1. **[20 points] Pressure and Boiling Points.**

   (a) **[5 points]** What is the equation of the regression line for predicting $PRES$ from $TEMP$?

   **Solution:** Here the $x$ variable is $TEMP$ in Fahrenheit and the $y$ variable is $PRES$ in inches of mercury. The sample means, standard deviations, and correlation for the data are

   $\bar{x} = 191.79, \ s_x = 8.74, \ \bar{y} = 20.03, \ s_y = 3.86, \ r = 0.996$.

   $b = r \frac{s_y}{s_x} = 0.0996 \times \frac{3.86}{8.74} = 0.440$,

   $a = \bar{x} - b\bar{x} = 20.03 - 0.440 \times 191.79 = -64.335$.

   So the regression line is,

   $\hat{PRES} = -64.335 + 0.440 \times TEMP$.

   (b) **[5 points]** What is the predicted barometric pressure for a temperature of 185 degrees Fahrenheit?

   **Solution:** For $TEMP= 185$ degrees Fahrenheit, the predicted barometric pressure (in inches of mercury) is

   $\hat{PRES} = -64.335 + 0.440 \times 185 = 17.065$.

   (c) **[5 points]** What is the fraction of variation that is explained by the least-squares regression of $PRES$ on $TEMP$?

   **Solution:** The fraction of variation explained by the least square regression is $r^2 = (0.996)^2 = 0.992$.

   (d) **[5 points]** Is the linear regression model appropriate for this data? Explain.

   **Solution:** No. Though the fraction of variation explained by least square regression is quite high, from the residual plot it is clear that for this data the linearity assumption is not valid. There is a quadratic trend in the residual plot. So a non-linear regression, such as a quadratic regression, will be more appropriate.

2. **[24 points] A Continuous Distribution.**

   Consider a continuous distribution defined by a density: $f(x) = \frac{x^3}{c}$ for $0 < x < 2$, and $f(x) = 0$ otherwise, where $c$ is a constant.

   (a) **[6 points]** Find the value of $c$.

   **Solution:** The density is given by

   $$f(x) = \begin{cases} \frac{x^3}{c} & \text{if } 0 < x < 2, \\ 0 & \text{otherwise.} \end{cases}$$
For a density curve \( \int_{-\infty}^{\infty} f(x) \, dx = 1 \). In this case,
\[
\int_{-\infty}^{\infty} f(x) \, dx = \int_{0}^{2} \frac{x^3}{4c} \, dx = \frac{x^4}{4c} \bigg|_{0}^{2} = \frac{16}{4c} = \frac{4}{c}.
\]
So
\[
\frac{4}{c} = 1 \Rightarrow c = 4.
\]

(b) [6 points] What proportion of the observations from the distribution fall in the interval \((1, 2)\)?

Solution: Proportions of observations that fall in the interval \((1, 2)\) is
\[
\int_{1}^{2} f(x) \, dx = \int_{1}^{2} \frac{x^3}{4} \, dx = \frac{x^4}{16} \bigg|_{1}^{2} = \frac{16 - 1}{16} = 0.9375.
\]

(c) [6 points] Find the median \(M\) of the distribution.

Hint: The median of a distribution is the value \(m\) such that the area under the density to the left of \(m\) is 0.5.

Solution: Suppose the median of the above distribution is \(a\). Then we have
\[
0.5 = \int_{-\infty}^{a} f(x) \, dx = \int_{0}^{a} \frac{x^3}{4} \, dx = \frac{a^4}{16}
\]
\[
\Rightarrow \quad a^4 = 16
\]
\[
\Rightarrow \quad a = \sqrt[4]{8} = 1.6818.
\]

(d) [6 points] Find the mean \(\mu\) of the distribution.

Solution: The mean of the distribution is
\[
\mu = \int_{-\infty}^{\infty} x f(x) \, dx = \int_{0}^{2} x \cdot \frac{x^3}{4} \, dx = \frac{x^5}{20} \bigg|_{0}^{2} = \frac{32}{20} = 1.6.
\]

3. [10 points] Monkeying Around. A monkey at a typewriter types each of the 26 letters of the alphabet exactly once, the order being random.

(a) [5 points] What is the probability that the word HAMLET appears somewhere in the string of letters?

Solution: Total number of strings in which each of the 26 letters of the alphabet appears exactly once is \(= 26! \). Now to calculate number of strings among the above in which the word “HAMLET” appears somewhere, we can consider “HAMLET” as a single letter and there are other 20 letters. Then the problem is to find the number of possible arrangements of \(20 + 1 = 21\) letters. So the number is \(21! \). Hence the probability that the word “HAMLET” appears somewhere in the string is
\[
p = \frac{21!}{26!} = \frac{1}{22 \times 23 \times 24 \times 25 \times 26} = 1.267 \times 10^{-7}.
\]
(b) [5 points] How many independent monkey typists would you need in order that the probability that the word HAMLET appears is at least 0.90?

Solution: Suppose we have $n$ independent monkey typists and $X$ is the number of times the word “HAMLET” appears. Then we know that

$$X \sim \text{Bin}(n, p)$$

where $p = 1.267 \times 10^{-7}$ is the probability that the word “HAMLET” appears for one monkey typist. So the probability that the word “HAMLET” appears in at least one of the strings is,

$$P(X > 0) = 1 - P(X = 0) = 1 - \binom{n}{0} p^0 (1 - p)^n \geq 0.90.$$ 

So $1 - (1 - p)^n \geq 0.90$ implies that

$$n \geq \frac{\log(0.10)}{\log(1 - 1.267 \times 10^{-7})} = 1.8174 \times 10^7.$$ 

So we need at least $1.8174 \times 10^7$ many independent monkey typists.

4. [22 points] Sugar Content in Cereal.

The sugar content in servings of a breakfast cereal is normally distributed with mean $10.42$ gr. and standard deviation $1.76$ gr.

(a) [6 points] What proportion of servings has sugar content at least $13.32$ gr.? 

Solution:

$$P(X > 13.32) = P \left( \frac{X - \mu}{\sigma} > \frac{13.32 - \mu}{1.76} \right) = P(Z > \frac{13.32 - 10.42}{1.76}) = P(Z > 1.648) = 1 - 0.9503 = 0.0497.$$ 

(b) [6 points] What proportion of servings has sugar content between $6.9$ gr. and $13.94$ gr.?

Solution:

$$P(6.9 \leq X \leq 13.94) = P \left( \frac{6.9 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{13.94 - \mu}{1.76} \right) = P(-2 \leq Z \leq 2) = 0.9772 - 0.0228 = 0.9544.$$ 

(c) [6 points] Find the first quartile $Q_1$ and the third quartile $Q_3$ of the distribution.

Solution: The first and third quartiles for standard normal distribution are $-0.675$ and $0.675$ respectively. Hence in this case $Q_1$ and $Q_3$ are

$$Q_1 = 10.42 + 1.76 \times (-0.675) = 9.232.$$ 

$$Q_3 = 10.42 + 1.76 \times (0.675) = 11.608.$$
(d) [4 points] What proportion of servings falls in the interval between $Q_1$ and $Q_3$?

**Solution:** 50%. Since 25% of servings will fall below $Q_1$ and 25% of servings will fall above $Q_3$.

5. [24 points] Coin Toss.

A biased coin is tossed 5 times, with the probability of a head being equal to 0.65 for a toss (and obviously the probability of a tail being equal to 0.35). Let $X$ denote the total number of heads on the first toss. Let $Y$ denote the total number of heads in the 5 tosses of the coin.

(a) [6 points] What is the probability that $X = 1$ and $Y = 1$?

**Solution:** The only way in which $(X = 1$ and $Y = 1)$ is possible is if the first toss results in head and all the remaining four tosses result in tails. Hence

$$P(X = 1 \& Y = 1) = 0.65 \times (1 - 0.65)^4 = 0.00975.$$ 

(b) [6 points] What is the probability that $Y \geq 3$?

**Solution:**

$$P(Y \geq 3) = P(Y = 3) + P(Y = 4) + P(Y = 5)$$

$$= \binom{5}{3}(0.65)^3(1 - 0.65)^2 + \binom{5}{4}(0.65)^4(1 - 0.65)^1 + \binom{5}{5}(0.65)^5(1 - 0.65)^0$$

$$= 0.3364 + 0.3124 + 0.1160$$

$$= 0.7648.$$ 

(c) [6 points] What is the probability that $X = 1$ given that $Y = 1$?

**Solution:**

$$P(X = 1 \& Y = 1|Y = 1) = \frac{P(X = 1 \& Y = 1)}{P(Y = 1)} = \frac{0.00975}{5 \times 0.65 \times (0.35)^4} = 0.2.$$ 

(d) [6 points] What is the expected value of $Y$?

**Solution:** $Y$ is Binomial$(5, 0.65)$. Hence $E(Y) = np = 5 \times 0.65 = 3.25$. 

4