

Math 341 Homework 6

Isaac Boaz

February 14, 2023

Problem 2

Let X be a random variable representing the number of shooting stars per hour. Assume X is Poisson distributed with $E[x] = \lambda$ (i.e: $X \sim \text{Poisson}(\lambda)$).

- a) Instead of a 1-hour interval, consider an interval of t hours where $t > 0$. If Y denotes the number of stars in t hours, what is its distribution and parameter value?

$$Y \sim \text{Poisson}(t\lambda)$$

- b) Calculate the probability that no shooting stars are observed in t hours (i.e: $P(Y = 0)$).

$$P(Y = y) = \frac{e^{-t\lambda} \cdot (t\lambda)^y}{y!}$$

$$\begin{aligned} P(Y = 0) &= \frac{e^{-t\lambda} \cdot (t\lambda)^0}{0!} \\ &= \frac{e^{-t\lambda}}{1} \\ &= e^{-t\lambda} \end{aligned}$$

- c) Suppose we measure the time (T) until the first shooting star. Explain why the event $A = \{T > t\}$ is equivalent to the event $B = \{Y = 0\}$

Event B represents the event that no shooting stars are observed in t hours.

Event A represents the event that the first shooting star is observed after t hours.

Since event A is defined as $T > t$, we know that at least t hours have passed before the first shooting star is observed.

- d) Using (c), compute the cdf of T . Then, state the distribution of T and its parameter value.

$$\begin{aligned} A = B &\implies P(A) = P(B) \\ P(T > t) &= P(Y = 0) \\ &= e^{-t\lambda} \end{aligned}$$

$$\begin{aligned} \text{CDF of } T: F(t) &= P(T \leq t) \\ &= 1 - e^{-t\lambda} \end{aligned}$$

T represents an exponential distribution with parameter λ .

Problem 7

X is a random variable for the % of degree holders where $\mu = 31.22$, $\sigma = 5.3$.

- a) Using a normal distribution, compute the proportion of countries where $\geq 40\%$ have a degree.

$$\begin{aligned}
 P(X \geq 40) \\
 X &\sim N(31.22, 5.3^2) \\
 Z &= (X - \mu)/\sigma \\
 P(X \geq 40) &= P\left(\frac{X - \mu}{\sigma} \geq \frac{40 - \mu}{\sigma}\right) \\
 P(Z \geq \frac{40 - 31.22}{5.3}) \\
 &= 1 - P(Z \leq 1.657) \\
 &= 1 - \text{pnorm}(1.657) \\
 &= 1 - 0.9512403 = 0.0487597
 \end{aligned}$$

- b) Let Y denote per capita income represented by $Y = 45000 + 1.7X$.

$$Y = aX + b, \mu = a\mu + b, \sigma = a^2\sigma^2$$

We know that $Y = aX + b$ can be represented as a Normal Distribution such that:

$$Y \sim N(a\mu + b, a^2\sigma^2)$$

Plugging in the values for a, b, μ, σ :

$$\begin{aligned}
 \mu_y &= 1.7 \cdot 31.22 + 45000 = 45053.074 \\
 \sigma_y^2 &= 1.7^2 \cdot 5.3^2 = 81.1801
 \end{aligned}$$

As such, Y is normally distributed with $\mu_y = 45053.074$ and $\sigma_y^2 = 81.1801$.

- c) Calculate c such that $P(|Y - 45053.07| < c) = 0.758$

$$\begin{aligned}
 P(-c < Y - 45053.07 < c) &= 0.758 \\
 &= P(-c + 45053.07 < Y < c + 45053.07) \\
 &= P\left(\frac{-c + 45053.07 - \mu_y}{\sigma_y} < \frac{Y - \mu_y}{\sigma_y} < \frac{c + 45053.07 - \mu_y}{\sigma_y}\right) \\
 &= P\left(\frac{-c}{\sigma_y} \leq Z < \frac{c}{\sigma_y}\right) \\
 &= 1 - 2P(Z \leq \frac{-c}{\sigma_y}) \\
 P(Z < \frac{-c}{\sigma_y}) &= \frac{1 - 0.758}{2} \\
 &= 0.121 \\
 Z &= \frac{c}{\sigma_y} \\
 P(Z \leq \frac{c}{\sigma_y}) &= 0.121 \\
 Z &= \text{invNorm}(0.121) = -1.1700 \\
 c &= -1.17 \cdot \sigma_y \\
 c &= -1.17 \cdot \sqrt{81.1801} = -10.54
 \end{aligned}$$

- d) For any N with μ and σ^2 , $Q_1 - 1.5IQR = \mu - 2.68\sigma$ and $Q_3 + 1.5IQR = \mu + 2.68\sigma$.

$$\begin{aligned}
 P(\mu - 2.68\sigma \leq X \leq \mu + 2.68\sigma) \\
 &= P(-2.68 \leq \frac{X - \mu}{\sigma} \leq 2.68) \\
 &= P(-2.68 \leq Z \leq 2.68) \\
 &= P(Z \leq 2.68) - P(Z \leq -2.68) \\
 &= 0.99632 - 0.00368 \\
 &= 0.99264
 \end{aligned}$$

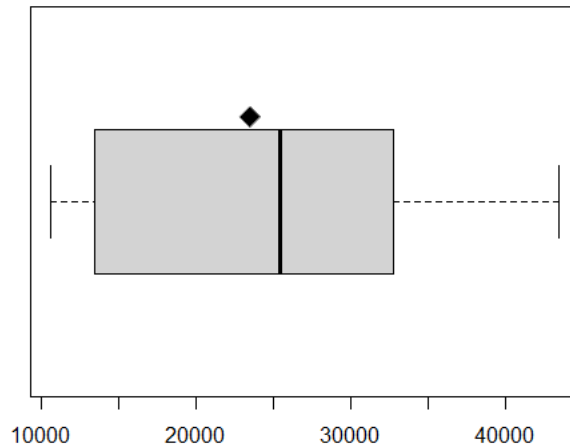
Problem 8

The following data represents the annual total releases of TRI-covered chemicals to the environment:

$$\begin{aligned} &\{33374, 32774, 24280, 26575, 26899, 13361, 13433, 43466, 13760, 10637\} \\ &\rightarrow \{10637, 13361, 13433, 13760, 24280, 26575, 26899, 32774, 33374, 43466\} \end{aligned}$$

- a) Compute the sample mean (\bar{x}) and sample standard deviation s .

$$\begin{aligned} \bar{x} &= \sum_{k=1}^n \frac{x_k}{n} \\ &= \frac{33374 + \cdots + 10637}{10} \\ \bar{x} &\approx 23856 \\ s^2 &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \\ s^2 &= \frac{(33374 - \bar{x})^2 + \cdots + (10637 - \bar{x})^2}{10 - 1} \\ s &\approx 10888.5 \end{aligned}$$



- b)
- c) Does the data appear to be symmetric and/or skewed?

- (i) Symmetric

Since the sample size is small, it would be very possible for more data to rebalance the Q1 and Q3 values. Additionally, the median line is visually close to the center.

- (ii) Skewed

The data appears to be skewed to the left going off of the Q1 and Q3 offsets.