1. Let $X$ denote the proportion of allotted time that a randomly selected student spends working on a certain aptitude test. Suppose that the pdf of $X$ is 

$$f(x; \theta) = \begin{cases} 
(\theta + 1)x^\theta & 0 \leq x \leq 1 \\
0 & \text{otherwise}
\end{cases}$$

where $\theta > -1$. A random sample of ten students yields data $x_1 = .92$, $x_2 = .79$, $x_3 = .90$, $x_4 = .65$, $x_5 = .86$, $x_6 = .47$, $x_7 = .73$, $x_8 = .97$, $x_9 = .94$, $x_{10} = .77$.

(a) Use the method of moments to obtain an estimator of $\theta$, and then compute the estimate for this data.

(b) Obtain the maximum likelihood estimator of $\theta$ and compute the estimate for the given data.

(c) If $X_1, \ldots, X_n$ are a random sample from Bernoulli($p$), show that the maximum likelihood estimator is $\hat{p} = n^{-1}(X_1 + \cdots + X_n)$, the sample proportion.

2. The US Army commissioned a study to assess how deeply a bullet penetrates a ceramic body armor. In the standard test, a cylindrical clay model is layered under the armor vest. A projectile is then fired, causing an indentation in the clay. The deepest impression in the clay is measured as an indication of survivability of someone wearing the armor. Here is data from one testing organization under particular experimental conditions; measurements (in mm) were made using a manually controlled digital caliper:

$$
22.4 \ 23.6 \ 24.0 \ 24.9 \ 25.5 \ 25.6 \ 25.8 \ 26.1 \ 26.4 \ 26.7 \ 27.4 \ 27.6 \\
28.3 \ 29.0 \ 29.1 \ 29.6 \ 29.7 \ 29.8 \ 29.9 \ 30.0 \ 30.4 \ 30.5 \ 30.7 \ 30.7 \\
31.0 \ 31.0 \ 31.4 \ 31.6 \ 31.7 \ 31.9 \ 31.9 \ 32.0 \ 32.1 \ 32.4 \ 32.5 \ 32.5 \\
32.6 \ 32.9 \ 33.1 \ 33.3 \ 33.5 \ 33.5 \ 33.5 \ 33.6 \ 33.6 \ 33.8 \ 33.9 \\
34.1 \ 34.2 \ 34.6 \ 34.6 \ 35.0 \ 35.2 \ 35.2 \ 35.4 \ 35.4 \ 35.4 \ 35.5 \ 35.7 \\
35.8 \ 36.0 \ 36.0 \ 36.0 \ 36.1 \ 36.1 \ 36.2 \ 36.4 \ 36.6 \ 37.0 \ 37.4 \ 37.5 \\
37.5 \ 38.0 \ 38.7 \ 38.8 \ 39.8 \ 41.0 \ 42.0 \ 42.1 \ 44.6 \ 48.3 \ 55.0
$$

These 83 data points have an average 33.37mm and SD 5.27mm.

(a) Construct a boxplot of the data and comment on interesting features.

(b) Construct a quantile-quantile plot for checking the normality of the data. Is it plausible that impression depth is normally distributed? Is a normal assumption needed in order to calculate a confidence interval or bound for the true average depth $\mu$ using the foregoing data? Explain briefly.
(c) The population mean is ____________, give or take ____________ or so.

(d) Construct 95% confidence interval for \( \mu \) and interpret the result.

(e) How many more measurements are needed in order to have 95% confidence that error bound is within 1mm?


(a) Calculate and interpret the 98% confidence interval for the proportion of all adult Americans who believe in astrology, using the tradition method.

(b) What sample size would be required so that the 95% error bound is 1%? (Use simplified formula)

(c) What sample size would be required so that the 95% error bound is 1%, irrespective of the value of \( \hat{p} \)? (Use simplified formula)

(d) Construct 98% confidence lower limit for the proportion of all adult Americans who believe in astrology, using the traditional method.

4. Repeated measurements are made of the age (based on Carbon 14 content) of a fossil. The average of 8 measurements is 2,050 years, with an SD of 90 years.

(a) The age of the fossil is about ____________ years, give or take ____________ years or so.

(b) Construct a 90% confidence interval for the age of the fossil.

(c) What is the 95% confidence upper bound?

(d) What assumptions do you make in making the above calculations?

5. Consider again the log-returns of SP500 index based on its adjusted closing prices from Jan. 1, 2000 to September 8, 2016.

(a) Estimate the daily volatility \( \sigma \) (the population SD), and its associated SE and 95% confidence interval by using the nonparametric bootstrap with \( B = 1000 \).

(b) If the daily returns were \( N(0, \sigma^2) \), how often do we expect to have a daily loss of SP500 index 5% or more? If the daily returns followed \( \frac{\sigma}{\sqrt{2}} t_4 \) (so that it has also SD \( \sigma \)), how often do we expect to have a daily loss of 5% on more?

(c) Construct 98% lower confidence limits (called Value at Risk in finance) based on the two models in (b)? Now, compute empirically what percents of days the losses actually exceed these limits. Which model gives a better confidence lower limit?

(d) Use the nonparametric bootstrap with \( B = 1,000 \) to estimate the 98% confidence lower limit. What percent of days does the SP 500 index lose more than this limit?