1. One hundred Iowa State students in a political science course were asked their political preference. The results of this survey are given in the table.

<table>
<thead>
<tr>
<th></th>
<th>Republican</th>
<th>Democrat</th>
<th>Independent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>22</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>Female</td>
<td>20</td>
<td>24</td>
<td>6</td>
</tr>
</tbody>
</table>

(a) [4] Find the probability that a student from this class is a Democrat.

\[
\frac{24}{100} = .24
\]

(b) [6] Find the probability that a female from this class is a Democrat.

\[
\frac{24}{50} = .48
\]

(c) [6] Find the probability that a student from this class is male or Republican

\[
\frac{22 + 20 + 8 + 20}{100} = .7
\]

(d) [6] Are political preference and gender independent? Justify your answer

No, they are not independent. For example,

\[
P[D | F] = .48 \quad \text{while} \quad P[D] = .44
\]

or

\[
P[F \text{ and } D] \neq P[F] \cdot P[D] = .5 \cdot .44 = .22
\]
2. Let \( X(\omega) \) denote the number of trucks owned by household, \( \omega \). The distribution of \( X \) is given by

\[
x_i = \text{number of trucks} \quad \begin{array}{c|c}
p_i = P(X(\omega) = x_i) \\
0 & .5 \\
1 & .4 \\
2 & .1 \\
\end{array}
\]

(a) [8] Find \( E(X) \) and \( Var(X) \).

\[
E(X) = 0(.5) + 1(.4) + 2(.1) = .6 \\
Var(X) = E(X^2) - (E(X))^2 = 0^2(.5) + 1^2(.4) + 2^2(.1) - (.6)^2 \\
= .8 - .36 = .44
\]

(b) [3] Assume that 100 households are sampled at random and let \( Y \) denote the number of households out of the 100 that have one truck. Find the distribution of \( Y \).

\( Y \) is binomial with \( n = 100 \) and \( p = .4 \)

(c) [4] Find \( E(Y) \) and \( Var(Y) \).

\[
E(Y) = np = 100(.4) = 40 \\
Var(Y) = np(1-p) = 100(.4)(.6) = 24
\]

(d) [6] Use the Binomial distribution to write an expression for \( P[36 \leq Y < 42] \).

\[
\sum_{y=36}^{41} \binom{100}{y} (.4)^y (.6)^{100-y}
\]

(e) [3] Justify the claim that \( Y \) is approximately normal.

\[
np(1-p) = 100(.4)(.6) = 24 > 10
\]
(f) Use the normal distribution to approximate \( P[36 < Y < 42] \).

\[
P[36 \leq Y < 42] = P \left[ 35.5 \leq Y \leq 41.5 \right] \\
= P \left[ \frac{35.5 - 40}{\sqrt{24}} \leq \frac{Y - 40}{\sqrt{24}} \leq \frac{41.5 - 40}{\sqrt{24}} \right] \\
= P \left[ -0.92 \leq Z \leq 0.31 \right] \\
= 0.6217 - 0.1788 \\
= 0.4429
\]

3. A certain jogger runs one hour per day and burns an average of 500 calories with a standard deviation of \( \sigma = 10 \). (Assume that calories burned follows a normal distribution.) Let \( X \) denote the calories burned by this jogger in a daily run.

(a) Find \( P[X > 505] \).

\[
P[X > 505] = P \left[ \frac{X - 500}{10} > \frac{505 - 500}{10} \right] \\
= P[Z > 0.5] \\
= 1 - 0.6915 = 0.3085
\]

(b) Find \( P[|X - 500| < 10] \).

\[
P[-10 < X - 500 < 10] = P[490 < X < 510] \\
= P \left[ \frac{-10}{10} < \frac{X - 500}{10} < \frac{10}{10} \right] \\
= P[-1 < Z < 1] \\
= 0.8413 - 0.1587 = 0.6826
\]

(c) Let \( X_7 \) denote the average number of calories burned per daily run over a 7 day week. Find \( P[X > 505] \).

\[
E X = \mu = 500 \\
\sigma X = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{7}} = 3.8
\]

\[
P[X > 505] = P \left[ \frac{X - 500}{10} > \frac{505 - 500}{10} \right] \\
= P[Z > 1.32] = 1 - 0.9066 \\
= 0.0934
\]
(d) [6] Find the 90th percentile for $X_r$.

4. The admissions director of a graduate program wants to know how well the Quantitative part of the Graduate Record Exam, $x_1$, and the undergraduate GPA, $x_2$, predict the student’s graduate school GPA, $y$. Data are collected on 51 graduate students and are analyzed in the attached output.

(a) [4] What is the mean of the response variable, $y$?

$$3.594118$$

(b) [4] What is the correlation between quantitative score and graduate GPA?

$$\sqrt{R^2} = \sqrt{.507382} = .712$$

(c) [6] If you wanted to predict a new student’s graduate GPA what equation would you use? Justify your answer.

I would use the 2-predictor equation, as its $R^2$ is substantially larger than those for the two 1-predictor equations.

$$\hat{y} = .262 + .4140(x_{1\text{GPA}}) + .00246(\text{quant})$$

(or use the equation only including $x_1$ since $(R^2$ is reasonable and the equation is simpler))

$$\hat{y} = x_1$$

$$\hat{y}$$

(d) [6] Using the whole model, what is the residual, $y_i - \hat{y}$, for the point (800, 3.65, 3.68).

$$y_i - \hat{y} = 3.68 - (\hat{y} = \hat{y})$$

$$= 3.68 - 3.74$$

$$= -.06$$