Statistical data analysis, Assignment 4

Problem 1. In 1986, a county in Texas, USA, with a population of approximately 2,942,000, reported 18 cases of pertussis. If the national incidence rate of pertussis is 1.2 cases per 100,000 people annually, evaluate whether this rate accurately reflects the situation in the county.

Hint: Assume that the number of pertussis cases follows a Poisson distribution. The national rate suggests a parameter $\lambda = 1.2 \times 29.42 = 35.304$. Use R to calculate the probability of observing 18 or fewer cases in the county with this λ. This involves the use of the `ppois` function.

Problem 2. R Practice: Complete the following table by calculating the specified probabilities using R. Assume the random variables X follow different distributions as indicated.

Distribution	$P(X < 16)$	$P(X \leq 16)$	$P(X = 25)$	$P(X > 20)$
$X \sim N(18.5, 5^2)$				
$X \sim Binom(50, 0.3)$				
$X \sim NB(5, 0.3)$				

Problem 3. R Practice: Complete the table below by calculating the specified quantiles x using R for the given distributions of the random variable X. If X is a discrete variable, round your results to the nearest integer. Use the `qnorm` function for normal distributions, `qbinom` for binomial distributions, and `qnbinom` for negative binomial distributions. The probabilities are given by $P(X < x)$, $P(X > x)$, and $P(X \geq x)$.

Distribution	$P(X < x) = 0.25$	$P(X > x) = 0.25$	$P(X \geq x) = 0.15$
$X \sim N(18.5, 5^2)$			
$X \sim Binom(50, 0.3)$			
$X \sim NB(5, 0.3)$			

Problem 4. R Practice: Comparing Binomial and Poisson Distributions.

Objective: Explore the approximation of the Binomial distribution by the Poisson distribution for a small probability of success (p) across various sample sizes (n).

Consider $p = 0.001$ and a set of n values ($n = 100, 500, 1000, 5000$). For each n:

1. Calculate $P(X = k)$ for $k = 0, 1, \ldots, 10$ using both:
   - The Binomial distribution $Bino(n, p)$.
   - The Poisson distribution with mean $\lambda = np$.
2. Plot both Probability Mass Functions (PMFs) up to $k = 10$ on the same graph. Ensure each distribution is clearly labeled to visualize their similarities and differences.

Problem 5. Textbook OpenIntro Statistics, 2019 Chapter 4 Exercises: 28, 32, 35, 36, 39