1. **Definition:** Let 𝑋 be the number of cracked lenses in a sample.  
   **Distribution:** 𝑋 ∼ Binomial (𝑛, 𝑝), where 𝑛=*n*= the sample size and 𝑝=0.0065 (probability of a lens being cracked).  
   1. **For a sample of 500 lenses:   
      Probability:** 𝑃(𝑋=7)  
        
      > prob <- dbinom(7, size=500, prob=0.0065)

> prob

[1] 0.02925457

Probability of exactly 7 cracked lenses in 500: 0.02925457

* 1. **For a sample of 3000 lenses:  
     Probability:** 𝑃(20 ≤ 𝑋 ≤ 25)   
       
     > prob <- sum(dbinom(20:25, size=3000, prob=0.0065))

> prob

[1] 0.3944213  
  
Probability of 20 to 25 cracked lenses in 3000: 0.3944213

1. **Definition:** Let *Y* be the number of significant earthquakes in a 10-year period  
   **Distribution:** *Y* ∼ Poisson(*λ*), where 𝜆=4.8×10=48*λ*=4.8×10=48 (since the average rate is 4.8 per year over 10 years).  
   **Probability:** 𝑃(𝑌=50)   
     
   > prob <- dpois(50, 48)

> prob

[1] 0.05405699

Probability of exactly 50 earthquakes in 10 years: 0.05405699

1. **Definition:** Let *Z* be the zinc content in a human hair sample.  
   **Distribution:** 𝑍 ∼ Normal (𝜇, 𝜎)*,* where *μ* = 159 and 𝜎 = 13.1.  
   **Probability:** 𝑃(160 ≤ 𝑍 ≤ 165)  
     
   > # Define parameters

> mu <- 159

> sigma <- 13.1

> # Calculate probability

> prob <- pnorm(165, mean=mu, sd=sigma) - pnorm(160, mean=mu, sd=sigma)

> prob

[1] 0.1461052

Probability of zinc content between 160 and 165 µg/g: 0.1461052

1. **Definition:** Let 𝑊 be the lifespan of a Plasmodium.  
   **Distribution:** *W* ∼ Gamma (*α*, *β*), where 𝛼=3.4 and 𝛽=2.8.  
   **Probability:** 𝑃(𝑊≤7)R code and output:  
   # Define parameters

alpha <- 3.4

beta <- 2.8

# Calculate probability

prob <- pgamma(7, shape=alpha, scale=beta)

prob  
[1] 0.3619324  
  
Probability of lifespan no more than 7 days: 0.3619324

**Definition:** Let 𝑋 be the number of letter’s addresses that the machine fails to correctly read.

**Distribution:** 𝑋 ∼ Binomial (𝑛=4000, 𝑝=0.0021)

**Probability:** 𝑃 (𝑋 ≥ 10)

1. **Justification:** The Poisson approximation is appropriate when 𝑛 is large, and 𝑝 is small such that 𝜆=𝑛𝑝 is moderate. Here, 𝑛=4000 and 𝑝=0.0021, giving 𝜆=4000×0.0021=8.  
   **Probability Calculation:** 𝑃 (𝑋 ≥ 10) ≈ 1−𝑃 (𝑋 ≤ 9)

lambda <- 4000 \* 0.0021

prob <- 1 - ppois(9, lambda=lambda)

prob

[1] 0.3340803

**Poisson Approximation:** 𝑃 (𝑋 ≥ 10) ≈ 0.3340803

1. The normal approximation is appropriate when np > 5 and n(1-p) > 5.

Checking: np = 4000 \* 0.0021 = 8.4 (which is > 5)

n(1-p) = 4000 \* 0.9979 = 3991.6 (which is > 5)

Therefore, the normal approximation is appropriate.

Mean (μ) = np = 8.4

Standard deviation (σ) = √(np(1-p)) = √ (8.4 \* 0.9979) ≈ 2.895230

With continuity correction, we calculate P (X ≥ 9.5) instead of P(X ≥ 10):  
  
> 1 - pnorm(9.5, mean = 8.4, sd = 2.895230)

> [1] 0.3519967

**Normal Approximation with Continuity Correction:** 𝑃 (𝑋 ≥ 9.5) ≈ 0.3519967

1. Using the same mean and standard deviation as in (b), but without continuity correction:  
     
   > 1 - pnorm(10, mean = 8.4, sd = 2.895230)

> [1] 0.2902573

**Normal Approximation without Continuity Correction:** 𝑃 (𝑋 ≥ 10) ≈ 0.2902573

1. Calculating the true probability using R's binomial CDF:  
   > 1 - pbinom(9, size = 4000, prob = 0.0021)

> [1] 0.3339988

**True Probability (Binomial CDF):** 𝑃 (𝑋 ≥ 10) = 0.3339988  
Therefore, by comparison of the values from a,b and c that the value from (a) or Poisson approximation (0.3340803) was the closest to the true value (d) (0.3339988).

1. **Seven days of data simulated using R:**  
     
   > set.seed(123) # for reproducibility

> one\_week\_data <- rpois(7, lambda = 4.3 \* 24)

> one\_week\_data

> [1] 97 115 86 104 120 107 90

1. **Simulate 200 days of data:**

> set.seed(123) # For reproducibility

> two\_hundred\_days\_data <- rpois(200, lambda=4.3 \* 24)  
> two\_hundred\_days\_data

1. A graph of fish passing through fps over 200 days

   Description automatically generatedThe histogram appears to be roughly symmetrical and bell-shaped, centered around 100-105 fish per day. The histogram also appears to be skewed to the right. This shape is consistent with what we'd expect from a Poisson distribution with a large lambda value. The spread of the data seems to range from about 70 to 140 fish per day, with most days falling between 90 and 120 fish.  
     
     
     
     
     
   R code for histogram:  
   hist(fish\_data\_200days, main = "Histogram of Fish Passing Through FPSs Over 200 Days", xlab = "Number of Fish", ylab = "Frequency", col = "skyblue",border = "black")
2. **Simulated Mean:**  
     
   > # Calculate the mean of the simulated data

> mean\_simulated <- mean(two\_hundred\_days\_data)

> mean\_simulated

> [1] 101.89  
  
**Theoretical Expected Mean:**

The theoretical expected number of fish during a 24-hour period is 𝜆×24=4.3×24=103.2.   
  
The simulated mean should be close to the theoretical mean of 103.2, but due to random variation, it may not be the same. In our case, the mean values are not the same since there is a difference of 1.31. The small difference between the two is due to the randomness in our simulation. `