Questions for STAT 2006 mid-term exam

1. A parameter is

   (A) a sample characteristic

   (B) a population characteristic

   (C) unknown

   (D) normally distributed

   Answer: a population characteristic

2. A statistic is

   (A) a function of samples

   (B) a population characteristic

   (C) unknown

   (D) normally distributed

   Answer: a function of samples

3. Which of the following denotes sample variance

   (A) $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$

   (B) $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$

   (C) $\frac{\sum_{i=1}^{n}x_i}{n}$

   (D) $\sum_{i=1}^{n}(x_i-\bar{x})^2$

   Answer: $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$

4. Since the population size is always larger than the sample size, the value of a sample statistic

   (A) can never be larger than the value of the corresponding population parameter

   (B) can never be equal to the value of the corresponding population parameter

   (C) can never be smaller than the value of the corresponding population parameter

   (D) None of the above answers is correct.

   Answer: None of the above answers is correct

5. The histogram below represents the lifespan of a random sample of a particular type of insect. Determine the relationship between the mean and median.

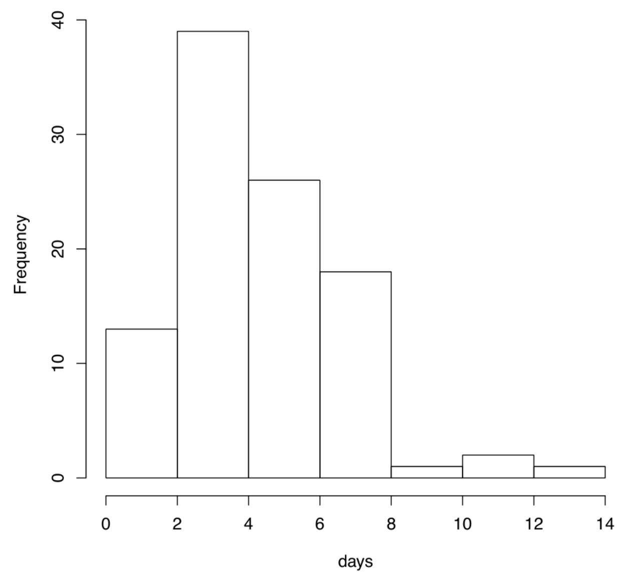
   (A) mean $=$ median

   (B) mean $\approx$ median

   (C) mean $<$ median

   (D) mean $>$ median

   Answer: mean $>$ median



6. Given the following box plot:

   What is the value of the third quartile?

   (A) 2
   (B) 4
   (C) 7
   (D) 11

   Answer: 11

7. Let W, X, Y be i.i.d. and follow $U(-1, 1)$. Evaluate $P(W > XY)$.

   (A) $\frac{1}{2}$
   (B) $\frac{3}{4}$
   (C) $\frac{2}{3}$
   (D) $\frac{4}{5}$

   Answer: $\frac{1}{2}$

   Since W, X, Y are independent,

$$f_{W,X,Y}(w, x, y) = \frac{1}{8}, \quad -1 \leq W, X, Y \leq 1.$$

   Then

$$P(W > XY) = \int\int\int_{w>xy} \frac{1}{8} dwdxdy = \frac{1}{8}\int_{-1}^{1}\int_{-1}^{1}\int_{xy}^{1} dwdxdy = \frac{1}{8}\int_{-1}^{1}\int_{-1}^{1}(1 - xy)dxdy$$
$$= \frac{1}{8}\int_{-1}^{1} 2dy = \frac{1}{2}.$$

8. Let X and Y be two jointly continuous random variables with joint PDF

$$f_{XY}(x,y) = \begin{cases} \frac{1}{2} & y+x \le 2, x>0, y>0 \\ 0 & \text{otherwise} \end{cases}$$

Find $Cov(X,Y)$.

(A) $-\frac{1}{36}$
(B) $-\frac{1}{12}$
(C) $-\frac{1}{9}$
(D) $-\frac{1}{6}$

Answer: $-\frac{1}{9}$

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy = \int_0^{2-x} \frac{1}{2}dy = \frac{1}{2}(2-x), \quad 0 \le x \le 2.$$

Similarly,

$$f_Y(y) = \frac{1}{2}(2-y), \quad 0 \le y \le 2.$$

Then

$$E(XY) = \int\int_{0\le x+y\le 2} \frac{1}{2}xydxdy = \int_0^2 \int_0^{2-x} \frac{1}{2}xydydx = \frac{1}{3}.$$

$$Cov(X,Y) = E(XY) - EXEY = -\frac{1}{9}.$$

9. Let $W \sim U(0,1), X \sim U(0,2), Y \sim U(0,3)$ and they are mutually independent. Let $A = \max(X,Y)$ and $B = \min(A,W)$. Find $P(B < \frac{2}{3})$.

(A) $\frac{56}{81}$
(B) $\frac{60}{81}$
(C) $\frac{64}{81}$
(D) $\frac{76}{81}$

Answer: $\frac{56}{81}$

$$\begin{aligned} P(B<\frac{2}{3}) &= P(\min(A,W)<\frac{2}{3}) = 1-P(\min(A,W)\ge\frac{2}{3}) = 1-P(A\ge\frac{2}{3})P(W\ge\frac{2}{3}) \\ &= 1-P(W\ge\frac{2}{3})P(\max(X,Y)\ge\frac{2}{3}) = 1-P(W\ge\frac{2}{3})[1-P(\max(X,Y)<\frac{2}{3})] \\ &= 1-P(W\ge\frac{2}{3})[1-P(X<\frac{2}{3})P(Y<\frac{2}{3})] \\ &= 1-\frac{1}{3}[1-\frac{1}{3}\times\frac{2}{9}] = \frac{56}{81}. \end{aligned}$$

10. The moment generating functions of two independent random variables K, L are:

$$M_K(t) = (1-2t)^{-2.5}, \quad M_L(t) = (1-2t)^{-4.5}.$$

Let $X = K + L$. Calculate $E(X^3)$.

(A) 224

(B) 2082

(C) 4032

(D) 8064

Answer: 4032

Due to independence,

$$M_X(t) = M_K(t)M_L(t) = (1-2t)^{-7}.$$
$$M'(t) = 14(1-2t)^{-8}.$$
$$M''(t) = 224(1-2t)^{-9}.$$
$$M'''(t) = 4032(1-2t)^{-10}.$$
$$E[X^3] = M'''(0) = 4032.$$

11. Let $X_1, X_2, \cdots, X_n \overset{i.i.d.}{\sim} U(0,1)$, find $\lim_{n\to\infty} \mathbb{P}\left(1 \le (X_1X_2\cdots X_n)^{n^{-1/2}} e^{n^{1/2}} \le 2\right)$. ($\Phi(\cdot)$ stands for the CDF of standard normal distribution)

(A) $\Phi(\ln(2)) - \Phi(0)$

(B) $\Phi(\ln(4)) - \Phi(0)$

(C) $\Phi(0)$

(D) $\Phi(\ln(3)) - \Phi(0)$

Answer: $\Phi(\ln(2)) - \Phi(0)$

$$\mathbb{E}[\ln X_1] = \int_0^1 \ln x dx = (x\ln x - x)\Big|_0^1 = -1.$$
$$\mathbb{E}[(\ln X_1)^2] = \int_0^1 (\ln x)^2 dx = \left[x(\ln x)^2 - 2(x\ln x - x)\right]\Bigg|_0^1 = 2.$$
$$Var(\ln X_1) = 2 - (-1)^2 = 1.$$
$$\begin{aligned}\mathbb{P}\left(1 \le (X_1X_2\cdots X_n)^{\frac{1}{\sqrt{n}}} e^{\sqrt{n}} \le 2\right) &= \mathbb{P}\left(0 \le \frac{1}{\sqrt{n}}\sum_{i=1}^n \ln X_i + \sqrt{n} \le \ln 2\right)\\ &= \mathbb{P}\left(0 \le \frac{\sum_{i=1}^n \ln X_i + n}{\sqrt{n}} \le \ln 2\right).\end{aligned}$$

Then by CLT,

$$\lim_{n\to\infty} \mathbb{P}\left(1 \le (X_1X_2\cdots X_n)^{n^{-1/2}} e^{n^{1/2}} \le 2\right) = \Phi(\ln 2) - \Phi(0).$$

12. Let X equal the weight (in pounds) of a 12-ounce can of buttermilk biscuits. Assume that the distribution of X is $N(\mu, \sigma^2)$. There are 18 random samples with summary statistics: $\bar{x} = 0.7639, s = 0.1577$. Find the a 90% one-sided confidence interval for μ that provides a lower bound for μ.

(A) $[0.714, \infty)$

(B) $[0.699, \infty)$

(C) $[0.685, \infty)$

(D) $[0.735, \infty)$

Answer: $[0.714, \infty)$

With $n = 18$ and $\alpha = 0.10$, the lower bound for μ is

$$\bar{x} - t_{17,0.1}\frac{s}{\sqrt{n}} = 0.7639 - 1.333 \cdot \frac{0.1577}{\sqrt{18}} = 0.714.$$

So the required confidence interval is $[0.714, \infty)$.

13. To determine the effect of 100% nitrate on the growth of pea plants, several specimens were planted and then watered with 100% nitrate every day. At the end of two weeks, the plants were measured. Here are the data on seven of them:

$$17.5, \quad 14.5, \quad 15.2, \quad 14.0, \quad 17.3, \quad 18.0, \quad 13.8$$

Assume that these data are observations from a normal distribution $N(\mu, \sigma^2)$. Give the 95% confidence interval for μ.

(A) $[14.100, 17.414]$

(B) $[14.441, 17.073]$

(C) $[14.473, 17.041]$

(D) $[14.155, 17.359]$

Answer: $[14.100, 17.414]$

Since σ^2 is unknown and the sample size n is small, we use the fact that $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ follows a $t(n-1)$ distribution. Then the $100(1-\alpha)\%$ confidence interval is

$$\left[\bar{x} - t_{\alpha/2}(n-1)\left(\frac{s}{\sqrt{n}}\right), \bar{x} + t_{\alpha/2}(n-1)\left(\frac{s}{\sqrt{n}}\right)\right]$$

In this case, $\bar{x} = 15.757, s = 1.792, \alpha = 0.1, n = 7, t_{\alpha/2}(n-1) = t_{0.025}(6) = 2.447$. Hence we have the 95% confidence interval

$$\left[15.757 - 2.447(\frac{1.792}{\sqrt{7}}), 15.757 + 2.447(\frac{1.792}{\sqrt{7}})\right] = [14.100, 17.414].$$

14. Let $X_1, X_2, \ldots, X_n$ be i.i.d. with one of two pdfs. If $\theta = 0$, then

$$f(x;\theta) = \begin{cases} 1, & \text{if } 0 < x < 1; \\ 0, & \text{otherwise,} \end{cases}$$

while if $\theta = 1$, then

$$f(x;\theta) = \begin{cases} 1/(2\sqrt{x}), & \text{if } 0 < x < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Find the MLE of θ (Hint: $\mathbf{1}_{\{\text{argument}\}}$: when argument is true, it is 1; when argument is false, it is 0).

(A) $1 - \mathbf{1}_{\{1 \geq \prod_{i=1}^{n} 1/(2\sqrt{X_i})\}}$

(B) $1 - \mathbf{1}_{\{1 \geq \prod_{i=1}^{n} 1/(\sqrt{X_i})\}}$

(C) $1 - \mathbf{1}_{\{1 \geq \sum_{i=1}^{n} 1/(\sqrt{X_i})\}}$

(D) $1 - \mathbf{1}_{\{1 \geq \sum_{i=1}^{n} 1/(2\sqrt{X_i})\}}$

Answer:$1 - \mathbf{1}_{\{1 \geq \prod_{i=1}^{n} 1/(2\sqrt{X_i})\}}$

The likelihood function is

$$L(\theta = 0; x_1, \ldots, x_n) = 1, \quad 0 < x_i < 1$$

$$L(\theta = 1; x_1, \ldots, x_n) = \prod_{i=1}^{n} 1/(2\sqrt{x_i}), \quad 0 < x_i < 1$$

Thus, the MLE of θ is 0 if $1 \geq \prod_{i=1}^{n} 1/(2\sqrt{x_i})$ and the MLE is 1 if $1 < \prod_{i=1}^{n} 1/(2\sqrt{x_i})$. That is,

$$\hat{\theta} = 1 - \mathbf{1}_{\{1 \geq \prod_{i=1}^{n} 1/(2\sqrt{X_i})\}}.$$

15. Assume that X has the cumulative distribution function

$$F(x) = \begin{cases} 0 & \text{for } x < 1 \\ \frac{x^2 - 2x + 2}{2} & \text{for } 1 \leq x < 2 \\ 1 & \text{for } x \geq 2. \end{cases}$$

Calculate the expectation of X.

(A) $\frac{7}{6}$

(B) $\frac{7}{3}$

(C) $\frac{5}{6}$

(D) $\frac{4}{3}$

Answer: $\frac{4}{3}$

First note that the pdf for X is

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } x = 1 \\ x - 1 & \text{if } 1 < x < 2 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$E(X) = \frac{1}{2} + \int_{1}^{2} x(x-1)dx = \frac{1}{2} + \left(\frac{1}{3}x^3 - \frac{1}{2}x^2\right)\Big|_1^2 = \frac{4}{3}$$

16. (a) A sample $x_1, x_2, \cdots, x_{10}$ is drawn from a distribution with probability density function:

$$\frac{1}{2}\left[\frac{1}{\theta}\exp(-\frac{x}{\theta}) + \frac{1}{\sigma}\exp(-\frac{x}{\sigma})\right], \quad 0 < x < \infty$$

(b) $\theta > \sigma$

(c) $\sum x_i = 150$ and $\sum x_i^2 = 5000$

Estimate θ by matching the first two sample moments to the corresponding population quantities.

(A) 10

(B) 15

(C) 20

(D) 22

Answer: 20

One can infer that the first two sample moments are 15 and 500 and the first two population moments are calculated to be

$$E(X) = 0.5(\theta + \sigma), \quad E(X^2) = \theta^2 + \sigma^2.$$

Then by method of moments, let

$$\begin{cases} 0.5(\theta + \sigma) = 15 \\ \theta^2 + \sigma^2 = 500. \end{cases}$$

One can solve that $\theta = 10, \sigma = 20$ or $\theta = 20, \sigma = 10$, by condition (b) we know that $\hat{\theta} = 20$.

17. (a) X follows a shifted exponential distribution with probability density function:

$$f(x) = \frac{1}{\theta} e^{-(x-\delta)/\theta}, \quad \delta < x < \infty$$

(b) A random sample of claim amounts $X_1, X_2, \cdots, X_{10}$:

$$5 \quad 5 \quad 5 \quad 6 \quad 8 \quad 9 \quad 11 \quad 12 \quad 16 \quad 23$$

(c) $\sum X_i = 100$ and $\sum X_i^2 = 1306$

Estimate δ using the method of moments.

(A) 3.5
(B) 4.0
(C) 4.5
(D) 5.0

Answer: 4.5

$$EX = \int_\delta^\infty \frac{x}{\delta} e^{-(x-\delta)/\theta} dx = \int_0^\infty \frac{y+\delta}{\theta} e^{-y/\theta} dy = \theta + \delta$$

$$EX^2 = \int_\delta^\infty \frac{x^2}{\theta} e^{-(x-\delta)/\theta} dx = \int_0^\infty \frac{y^2 + 2y\delta + \delta^2}{\theta} e^{-y/\theta} dy = 2\theta^2 + 2\theta\delta + \delta^2.$$

By method of moments,

$$\begin{cases} \theta + \delta = 10 \\ 2\theta^2 + 2\theta\delta + \delta^2 = 130.6 \end{cases}$$

One can then solve that $\hat{\delta} = 4.468$ (the other possible solution is not reasonable as $\delta < x < \infty$).

18. You are given the following three observations:

$$0.74 \qquad 0.81 \qquad 0.95$$

You fit a distribution with the following density function to the data:

$$f(x) = (p+1)x^p, \quad 0 < x < 1, \quad p > -1.$$

Calculate the maximum likelihood estimate of p.

(A) 4.1

(B) 4.2

(C) 4.3

(D) 4.4

Answer: 4.3

The likelihood function is

$$L(p) = f(0.74)f(0.81)f(0.95) = (p+1)0.74^p(p+1)0.81^p(p+1)0.95^p = (p+1)^3(0.56943)^p.$$

The log-likelihood function is thus

$$l(p) = \ln L(p) = 3\ln(p+1) + p\ln(0.56943).$$

Let

$$l'(p) = \frac{3}{p+1} - 0.563119 = 0$$

and solve $\hat{p} = 4.32747$.

19. Let $X_1, X_2, \cdots, X_n$ be random samples from distribution with pdf

$$f(x;\theta) = \frac{\theta^4}{6}x^3e^{-\theta x}, \quad 0 < x < \infty, \quad 0 < \theta < \infty.$$

Find the maximum likelihood estimator $\hat{\theta}$ (Denote $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$).

(A) $4/\bar{X}$

(B) $2/\bar{X}$

(C) $1/\bar{X}$

(D) $\bar{X}$

Answer: $4/\bar{X}$

$$L(\theta; X_1, \dots, X_n) = \prod_{i=1}^n \frac{\theta^4}{6}X_i^3e^{-\theta X_i} = \frac{\theta^{4n}}{6^n}\left(\prod_{i=1}^n X_i^3\right)e^{-\theta\sum_{i=1}^n X_i}.$$

$$l(\theta; X_1, \dots, X_n) = -n\ln 6 + 4n\ln\theta + 3\sum_{i=1}^n \ln X_i - \theta\sum_{i=1}^n X_i.$$

$$\frac{\partial l}{\partial \theta} = 0 \Rightarrow \hat{\theta} = \frac{4}{\bar{X}}.$$

$$\left.\frac{\partial^2 l}{\partial \theta^2}\right|_{\theta=\hat{\theta}} = \frac{-4n}{\hat{\theta}^2} < 0.$$

Therefore, $\hat{\theta}_{\text{MLE}} = \dfrac{4}{\bar{X}}$.

20. (a) X follows an exponential distribution with mean θ .

(b) Y follows an exponential distribution with mean 2θ .

(c) Z follows an exponential distribution with mean 3θ.

(d) No samples from X are observed.

(e) Three samples from Y are observed, of values 1, 2 and 3.

(f) One sample from Z is observed, of value 15.

Calculate the maximum likelihood estimate of θ .

(A) 1

(B) 2

(C) 3

(D) 4

Answer: 2

The likelihood function is

$$\frac{e^{-1/(2\theta)}}{2\theta}\frac{e^{-2/(2\theta)}}{2\theta}\frac{e^{-3/(2\theta)}}{2\theta}\frac{e^{-15/(3\theta)}}{3\theta} = \frac{e^{-8/\theta}}{24\theta^4}$$

The log-likelihood function is

$$-\ln(24) - 4\ln(\theta) - 8/\theta$$

Differentiate it with respect to θ and let the result equal to 0, we get

$$-\frac{4}{\theta} + \frac{8}{\theta^2} = 0,$$

which means $\hat{\theta} = 2$.