Math 32 Written Homework #12 Fall 2021

Complete the following tasks. Some answers have been provided. Assemble your work into one PDF document and upload the PDF back into our CatCourses page.

1. A Poisson distribution with rate parameter λ has probability mass function

$$f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \geq 0$$

Show that all probabilities do indeed add up to 100 percent. (Hint: this is a quick proof. You do not have to do a proof within a proof.)

2. In a battle arena, the number of enemies that are procedurally generated follow a Poisson distribution with a mean of 4.10 enemies. Compute the probability that at most 2 enemies will be generated.

3. The music playlist for Bio 184 features a variety of songs. “Baby Shark” tends to evoke reactions. We tend to play the kids’song minimally, so let’s model the number of times the song is played during our class session with a Poisson distribution. If the average number of times the YouTube playlist plays “Baby Shark” during a class session is 0.62, compute the probability that “Baby Shark” plays at least once during our class session.

4. Here $X_i \sim Bin(n,p)$ are a set of m i.i.d. random variables, where n is the sample size. Given a data set $\{x_1, x_2, ..., x_m\}$ of m observations, assume an $Bin(n,p)$ distribution. Compute the value of parameter p that maximizes the likelihood of the data set.[1]

5. Given at bivariate data set $\{x_i, y_i\}_{i=1}^n$, we will form the best-fit *linear regression model*

$$\hat{y} = a + bx$$

by the Method of Least Squares. Recall that the total error is

$$E = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - a - bx_i)^2$$

Derive[2] that the coefficients are

$$a = \frac{(\sum y_i)(\sum x_i^2) - (\sum x_i)(\sum x_i y_i)}{n \sum x_i^2 - (\sum x_i)^2}, \quad b = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

[1] Hint: $0 \leq p \leq 1$

[2] Hint: if you have at least a modicum of knowledge of linear algebra, use the matrix inverse of a 2-by-2 matrix.

6. Given a bivariate data set $\{x_i, y_i\}_{i=1}^n$, show that the best-fit linear regression line has to go through the center of gravity $(\bar{x}, \bar{y})$. Hint: there is a quick proof that does not need a summation symbol.

7. InnerSloth continue to update their popular video game *Among Us*. Scraping their server logs, we have the sample data below. For the number of players, the sample statistics include a mean of about 7.7778 players and a standard deviation of about 1.3017 players. For the game times, the sample statistics include a mean of about 407.5556 seconds and a standard deviation of about 68.4235 seconds. The correlation between the number of players and the game time is about 0.9012.

   (a) Build a linear regression model for the sample data.

   (b) Use the model to predict the length of a game that has nine players.

8. "Will someone please think of the salmon?", cries out an enviromentalist. We gather some data about local salmon fisheries[3], and the recent data selection is shown below. If we treat the years as the $\{x_i\}$ data and the Chinnook salmon population as the $\{y_i\}$ data, then some sample statistics include $\bar{x} = 2002$, $\bar{y} \approx 356.1429$ thousand salmon, $s_x \approx 6.4807$, $s_y \approx 271.2320$ thousand salmon, and the data have a correlation of $r \approx -0.2352$. Use this information to predict the population amount of Chinnook salmon in local California fisheries in the year 2020.

Year	**Chinook Salmon (in thousands)**
1993	225
1996	350
1999	400
2002	880
2005	450
2008	66
2011	122

[3] Source: `https://www.fishwildlife.org/application/files/5815/3556/8600/California_Chinook_Salmon_Population_Data.pdf`

Here are some of the answers. Note that numbers may slightly vary depending on when and where the rounding took place.

1.
2. 0.2238
3. 0.4621
4. $p^* = \dfrac{\bar{x}_m}{n}$
5.
6.
7. (a)
   (b) 465.4528 seconds
8. (a)
   (b) 178.9563 thousand salmon