innovate achieve lead

MATH F113
Probability and Statistics

Dr. Rakhee
Department of Mathematics

BITS Pilani
Pilani Campus

Section 5.3-5.5

Tutorial Sheet 2
Question 1

Let X denote the gasoline mileage obtained in tests on a newly designed SUV (sport utility vehicle). A sample of size 22 obtained are as

5	16	17	18	17	20	16	
17	18	20	18	18	19	19	17
21	17	19	18	17	42	50	

a. Find the sample mean, sample median, sample variance and sample slandered deviation for these data.

b. There are outliers in the data, to see the effect of outliers, drop them from data set and calculate sample mean, sample median, sample variance and sample slandered deviation.

Solution 1

innovate achieve lead

a. Sample mean $\bar{X} = \frac{\sum_{i=1}^{22} X_i}{n} = \frac{439}{22} = 19.9545$

Sample Median = $\tilde{X}$ = 18

Sample Variance $S^2 = \frac{n\sum_{i=1}^{22} x_i^2 - (\sum_{i=1}^{22} x_i)^2}{n-1}$ = 81.85

Sample slandered deviation = S = 9.0473

Solution 1

5	16	17	18	17	20	16	
17	18	20	18	18	19	19	17
17	21	17	19	18	17	42	50

b. Outliers are 5, 42, 50, after dropping the outliers

Sample mean $\bar{X} = \frac{\sum_{i=1}^{19} X_i}{n} = \frac{342}{19} = 18$

Sample Median = $\tilde{X}$ = 18

Sample Variance $S^2 = \frac{19\sum_{i=1}^{19} x_i^2 - \left(\sum_{i=1}^{19} x_i\right)^2}{19-1} = 1.8889$

Sample slandered deviation = S = 1.3744

Question 2

A manufacturing company produces electric insulators. If the insulators break when in use, a short circuit is likely. To test the strength of the insulators, you carry out destructive testing to determine how much force is required to break the insulators. You measure force by observing how many pounds are applied to the insulator before it breaks. The following table lists 30 values from this experiment.

1,870	1,728	1,656	1,610	1,634	1,784	1,522	1,696	1,592	1,662
1,866	1,764	1,734	1,662	1,734	1,774	1,550	1,756	1,762	1,866
1,820	1,744	1,788	1,688	1,810	1,752	1,680	1,810	1,652	1,736

a. Find the sample mean, sample standard deviation, and sample range for the above data.

b. Assuming that the distribution is normal, can you provide an estimate of σ by using the sample range calculated in part (a)?

c. If the distribution is not normal, can you provide an estimate of σ by using the sample range calculated in part (a)?

Solution 2

Sample mean = 1723.4

Sample standard deviation = 89.55 (use calculator)

Sample range = 1870 – 1522 = 348

(Approximation of σ via range) :

If X is normal, $P[-2\sigma < X-\mu < 2\sigma] = 0.95$ (close to 1)

So we can assume 4σ to be the whole range, or

Approximately, estimate of $\hat{\sigma}$ = (estimated range)/4 = 87

If X is not normal, by Chebyshev's inequality :

$$P[-3\sigma < X-\mu < 3\sigma] \geq 0.89,$$

So we can approximately take estimate of $\hat{\sigma}$

= (estimated range)/6 = 58

Question 3

Let X be the number of packages being mailed by a randomly selected customer at a certain shipping facility. Suppose the distribution of X is as follows:

x	1	2	3	4
$p(x)$	0.4	0.3	0.2	0.1

a. Consider a random sample of size $n = 2$ (two customers), and let $\bar{X}$ be the sample mean number of packages shipped. Obtain the probability distribution of $\bar{X}$.

b. Refer to part (a) and calculate $P(\bar{X} \leq 2.5)$.

Question 3

c. Again consider a random sample of size $n = 2$, but now focus on the statistic R = the sample range (difference between the largest and smallest values in the sample). Obtain the distribution of R. [*Hint:* Calculate the value of R for each outcome and use the probabilities from part (a).]

d. If a random sample of size $n = 4$ is selected, what is $P(\bar{X} \leq 1.5)$?

(*Hint:* You should not have to list all possible outcomes, only those for which $\bar{x} \leq 1.5$.)

Solution 3

a.

Outcome	1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4
Probability	0.16	0.12	0.08	0.04	0.12	0.09	0.06	0.03
$\bar{x}$	1	1.5	2	2.5	1.5	2	2.5	3
r	0	1	2	3	1	0	1	2

Outcome	3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4
Probability	0.08	0.06	0.04	0.02	0.04	0.03	0.02	0.01
$\bar{x}$	2	2.5	3	3.5	2.5	3	3.5	4
r	2	1	0	1	3	2	1	0

Probability Distribution of $\bar{X}$

$\bar{x}$	1	1.5	2	2.5	3	3.5	4
$p(\bar{x})$	0.16	0.24	0.25	0.20	0.10	0.04	0.01

b. $$P(\bar{x} \le 2.5) = p(1) + p(1.5) + p(2) + p(2.5)$$
$$= 0.16 + 0.24 + 0.25 + 0.20 = 0.85$$

Solution 3

c. random sample of size $n = 2$,

statistic R = the sample range

r	0	1	2	3
$p(r)$	0.30	0.40	0.22	0.08

d. With $n = 4$, there are numerous ways to get a sample average of at most 1.5, since $\bar{X} \le 1.5$ iff the sum of the X_i is at most 6. Listing out all options, $P(\bar{X} \le 1.5) = P(1,1,1,1) + P(2,1,1,1) + \ldots + P(1,1,1,2) + P(1,1,2,2) + \ldots + P(2,2,1,1) + P(3,1,1,1) + \ldots + P(1,1,1,3)$
$= (.4)^4 + 4(.4)^3(.3) + 6(.4)^2(.3)^2 + 4(.4)^2(.2)^2 = .2400.$

Question 4

Suppose your waiting time for a bus in the morning is uniformly distributed on [0,5], whereas waiting time in the evening is uniformly distributed on [0,10] independent of morning waiting time.

a. If you take the bus each morning and evening for a week, what is your total expected waiting time? [*Hint:* Define random variables $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_{10}$ and use a rule of expected value.]

b. What is the variance of your total waiting time?

c. What are the expected value and variance of the difference between morning and evening waiting times on a given day?

d. What are the expected value and variance of the difference between morning waiting time and total evening waiting time for a particular week?

Solution 4

Let $X_1, X_2, \ldots, X_5$ denote morning times and $X_6, \ldots, X_{10}$ denote evening times.

a. $E(X_1, X_2, \ldots, X_{10}) = E(X_1) + \ldots + E(X_{10}) = 5E(X_1) + 5E(X_6) = 5(2.5) + 5(5) = 37.5$

b. $Var(X_1, X_2, \ldots, X_{10}) = Var(X_1) + \ldots + Var(X_{10}) = 5Var(X_1) + 5Var(X_6)$

$$= 5\left[\frac{25}{12} + \frac{100}{12}\right] = \frac{625}{12} = 52.083$$

c. $E(X_1 - X_6) = E(X_1) - E(X_6) = 2.5 - 5 = -2.5$

$Var(X_1 - X_6) = Var(X_1) + (-1)^2 Var(X_6) = \frac{25}{12} + \frac{100}{12} = \frac{125}{12} = 10.417$

d. $E(X_1 + \ldots + X_5) - E(X_6 + \ldots + X_{10}) = 5(2.5) - 5(5) = -12.5$

$Var[(X_1 + \ldots + X_5) - (X_6 + \ldots + X_{10})] = Var(X_1 + \ldots + X_5) + (-1)^2 Var(X_6 + \ldots + X_{10})$

$$= 52.083$$

Question 5

The daily amount of coffee, in litres, dispensed by a coffee vending machine located at New Delhi airport lobby is equally likely to lie between 7 to 10 litres. If daily amount of coffee from this machine is collected for a random sample of 100 days, what would be the distribution of the sample mean?

Solution:

As we have a random sample of size 100 (sufficiently large n) from uniform dist.,

we can use CLT to find the distribution of $\bar{X}$, that is, $\bar{X} \sim N\left(E(X), \dfrac{\sigma(X)}{\sqrt{100}}\right)$

$\Rightarrow \bar{X} \sim N(8.5, 0.087)$

Question 6

The breaking strength of a rivet has a mean value of 10,000 psi and a standard deviation of 500 psi.

a. What is the probability that the sample mean breaking strength for a random sample of 40 rivets is between 9950 and 10,250?

b. If the sample size had been 15 rather than 40, could the probability requested in part (a) be calculated from the given information?

Solution 6

a. $\mu = 10000\ psi,\ \sigma = 500\ psi\ and\ n = 40$, using C.L.T.

$$P(9{,}950 \le \bar{X} \le 10{,}250) \approx P\left(\frac{9{,}950 - 10{,}000}{500/\sqrt{40}} \le Z \le \frac{10{,}250 - 10{,}000}{500/\sqrt{40}}\right)$$

$$= P(-.63 \le Z \le 3.16) = \Phi(3.16) - \Phi(-.63)$$

$$= 0.9992 - 0.2643 = 0.7349$$

b. According to the Rule of Thumb given in your text, n should be greater than 30 in order to apply the C.L.T., thus using the same procedure for $n = 15$ as was used for $n = 40$ would not appropriate.

Question 7 (Solution: trivial)

a. Let X_1 and X_2 be independent Poisson random variable with parameters λ_1 and λ_2 respectively.

Let $Y = X_1 + X_2$. What is the distribution of Y?

b. Let $X_1 \ldots X_n$ be a random sample from an exponential distribution with parameter λ.

Let $Y = X_1 + \ldots + X_n$. What is the distribution of Y?

Solution 7

a. As X_1 and X_2 are Poisson distributed

$$m_{X_1}(t) = e^{\lambda_1(e^t-1)} \quad and \; m_{X_2}(t) = e^{\lambda_2(e^t-1)}$$

$$Y = X_1 + X_2$$

$$m_Y(t) = \prod_{i=1}^{2} m_{X_i}(t) = \prod_{i=1}^{2} e^{\lambda_i(e^t-1)}$$

$$= e^{(\lambda_1+\lambda_2)(e^t-1)}$$

Thus, Y is Poisson r.v. with parameter

$$\lambda = \lambda_1 + \lambda_2$$

Solution 7

b. As X_1 ..., X_n are exponential r.v.

$$\mathrm{m}_{X_i}(\mathrm{t}) = (1 - \lambda t)^{-1} \quad i = 1, \dots, n$$

$$Y = X_1 + \dots + X_n$$

$$\mathrm{m}_Y(\mathrm{t}) = \prod_{i=1}^{n} m_{X_i}(t) = \prod_{i=1}^{n} (1 - \lambda t)^{-1}$$

$$= (1 - \lambda\, t)^{-n}$$

We know that mgf for gamma random variable with parameter α and β is

$$\mathrm{m}_X(\mathrm{t}) = (1 - \beta\, t)^{-\alpha}$$

Thus, Y is gamma r.v. with parameter n and λ.