Module 5: Homework (13 Questions)

Question 1

A manufacturer receives a shipment of 100 parts from a vendor. The shipment will be unacceptable if more than five of the parts are defective. The manufacturer is going to randomly select K parts from the shipment for inspection, and the shipment will be accepted if no defective parts are found in the sample.

How large does K have to be to ensure that the probability that the manufacturer accepts an unacceptable shipment is less than 0.1?

Hint: We recommend using R to plug in different values of K.

- ◯ 42
- ◯ 22
- ◉ 32
- ◯ 12

✔

Explanation

Let's denote by X the number of defective parts in the sample. Then, we have that $X \sim \textbf{hypergeometric}\,(N = 100, M, K)$ where M is the number of defectives in the shipment and K equals the sample size chosen by the manufacturer. If there are 6 or more defectives in the shipment, the the probability that the shipment is accepted ($X = 0$) is at most:

$$P\,(X = 0|N = 100, M = 6, K) = \frac{\binom{6}{0}\binom{94}{K}}{\binom{100}{K}} = \frac{(100-K)\cdots(100-K-5)}{100\cdots95}$$

You can simulate this in R and vary the number of K. We find that $P\,(X = 0) = 0.10056$ for $K = 31$ and $P\,(X = 0) = 0.09182$ for $K = 32$. Then, the sample size must be at least 32.

Module 5: Homework (13 Questions)

Question 2

Now suppose that the manufacturer decides to accept the shipment if there is at most one defective part in the sample. How large does K have to be to ensure that the probability that the manufacturer accepts an unacceptable shipment is less than 0.1? As above, a shipment is unacceptable if there are more than 5 defective parts.

51 ✔ **Answer:** 51

51

Explanation

Now we have that $P\left(accept\ shipment\right) = P\left(X = 0\ or\ 1\right)$, and, for 6 or more defectives, the probability is at most:

$$P\left(X = 0\ or\ 1 | N = 100, M = 6, K\right) = \frac{\binom{6}{0}\binom{94}{K}}{\binom{100}{K}} + \frac{\binom{6}{1}\binom{94}{K-1}}{\binom{100}{K}}$$

We can simulate this in R and vary K and we have that $P\left(X = 0\ or\ 1\right) = 0.10220$ for $K = 50$ and $P\left(X = 0\ or\ 1\right) = 0.09331$ for $K = 51$. Then, the sample size must be at least 51.

Question 3

A man with n keys wants to open his door and tries the keys at random. Exactly one key will open the door.

What is the expected value of the number of trials needed to open the door if unsuccessful keys are not eliminated for further selections?

- (●) n
- ( ) $n - 2$
- ( ) $\frac{n-1}{n+1}$
- ( ) $n - 1$

✔

Explanation

To model the problem, compute the probability of trials up to and including the first success. In this case, $\mathbb{E}\left[X\right] = n$.

Module 5: Homework (13 Questions)

Question 4

Let the number of chocolate chips in a certain type of cookie have a Poisson distribution. We want the probability that a randomly chosen cookie has at least two chocolate chips to be greater than 0.99. For which of the following values of the **mean** of the distribution is this condition assured? (Please select all that apply!)

Hint: You may wish to try different values in R when solving this problem if you have trouble solving the relevant equations.

- [ ] 6
- [x] 7
- [x] 8
- [x] 9

✔

Explanation
We have that $X \sim Poisson(\lambda)$. We want $P(X \geq 2) \geq 0.99$, that is:

$$P(X \leq 1) = e^{-\lambda} + \lambda e^{-\lambda} \leq 0.01.$$

If we simulate this in R we find that this condition is assured for $\lambda \geq 6.64$.
Note: You can verify this in R using the `ppois` command. For example, if $\lambda = 7$, we can type in `ppois(1, lambda = 7, lower = FALSE)` which will give $P(S \leq 1)$ where X is Poisson with parameter $\lambda = 7$ and this is 0.0073.

You decide to move out of your college's dorms and get an apartment, and you want to discuss the budget with your roommate. You know that your monthly grocery bill G will depend on a number of factors, such as whether you are too busy to cook, whether you invite guests for meals frequently, how many special holiday meals you will cook, etc. In particular, G will have an approximate normal distribution with a variance of 2500 and a mean:

$$\mu = 300 + 10M - 100B + 50H$$

Where M is the number of meals to which you invite guests, and $\mathbb{E}[M] = 8$. B is a measure for how busy you are with 14.310X problem sets and assume it is $U[0, 1]$. H is a variable that takes on the value 1 for holiday months of November, December, and January and 0 otherwise.

Module 5: Homework (13 Questions)

Question 5

What is the mean of G in a November, where $M = 10$ and $B = 0.5$?

400

✔ **Answer:** 400

400

Explanation

We have that:

$$\mathbb{E}\left[G|M, B, H = 1\right] = 300 + 10 * 10 - 100 * 0.5 + 50$$

$$\mathbb{E}\left[G|M, B, H = 1\right] = 300 + 100 - 50 + 50 = 400$$

Question 6

For a month chosen at random what is $\mathbb{E}\left[G\right]$? (Select al that apply)

- [ ] $\mathbb{E}\left[300 + 10M - 100B + 50\right]$
- [ ] $\mathbb{E}\left[300 + 10M - 100B\right]$
- [ ] $300 + 10\mathbb{E}\left[M\right] - 100\mathbb{E}\left[B\right] + 50 * \mathbb{E}\left[H\right]$ ✔
- [x] $312.5 + 10\mathbb{E}\left[M\right] - 100\mathbb{E}\left[B\right]$ ✔
- [ ] $\mathbb{E}\left[300 + 10M - 100B + 50 * \frac{3}{4}\right]$
- [ ] $\mathbb{E}\left[300 + 10M - 100B + 50 * \frac{1}{4}\right]$ ✔

Explanation

We have that

$$\mathbb{E}\left[G\right] = \mathbb{E}\left[G|M, B\right] = \mathbb{E}\left[300 + 10M - 100B + 50H\right] = 300 + 10\mathbb{E}\left[M\right] - 100\mathbb{E}\left[B\right] + 50 * \mathbb{E}\left[H\right] = 300 + 10\mathbb{E}\left[M\right] - 100\mathbb{E}\left[B\right] + 50 * \frac{1}{4} = 312.5 + 10\mathbb{E}\left[M\right] - 100\mathbb{E}\left[B\right]$$

Module 5: Homework (13 Questions)

Question 7

What is $E(G)$?

342.5

✔ **Answer:** 342.5

342.5

Explanation
We have that:

$$\mathbb{E}[G|M,B] = 312.5 + 10\mathbb{E}[M] - 100\mathbb{E}[B]$$

$$\mathbb{E}[G|M,B] = 312.5 + 10*8 - 100*0.5$$

$$\mathbb{E}[G|M,B] = 312.5 + 80 - 50$$

$$\mathbb{E}[G|M,B] = 342.5$$

Now we are going to perform some simulations in R. We are going to follow Sara's example in the lecture where we imagine a case where the x_i follow a uniform distribution between 0 and θ $(U[0,\theta])$, and two researchers are trying to figure out the value of θ. (We will set $\theta = 5$). We are going to simulate different random samples from this distribution with a sample size of 100 observations each. These samples will be available to the two researchers, and we are going to plot how $\hat{\theta}$ is distributed for different estimators.

There are two types of researchers in this world. Researcher A uses as an estimator for θ, $\hat{\theta}_A = 2 * \overline{x}$, where $\overline{x}$ corresponds to the sample mean of the sample he receives from us. Researcher B uses as an estimator $\hat{\theta}_B = 2 * \text{median}(x)$, where $\text{median}(x)$ corresponds to the median of the sample he receives from us.

We have provided you with this R code that has some information missing in case you need help for this exercise.

Module 5: Homework (13 Questions)

Question 8

What would be the mean of this distribution of $\overline{x}$?

2.5 ✔ **Answer:** 2.5

2.5

Explanation

As it was discussed in the lecture, we have that $\overline{x} = \frac{\sum x_i}{n}$, then:

$$\mathbb{E}\left[\overline{x}\right] = \frac{1}{n}\sum\mathbb{E}\left[x_i\right] = \frac{n\mu}{n} = \mu$$

In this case since θ is equal to 5, we know that $\mu = \frac{\theta}{2} = 2.5$.

Question 9

What would be the variance of the distribution of $\hat{\theta}_A$? Please enter the numerical value of the variance.

Note: Please review our guidelines on precision regarding rounding answers here.

Answer: 1/12

Explanation

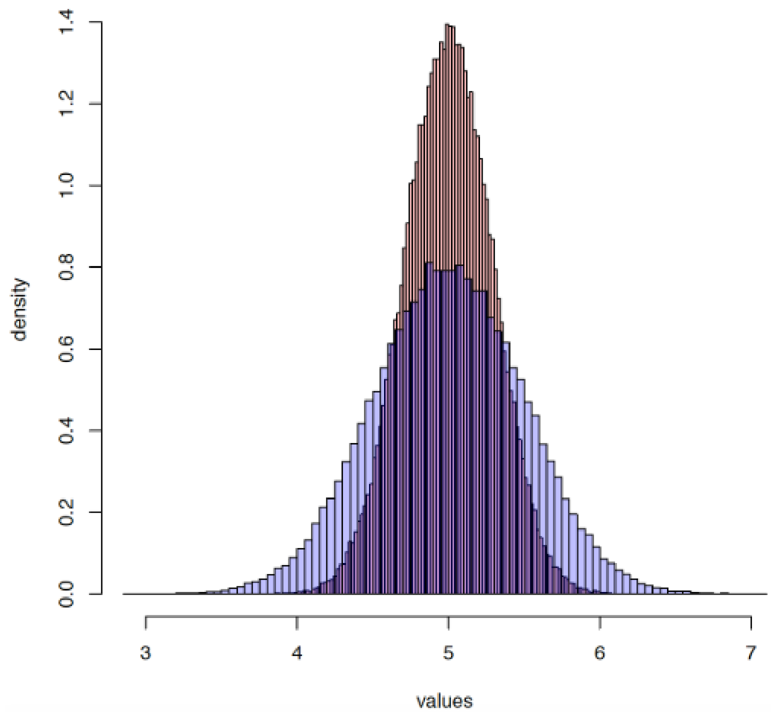
We know that the variance of $\overline{x} = \frac{\sigma^2}{n}$, where σ^2 corresponds to the variance of x_i and n to the sample size. For a uniform distribution between 0 and θ, we have that the variance is given by $\frac{\theta^2}{12}$. In this case, then we know that $var\left(\overline{x}\right) = \frac{25}{12*100} = \frac{25}{1200}$. Then, we have that:

$$var\left(\hat{\theta}_A\right) = var\left(2*\overline{x}\right) = 4*var\left(\overline{x}\right) = \frac{4*25}{12*100} = \frac{1}{12} \text{ or } 0.08333$$

Module 5: Homework (13 Questions)

If you haven't already, please fill in the R code provided earlier.

We have run our simulations, simulating 100,000 different samples of size 100. We have provided 200,000 researchers (A and B), each with one of these samples. They have sent us their estimators for $\hat{\theta}$. The following plot shows a histogram of their estimators (Figure 1).



Question 10

Does the blue histogram correspond to the estimator of researcher A or researcher B?

- ○ Researcher A
- ● Researcher B

✔

Explanation
When you fill in the missing parts of the code we sent you and run the simulations, you would realize that the blue distribution corresponds to the distribution of the median of the 100,000 simulations we have run for this exercise.

Module 5: Homework (13 Questions)

Question 11

Since both of the estimators are centered around the real value of the parameter θ, you should use the estimator with the lowest variance. Which estimator should you use?

- (selected) $\hat{\theta}_A$
- $\hat{\theta}_B$

✔

Explanation
The estimator with the lowest variance two corresponds to $\hat{\theta}_A$.

Question 12

Now, let's increase the sample size to 1000. As an exercise try to use the provided code to code this yourself in R. What would be the new variance of the estimator $\hat{\theta}_A$?

Note: Please review our guidelines on precision regarding rounding answers here.

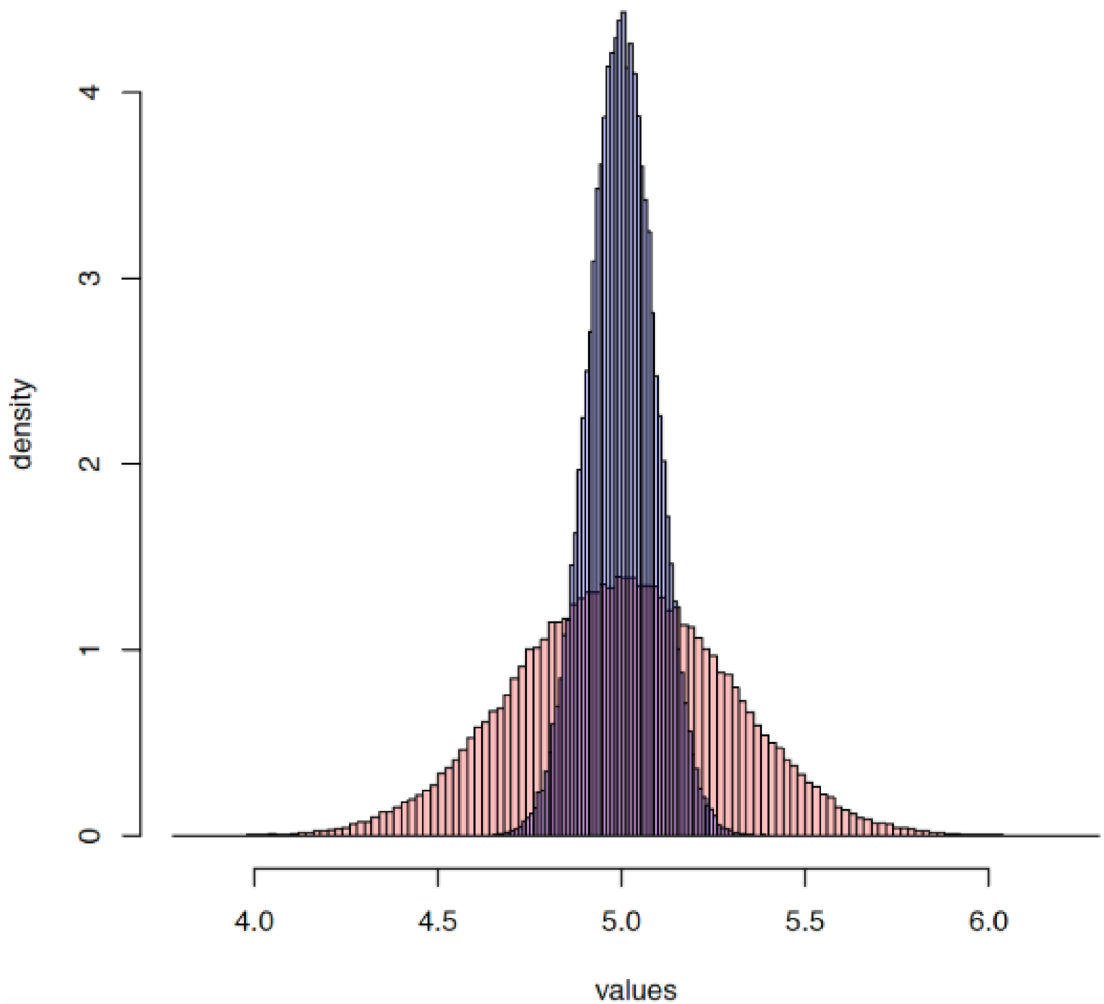
Answer: 1/120

Explanation
We know that the variance of $\overline{x} = \frac{\sigma^2}{n}$, where σ^2 corresponds to the variance of x_i and n to the sample size. For a uniform distribution between 0 and θ, we have that the variance is given by $\frac{\theta^2}{12}$. In this case, then we know that $var\left(\overline{x}\right) = \frac{25}{12*1000} = \frac{25}{12000}$. Then, we have that:

$$var\left(\hat{\theta}_A\right) = var\left(2*\overline{x}\right) = 4*var\left(\overline{x}\right) = 4*\frac{25}{12000} = \frac{1}{120}$$

Module 5: Homework (13 Questions)

The following figure shows the distribution for $\hat{\theta}_A$ for $n = 100$, and $n = 1000$



Question 13

Does the blue histogram correspond to a sample size of 100 or of 1000?

- ◯ 100
- ◉ 1000

✔

Explanation
We know that the variance of the estimator when $n = 1000$ is lower. Then, it must correspond to the blue histogram.