Name: Shashank Shekhar
Student Id: 200262327

FDS - Assignment 2 (HW2)

Ans 1-

a) Mean, $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

Mode is defined as the number with highest frequency.

Variance (population) denoted by σ^2 is given by

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

Variance (sample) denoted by s^2 is given by

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

Since, $S.D = \sqrt{var}$

Population S.D denoted by $\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$

Sample S.D denoted by $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$

Yes, sample variance (s^2) is an unbiased estimator of population variance.

Proof :- To prove this we need to show that

$$E(s^2) = \sigma^2$$

Taking L.H.S

$$E(s^2) = E\left[\frac{1}{n-1}\sum(x_i-\bar{x})^2\right]$$

$$= E\left[\frac{1}{n-1}\sum(x_i^2+\bar{x}^2-2x_i\bar{x})\right]$$

$$= E\left[\frac{1}{n-1}\sum x_i^2+\bar{x}^2\sum 1-2\bar{x}\sum x_i\right]$$

$$= E\left[\left(\frac{1}{n-1}\right)\times\left(\sum x_i^2+n\bar{x}^2-2\bar{x}\sum x_i\right)\right]$$

Now, we know that $E(cX) = c\,E(X)$ and $E(\Sigma x_i) = \Sigma(E(x_i$

So,

$$E(s^2) = \left(\frac{1}{n-1}\right)E\left(\sum x_i^2+n\bar{x}^2-2\bar{x}\sum x_i\right)$$

Also, $\bar{x} = \frac{\sum x_i}{n}$

Hence, $\sum x_i = n\bar{x}$

Therefore, $E(s^2) = \frac{1}{n-1}E\left(\sum x_i^2+n\bar{x}^2-2n\bar{x}^2\right)$

$$= \frac{1}{n-1}E\left(\sum x_i^2-n\bar{x}^2\right)$$

$$E(s^2) = \frac{1}{n-1}\left[\sum E(x_i^2) - n\,E(\bar{x}^2)\right] \quad \text{—— } (0)$$

Also, we know that

$$\operatorname{var}(X) = E(X^2) - (E(X))^2 \quad \text{—— } (1)$$

and

$$\operatorname{var}(\bar{X}) = E(\bar{X}^2) - (E(\bar{X}))^2 \quad \text{—— } (2)$$

From ①,

$$E(x^2) = var(X) + (E(X))^2$$
$$E(x^2) = \sigma^2 + \mu^2 \quad \text{——— ③}$$

From ②,

$$E(\bar{x}^2) = var(\bar{x}) + (E(\bar{x}))^2$$

Also, $var(\bar{x}) = \frac{\sigma^2}{n}$

and on average, the sample mean = population mean

Therefore, $E(\bar{x}^2) = \frac{\sigma^2}{n} + \mu^2$ ——— ④

Replacing ③ and ④ in Equation ⓪

$$E(s^2) = \frac{1}{n-1}\left[\Sigma(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right]$$

$$= \frac{1}{(n-1)}\left[(\sigma^2+\mu^2)\Sigma 1 - \sigma^2 - n\mu^2\right]$$

$$= \left(\frac{1}{n-1}\right)\left[n\sigma^2 + \cancel{n\mu^2} - \sigma^2 - \cancel{n\mu^2}\right]$$

$$= \left(\frac{1}{\cancel{n-1}}\right)\cancel{(n-1)}\,\sigma^2$$

$$= \sigma^2$$

$= R.H.S$

Hence, LHS = RHS

& hence proved.

b) Central Limit Theoram states that given any population distribution, the sampling distribution of the sample means approaches a normal distribution as the sample size gets larger.

Assumptions:-

Assumption ①: In order for this approximation to be accurate, the sample size should be sufficiently large.

Assumption ②: The data must be sampled randomly and a good sampling technique must be used.

Assumption ③: The sample values drawn must be independent of each other. ie. occurance of an event should not influence another event.

Assumption ④: If the sample is drawn without replacement, the sample size should be no more than ~~the~~ 10% of sample population.

Ans 2-

a) Expected value of a discrete random variable is the measure of centre for the distribution of the random variable.

b) Tossing of a fair coin is a Binomial experiment whose expectation is given by

$$E(X) = np$$

where n is the no. of times the experiment was conducted and p is the probability of success.

So, $E(X) = 4 \times \frac{1}{2} = 2$

c)

$$E(X = S_n) = \sum_{i=0}^{n} x_i p_i$$

$$= \sum_{i=0}^{n} {}^nC_i \, p^i (1-p)^{n-i}$$

$$= \sum_{i=0}^{n} i \times \frac{n!}{(n-i)!\,i!} \, p^i (1-p)^{n-i}$$

$$= \sum_{i=0}^{n} \frac{n(n-1)!}{(n-i)!\,(i-1)!} \times p^i (1-p)^{n-i}$$

$$= \sum_{i=0}^{n} n \binom{n-1}{i-1} \times p^i (1-p)^{n-i}$$

Since $i = 0^{th}$ term will be zero, hence

$$= np \sum_{i=1}^{n} \binom{n-1}{p-1} p^{(i-1)} (1-p)^{n-i}$$

$$= np$$

$\left\{ \text{Since, } \sum_{i=1}^{n} \binom{n-1}{p-1} p^{i-1} (1-p)^{n-i} = [p + (1-p)]^{n-1} = 1 \right.$

d)

X_1	0	1
$P(X_1)$	$1/2$	$1/2$

X_2	0	1
$P(X_2)$	$1/2$	$1/2$

where X_1 and X_2 are the values obtained from tossing a coin twice, both of which can take values 0 & 1.

$$E(X_1 \cdot X_2) = E(X_1) \cdot E(X_2)$$

$$= \sum x_i p_i \times \sum x_i p_i$$

$$= \left[\left(0 \times \frac{1}{2}\right) + \left(1 \times \frac{1}{2}\right)\right] \times \left[\left(0 \times \frac{1}{2}\right) + \left(1 \times \frac{1}{2}\right)\right]$$

$$= 1/4$$

c) $P(X = S_n) = {}^nC_{S_n}\, p^{S_n} (1-p)^{n-S_n}$

Now, $E(X) = \sum_{i=0}^{n} x_i p_i$

$$= \sum_{i=0}^{n} x_i \, {}^nC_{S_n}\, p^{S_n} (1-p)^{n-S_n}$$

e) We know that

$$
\begin{aligned}
V(X) &= E\left((X - E(X))^2\right) \\
&= \sum (X - E(X))^2 p(x) \\
&= \sum \left(x^2 + E(X)^2 - 2x\,E(X)\right) \cdot p(x) \\
&= \left(\sum x^2 + E(X)^2 \sum 2x\,E(X)\right) \cdot p(X) \\
&= \sum x^2 p(x) + E(x)^2 \sum p(x) - 2E(x) \sum x p(x) \\
&= E(x^2) + E(x)^2 - 2E(x)^2 \\
&= E(x^2) - E(x)^2
\end{aligned}
$$

Hence, proved.

Ans 3 -

a)
$$E(x) = \int_{-\infty}^{\infty} x f(x)\, dx$$

Given $f_x(x) = \lambda e^{-\lambda x}$

Therefore,
$$E(X) = \int_0^{\infty} x \lambda e^{-\lambda x}\, dx$$

$$E(X) = \lambda \int_0^{\infty} x e^{-\lambda x}\, dx$$

Using integration by parts to solve the above

Assuming, $u = x$ & $v' = e^{-\lambda x}$

Then, $\int u v' = uv - \int u' v$

$$E(X) = \lambda \left\{ \left[\frac{x e^{-\lambda x}}{-\lambda} \right]_0^{\infty} - \int_0^{\infty} \frac{e^{-\lambda x}}{-\lambda}\, dx \right\}$$

$$E(x) = \int_0^{\infty} e^{-\lambda x}\, dx - \left[x e^{-\lambda x} \right]_0^{\infty}$$

$$E(x) = \left[\frac{e^{-\lambda x}}{-\lambda} \right]_0^{\infty} - \left[x e^{-\lambda x} \right]_0^{\infty}$$

$$E(X) = -\frac{1}{\lambda} \left(e^{-\infty} - e^{0} \right) - \left[x e^{-\lambda x} \right]_0^{\infty}$$

$$E(X) = \frac{1}{\lambda} - \left[x e^{-\lambda x} \right]_0^{\infty}$$

$$E(X) = \frac{1}{\lambda} - (0 - 0) = \frac{1}{\lambda}$$

$$E(X) = 1/\lambda$$

$$Var(X) = E(X - E(X))^2$$
$$= E(X - 1/\lambda)^2$$

$$Var(X) = \int_0^\infty \left(x - \frac{1}{\lambda}\right)^2 \lambda e^{-\lambda x} dx$$

$$= \int_0^\infty x^2 \lambda e^{-\lambda x} dx + \int_0^\infty \frac{1}{\lambda} e^{-\lambda x} dx - \int_0^\infty \frac{2x}{\lambda} \lambda e^{-\lambda x} dx$$

$$var(X) = \int_0^\infty x^2 \lambda e^{-\lambda x} dx - \frac{\left[e^{-\lambda x}\right]_0^\infty}{\lambda^2} - 2\int_0^\infty x e^{-\lambda x} dx$$

Now, integrating $\int_0^\infty x^2 \lambda e^{-\lambda x} dx$

Let $I' = \lambda \int_0^\infty x^2 e^{-\lambda x} dx$

$$= \lambda \left[-\frac{\left[x^2 e^{-\lambda x}\right]_0^\infty}{\lambda} + \frac{2}{\lambda} \int_0^\infty x e^{-\lambda x} dx \right]$$

$$= 2\int_0^\infty x e^{-\lambda x} dx - \left[x^2 e^{-\lambda x}\right]_0^\infty$$

$$= 2\int_0^\infty x e^{-\lambda x} dx - 0 = 2\int_0^\infty x e^{-\lambda x} dx \quad \text{——— (1)}$$

Replacing (1) in original eq^n.

Therefore, $var(X) = 2\int_0^\infty x e^{-\lambda x} dx - \frac{\left[e^{-\lambda x}\right]_0^\infty}{\lambda^2} - 2\int_0^\infty x e^{-\lambda x} dx$

$$= -\frac{\left[e^{-\lambda x}\right]_0^\infty}{\lambda^2}$$

$$= -\frac{(0-1)}{\lambda^2} = \frac{1}{\lambda^2}$$

b) Let X be the random variable denoting company profit and let y be the amount (premium) the company should charge to make profit.

Then,

X	y	-500
p(X)	1	0.01

Now, in order to make profit, expected value of profit (X) must be greater than or equal to zero.

$$E(X) = \Sigma x_i p_i$$
$$0 = 0.01 \times (-500) + (y \times 1)$$
$$0 = -5 + y$$
$$y = 5$$

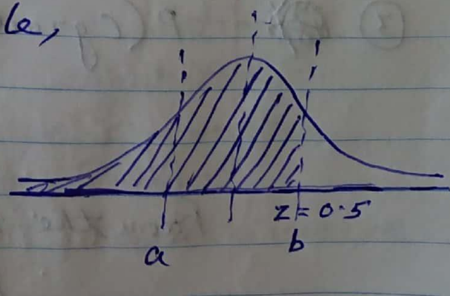
Hence, the premium to be charged is \$5.

Ans 4- a) Converting from normal to standard normal distribution

$$Z = \frac{X-\mu}{\sigma}$$ Given, $\mu = 10$ & $\sigma = 2$

$$Z = \frac{11-10}{2} = \frac{1}{2} = 0.5$$

Calculating the value using z-table, we get $P(X<b) = 0.6915$



So, $P(a<X<b) = (0.6915-0.5) \times 2$
$= 0.1915 \times 2$
$= 0.3830$

So, the probability that the pen drive will fail between 9 and 11 yrs is 0.3830

b) i) Converting from normal to standard normal distribution.

$Z = \frac{X-\mu}{\sigma}$

① $P(A\ grade) = P(X \geq 90)$

$\Rightarrow Z = \frac{90-75}{15} = 1$

Finding the value from table, we get

$P(A\ grade) = (1-0.8413) = 0.1587$

② $P(B\ grade) = P(80 \leq X \leq 90)$

$\Rightarrow Z = \frac{80-75}{15} = \frac{1}{3} = 0.33$

Finding value from the table,

$P(B\ grade) = (0.8413 - 0.6293)$
$= 0.2120$

③ $P(C\ grade) = P(70 \leq X < 80)$

$\Rightarrow Z = \frac{70-75}{15} = \frac{-1}{3} = -0.33$

From the z-table,

$P(C\ grade) = 0.6293 - 0.3707$
$= 0.2586$

④ $P(D \text{ grade}) = P(60 \leq X < 70)$

$$\Rightarrow z = \frac{60-75}{15} = -1$$

Using z-table

$$P(D \text{ grade}) = 0.3707 - 0.1587$$
$$= 0.2120$$

⑤ $P(F \text{ grade}) = P(X < 60) = 0.1587$ (as calculated above)

ii) Expected Proportion is same as probability calculated in previous part.

Ans 5-

a) MLE procedure for single parameter:

① Write down the likelihood function, $L(\theta)$

② In order to easily find the parameter estimate, we use the loglikelihood since logarithmic function is a monotonically increasing function.

③ Differentiate $\ln(L(\theta))$ with respect to θ.

④ In order to find the maxima, set the derivative to 0 and solve for θ.

⑤ Check for maximum by taking the 2nd derivative and finding the sign of the resulting value. Negative sign of the 2nd derivative indicates a maxima

b) $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} P(X_i;\theta)$

$$L(\theta) = \prod_{i=1}^{n} (\theta-1)\, x^{-\theta}$$

Finding Log Likelihood,

$$\ln L(\theta) = \ln \prod_{i=1}^{n} (\theta-1)\, x_i^{-\theta}$$

$$\ln L(\theta) = \sum_{i=1}^{n} \ln\left[(\theta-1)\, x_i^{-\theta}\right]$$

We know, $\ln(a \cdot b) = \ln a + \ln b$

Therefore, $\ln L(\theta) = \sum_{i=1}^{n} \left[\ln(\theta-1) + (-\theta)\ln x_i\right]$

$$= \ln(\theta-1)\sum_{i=1}^{n} 1 - \theta \sum_{i=1}^{n} \ln x_i$$

$$\ln L(\theta) = n\ln(\theta-1) - \theta \sum_{i=1}^{n} \ln x_i$$

Now, taking derivative and equating to zero to find the maximum log likelihood:

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left[n\ln(\theta-1) - \theta \sum_{i=1}^{n} \ln x_i\right]$$

$$0 = \left(\frac{n}{\theta-1}\right) - \sum_{i=1}^{n} \ln x_i$$

$$\Rightarrow \frac{n}{\theta - 1} = \sum_{i=1}^{n} \ln x_i$$

$$\theta - 1 = \frac{n}{\sum_{i=1}^{n} \ln x_i}$$

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \ln x_i} + 1 \Rightarrow \theta = \frac{1}{\ln x} + 1$$

Confirming that this value is the required maxima, finding the 2^{nd} derivative to check the sign.

We have

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{n}{\theta - 1} + C \quad , \text{where } C = \text{constant}$$

$$\frac{\partial^2 \ln L(\theta)}{\partial \theta} = -n(\theta - 1)^{-2}$$

Since, $\theta > 2$, we have $-n(\theta-1)^{-2} < 0$.

Hence,

$$\left(\frac{n}{\sum_{i=1}^{n} \ln x_i} + 1 \right)$$ is the required MLE for θ

Ans 6-

a) Confidence Interval is the range of values we are sure that our true value lies in. This is given with a fixed confidence level. (Say 95%)

b) Procedure for finding C.I

1. Find out the sample statistic for which the population confidence interval is to be calculated.
2. Select the required confidence level.
3. Compute the margin of error given by

$$\varepsilon = z^* \frac{\sigma}{\sqrt{n}}$$

where σ is the population standard deviation, n is sample size and z^* is the appropriate z^*-value from the standard normal distribution for the desired confidence level.

4. Finally,

$$CI = \text{sample statistic} \pm \varepsilon$$

c) We need to compute 97% CI for the point estimate of mean given that the sample size is 40.

Following above steps:-

i) Required Sample statistic for which CI needs to be calculated is mean.

ii) CI given is 97%

iii)

$$\varepsilon = z^* \frac{\sigma}{\sqrt{n}}$$

Finding the z^* value for (97% + 1.5%) from the table

$\Rightarrow z^* = 2.17$

Now, finding the value of σ

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

$$\sigma = \sqrt{\frac{1}{40}\sum_{i=1}^{40}(x_i - 41.925)}$$

$$\sigma = 14.93$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\bar{x} = \frac{1677}{40}$$

$$= 41.925$$

Now, $E = 2.17 \times \frac{14.93}{\sqrt{40}}$

$$= 5.12$$

Therefore, $CI = \bar{x} \pm E$

$$= 41.925 \pm 5.12$$

$$CI\ (36.805,\ 47.045)$$

Ans 7 - Hypothesis Testing Steps :-

i) State null (H_0) and Alternate (H_1) hypothesis.
ii) Choose level of significance (α)
iii) Find critical values.
iv) Find test statistic.
v) Draw conclusion.

Following these steps:-

i) $H_0 : \mu = 80$
$H_1 : \mu < 80$

ii) & iii) Given $\alpha = 0.1$, this being a left tailed test, finding the z value from table,

$$z = -1.28$$

Hence, the critical value is -1.28

iv) Now, calculating the test statistic,

$$\mu = \bar{X} \pm E$$
$$\mu = \bar{X} \pm z \frac{\sigma}{\sqrt{n}}$$

$$z = \frac{(\bar{X} - \mu)}{\sigma / \sqrt{n}} \qquad \left\{ \bar{x} = 75 \right\}$$

$$= \frac{75 - 80}{19.2/\sqrt{36}} = \frac{-5}{19.2} \times 6 = \frac{-300}{192}$$
$$= -1.56$$

v) Since the calulated ~~sample~~ test statistic is $-1.56 < -1.28$ (critical value), we can reject the null hypothesis and accept the alternate hypothesis ie. the average cost of an engineering book is less than \$80.