STATISTICS WORKBOOK

STATISTICS 2

Topic: Stat Paper 2 Complete

INSTRUCTIONS

- Carry out every instruction in each task.
- Answer **all** questions.
- Use a black or dark blue pen.
- You may use an HB pencil for any diagram, graphs or rough working.
- **Calculator Allowed**.
- Show your workings if relevant.

INFORMATION

- The number of marks for each question or part question is shown in brackets [ ].

bishaltwr@gmail.com

Poisson Distribution

1. Patients arrive at a hospital accident and emergency department at random at a rate of 5 per hour.

(a) Find the probability that, during any 90 minutes period, the number of patients arriving at the hospital accident and emergency department is
(i) exactly 7
(ii) at least 10 [5]

(b) A patient arrives at 11:30. Find the probability that the next patient arrives before 11:45 a.m. [3]

2. An online shop sells a computer game at an average rate of 1 per day.

(a) Find the probability that the shop sells more than 10 games in a 7 day period. [3]

(b) Find the probability that the shop sells at least 2 games in a 3 day period [2]

3. The number of errors on a webpage article can be modelled by a Poisson Distribution with a mean of 0.5 errors per page.

(a) Find the probability that a randomly selected page has no errors. [2]

(b) A section of the website has 5 webpages. Find the probability that there are no more than 2 errors in that section. [3]

(c) The overall website has 30 webpages. Find the probability that there are exactly 8 errors altogether [3]

4. $X \sim \text{Po}(\lambda)$ and $P(X=4) = 1.2 \times P(X=3)$.

(a) Find the value of λ. [3]

(b) Find the mode of X. [2]

5. X is the number of customers arriving at a bank in a 10 minute period. X may be modelled by a Poisson Distribution with parameter 6.

(a) Find the mean and standard deviation of X. [3]

(b) Find $P(X \geq \mu)$, where μ = E (X). [3]

(c) Find $P(X \geq \mu + 2 \times \sigma)$. [3]

(d) Find $P(X \leq \mu - 2 \times \sigma)$. [3]

6. Cars travelling south on a rural road pass a particular point randomly and independently at an average rate of 2 cars every three minute.

(a) Find the probability that exactly 3 cars travel south past that point in a 5 minute time period. [3]

Cars travelling north on that road pass the same point randomly and independently at an average rate of 1 car every minute.

(b) Find the probability that a total of fewer than 4 cars cross that point in a-3 minute time period. [3]

7. $X \sim \text{Po}(\lambda)$ and $P(X=2) = 3P(X=1)$.

(a) Find the value of λ [3]

(b) Find the mode of X [2]

(c) What would be the mode if λ were 0.5 and if λ were 4.5? [2]

Approximation involving the Poisson Distribution

1. A new air-borne virus afftects 1 in 100 people. In a gathering of 200 people:

(a) Find the probability that at least 3 people are affected by the virus. [3]

(b) Use normal approximation to find the probability that at least 3 people are affected by the virus. [3]

(c) Is the approximation in (b) appropriate? [2]

2. A shopkeeper is trying to sell his shop and claims that he get an average of 2 customers every 5 minutes during peak hours. A potential buyer observes the shop for 30 minutes.

Assuming the shopkeeper's claim is true, find the probability that:

(a) 10 customers comes to the shop [2]

(b) at least 20 customers come to the shop [3]

(c) If the potential buyer only notices 8 customers in that time. Can he be sure that the shopkeeper was lying? [3]

3. A manufacturer produces low tolerance resistors. The number of defective resistors produced in a day is modelled by a Poisson Distribution with a mean of 1 per 5000 error.

(a) Use a suitable approximation to find the probability that in a batch of 5000 resistors, there are no defective resistors. [3]

4. A rare medical condition affects 1 in 150 sheep.

(a) In a small farm holding with a flock of 180 sheep, what is the probability that exactly one sheep has the condition?

(b) b. A large farm has a flock of 500 sheep. Use an appropriate approximate distribution to find the probability that there are fewer than 5 sheeps with the condition.

5. The number of letters delivered to a house on a day may be modelled by a Poisson distribution with parameter 1.5.

(a) Find the probability that there are 2 letters delivered on a particular day.

The home owner is away for 3 weeks.

(b) Find the probability that there will be more than 25 letters waiting for him when he gets back.

Linear Combination of Random Variables

1. X and Y are independent random variables. $X \sim N(2, 4)$ and $Y \sim N(3, 9)$.

(a) Find the mean and variance of 2X + 3Y. [3]

(b) Find the probability that 2X + 3Y is less than 10. [3]

2. A discrete random variable X has the following probability distribution:

x	1	2	3	4	5
P(X=x)	0.1	0.2	0.3	0.2	0.2

Find

(a) P (X = 8) **(b)** E (X) **(c)** Var (X) **(d)** E (5 -2X).

3. The random variable X has probability function:

$P(X = x) = \frac{(11-2x)}{25}$ for $x = 1,2,3,4,5$

(a) Construct a table giving the probability distribution of X [3]

Find

(b) P (2 < X < 5)

(c) E (X)

(d) Var (X)

(e) Var $(3x - 2)$.

4. The masses, in grams, of large and small packets of sugar are random variables X and Y respectively. X follows a normal distribution with mean 500 and variance 100. Y follows a normal distribution with mean 250 and variance 25.

(a) Find the probability that a randomly choosen large packet has a mass that is more than double the mass of a randomly chosen small packet. [5]

The packets are placed in boxes. The boxes are identical in appearance. 60% of the boxes contain exactly 10 randomly chosen large packets. 40% of the boxes contain exactly 20 randomly chosen small packets.

(b) Find the probability that a randomly chosen box contains packets with a total mass of more than 400 grams. [3]

5. The number of vehicles passing a point on a motorway heading east has a Poisson distribution with mean 10 per minute. The number of vehicles passing the same point heading west has a Poisson distribution with mean 8 per minute.

(a) Find the probability that in a 5 minute period, the number of vehicles heading east is more than double the number of vehicles heading west. [5]

6. A bach of capacitor has capacitance which are normally distributed with a mean of 10 microfarads and a standard deviation of 0.5 microfarads. Another batch of capacitors has capacitance which are normally distributed with a mean of 12 microfarads and a standard deviation of 0.6 microfarads. An engineer plans on combining the capacitors in parallel to form a new capacitor whose capacitance is sum of individual capacitance. He plans to use 3 capacitors from first batch and 2 capacitance from second batch.

He needs the total capacitance to be between 50 and 60 microfarads.

If he makes 1000 such circuits. How many of them will be useful to him? 5

7. A piston ring has a diameter which is normally distributed with a mean of 10 cm and a standard deviation of 0.1 cm. A cylinder has a diameter which is normally distributed with a mean of 10.1 cm and a standard deviation of 0.2 cm. The ring and the cylinder are to be combined to form a piston. Error of upto 0.1 cm is allowed in the diameter of the piston.

(a) Find the probability that a randomly selected piston assembly will have acceptable error. [5]

8. The random varibles X $\sim$ Po (2) and Y $\sim$ Po (3) represent number of two different type of circuits per 1000 that gets damaged during shipping.

(a) What is the distribution of X + Y? [3]

(b) For a batch of 2000 circuits with equal number of first and second circuit. What is the probability that a total of 7 circuit gets damaged? [3]

500 circuits of each type was shipped. Assuming P(X $\geq$ 5) and P(Y $\geq$ 7) are negligible.

(c) What is the probability that number of damaged first circuit is greater than number of damaged second circuit. [5]

Continuous Random Variables

1. The random variable X has a probability density function given by:

$$f(x) = \begin{cases} kx^2 & \text{for } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value of k [3]

(b) Find the cumulative distribution function of X [3]

(c) Find P (X > 1) [3]

2. The random variable X has a probability density function given by:

$$f(x) = \begin{cases} k(1-x) & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value of k [3]

(b) Find the cumulative distribution function of X [3]

(c) Find P (X > 0.5) [3]

(d) Find E (X) [3]

(e) Find Var (X) [3]

(f) Find P (0.25 < X < 0.75) [3]

(g) Find the median of X [3]

(h) Find the mode of X [3]

3. What are the requirements for a probability density function? [5]

4. You are planning for investing in a mutual fund. One fund provides profit modelled by a random variable X with probability density function given by:

$$f(x) = \begin{cases} 2x & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

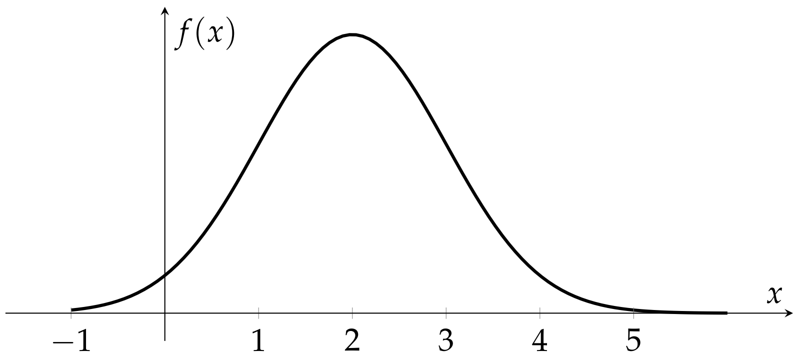
Another fund provides profit modelled by a random variable Y with probability density function given by:

$$g(y) = \begin{cases} 3(1-y)^2 & \text{for } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) If you were looking for highest profit and didn't care about risk. Which fund would you choose? [3]

(b) If you were looking for lowest risk. Which fund would you choose? [3]

5. The probability function whose graph is given below is symmetric about x = 2.



It is given that $P(2 \leq X \leq 5) = \frac{117}{256}$.

(a) Using only this information, show that $P(X \geq -1) = \frac{245}{256}$ [2]

Sampling and Estimation

Note: *A samples mean is denoted by $\bar{x}$ and population mean is denoted by μ. Sample variance is denoted by s^2 and population variance is denoted by σ^2. Mean of sample and population is same but the variance is reduced for sample by a factor of n*

1. Three quarters of the members of a fitness club are over 40 years old. A random sample of 20 members is taken. The random variables X_i is the age of the i-th member in the sample are defined as:

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th member is over 40 years old} \\ 0 & \text{otherwise} \end{cases}$$

(a) Write down the distribution of X_i [2]

(b) Find the mean and variance of X_i [3]

(c) Find P $(X_i = 2)$ [3]

2. A video game machine takes token for 1 game and for 3 games. 20% of the token used are for 1 game.

(a) Find the mean for the number of games per token. [3]

A ramdom sample of 3 tokens is taken when the machine is emptied.

(b) List all possible samples. [3]

3. The content of a box of chocolates is normally distributed with a mean of 250 grams and a standard deviation of 10 grams.

(a) Give the distribution of the mean weight of a random sample of 4 boxes. [3]

(b) Find the probability that the mean weight of a random sample of 4 boxes is less than 245 grams. [3]

4. $X \sim \text{Po}(5)$. A random sample of 3 values of X is taken.

(a) State the approximation distribution of the sample mean. [3]

(b) Find the probability that the sample mean is less than 3. [3]

5. Calculate unbiased estimate of the population mean and population variance from the following sample data: n = 50, $\sum x_i = 2500$, $\sum x_i^2 = 125000$ 5

6. For the summary statistics: n=10, $\sum x_i = 50$, $\sum x_i^2 = 300$

(a) Find the unbiased estimate of the population variance. [3]

(b) Give 90% confidence interval for the population mean. [5]

7. X B (10, 0.3). A random sample of 50 values of X is taken.

(a) State the appropriate distribution of the sample mean, $\bar{X}_{50}$. [2]

(b) Find the probability that $\bar{X}_{50}$ is greater than 5.25. [5]

8. Give 90% confidence interval for the population mean from the following sample data:

(a) n = 100, $\bar{x} = 25$, $s = 5$ [5] **(b)** n=50, $\sum x_i = 2500$, $\sum x_i^2 = 125000$ [5]

9. The contents of bottles of water are normally distributed with mean 600ml and standard deviation 7.2ml.

(a) Give the distribution of the mean content of a random sample of 6 bottles. [3]

(b) Find the probability that the mean content of a random sample of 6 bottles is less than 597ml. [3]

10. X $\sim$ Po (7). A random sample of 72 observations of X is taken.

(a) State the appropriate distribution of the sample mean, $\bar{X}_{72}$. [2]

(b) Find the probability that $\bar{X}_{72}$ is greater than 7.5. [5]

11. A certain train journey takes place every day throughout the year. The time taken, in minutes, for the journey is normally distributed with variance 11.2.

(a) The mean time for a random sample of n of these journeys was found. A 94% confidence interval for the population mean time was calculated and was found to have a width of 1.4076 minutes, correct to 4 decimal places. Find the value of n. [3]

(b) A passenger noted the times for 50 randomly chosen journeys in January, February, and March.
Give a reason why this sample is unsuitable for use in finding a confidence interval for the population mean time. [1]

(c) A researcher took 4 samples and a 94% confidence interval for the population mean was found from each sample.
Find the probability that exactly 3 of these confidence intervals contain the true value of the population meean. [2]

Hypothesis Testing

1. A financial advisor's publicity claims that at least 90% of their customers are satisfied with their performance.

(a) Is a 5% or a 10% test more likely to conclude that the financial advisor has overstated their performance? [1]

In a random sample of 20 customers, 4 said that they were not satisfied.

(b) Carry out a 5% test to determine whether the financial advisor has overstated their performance. [5]

(c) Could a Type II error have been made in this situation? [1]

2. In a hotel, 25% of people take longer than 10 minutes to register on arrival. The management install a new computer system which they claim will reducee the time to register. It is decided to accept the claim if, in a random sample of 24 people, the number taking longer than 10 minutes to register is no more than 2.

(a) Calculate the significance level of the test. [2]

(b) State the probability that a Type I error occurs. [1]

(c) Calculate the probability that a Type II error occurs if the probability of a person taking longer than 10 minutes to register is now 10%. [3]

3. A single obserevation x is to be taken from a Poisson distribution with parameter λ. This obserevation is to be used to test $H_0 : \lambda = 2.5$ against $H_1 : \lambda$ 2.5.

(a) Using a 5% significance level, find the critical region for the test. [3]

The actual value of x obtained was 5.

(b) State the conclusion that can be drawn based on this value. [2]

4. For the following Hypothesis tests, state whether you would use a one- or two-tail test, and find the critical region , the exact significance level it will give and probability of Type I and Type II error.

(a) For a binomial with n = 10, testing H0: p = 0.5 against H1: p > 0.5 at the 5% level of significance. It wasa later found that p is actually 0.13. [7]

(b) For a Poisson, testing H0: $\lambda = 6.5$ against H1: $\lambda \neq 6.5$ at the 10% level of significance. It was later found that $\lambda = 4.5$. [7]

5. Jars of jam are filled by a machine. It has been found that the quantity of jam in each jar is normally distributed with a mean of 351.2 g and a standard deviation of 4.1g. It is believed that the settings of the mean amount on the machine have been altered accidenttally. A random sample of 40 jars is taken and the mean quantity per jar is found to be 349.9 g. Assuming that the standard deviation has not been altered, state suitable null and alternative hypotheses and carry out a test Using 5% level of significance. [10]

6. A political candidate has consistently polled at 35% throughout a campaign. As a result of neegative newspaper article, the candidate believes his support has decreased. A mini survey is conducted of 14 voters and just 3 support the candidate.

(a) Carry out the test at 5% significance level. [5]

(b) Find the probability of Type I error in this case. [2]

(c) Find the probability of Type II error if the true level of support is 20%. [3]

7. A manufacturer of light bulbs claims that the mean life of their bulbs is 1000 hours. A consumer group believes that the mean life is less than 1000 hours. The claimed standard deviation is 50 hours

(a) State the null and alternative hypotheses for the test. [2]

(b) Find the critical region at 5% significance level [3]