Naman Jain Vimal Kumar
1001670153

Question 4.24

Installation of some software package requires downloading 82 files. On the average, it takes 15 sec to download one file, with a variance of 16 sec2. What is the probability that the software is installed in less than 20 minutes?

Solution

Given:

$$n = 82$$
$$\mu = 15\ sec$$
$$\sigma^2 = 16\ sec^2$$

To find:

$$P(S_n < 20mins) = P(S_n \leq 1200\ sec)$$ [As it is continuous]

We can apply Central Limit Theorem for the same:

$$P(S_n \leq z) = P\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right\} = \emptyset(z)$$

Therefore,

$$P(S_n \leq 1200) = P\left\{Z_n \leq \frac{1200 - (82 * 15)}{\sqrt{16 * 82}}\right\}$$
$$P(S_n \leq 1200) = P\left\{Z \leq \frac{1200\ - 1230}{\sqrt{1312}}\right\}$$
$$P(S_n \leq 1200) = P\left\{Z \leq \frac{-30}{36.2215}\right\}$$
$$P(S_n \leq 1200) = P(Z \leq -0.8282)$$
$$P(S_n \leq 1200) = \emptyset(-0.828) \approx \emptyset(-0.83)$$
$$P(S_n \leq 1200) = 0.2033$$

Question 4.28

Seventy independent messages are sent from an electronic transmission center. Messages are processed sequentially, one after another. Transmission time of each message is Exponential with parameter $\lambda = 5\ min^{-1}$. Find the probability that all 70 messages are transmitted in less than 12 minutes. Use the Central Limit Theorem.

Solution

Given:

$$n = 82$$
$$\lambda = 5\ min^{-1}$$ [Exponential Distribution]

Therefore,

$$\mu = \frac{1}{\lambda} = \frac{1}{5} = 0.2\ min$$
$$\sigma^2 = \frac{1}{\lambda^2} = \frac{1}{25} = 0.04\ min^2$$

We can now apply Central Limit Theorem for the same:

$$P(S_n \leq z) = P\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right\} = \emptyset(z)$$

Thus,

$$P(S_n \leq 12) = P\left\{Z_n \leq \frac{12 - (70 * 0.2)}{0.2 * \sqrt{70}}\right\}$$
$$P(S_n \leq 1200) = P\left\{Z \leq \frac{12 - 14}{0.2 * 0.3667}\right\}$$
$$P(S_n \leq 1200) = P\left\{Z \leq \frac{-2}{1.6733}\right\}$$
$$P(S_n \leq 1200) = P(Z \leq -1.1952)$$
$$P(S_n \leq 1200) = \emptyset(-1.1952) \approx \emptyset(-1.20)$$
$$P(S_n \leq 1200) = 0.1151$$

Question 8.9

The following data set represents the number of new computer accounts registered during ten consecutive days.

43, 37, 50, 51, 58, 105, 52, 45, 45, 10.

(a) Compute the mean, median, quartiles, and standard deviation.
(b) Check for outliers using the 1.5(IQR) rule.
(c) Delete the detected outliers and compute the mean, median, quartiles, and standard deviation again.
(d) Make a conclusion about the effect of outliers on basic descriptive statistics.

Solution

Given:

n = 10

Sorting the current list for better understanding.

10, 37, 43, 45, 45, 50, 51, 52, 58, 105

[Note: Calculations for Mean, Variance and Standard Deviation are in Part A of hw2_problem_8.9.xlsx]

a. *Finding mean*: We know that $\bar{X} = \sum_i^n X_i/n$

$$\bar{X} = \frac{(10 + 37 + 43 + 45 + 45 + 50 + 51 + 52 + 58 + 105)}{10}$$
$\bar{X}$ = 496/10 = 49.6

Finding median:
Since 'n' is even, we know that the median is the number between $(n/2)^{th}$ and $(n+2/2)^{th}$ elements.

$$\hat{M} = (S[5] \leq x \leq S[6])$$
$$\hat{M} = (45 \leq x \leq 50)$$
$$\hat{M} = 47.5$$

Finding quartiles:
Q1 – For p = 0.25, that is 25% of sample space would be 2.5^{th} element

$$\widehat{Q1} = (S[2] \leq x \leq S[3])$$
$$\widehat{Q1} = (37 \leq x \leq 43)$$
$$\widehat{Q1} = 40.0$$

Q2 – We know that second quartile is equal to the median. Therefore,
$\widehat{Q2} = 47.5$

Q3 – For p = 0.75, that is 75% of sample space would be 7.5th element
$\widehat{Q3} = (S[7] \leq x \leq S[8])$
$\widehat{Q3} = (51 \leq x \leq 52)$
$\widehat{Q3} = 51.5$

Finding Standard Deviations:
We know that $s = \sqrt{s^2}$ and
$s^2 = \frac{1}{n-1}\sum_i^n (X_i - \bar{X})^2$

$$s^2 = \frac{4960.4}{9} = 551.1556$$
$$s = 23.4767$$

b. Checking for Outliers:
As we know, the dataset should lie in $\left[\widehat{Q_1} - 1.5\left(\widehat{IRQ}\right), \widehat{Q_3} + 1.5\left(\widehat{IRQ}\right)\right]$.

Computing, $\widehat{IRQ} = \widehat{Q_3} - \widehat{Q_1} = 51.5 - 40.0 = 11.5$.

Therefore, the dataset would lie between (40.0 – 11.5) to (51.5 + 11.5). That is [28.5, 63.0].

With respect to 1.5(IQR) rule, we get 2 outliers when X = [10, 105].

c. Removing the outliers and computing operations again:

[Note: Calculations for Mean, Variance and Standard Deviation are in Part B of hw2_problem_8.9.xlsx]

Thus, n = 8
And Series is 37, 43, 45, 45, 50, 51, 52, 58

Finding mean:

$$\bar{X} = \frac{(37 + 43 + 45 + 45 + 50 + 51 + 52 + 58)}{8}$$
$$\bar{X} = \frac{381}{8} = 47.625$$

Finding median:
Since 'n' is even, we know that the median is the number between (n/2)th and (n+2/2)th elements.
$\widehat{M} = (S[4] \leq x \leq S[5])$
$\widehat{M} = (45 \leq x \leq 50)$
$\widehat{M} = 47.5$

Finding quartiles:
Q1 – For p = 0.25, that is 25% of sample space would be 2[th] element

$$\widehat{Q1} = (S[2] \leq x \leq S[3])$$
$$\widehat{Q1} = (43 \leq x \leq 45)$$
$$\widehat{Q1} = 44.0$$

Q2 – We know that second quartile is equal to the median. Therefore,

$$\widehat{Q2} = 47.5$$

Q3 – For p = 0.75, that is 75% of sample space would be 6[th] element

$$\widehat{Q3} = (S[6] \leq x \leq S[7])$$
$$\widehat{Q3} = (51 \leq x \leq 52)$$
$$\widehat{Q3} = 51.5$$

Finding Standard Deviations:

We know that $s = \sqrt{s^2}$ and

$$s^2 = \frac{1}{n-1}\sum_i^n (X_i - \bar{X})^2$$

$$s^2 = \frac{291.875}{7} = 41.6964$$
$$s = 6.4573$$

d. Effect of outliers:
- Though there is very small difference in the mean, we can see a huge difference with respect to standard deviation and variance.
- With outliers, it could spread out the gaussian graph, which could result to inaccurate results.

Question 9.4

A sample of 3 observations (X1 = 0.4, X2 = 0.7, X3 = 0.9) is collected from a continuous distribution with density

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

Estimate θ by your favorite method.

Solution by method of moments:

We know that $\mu_1 = E(X) = \int_{-\infty}^{\infty} x * f(x)\, dx$

Therefore,

$$\mu_1 = \int_{-\infty}^{0} x * f(x)\, dx + \int_{0}^{1} x * f(x)\, dx + \int_{1}^{\infty} x * f(x)\, dx$$
$$\mu_1 = 0 + \int_{0}^{1} x * \theta x^{\theta-1}\, dx + 0$$
$$\mu_1 = \theta \int_{0}^{1} x^{\theta}\, dx$$
$$\mu_1 = \left.\frac{\theta x^{\theta+1}}{\theta+1}\right]_{x=0}^{x=1}$$

$$\mu_1 = \frac{\theta}{\theta + 1} \quad \ldots 1$$

From the observations,

$$m_1 = \bar{X} = \frac{\sum_i^n X_i}{n}$$

Given:

n = 3 and X = [0.4, 0.7, 0.9]

Therefore:

$$m_1 = \frac{0.4 + 0.7 + 0.9}{3} = \frac{2}{3} = 0.6667 \quad \ldots 2$$

Equating 1 and 2:

$\frac{\theta}{\theta+1} = \frac{2}{3} => \hat{\theta} = 2$ [As it is a normalized vector]

Solution by maximum likelihood equation:

Given:

$$f(X_1, X_2, X_3, \ldots., X_n) = \prod_{i=1}^{n} \theta x_i^{\theta-1}$$

Taking logarithmic on both the sides:

$$\ln f(X_1, X_2, X_3, \ldots., X_n) = \ln \prod_{i=1}^{n} \theta x_i^{\theta-1}$$

$$\ln f(X_1, X_2, X_3, \ldots., X_n) = \sum_{i=1}^{n} \ln \theta + \sum_{i=1}^{n} \ln {x_i}^{\theta-1}$$

$$\ln f(X_1, X_2, X_3, \ldots., X_n) = (n * \ln \theta) + \left((\theta - 1) * \sum_{i=1}^{n} \ln x_i \right)$$

Taking derivative with respect to θ

$$\frac{d \ln f(X_1, X_2, X_3, \ldots., X_n)}{d\theta} = \frac{d}{d\theta}(n * \ln \theta) + \left((\theta - 1) * \sum_{i=1}^{n} \ln x_i \right)$$

$$0 = \frac{n}{\theta} + \sum_{i=1}^{n} \ln x_i$$

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

Substituting the given sample,

$$\text{Ln}(0.4) + \ln(0.7) + \ln(0.9) = -\frac{3}{\theta}$$

$$\theta = \frac{-3}{-(0.9163 + 0.3567 + 0.1054)}$$

$$\theta = \frac{3}{1.3784}$$

$$\hat{\theta} = 2.1766$$

Question (5)

The following are samples obtained from a population that follows a normal distribution

69, 47, 175, 70, 53, 64, 74, 52, 58, 45, 67, 44, 58, 64, 49,
70, 65, 70, 48, 16, 67, 55, 42, 72, 61, 65, 77, 70, 60, 39

1. *Find the Sample mean, variance and standard deviation*
2. *Use them to estimate the parameters of the normal distribution (μ and σ)*
3. *Find and eliminate any possible outliers.*
    a. *Recalculate the sample mean, variance and standard deviation.*
    b. *Use the new values to recalculate the parameters of normal distribution.*
    c. *If the population followed Normal (60, 10), did eliminating outliers improve the accuracy of your estimate?*

Solution

Given:

n = 30

Note: All below parameters are solved in an excel named as 'hw2_problem5.xlsx' which is attached with the same.

1. Finding the Sample mean, variance and standard deviation.

$$\bar{X} = \frac{\sum_i^n X_i}{n}, \quad s^2 = \frac{\sum_i^n (X_i - \bar{X})^2}{n-1}$$

Therefore,

$\bar{X} = 62.2$
$s^2 = 625.2690$
$s = 25.0054$

2. Estimating the parameters of the normal distribution (μ and σ)

$\mu = \bar{X} = 62.2$
$\sigma = s = 25.0054$

3. Find and eliminate any possible outliers.

Sorted List:

16, 39, 42, 44, 45, 47, 48, 49, 52, 53,
55, 58, 58, 60, 61, 64, 64, 65, 65, 67,
67, 69, 70, 70, 70, 70, 72, 74, 77, 175

Finding quartiles:
Q1 – For p = 0.25, that is 25% of sample space would be 7.5th element

$\widehat{Q1} = (S[7] \leq x \leq S[8])$
$\widehat{Q1} = (48 \leq x \leq 49)$
$\widehat{Q1} = 48.5$

Q2 – For p = 0.5, that is 50% of sample space would be 15th element

$\widehat{Q2} = (S[15] \leq x \leq S[16])$
$\widehat{Q2} = (61 \leq x \leq 64)$
$\widehat{Q2} = 62.5$

Q3 – For p = 0.75, that is 75% of sample space would be 22.5^{th} element
$\widehat{Q3} = (S[22] \leq x \leq S[23])$
$\widehat{Q3} = (69 \leq x \leq 70)$
$\widehat{Q3} = 69.5$

As we know, the dataset should lie in $\left[\widehat{Q_1} - 1.5\left(\widehat{IRQ}\right), \widehat{Q_3} + 1.5\left(\widehat{IRQ}\right)\right]$.

Computing, $\widehat{IRQ} = \widehat{Q_3} - \widehat{Q_1} = 69.5 - 48.5 = 11.$

Therefore, the dataset would lie between (48.5 – 11) to (69.5 + 11). That is [37.5, 80.5].

With respect to 1.5(IQR) rule, we get 2 outliers when X = [16, 175].
And also with normalization, we get the same outliers.

a. Recalculate the sample mean, variance and standard deviation.
$\bar{X} = 59.8214$
$s^2 = 115.4114$
$s = 10.7430$

b. Use the new values to recalculate the parameters of normal distribution.

$\mu = \bar{X} = 59.8214$
$\sigma = s = 10.7430$

c. If the population followed Normal (60, 10), did eliminating outliers improve the accuracy of your estimate?

Yes, removing the outliers improves the accuracy of our estimate. We can see a huge difference with respect to the standard deviation, that makes our graph more specific to the current data eliminating the outliers. It is followed by Normal (59.82, 10.74)