Student Information

Name : Yarkın Özcan

ID : 2580835

Answer 1

a)

Since X and Y are both continuous variables, the total area under the joint probability density function should be 1. Using this fact we can find k.

$$\int_y \int_x f_{X,Y}(x,y)\,dx\,dy = 1$$
$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$$
$$\int_0^1 \int_0^1 f_{X,Y}(x,y)\,dx\,dy = 1$$
$$\int_0^1 \int_0^1 x + ky^3\,dx\,dy$$
$$\int_0^1 [\frac{x^2}{2} + kxy^3]_0^1\,dy = 1$$
$$\int_0^1 \frac{1}{2} + ky^3\,dy = 1$$
$$[\frac{y}{2} + \frac{ky^4}{4}]_0^1 = 1$$
$$\frac{1}{2} + \frac{k}{4} = 1$$
$$k = 2$$

b)

Since X and Y are both continuous variables, the probability of (X = $\frac{1}{2}$) is 0 and we can not talk about probability mass function of a continuous variable.

c)

We can use the joint probability density function $f_{X,Y}(x, y)$ in order to find

$$P\{0 \leq X \leq 1, 0 \leq Y \leq 1\}$$

$$\int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} x + 2y^3 \, dx \, dy$$

$$\int_0^{\frac{1}{2}} [\frac{x^2}{2} + 2xy^3]_0^{\frac{1}{2}} \, dy$$

$$\int_0^{\frac{1}{2}} \frac{1}{8} + y^3 \, dy$$

$$[\frac{y}{8} + \frac{y^4}{4}]_0^{\frac{1}{2}}$$

$$\frac{1}{16} + \frac{1}{64}$$

$$\frac{5}{64}$$

Answer 2

a)

In order to find the marginal PDF of Y, we need to calculate $\int_X f_{X,Y}(x, y) \, dx$

$$\int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx$$

$$\int_0^{\infty} \frac{e^{-y-\frac{x}{y}}}{y} \, dx$$

$$\lim_{m \to \infty} [-e^{-y-\frac{x}{y}}]_0^m$$

$$0 - (-e^{-y}) = e^{-y}$$

So $f_y(y) = e^{-y}$. From this equation we can infer that Y follows exponential distribution family because the general exponential PDF formula is $\lambda e^{-\lambda t}$ where λ is the average number of events in a time unit. Since the PDF of Y is e^{-y}, we can say that Y is a exponential distribution random variable with $\lambda = 1$.

b)

From part a we found that Y is a exponential distribution random variable with $\lambda = 1$. And we also know that expected value of exponential distribution variable is $\frac{1}{\lambda}$ which is $\frac{1}{1} = 1$ in our case.

Answer 3

a)

If we want at least %9 of the soldiers belong to the naval forces, then we want at least 90 soldiers out of 1000 to belong to the naval forces. At this points, we can use Normal approximation to Binomial distribution since n is large enough. The number X of soldiers from naval forces has Binomial Distribution with n = 1000, $p = 0.1$, $\mu =$ n*p = 100 and $\sigma = \sqrt{np(1-p)} = 9.49$. Applying the Central Limit Theorem with the continuity correction, we get;

$$P\{X \geq 90\} = 1 - P\{X < 90\} = 1 - P\{X < 89.5\}$$

$$1 - P\{\frac{X - 100}{9.49} < \frac{89.5 - 100}{9.49}\} = 1 - P\{Z < -1.106\}$$

Where Z is the Standart Normal Distribution random variable. At this point, we can use MATLAB in order to find 1 - Φ(-1.106) where Φ is the Standart Normal Distribution CDF. We can use the code 1-normcdf(-1.106) and eventually get the correct result 0.866.

b)

Again we can use the same method we used in part a except this time n = 2000. So the new $\mu =$ n*p becomes 200 and standart deviation becomes $\sigma = \sqrt{np(1-p)} = 13.416$. Also, this time we want the number of soldiers from naval forces X to be at least 180 (because %9 of 2000 is 180). So we need to find $P\{X \geq 180\}$ = 1 - $P\{X < 180\}$. Now we need to standardize the variable X.

$$1 - P\{\frac{X - 200}{13.416} < \frac{179.5 - 200}{13.416}\} = 1 - P\{Z < -1.528\}$$

Where Z is the Standart Normal Distribution random variable. Again we can use MATLAB in order to find 1 - Φ(-1.528) using the code 1 - normcdf(-1.528) and get the result 0.937. We can see that the probability has increased. This is because increasing the sample size reduces the effect of randomness and tends to make observed percentage closer to the true percentage (%10 in this case).

Answer 4

a)

We can represent the lifetime of a randomly selected elephant with continuous random variable X. It is given that the mean value of X is 65 and standart deviation 6. We need to find $P\{60 < X < 75\}$. In order to make it easier to calculate, we can standardize the normal random variable X to Z where Z = $\frac{X-\mu}{\sigma}$.

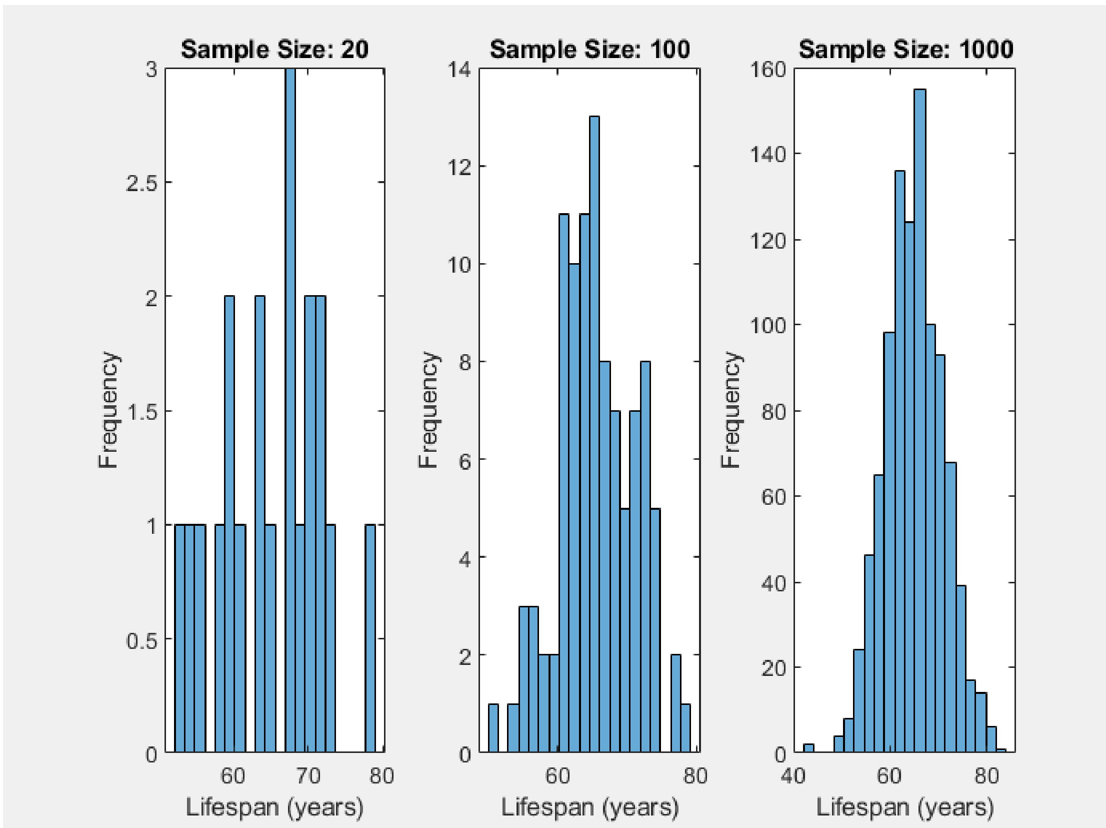
$$P\{60 < X < 75\} = P\{\frac{60 - \mu}{\sigma} < Z < \frac{75 - \mu}{\sigma}\} = P\{\frac{60 - 65}{6} < Z < \frac{75 - 65}{6}\}$$

$$P\{\frac{-5}{6} < Z < \frac{10}{6}\} = \Phi(\frac{10}{6}) - \Phi(-\frac{5}{6})$$

At this point we can use MATLAB in order to calculate the standart normal CDFs $\Phi(\frac{10}{6})$ and $\Phi(-\frac{5}{6})$ by using the code normcdf($\frac{10}{6}$) - normcdf(-$\frac{5}{6}$) and eventually get the result 0.750.

b)

For this question, I have used MATLAB where the X-axis represents the lifespan (in years) and y axis represents the frequency. Here is the histogram;



And here is the code that I used to produce this histogram;

```
>> % Parameters
mu = 65; % Mean
sigma = 6; % Standard deviation (square root of variance)

% Sample sizes
sample_sizes = [20, 100, 1000];

% Create figure
figure;

% Loop through sample sizes
for i = 1:length(sample_sizes)
    % Generate random sample
    sample = normrnd(mu, sigma, sample_sizes(i), 1);

    % Plot histogram
    subplot(1, 3, i);
    histogram(sample, 20); % You can change the number of bins here
    xlabel('Lifespan (years)');
    ylabel('Frequency');
    title(['Sample Size: ', num2str(sample_sizes(i))]);
end
```

c)

From part A we know that the probability that a randomly selected elephant will live between 60 and 75 years long is %75. So we expect that in most of the simulations the percentage should be around %75 which is higher than 70 and less than 85. In order to simulate that we can use MATLAB. Here is the code I used;

```
>> % Parameters
mu = 65; % Mean
sigma = 6; % Standard deviation (square root of variance)
sample_size = 100; % Sample size
num_iterations = 1000; % Number of iterations
range_min = 60; % Minimum lifespan
range_max = 75; % Maximum lifespan

% Initialize counter variables
count_70_percent = 0;
count_85_percent = 0;

% Perform simulations
for iter = 1:num_iterations
    % Generate random sample
    sample = normrnd(mu, sigma, sample_size, 1);

    % Count number of elephants in the given range
    count_in_range = sum(sample >= range_min & sample <= range_max);

    % Calculate percentage of elephants in the given range
    percentage_in_range = count_in_range / sample_size * 100;

    % Check if at least 70% of elephants are in the given range
    if percentage_in_range >= 70
        count_70_percent = count_70_percent + 1;
    end

    % Check if at least 85% of elephants are in the given range
    if percentage_in_range >= 85
        count_85_percent = count_85_percent + 1;
    end
end

% Display results
disp(['Number of simulations with at least 70% of elephants in the given range: ', num2str(count_70_percent)]);
disp(['Number of simulations with at least 85% of elephants in the given range: ', num2str(count_85_percent)]);
Number of simulations with at least 70% of elephants in the given range: 897
Number of simulations with at least 85% of elephants in the given range: 13
```

As we can see, in 897 simulations at least %70 of elephants were in given range whereas in only 13 simulations at least %85 of elephants were in given range.
So as we expected, in most of the simulations the percentage is more than 70 and less than 85. This is because for a single randomly selected elephant, the probability that its lifetime is between 60 and 75 is %75 as we calculated in part a. So in most of the simulations, the percentage is around %75 which is higher than 70 and lower than 85.