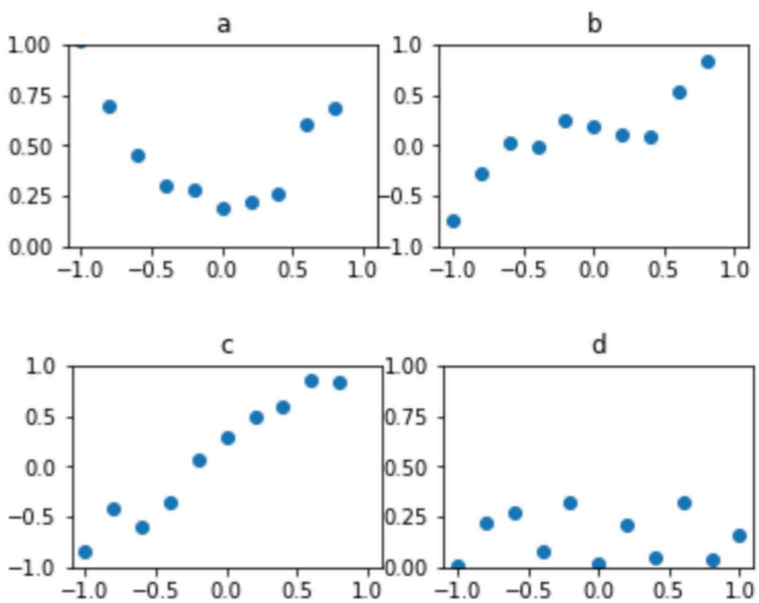
DDA2001: Assignment 3

1. The assignment is due at Monday **11:59 pm, March 21, 2022**.
2. Please submit your solution in **PDF** form. **Any other forms of solution will not be accepted and will be graded as 0.** Please leave enough time to make sure you have uploaded your solution as requirement before due.
3. If you submit the assignment late, you will get 0 for this assignment. **No excuses will be accepted for any late submission.**
4. **Please make sure that your file could be downloaded successfully from BB after uploading your solution file.**

---

1. **(5 points)**Which of following 4 graphs is the most suitable to be represented by a simple linear regression model.



Solution: C □

2. **(5 points)** Which of the following factors do not affect the *width* of the confidence interval for the population mean?

A. Sample mean

B. Population variance

C. Sample size

D. Confidence level

Solution: A $\square$

3. **(15 points)** Assume a discrete random variable that takes values in $\{0, 1, 2, 3, 4\}$. We would like to examine three candidate models for the distribution of this random variable, each model is represented using a parameter $\theta \in \{1, 2, 3\}$, and the PMF for each model $f(x|\theta)$ is shown in the following table. Assume we only have one observation X, i.e., the sample size is 1. Find the Maximum Likelihood Estimate (MLE) of θ for each observation value X. (What is the MLE for θ if (i) $X = 0$; (ii) $X = 1$; (iii) $X = 2$; (iv) $X = 3$; (v) $X = 4$?)

x	f(x|1)	f(x|2)	f(x|3)
0	1/3	1/4	0
1	1/3	1/4	0
2	0	1/4	1/4
3	1/6	1/4	1/2
4	1/6	0	1/4

Table 1: observations of X with $f(x|\theta)$.

Solution: For each value of x, the MLE $\hat{\theta}$ of θ is that maximizes $f(x|\theta)$ as shown in the following table. Note that when $x = 2$, $f(x|\theta = 2) = f(x|\theta = 3) = 1/4$ are both maxima, so both $\theta = 2$ and $\theta = 3$ are MLEs for $x = 2$. $\square$

x	0	1	2	3	4
$\hat{\theta}$	1	1	2/3	3	3

4. **(20 points)** It is not known what proportion p of the purchases of a certain brand of breakfast cereal are made by women and what proportion are made by men. In a random sample of 70 purchases of this cereal, it was found that 58 were made by women and 12 were made by men. Find the Maximum Likelihood Estimate (MLE) of p.

Solution: Suppose the observed samples are $x_1, x_2, \ldots, x_{70}$ and x_i is either 0 (by men) or 1 (by women). Then the likelihood function is

$$L(x|p) = \prod_{i=1}^{70} p^{x_i}(1-p)^{1-x_i}.$$

The log-likelihood is

$$l(p) = (\sum_{i=1}^{70} x_i) \log p + (70 - \sum_{i=1}^{70} x_i) \log(1-p)$$

which is maximized at $p = \frac{1}{n}\sum_{i=1}^{70} = 58/70 = 29/35$. So the M.L.E. of p is 29/35. $\square$

5. **(20 points)**The number of threes made during an NBA game is Poisson distribution. Last Saturday, the number of threes made were 14, 26, 25, and 13. Calculate the maximum likelihood estimate for the parameter λ.

Solution: The likelihood function is

$$\begin{aligned} L(\lambda) &= P(14, 26, 25, 13|\lambda) \\ &= \frac{\lambda^{14}e^{-\lambda}}{14!} \cdot \frac{\lambda^{26}e^{-\lambda}}{26!} \cdot \frac{\lambda^{25}e^{-\lambda}}{25!} \cdot \frac{\lambda^{13}e^{-\lambda}}{13!} \\ &= \frac{\lambda^{78}e^{-4\lambda}}{14!26!25!13!}. \end{aligned} \tag{1}$$

Taking the derivative and setting it equal to 0 gives

$$\lambda^{78}(-4e^{-4\lambda}) + 78\lambda^{77}e^{-4\lambda} = \lambda^{77}e^{-4\lambda}(-4\lambda + 78) = 0. \tag{2}$$

Thus,

$$\lambda = 0, \quad or \quad \lambda = 78/4 = 19.5. \tag{3}$$

Since $\lambda = 0$ doesn't make sense and 19.5 is a local maximum, we have $\lambda = 19.5$.

□

6. **(20 points)**Assume that baby weights are normally distributed with mean μ and variance σ^2. After knowing three babies weighing 7, 8, 9 ounces, what is the maximum likelihood estimate for μ?

Solution: The pdf for a normal distribution is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}). \tag{4}$$

So, we want to maximize

$$\frac{1}{\sqrt{2\pi}^2\sigma^3} \exp(-\frac{(7-\mu)^2}{2\sigma^2} - \frac{(8-\mu)^2}{2\sigma^2} - \frac{(9-\mu)^2}{2\sigma^2}) \tag{5}$$

for μ.

We can take the logarithmic derivative and set that equal to 0. Doing so gives us

$$-2(7-\mu) - 2(8-\mu) - 2(9-\mu) = 0 \Rightarrow \mu = 8 \tag{6}$$

□

7*. **(5×3 points)**To estimate how many fish there are in the lake, we take 1,000 of them, mark them, put them back in the lake, and then we take another 150 fish where you find 10 of them are marked.

(a) Suppose there are N fishes in total, what is the probability of having 10 marked fish out of 150? [Just give the formula, no need to simplify it]

(b) Denote the probability in (a) as $L(N, 10)$, and denote the ratio $\frac{L(N,10)}{L(N-1,10)}$ as $A(N, 10)$. Within what range should N be to make $A(N, 10) \geq 1$?

(c) Based on the above results, please find out how many fishes are there in the lake in order to maximize the probability of having 10 marked fish out of 150?

Solution: We let the number of fish with mark in the second trail be X, and we can calculate the probability of having 10 marked fish out of 150 as follows:

$$P(X = 10) = \frac{\binom{1000}{10}\binom{N-1000}{140}}{\binom{N}{150}}$$

where N denote the numer of fish in the lake, which is an unknown parameter, and the liklihood funtion is:

$$L(N; 10) = \frac{\binom{1000}{10}\binom{N-1000}{140}}{\binom{N}{150}}$$

We consider the ratio:

$$\begin{aligned} A(N, 10) = \frac{L(N; 10)}{L(N-1; 10)} &= \frac{(N-1000)(N-150)}{N(N-1000-140)} \\ &= \frac{N^2 - 1150N + 150000}{N^2 - 1150N + 10N} \end{aligned}$$

If and only if $N < 15000$, $A(N, 10) > 1$; If and only if $N > 15000, A(N, 10) < 1$. Therefore, only when $N = 15000$, $L(N; 10)$ reach the maximum value, which means $N = 15000$ is what we want.

□