Started on	Friday, 22 April 2022, 5:43 PM
State	Finished
Completed on	Saturday, 23 April 2022, 12:29 PM
Time taken	18 hours 46 mins

Question **1**

Correct

Marked out of 1.00

The mean and variance of random variable X are 50 and 4, respectively. Evaluate the standard deviation of

$$Y = -X$$

Select one:

- a. 1.414
- b. -2.137
- c. 2 ✔
- d. -1.414

The correct answer is: 2

Question **2**

Correct

Marked out of 1.00

Version of Section 5.4.2, exercise 4.

The number of calories burn by a biker on a biking day depends on the number of hours biking plus the fixed amount burnt by the regular functioning of the body to stay alive. Based on past experience it is known that the calories burnt by a biker follows this function

$$Calories = 1000 + 200X$$

where X is the number of hours biking. If X is a Poisson random variable with parameter λ=3, what is the expected number of calories burn and the standard deviation of the calories?

Select one:

- a. 1600 and 1600, respectively
- b. 1600 and 447.2136, respectively
- c. 1600 and 346.4102, respectively ✔
- d. 2000 and 1000, respectively

The correct answer is: 1600 and 346.4102, respectively

Question 3

Correct

Marked out of 3.00

The manager of a cosmetic products stand in a department store knows that the daily demand for the most expensive item in the stand, the "dramatically beautifying moisturizing lotion" has the following probability mass function:

Probability mass function of daily demand for expensive cosmetic item

Quantity demanded	0	1	2
Probability	0.1	0.5	0.4

Suppose that the bonus is $10 each time an item is used.

Let X denote daily demand and Y denote daily bonus

Match the following results

$E(Y^2)$	210
$E[(X-5)^2]$	14.1
$E(X^2)$	2.1
Expected daily bonus	13
Probability that in each and all three randomly chosen days we observe a demand of at least one item	0.729
Standard deviation of the daily bonus	6.4
Variance of X	0.41
Expected daily demand	1.3

The correct answer is:
$E(Y^2)$ → 210,

$E[(X-5)^2]$ → 14.1, $E(X^2)$ → 2.1, Expected daily bonus → 13,

Probability that in each and all three randomly chosen days we observe a demand of at least one item → 0.729, Standard deviation of the daily bonus → 6.4, Variance of X → 0.41, Expected daily demand → 1.3

Question **4**

Correct

Marked out of 1.00

Let X be a random variable. What is

$$E[(X(X-1))] + E(X) - [E(X)]^2$$

equal to ?

Select one:

- a. $$\sigma_x$$
- b. E(X)
- c. The variance of X ✔
- d. $$E(X^2)$$

The correct answer is: The variance of X

Question **5**

Correct

Marked out of 1.00

Chapter 5, end of chapter miniquiz

Which of the following does **NOT** equal the variance of X?

Select one:

- a. $$\sum_x x^2 P(x) + \sum_x (\mu_x)^2 P(x) - \mu \sum_x 2xP(x)$$
- b. $$\sum_x x^2 P(x) + \sum_x (E(X))^2 P(x) - \sum_x 2xE(X)P(x)$$
- c. $$E(\mu^2) + (E(X))^2 - 2(E(X))^2$$ ✔

The correct answer is:

$$E(\mu^2) + (E(X))^2 - 2(E(X))^2$$

Question **6**

Correct

Marked out of 1.00

A students knows that a random variable has expected value 5 and standard deviation 2. The student is given the following formula.

$$\sum_x [(x^2 + (E(X))^2 - 2xE(X))P(X = x)]$$

What is that expression equal to?

Select one:

- a. 1
- b. 4
- c. 29
- d. 3
- e. 16

The correct answer is: 4

Question **7**

Correct

Marked out of 1.00

Chapter 5-Section 5.14.1 Exercise 1

In this question, we review characteristics of the Poisson model.

Almost every year, there is some incidence of volcanic activity on the island of Japan. In 2005 there were 5 volcanic episodes, defined as either eruptions or sizable seismic activity. Suppose the mean number of episodes is 2.4 per year. Let X be the number of episodes in the next two years. An appropriate model to use for X is [Poisson] ✔ . The member of the Poisson family that we would use has expected value [4.8] ✔ by the [[3]. According to this model, the probability that there will be no volcanic episodes in the next two years is approximately [0.00823] ✔ . On the other hand, the probability that there will be more than three episodes in the next two years is approximately [0.7058.] ✔ . Considering that each episode costs the island 1 million dollars, the expected cost of volcanic activity in the next two years is [[6]] with a standard deviation of [$2190.89] ✔ .

Poisson extended period

4.8 milllion dollars | $61200 | 48 million dollars | log normal

exponential

For the standard deviation, there is a typo. Everybody will get 0.14 added to their grade in myucla by making the total max out of 14.86 instead of 15

Var(Cost) = (1000000^2) 4.8

sd(cost) 1000000 sqrt(4.8) =2190890

The correct answer is:

Chapter 5-Section 5.14.1 Exercise 1

In this question, we review characteristics of the Poisson model.

Almost every year, there is some incidence of volcanic activity on the island of Japan. In 2005 there were 5 volcanic episodes, defined as either eruptions or sizable seismic activity. Suppose the mean number of episodes is 2.4 per year. Let X be the number of episodes in the next two years. An appropriate model to use for X is [Poisson]. The member of the Poisson family that we would use has expected value [4.8] by the [[3]. According to this model, the probability that there will be no volcanic episodes in the next two years is approximately [0.00823]. On the other hand, the probability that there will be more than three episodes in the next two years is approximately [0.7058.]. Considering that each episode costs the island 1 million dollars, the expected cost of volcanic activity in the next two years is [[6]] with a standard deviation of [$2190.89].

Question **8**

Correct

Marked out of 1.00

A random variable X has expected value

$$\mu_X$$

and variance

$$\sigma_X^2$$

.

Parameters like

$$\mu$$

and

$$\sigma$$

are constants. What is the expected value and standard deviation of the following random variable?

$$W = \frac{X - \mu_x}{\sigma_X}$$

?

Select one:

- a.

  $$\mu_X$$

  and

  $$\frac{\sigma_X}{\sqrt{n}}$$

  , respectively.

- b.

  $$\mu_x$$

  and 1, respectively

- c. 0 and 1, respectively. ✔

- d.

  $$\frac{\mu_X}{\sigma_X}$$

  and 0, respectively

$$E(\frac{X - \mu_X}{\sigma_X} = E(\frac{1}{\sigma}X - \frac{\mu_X}{\sigma_X}) = \frac{\mu_X}{\sigma_X} - \frac{\mu_X}{\sigma_X} = 0$$

$$Var(\frac{X - \mu_X}{\sigma_X} = \frac{1}{\sigma^2}Var(X) = \frac{\sigma^2}{\sigma^2} = 1$$

The correct answer is: 0 and 1, respectively.

Question **9**

Correct

Marked out of 1.00

The time, X in seconds, that it takes the Ticket counter to sell a Universal Studios pass has been found to follow the following probability model

$$f(X) = \frac{1}{100} e^{-\frac{1}{100}X}, \qquad X \geq 0$$

That time changes if the person also wants to purchase Six Flags tickets and recharge the EZ bus pass. It has been found that the time changes according to the following function:

$$Y = \frac{X}{12} + 1$$

You are next in line and want to do all of the above. How long should you expect to be at the ticket counter? By how much could the time differ, on average, from this expectation of yours?

Select one:

a. $$\mu_y = 1199; \quad \sigma_y = 1200$$

b. $$\mu_y = 9.333; \quad \sigma_y = 8.3333$$

c. $$\mu_y = 19; \quad \sigma_y = 3.21$$

d. $$\mu_y = 11.30; \quad \sigma_y = 10$$

E(X) =100

Var(X) =10000

E(Y) = (100/12) +1 =9.333

Var(Y) = (1/144)(10000)

sd(Y) =100/sqrt(144) =8.3333

The correct answer is:

$$\mu_y = 9.333; \quad \sigma_y = 8.3333$$

Question **10**

Correct

Marked out of 1.00

The following expression

$$\int_x 2\mu_x^2 x f(x) dx$$

, where f(x) is a density function and the integration is over all the domain of the random variable X,

equals

Select one:

- a. 1
- b. $$2\mu_x^3$$
- c. $$\sigma_x^2$$
- d. $$2\mu_x^2$$

The correct answer is:

$$2\mu_x^3$$

Question **11**

Correct

Marked out of 1.00

Let f(x) be the density function of X, $0 \le x \le 1$. The expression

$$\int_0^1 (20 + 30x + 10x^2) f(x) dx$$

equals

Select one:

- a. $$20 + 30E(X) + 10E(X^2)$$
- b. Var(10+5X)
- c. Var(10-5X)
- d. $$E(X + 10X)^2$$

The correct answer is:

$$20 + 30E(X) + 10E(X^2)$$

Question **12**

Complete

Marked out of 2.00

Show how you reach the conclusion you reach in the following problem. Do not attach any files. Use the editor in this page. Also, show all your work in detail and justify your answer. If you are going to use some result discussed this week, prove that result as well, either using definitions pertaining to either a discrete or a continuous random variable.

Let X be a random variable. What is

$$E[(X(X-1))]+E(X)-[E(X)]^2$$

equal to ?

By distributing, the given expression is equal to:
E(X^2 - X) + E(X) - [E(X)]^2

Using the discrete definition of E(g(x)), we can say E(X^2 - X) equals:

Σ(X^2 - X)f(x)dx

Using properties of the summation operator, we know this equals:

Σ(X^2)f(x)dx - ΣXf(x)dx

Again using the discrete definition of E(g(x)), we can say this equals:

E(X^2) - E(X)

Plugging back into the overall expression, we see:

E(X^2) - E(X) + E(X) - [E(X)]^2

Simplifying, we get:

E(X^2) - [E(X)]^2

We know the above is the definition of Var(X), therefore, the given expression is equal to the variance of X

Question 13 (2 points)

Let X be a random variable. What is

$$\mathbb{E}[(X(X-1))]+\mathbb{E}(X)-[\mathbb{E}(X)]^2$$

Soln:

$= E(x^2 - x) + E(x) - [E(x)]^2$ 0.5 linearity of expectation.

$= E(x^2) - E(x) + E(x) - [E(x)]^2$ 0.5

$= E(x^2) - [E(x)]^2$ 0.5

$= Var(x)$ 0.5 See learning class lecture for alternative formula for the variance.

Question **13**

Correct

Marked out of 1.00

CHAPTER 5-TEXTBOOK-section 5.4.2., Exercise 3

Weekly downtime of internet services from an internet service provider (in hours) has expected value 0.5 and variance 0.25. Based on past experience, the data scientist of a retailer store has calculated the loss function to the store from the downtime as

$$C = 30X + 2X^2$$

where

$$X$$

is the amount of weekly downtime and C is cost. Find the expected cost.

Select one:

- a. 15.5
- b. 16
- c. 21
- d. 17

$E(C) = 30E(X) + 2E(X^2) = 30(0.5) + 2[0.25 + 0.5^2] = 16$

Notice that we used the known result: $Var(X) = E(X^2) - [E(X)]^2$

The correct answer is: 16

Question **14**

Complete

Marked out of 5.00

CHapter 5, problem 9 end of chapter.

This problem requires you to show work. If you would like to see the rubric that will be used, more or less, you may look at the supplementary reading in Module 4 "fitting a Poisson model to the counts of births per time interval.

--------------------Problem

Do extinctions occur randomly through the long fossil's record of Earth's history?, or are there periods in which extinction rates are unusually high ("mass extinctions") compared with background rates? Whitlock and Schluter (2009) give data on the number of extinctions of marine invertebrate families in 76 blocks of time of similar duration.

0 extinctions happened in none of the blocks

1 extinction happened in 13 blocks

2 extinctions happened in 15 blocks

3 extinctions happened in 16 blocks

4 extinctions happened in 7 blocks

5 extinctions happened in 10 blocks

6 extinctions happened in 4 blocks

7 extinctions happened in 2 blocks

8 extinctions happened in 1 block

9 extinctions happened in 2 blocks

$$\geq 10$$

extinctions happened in 6 blocks

Estimate the expected number of extinctions per block given the data above and compare the proportion of blocks predicted by the model for each of the above extinction numbers with the observed ones. Is there much difference between the two?

After providing those in a nice table, do the chi-square test that is being done in the supplementary document on births, by adding the counts to the table, both the counts observed and the predicted ones.

Hint: There is a document posted in module 4 near the Poisson lecture illustrating how we fitted a Poisson to the babies data. Follow the discussion there.

You must show very detailed work and explanations, providing intermediate work, what you calculate and final numerical answers. There is a rubric included in the babies document posted in the lectures folder. We will use similar rubric here. The table for the extinctions was not completed in class, so you need to include that.

You may attach a pdf file with your work.

Given the data, the expected number of extinctions per block can be estimated by finding the average extinctions per block. To do this we assume all blocks with >= 10 extinctions had 10 extinctions (for estimation purposes) and we calculate: (0 + 1(13) + 2(15) + 3(16) + 4(7) + 5(10) + 6(4) + 7(2) + 8(1) + 9(2) + 10(6)) / 76 blocks = 3.855 extinctions/block. This value is our λ.

Again, using the values in the given table, we can calculate the empirical probability of each value by dividing the number of blocks the extinctions occurred in by the number of total blocks. For instance, for the empirical probability of 1 extinction, we calculate 13/76 = 0.171. The rest of the calculations are done in the attached PDF.

We can then calculate theoretical probabilities by fitting the data to a Poisson distribution, using λ = 3.855. For instance, for the probability of 1 extinction, we calculate (3.855^1 * e^-3.855)/1! = 0.082. The rest of the calculations are done in the attached PDF.

We then calculate the expected number of blocks according to the Poisson distribution by multiplying the theoretical probability by the total number of blocks. In the case of 1 extinction, this would be 0.082 * 76 = 6.202. The rest of the calculations are done in the attached PDF.

We can then calculate the value of the Chi-square statistics by first calculating the intermediate step (O - E)^2, where O is the observed number of blocks and E is the expected number of blocks. For 1 extinction, this value is (13 - 6.202)^2 = 2.588. This value is then divided by E. Again, for 1 extinction, this results in 2.588 / 1.609 = 1.609. The rest of the calculations are done in the attached PDF.

We then sum up these values to get the value of the Chi-square statistic: 81.479. Plugging this into the provided app, we see that P(X > 81.479) = 0, which tells us this dataset is not a good fit for the Poisson model. This seems to imply extinctions do not occur randomly, and that there are periods of mass extinctions in Earth's fossil record.

STATS 100A Quiz 2 - Sheet1.pdf

Solution

This problem illustrates one of the uses of probability in Statistics. The theoretical Poisson model is first assumed for the number of extinctions per block. Call this random variable X. We do not know the λ parameter, but a statistician would estimate it given the data. A good estimate would be the average number of extinctions per block.

$$\texttt{Estimate of } \lambda) = \frac{(1)(13) + (2)(15) + 3(16) + \cdots + 9(2) + 10(6)}{76} = 3.855263$$

(1 point)

If a data set was generated by a Poisson model with parameter

$$\lambda = 3.55263$$

, the probabilities of observing the values of X observed above would be given by the following table. These probabilities represent the proportion of blocks that have that number of extinctions. Multiplying the probability by the 76 gives you the number of blocks predicted by the model for that value of X. For each X=x,

$$P(X = x) = \frac{3.855263^x e^{-3.855263}}{x!}$$

. Probabilities are rounded to 3 decimals.

x	0	1	2	3	4	5	6	7	8	9	≥ 10
P(X=x) with Poisson model	0.021	0.082	0.157	0.202	0.195	0.15	0.096	0.053	0.026	0.011	1-0.993
#blocks predicted wit Poisson **(1 point)**	1.61=0.021*76	6.2= 0.082*76	11.95	15.36	14.8	11.42	7.34	4.04	1.95	0.83	0.532
# blocks observed **(1 point)**	0	13	15	16	7	10	4	2	1	2	6

We can see from the table of probabilities that number of blocks observed do not match very closely the number of blocks predicted by the model. There is some evidence that the Poisson model with $\lambda = 3.855263$ is not a good fit for these data. However, statisticians would go one step further in using probability to determine how bad a fit it is. They design methods to determine the accuracy of the estimate and the strength of evidence in favor or against the model. Statisticians let their statistical inference methods determine their conclusions, not a simple comparison like the one we did above. But that simple comparison is always the first step. There has to be the Chi-square test. See more below.

1 point for calculation of chisquare

X	Freq/Observed	P(X=x)	No of Blocks Predicted	$\frac{(E-O)^2}{E}$
0	0	0.0212	$(0.0212)(76) = 1.6088$	$\frac{(1.6088-0)^2}{1.6088} = 1.608770$
1	13	0.0816	$(0.0816)(76) = 6.2022$	$\frac{(6.2022-13.0000)2}{6.2022} = 7.450482$
2	15	0.1573	$(0.1573)(76) = 11.9556$	$\frac{(11.9556-15.0000)^2}{11.9556} = 0.775221$
3	16	0.2022	$(0.2022)(76) = 15.3640$	$\frac{(15.3640-16.0000)^2}{15.3640} = 0.026326$
4	7	0.1948	$(0.1948)(76) = 14.8081$	$\frac{(14.8081-7.0000)^2}{14.8081} = 4.117089$
5	10	0.1502	$(0.1502)(76) = 11.4178$	$\frac{(11.4178-10.0000)^2}{11.4178} = 0.176058$
6	4	0.0965	$(0.0965)(76) = 7.3364$	$\frac{(7.3364-4.0000)^2}{7.3364} = 1.517339$
7	2	0.0532	$(0.0532)(76) = 4.0406$	$\frac{(4.0406-2.0000)^2}{4.0406} = 1.030523$
8	1	0.0256	$(0.0256)(76) = 1.9472$	$\frac{(1.9472-1.0000)^2}{1.9472} = 0.460742$
9	2	0.0110	$(0.0110)(76) = 0.8341$	$\frac{(0.8341-2.0000)^2}{0.8341} = 1.629696$
10	6	0.0042	$(0.0042)(76) = 0.3216$	$\frac{(0.3216-6.0000)^2}{0.3216} = 100.273392$
		$\sum_{x=0}^{11} P(X = x) \approx 1$	$\sum_{x=0}^{11} xP(X = x) \approx 76$	$\sum_{x=0}^{11} \frac{(E-O)^2}{E} = 119.065638$

Question **15**

Correct

Marked out of 1.00

CHAPTER 7 SECTION 7.2.1 Ex 2 TEXTBOOK

Let X be the time that it takes to drive between point A and point B during the afternoon rush hour period on highway 4005. The density function of X is

$$f(x) = \frac{1}{2}x, \qquad 0 \leq x \leq 2$$

Calculate the interquartile range	0.732	✔
Find the median	1.414	✔
Calculate $P(0.5 \leq x \leq 1.5)$	0.5	✔
Calculate the value of the 70th percentile	1.67332	✔

$f(x) = \dfrac{1}{2}x, \qquad 0 \leq x \leq 2$

$\int_0^2 \frac{1}{2}xdx = \frac{x^2}{4}\Big|_0^c = \frac{c^2}{4} = 0.7 \Rightarrow c^2 = 2.8 \Rightarrow c = 1.67332$\\ so 70^{th} percentile is x = 1.67332

$\int_0^2 \frac{1}{2}xdx = \frac{c^2}{4} = 0.25 \Rightarrow c^2 = 1.00 \Rightarrow c = 1.$\\
First quartile(Q1) is x = 1 \\

$\int_0^2 \frac{1}{2}xdx = \frac{c^2}{4} = 0.75 \Rightarrow c^2 = 3.00 \Rightarrow c = 1.732$ \\
Third quartile(Q3) is x = 1.732 \\

IQR = Q3 - Q1 = 1.732 - 1 = 0.732\\

$P(0.5 \leq x \leq 1.5) = \int_{0.5}^{1.5} \frac{1}{2}xdx = \frac{1}{2}\frac{x^2}{2}\Big|_{0.5}^{1.5} = \frac{1}{4}[1.5^2 - 0.5^2] = 0.5$

$\int_0^2 \frac{1}{2}xdx = \frac{c^2}{4} = 0.5 \Rightarrow c^2 = 2 \Rightarrow c = 1.414$\\
The median is 1.414

The correct answer is: Calculate the interquartile range → 0.732, Find the median → 1.414, Calculate $P(0.5 \leq x \leq 1.5)$

→ 0.5, Calculate the value of the 70th percentile → 1.67332

Question **16**

Correct

Marked out of 1.00

CHAPTER 7 SECTION 7.2.1 Ex 5 TEXTBOOK

A target is located at the point 0 on the horizontal axis. Let X be the landing point of a shot aimed at the target, a continuous variable with density function

$f(x) = 1.5(1 - x^2), \quad 0 \le x \le 1$

$1.5\left(x - \frac{x^3}{3}\right)$ is the	Cumulative distribution function of X ✔
Calculate the expected landing point	0.375 ✔
What is the probability that the landing point is before 0.4?	0.568 ✔
What is the standard deviation of X?	0.2436 ✔

$f(x) = 1.5(1 - x^2), \qquad 0 \le x \le 1$

$E(x) = \int_0^1 [1.5x(1-x^2)]dx = \int_0^1 (1.5x - 1.5x^3)dx = 1.5(\frac{x^2}{2} - \frac{x^4}{4})_0^1 = 1.5(\frac{1}{2} - \frac{1}{4}) = 0.375$

$P(x \le 0.4) = \int_0^{0.4} [1.5(1-x^2)]dx = (1.5x - 1.5\frac{x^3}{3})_0^{0.4} = 1.5\left(0.4 - \frac{0.4^3}{3}\right) = 0.568$

$F(x) = P(X \le x) = \int_0^{0.4} [1.5(1-t^2)]dx = 1.5\left(t - \frac{t^3}{3}\right)_0^x = 1.5\left(x - \frac{x^3}{3}\right), \qquad 0 \le x \le 1$

$E(x^2) = \int_0^1 x^2[1.5(1-x^2)]dx = \int_0^1 [1.5x^2 - 1.5x^4]dx = 1.5\left[\frac{x^3}{3} - \frac{x^5}{5}\right]_0^1 = 1.5\left[\frac{1}{3} - \frac{1}{5}\right] = 0.2$

$$Var(x) = 0.2 - 0.375^2 = 0.05937 \qquad \sigma_x = \sqrt{0.05937} = 0.2436$$

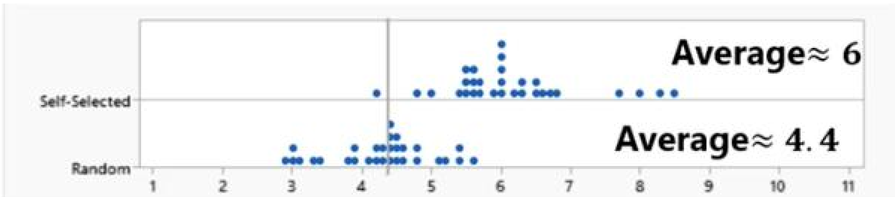
The correct answer is: $1.5\left(x - \frac{x^3}{3}\right)$ is the

→ Cumulative distribution function of X, Calculate the expected landing point → 0.375, What is the probability that the landing point is before 0.4? → 0.568, What is the standard deviation of X? → 0.2436

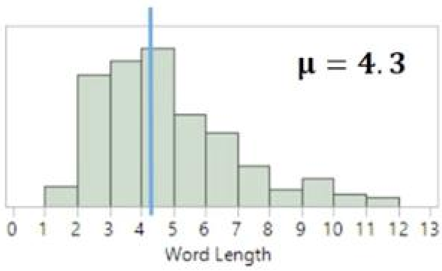
Question **17**

Complete

Marked out of 2.00



Average word length. Each dot is the average found by a single student from a sample of 10 words. Several dots superimposed means several students found that average in their sample of 10 words.



(a) Based on the illustration above about sampling of words from the Gettysburgh address, what is an advantage of random sampling?

(b) Explain how the distribution corresponding to the self-selected sample and the one for the random sample were obtained.

a) Given that the vertical line in the diagram is the true average of word length in the Gettysburg Address (4.3), it's clear that the randomly selected samples estimated the average much more accurately than the self-selected samples, as the majority of the averages found using random sampling are much closer to those found using self-selection.

b) Each sample (self-selected and random) selected 10 words and took their average length, but differed in the selection process. Students using self-sampling picked the words directly out of the text themselves, and clearly, tended to prefer words that ended up being larger than average. Students using random sampling used random.org, which picked 10 random numbers from 1 to 268 to randomly select words from the text. For example, if the random.org app selected 137, the student would use word #137 in the text as one of their 10 sample words.

a) Advantage of random sampling:

1

Equal chance of selection,

A good representation of the population is obtained

Reduces biasness

In the context of the image above,

"The random sample gave us an estimate of the population mean closer to the true population mean. The random sample gave mu=4.4, the nonrandom sample gave mu=6, and the true mu is 4.3. So obviously, the estimate of the mean is closer to the truth when we do random sampling. The estimate of the mean when using non-random sample is biased.

Part b) How the distribution was obtained. 1

Each student selected 10 words- whether self selected or using a rgn. The average length of the ten words obtained by each student was then computed.These averages were then plotted(dot/histogram plots) to show the distribution

- Not talking about each student (-0.3).

Question **18**

Correct

Marked out of 1.00

That they found that the counts of deaths by horse kick per cavalry unit followed a Poisson probability model means that

- [x] a. That death by horse kick is something that happens by chance (aka as luck, fate) ✔
- [ ] b. That death by horse kick can be perfectly predictable, no chance involved
- [ ] c. that death by horse kick always happens when the horse rider is not careful
- [x] d. That a death by horse kick is a Poisson process ✔

The correct answers are:
That a death by horse kick is a Poisson process,

That death by horse kick is something that happens by chance (aka as luck, fate)