with X, we will get the estimator $\hat{p} = \frac{16}{16+X}$

(d) X has negative binomial distribution with probability of p,

$$
\begin{aligned}
X &\sim \text{Negative Binomial}(16, p) \\
H_0 &: p = 0.5 \\
H_1 &: p \neq 0.5 \qquad \text{(two-sided test)}
\end{aligned}
$$

(e) $f_X(x) = \binom{16+x-1}{x} p^{16}(1-p)^{x}$, under null hypothesis, $p = 0.5$
$p_1 = \mathbb{P}(X = 9) = \binom{16+9-1}{x} p^{16}(1-p)^{9} = 0.0390$
$p_2 = \mathbb{P}(X = 25) = \binom{16+25-1}{x} p^{16}(1-p)^{25} = 0.0183$

(f) p-value is the probability of getting a result at least as extreme as $X = 9$. And results at least as extrem as $X = 9$ are $X \leq 9$ and the equal probability in the opposite direction.
We know that $\mathbb{P}(X \leq 9) = 0.1148$, so the p-value = $2 \times \mathbb{P}(X \leq 9) = 0.2296$
The data is compatible with H_0, we don't have evidence to against our null hypothesis which is that the probability of each customer switching the the chocolate fish is the same as not switching. Catriona's suspicion is not been justified.

3. (a) Each variable Y_i is independent, and each variable has distribution $Y_i \sim$ Poisson$(\beta \log(x_i))$
$\mathbb{E}(Y_i) = \beta \log(x_i)$
Var$(Y_i) = \mathbb{E}(Y_i) = \beta \log(x_i)$

(b) Using Matiu's model, $Y_1, Y_2, Y_3 Y_4$ are independent, then we have

$$\mathbb{P}(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3, Y_4 = y_4) = \mathbb{P}(Y_1 = y_1)\mathbb{P}(Y_2 = y_2)\mathbb{P}(Y_3 = y_3)\mathbb{P}(Y_4 = y_4)$$

So, we could write

$$\mathbb{P}(Y_1 = 329, Y_2 = 437, Y_3 = 590, Y_4 = 630) = \mathbb{P}(Y_1 = 329)\mathbb{P}(Y_2 = 437)\mathbb{P}(Y_3 = 590)\mathbb{P}(Y_4 = 630)$$

(c) The likelihood function is

$$\begin{aligned}
L(\beta; y, x) &= \mathbb{P}(y_1, y_2, y_3, y_4 | x_1, x_2, x_3, x_4; \beta) \\
&= \prod_{i=1}^{4} \mathbb{P}(y_i | x_i; \beta) \qquad \text{(by independence)} \\
&= \prod_{i=1}^{4} \frac{(\beta \log(x_i))^{y_i}}{y_i!} e^{-\beta \log(x_i)} \\
&= \prod_{i=1}^{4} (\frac{\log(x_i)^{y_i}}{y_i!}) \beta^{y_i} e^{-\beta \log(x_i)} \\
&= K \beta^{y_1+y_2+y_3+y_4} e^{-\beta(\log(x_1)+\log(x_2)+\log(x_3)+\log(x_4))} \\
&= K \beta^{y_1+y_2+y_3+y_4} e^{-\beta \log(x_1 x_2 x_3 x_4)} \quad \text{where K is a constant not depend on } \beta \\
&\text{for } 0 < \beta < \infty
\end{aligned}$$

(d) In order to find the maximum value for β, we differentiate likelihood function and set to 0 for MLE.

$$\begin{aligned}
0 &= \frac{d}{d\beta} L(\beta; y_1, y_2, y_3, y_4) \\
&= \frac{d}{d\beta} \{K \beta^{y_1+y_2+y_3+y_4} e^{-\beta \log(x_1 x_2 x_3 x_4)}\} \\
&= K((y_1 + y_2 + y_3 + y_4) \beta^{y_1+y_2+y_3+y_4-1} e^{-\beta \log(x_1 x_2 x_3 x_4)} \\
&- \beta^{y_1+y_2+y_3+y_4} \log(x_1 x_2 x_3 x_4) e^{-\beta \log(x_1 x_2 x_3 x_4)}) \\
&= K \beta^{y_1+y_2+y_3+y_4-1} e^{-\beta \log(x_1 x_2 x_3 x_4)} ((y_1 + y_2 + y_3 + y_4) - \beta \log(x_1 x_2 x_3 x_4)) \\
\implies 0 &= (y_1 + y_2 + y_3 + y_4) - \beta \log(x_1 x_2 x_3 x_4) \quad \text{or } \beta = 0, \infty \\
\implies \hat{\beta} &= \frac{y_1 + y_2 + y_3 + y_4}{\log(x_1 x_2 x_3 x_4)} \quad \text{assuming a unique maximum in } 0 < \beta < \infty
\end{aligned}$$

To obtain maximum likelihood estimator, just replace y_i with variable Y_i

$$\hat{\beta} = \frac{Y_1 + Y_2 + Y_3 + Y_4}{\log(x_1 x_2 x_3 x_4)}$$

ID: 107664618
Joy HU
UPI: mhu138

(e) The MLE is

$$\begin{aligned}\hat{\beta} &= \frac{329 + 437 + 590 + 630}{\log(50 \times 100 \times 150 \times 200)} \\ &= 105.4916\end{aligned}$$

(f)

$$\begin{aligned}\mathbb{E}(\hat{\beta}) &= \mathbb{E}\left(\frac{Y_1 + Y_2 + Y_3 + Y_4}{\log(x_1 x_2 x_3 x_4)}\right) \\ &= \frac{1}{\log(x_1 x_2 x_3 x_4)}\mathbb{E}(Y_1 + Y_2 + Y_3 + Y_4) \\ &= \frac{1}{\log(x_1 x_2 x_3 x_4)}\left(\mathbb{E}(Y_1) + \mathbb{E}(Y_2) + \mathbb{E}(Y_3) + \mathbb{E}(Y_4)\right) \\ &= \frac{1}{\log(x_1 x_2 x_3 x_4)}\left(\beta\log(x_1) + \beta\log(x_2) + \beta\log(x_3) + \beta\log(x_4)\right) \\ &= \beta\frac{1}{\log(x_1 x_2 x_3 x_4)} \times \log(x_1 x_2 x_3 x_4) \\ &= \beta\end{aligned}$$

We derived our case that $\mathbb{E}(\hat{\beta}) = \beta$, so $\hat{\beta}$ is an unbiased estimator.

(g)

$$\begin{aligned}\mathbb{E}(Z_i) &= \mathbb{E}(Y_i - x_i) \\ &= \mathbb{E}(Y_i) - \mathbb{E}(x_i) \\ &= \beta\log(x_i) - x_i\end{aligned}$$