a) **Definition:** Let X be the number of cracked lenses in a sample.
Distribution: $X \sim$ Binomial (n, p), where n=n= the sample size and p=0.0065 (probability of a lens being cracked).

i. **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

ii. **For a sample of 3000 lenses:**
Probability: $P(20 \leq X \leq 25)$

```
> 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

b) **Definition:** Let Y be the number of significant earthquakes in a 10-year period
Distribution: $Y \sim$ Poisson(λ), where λ=4.8×10=48λ=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

c) **Definition:** Let Z be the zinc content in a human hair sample.
Distribution: $Z \sim$ Normal (μ, σ), where μ = 159 and σ = 13.1.
Probability: $P(160 \leq Z \leq 165)$

```
> # 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

d) **Definition:** Let W be the lifespan of a Plasmodium.
Distribution: $W \sim$ Gamma (α, β), where α=3.4 and β=2.8.
Probability: $P(W \leq 7)$

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: $X \sim$ Binomial (n=4000, p=0.0021)
Probability: P ($X \geq 10$)

a) **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 ($X \geq 10$) ≈ 1−P ($X \leq 9$)

```
lambda <- 4000 * 0.0021
prob <- 1 - ppois(9, lambda=lambda)
prob
[1] 0.3340803
```

Poisson Approximation: P ($X \geq 10$) ≈ 0.3340803

b) 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

$$\sqrt{np(1-p)} = \sqrt{(4000 * 0.0021) * (1-0.0021)}$$

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 ($X \geq 9.5$) ≈ 0.3519967

c) 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 ($X \geq 10$) ≈ 0.2902573

d) 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 ($X \geq 10$) = 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).

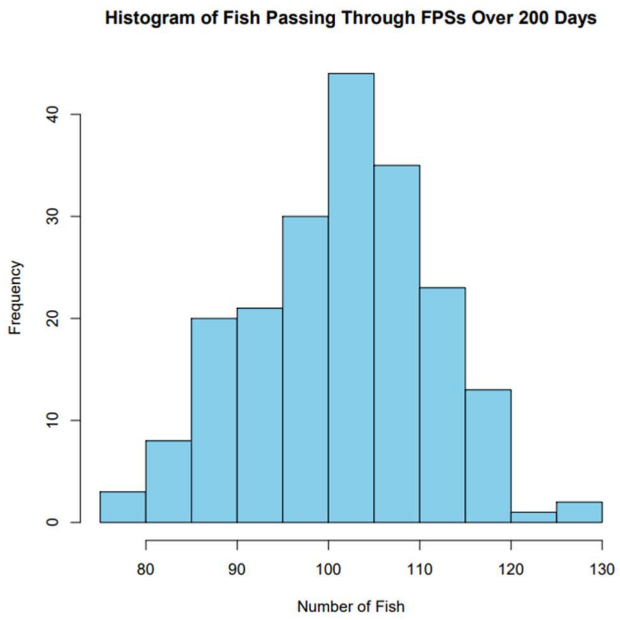
a) **Seven days of data simulated using R:**

```
> 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
```

b) **Simulate 200 days of data:**

```
> set.seed(123) # For reproducibility
> two_hundred_days_data <- rpois(200, lambda=4.3 * 24)
> two_hundred_days_data
```

c)



The histogram appears to be roughly symmetrical and bell-shaped, centered around 100-105 fish per day. 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")
```

d) **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 $\lambda \times 24 = 4.3 \times 24 = 103.2$.

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. `