Lab 4 Homework: Distributions

1 Overview

This lab explores distributions of random variables. The normal distribution is an example of a continuous random variable, whereas the binomial distribution represents a discrete random variable. We will analyze data to obtain empirical estimates of parameters and use these estimates to explore the binomial and normal distributions, calculating probabilities of certain events occurring based on these distributions.

2 The Data

The dataset contains observations on five groups of 25 male fruitflies from an experiment designed to test whether increased reproduction reduces longevity for male fruitflies. The name of the dataset is "fruitfly.csv".

No	serial number (1-25) within each group of 25
type	Type of experimental assignment
	1 = no females
	2 = 1 newly pregnant female
	3 = 8 newly pregnant females
	4 = 1 virgin female
	5 = 8 virgin females
lifespan	lifespan (days)
thorax	length of thorax (mm)
sleep	percentage of each day spent sleeping

3 Practice

3.1 Comparing the Distribution of Sleep Percentage Among Experimental Groups

(a) Produce an appropriate figure to compare the distribution of the percentage of each day spent sleeping among the five experimental groups. What figure did you produce?

(b) Identify the group with the shortest average percentage of sleep and report the mean and standard deviation of sleep percentage for this group.

3.2 Normal Distribution Approximation

We compare the distribution of sleep percentage between the group supplied with 8 virgin females and the group supplied with 8 newly pregnant females using the normal distribution.

(a) Using the normal distribution, complete the table below by calculating the probability of sleeping within the given percentage range. Round probabilities to two decimal places.

Sleep Percentage	8 Newly Pregnant Females $N(25.16, 19.8)$	8 Virgin Females $N(\mu_2, \sigma_2)$
$\leq 10\%$	0.22	
$10\% - 25\%$	0.27	
$25\% - 29\%$	0.08	
$> 29\%$	0.42	

Table 1: Probability of Sleep Percentage in Two Groups

(b) Suppose three fruitflies escaped their experimental conditions and were observed spending 15%, 28%, and 35% of their day sleeping. Do you think they came from the 'supplied with 8 newly pregnant females' group or the 'supplied with 8 virgin females' group and why?

3.3 Quantile Comparison

Submit the following code to create a dataset that only contains the group of fruitflies with the shortest average sleep percentage. Be sure to enter the number corresponding to the type of fruitflies you identified in (1b) after the double equal sign.

```
fruitflysubset <- subset(fruitfly, type == )
```

Complete the table comparing theoretical and observed quantiles for the two groups:

Percentile	8 Newly Pregnant Females $N(25.6, 19.8)$		8 Virgin Females $N(--, --)$	
	Theoretical	Observed	Theoretical	Observed
10th	-0.247	9.0%	–	–
25th	11.78	14.0%	–	–
50th	25.16	23.0%	–	–
75th	38.53	26.0%	–	–
90th	50.56	44.2%	–	–

Table 2: Comparison of Theoretical and Observed Sleep Quantiles

3.4 Binomial Distribution Approximation

(a) Calculate the proportion of fruitflies that spent at least 30% of their day sleeping among those supplied with 8 virgin females using the `fruitflysubset` dataset.

(b) Consider a new set of 10 fruitflies in each of the groups supplied with 8 virgin females and 8 pregnant females. Use the binomial distribution to calculate probabilities and complete the table below. Round probabilities to two decimal places.

	8 Newly Pregnant Females $Bin(10, 0.16)$	8 Virgin Females $Bin(10, p_2)$
0	0.18	
1	0.33	
2	0.29	
3	0.14	
4	0.05	
5	0.01	
6	0.0	
7	0.0	
8	0.0	
9	0.0	
10	0.0	

Table 3: Probability of Fruitflies Sleeping at Least 30% of the Day

(c) Report the probability of exactly 6 fruitflies sleeping at least 30% of the day in each group. Which group has a higher probability of exactly 6 fruitflies spending at least 30% of their day sleeping?

(d) For each group, report the most likely number of fruitflies out of 10 to spend at least 30% of their day sleeping. Which group is expected to have more?

(e) Which group is more likely to have at least 5 fruitflies spending at least 30% of their day sleeping? Calculate this probability for each group.