1. (10 points) A sample of size 100, which has the sample mean \( \bar{X} = 400 \), was drawn from a population with an unknown mean \( \mu \) and the standard deviation \( \sigma = 60 \).

   a) What is the probability that the population mean will be in the interval \((410, 420)\)?

   Answer: \( Z = \frac{\mu - 400}{60/\sqrt{100}} \) follows normal distribution by approximation. Hence,
   \[
P(410 < \mu < 420) = P \left( \frac{410 - 400}{60/\sqrt{100}} < Z < \frac{500 - 400}{60/\sqrt{100}} \right) = P(1.67 < Z < 3.33)
   \]
   \[
   = 0.047025
   \]

   b) Give the 95% confidence interval for the population mean.

   Answer: \( P(400 - z_{0.025} \frac{60}{\sqrt{100}} < Z < 400 + z_{0.025} \frac{60}{\sqrt{100}}) = 0.95 \) where \( z_{0.025} = 1.96 \).

   Therefore, the interval is \([388.24, 411.76]\).

2. (10 points) You play on \( n \) identical arcade games. On each game, you play the game until you win it once and record as \( x_i \) the number of times you had to play until you won it. Find the maximum likelihood estimate for the probability of winning a game on one of these arcade games, \( p \) using the MLE method. You can leave your answer in terms if \( n \) and \( x_i \).

   Answer:

   This problem is described by the Geometric distribution with PMF given by
   \[
f(X = x) = p(1 - p)^{x-1}
   \]

   The likelihood function is
   \[
   L(p) = p(1 - p)^{x_1-1} \times p(1 - p)^{x_2-1} \times \cdots \times p(1 - p)^{x_n-1}
   = \prod_{i=1}^{n} p(1 - p)^{x_i-1} = (1 - p)^{-n + \sum x_i} p^n
   \]
   \[
   \ln L(p) = n \ln(p) + (-n + \sum x_i) \ln(1 - p)
   \]
   \[
   \frac{d \ln L(p)}{dp} = \frac{n}{p} - \frac{(-n + \sum x_i)}{1 - p} = 0
   \]
3. (12 points) All cigarettes presently on the market have an average nicotine content of 1.6 mg per cigarette. A company that produces cigarettes want to test if the average nicotine content of a cigarette is 1.6 mg. To test this, a sample of 36 of the company’s cigarettes were analyzed.

a) If it is known that the standard deviation of a cigarette’s nicotine content is 0.3 mg, what conclusions can be drawn, at the 1 percent level of significance, if the average nicotine content of the 36 cigarettes is 1.475?

Answer: Two-tailed test \( H_0 : \mu = 1.6 \) \( H_1 : \mu \neq 1.6 \). The z-statistic is \( z = \frac{1.475 - 1.6}{0.3 / \sqrt{36}} = -2.5 \).

The rejection region is located between \( [z_{-a/2}, z_{a/2}] = [-2.58, 2.58] \). Since \( z_{-a/2} < z < z_{a/2} \), we cannot reject the null hypothesis at the 1% significance level.

b) What is the P-value for the hypothesis test in (a)?

Answer: \( P \text{ value} = 2 \times P(z < -2.5) = 0.0124 > 0.01 \)

4. (10 points) The true mean height of adult women is 63 inches with a standard deviation of 2.2, and the mean height of men is 69 inches with a standard deviation of 2.5. Random samples of sizes 20 and 10 correspondingly are taken, find the probability that the \( \bar{X}_{men} - \bar{X}_{women} \geq 7 \).

Answer: Let \( \Delta X = \bar{X}_{men} - \bar{X}_{women} \) denote the sample difference, which approximately follows normal distribution with mean equal to \( 69-63=6 \) and standard deviation \( \sigma = \sqrt{\frac{2.2^2}{20} + \frac{2.5^2}{10}} = 0.9311 \). Using z-table, we could find \( P(\Delta X \geq 7) = P\left(z = \frac{\Delta X - 6}{0.9311} \geq -1.074\right) = 0.14141 \).

5. (12 points) A lab measures the viral load (virions per milliliter) of SARS-CoV-2 of infected patients:

\[
\begin{align*}
6.48366553e+05, & \quad 5.85064552e+03, & \quad 6.09144634e+05, & \quad 5.86114118e+03, \\
8.12354732e+08, & \quad 1.92888061e+06, & \quad 2.43946293e+07, & \quad 4.48598119e+06, \\
7.33095635e+03, & \quad 4.58773594e+06, & \quad 2.60489048e+04, & \quad 1.52270429e+07, \\
3.21471724e+05, & \quad 6.79572147e+08, & \quad 1.21258820e+05, & \quad 6.35426652e+07, \\
5.24408559e+06, & \quad 1.11827164e+06, & \quad 1.31108135e+06, & \quad 9.29018085e+05
\end{align*}
\]
(a) Find a point estimate of the mean log10 viral load. You can use a computer for this.

Answer: 6.1286

(b) Find a point estimate of the standard deviation of the log10 viral load. You can use a computer for this.

Answer: 1.4940

c) What is approximately the standard error of the estimate of the mean log10 viral load number obtained in part a)

Answer: 0.3341

d) Find a point estimate for the proportion of readings that are less than 10000000.

Answer: Estimate = \frac{\# \text{ of readings smaller than 5000}}{25} = \frac{15}{20}

e) Find 95% confidence intervals for the point estimate in part d)

Answer: The proportion of readings follows the Bernoulli distribution with p=0.75. Using normal approximation, the z-variable is constructed by

\[ z = \frac{\mu - 0.75}{\sqrt{0.75 \cdot (1 - 0.75)/20}} = \frac{\mu - 0.75}{0.09682} \]

Hence, the interval given by [0.560, 0.940].

Note: the standard error of the sample mean is \(\sqrt{p(1-p)/n}\).

f) Use a computer to plot the histogram and the box-and-whisker for the sample and the log base 10 of the sample.
g) What can you observe in comparing the sample to the log of the sample?

The log of the sample appears to be closer to a normal distribution. This implies that the viral load is drawn from a log-normal distribution.