Expectation and variance of the sample mean

Starter

1. **(Review of last lesson)** Independent random variables X and Y are such that $E(X^2) = 14$, $E(Y^2) = 20$, Var$(X) = 10$ and Var$(Y) = 11$. Find:
   (a) the possible values of $E(3X - 2Y)$
   (b) Var$(5X - 2Y)$

Notes

In the A2 mathematics course, we looked at the distribution of the sample mean.

When a number of random samples of size n are taken from a normal distribution with mean μ and variance σ^2 such that $X \sim \text{N}(\mu, \sigma^2)$, then the distribution of the sample means of the samples will be normally distributed such that $\bar{X}_n \sim \text{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$. The standard deviation of this distribution, $\dfrac{\sigma}{\sqrt{n}}$, is called the ***standard error of the mean***.

This can be generalised for all random variables with random samples of size n:

$$E(\bar{X}) = E(X) \qquad \text{and} \qquad \text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n}.$$

Binomial:	$X \sim \text{B}(n, p)$	Mean:	$\mu = E(X) = np$
		Variance:	$\sigma^2 = \text{Var}(X) = np(1-p)$
Geometric:	$X \sim \text{Geo}(p)$	Mean:	$E(X) = \dfrac{1}{p}$
		Variance:	$\text{Var}\,(X) = \dfrac{1-p}{p^2}$
Poisson:	$X \sim \text{Po}(\lambda)$	Mean:	$\mu = E(X) = \lambda$
		Variance:	$\sigma^2 = \text{Var}(X) = \lambda$

N.B. With the binomial distribution, the n above refers to the number of trials, whereas the n in the formula for Var$(\bar{X})$ refers to the size of the sample.

There is no need to copy this part.

Proof

Let the mean and variance of the population of random variable X be $\mu = E(X)$ and $\sigma^2 = \text{Var}(X)$ respectively. A random sample of n values is taken from the population. The sample mean, $\bar{x}$, is given by:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

This is an estimate for the population mean, $E(X)$.

Each of the sample values $x_1 + x_2 + x_3 + \ldots + x_n$ can be thought of as a value from the independent variables $X_1, X_2, X_3, \ldots, X_n$. These variables have the same distribution as the population so $E(X_i) = E(X)$ and Var$(X_i) = \text{Var}(X)$.

So the sample mean is a value of the of the random variable $\bar{X}$ given by:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \frac{1}{n}X_3 + \ldots + \frac{1}{n}X_n$$

Therefore, $E(\bar{X}) = \frac{1}{n}E(X_1) + \frac{1}{n}E(X_2) + \frac{1}{n}E(X_3) + \ldots + \frac{1}{n}E(X_n)$

$$E(\bar{X}) = \frac{1}{n}E(X) + \frac{1}{n}E(X) + \frac{1}{n}E(X) + \ldots + \frac{1}{n}E(X)$$

$= E(X)$ i.e. the population mean

Given the fact that $X_1, X_2, X_3, \ldots, X_n$ are independent,

$$\begin{aligned}\text{Var}(\bar{X}) &= \text{Var}\left(\frac{X_1}{n}\right) + \text{Var}\left(\frac{X_2}{n}\right) + \text{Var}\left(\frac{X_3}{n}\right) + \ldots + \text{Var}\left(\frac{X_n}{n}\right)\\ &= \frac{1}{n^2}\text{Var}(X_1) + \frac{1}{n^2}\text{Var}(X_2) + \frac{1}{n^2}\text{Var}(X_3) + \ldots + \frac{1}{n^2}\text{Var}(X_n)\\ &= \frac{1}{n^2}\text{Var}(X) + \frac{1}{n^2}\text{Var}(X) + \frac{1}{n^2}\text{Var}(X) + \ldots + \frac{1}{n^2}\text{Var}(X)\\ &= \frac{1}{n}\text{Var}(X)\end{aligned}$$

Please start copying again.

The distribution of the sample means is called the ***sampling distribution of the means*** or just ***sampling distribution***.

As n increases, the variance of the sample decreases. In other words, the value of $\bar{x}$ is more reliable when it is calculated from a large sample which is logical.

E.g. 1 Find the expected value and the variance of the sample mean:

(a) $E(X) = 10$, $\text{Var}(X) = 1.6$, $n = 20$

(b) $X \sim \text{N}(120, 5^2)$, $n = 8$

(c) $X \sim \text{B}(9, 0.4)$, $n = 15$

(d) $X \sim \text{Po}(8.5)$, $n = 30$

(e) $X \sim \text{Geo}(0.25)$, $n = 32$

Working: (a) $E(\bar{X}) = E(X) = 10$

$$\text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n} = \frac{1.6}{20} = \frac{2}{25} = 0.08$$

E.g. 2 A machine fills cans of drink with a mean liquid content of 355 ml and standard deviation 17 ml. A sample of 30 cans is taken. Calculate the expectation and variance of the sample mean of the 30 cans.

E.g. 3 Bananas are sold in bags of 5 with the mass of the bag being exactly 37 g. The mass of one banana has mean mass 180 g and standard deviation 16.4 g. Find the mean and standard deviation of a bag of bananas.

Comparing this question to ***E.g. 4*** from the previous lesson and may wonder what is the difference. Here is the question from the previous lesson:

“A crane is lifting a crate with 4 large boxes and 5 small boxes. The large boxes have mean mass 18 kg and standard deviation 3 kg while the small boxes have mean mass 12 kg and standard deviation 1.5 kg. Given that the crate has mass 25 kg, calculate the expectation and standard deviation of the total mass of the crate with 4 large boxes and 5 small boxes loaded on it.”

The answer is that there is no real difference in the question and similar working for one will work for the other.

Here is ***sample mean*** working for the crane question:

$$E(\bar{L}) = E(L) = 18 \quad \text{and} \quad \text{Var}(\bar{L}) = \frac{\text{Var}(L)}{n} = \frac{3^2}{4} = 2.25$$

$$E(\bar{S}) = E(S) = 12 \quad \text{and} \quad \text{Var}(\bar{S}) = \frac{\text{Var}(S)}{n} = \frac{1.5^2}{4} = 0.45$$

$$\begin{aligned} E(4\bar{L} + 5\bar{S} + 25) &= 4E(\bar{L}) + 5E(\bar{S}) + 25 \\ &= 4E(L) + 5E(S) + 25 \\ &= 4 \times 18 + 5 \times 12 + 25 \\ &= 157 \end{aligned}$$

$$\begin{aligned} \text{Var}(4\bar{L} + 5\bar{S} + 25) &= 4^2\text{Var}(\bar{L}) + 5^2\text{Var}(\bar{S}) \\ &= 16 \times 2.25 + 25 \times 0.45 \\ &= 47.25 \end{aligned}$$

Standard deviation is $\frac{82\sqrt{5}}{5} \approx 36.7.$

The mean and standard deviation are 937 g and 36.7 g (3 s.f.)

Here is the ***independent observations*** working for the bananas question:

$$E(B) = 180 \quad \text{and} \quad \text{Var}(B) = 16.4^2$$

$$\begin{aligned} E(5 \text{ bananas plus bag}) &= E(B_1 + B_2 + \ldots B_5 + 37) \\ &= E(B_1) + \ldots + E(B_5) + 37 \\ &= 5 \times 180 + 37 \\ &= 937 \end{aligned}$$

$$\begin{aligned} \text{Var}(5 \text{ bananas plus bag}) &= \text{Var}(B_1 + B_2 + \ldots + B_5) \\ &= \text{Var}(B_1) + \text{Var}(B_2) + \ldots + \text{Var}(B_5) \\ &= \text{Var}(B) + \text{Var}(B) + \text{Var}(B) + \text{Var}(B) + \text{Var}(B) \\ &= 5\text{Var}(B) \\ &= 5 \times 16.4^2 \\ &= 1344.8 \end{aligned}$$

Video: Deriving the mean and variance of the sample mean

Video: Expectation and variance of the sample mean

Solutions to Starter and E.g.s

Exercise

p155 8B Qu 1i, 2-5, (7-8 red)

Summary

$E(\bar{X}) = E(X)$

$\text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n}$