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>20</td>
<td>22</td>
<td>8</td>
</tr>
<tr>
<td>Female</td>
<td>24</td>
<td>20</td>
<td>6</td>
</tr>
</tbody>
</table>

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

\[
\frac{42}{100} = .42
\]

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

\[
\frac{20}{50} = .40
\]

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

\[
\frac{20 + 22 + 8 + 24}{100} = .74
\]

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

No, they are not independent — For example

\[P[D|F] = .40 \text{ while } P[D] = .42\]

or

\[P[F \text{ and } D] \neq P[F] \cdot P[D] = .5 \cdot (.42) = .21\]
2. Let $X(\omega)$ denote the number of trucks owned by household, $\omega$. The distribution of $X$ is given by

<table>
<thead>
<tr>
<th>$x_i$ = number of trucks</th>
<th>$p_i = P(X(\omega) = x_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.6</td>
</tr>
<tr>
<td>1</td>
<td>.3</td>
</tr>
<tr>
<td>2</td>
<td>.1</td>
</tr>
</tbody>
</table>

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

$$E(X) = 0(.6) + 1(.3) + 2(.1) = .5$$

$$Var(X) = E(X^2) - (E(X))^2 = 0^2(.6) + 1^2(.3) + 2^2(.1) - (.5)^2$$

$$= .7 - .25 = .45$$

(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 $p = .3$

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

$$E(Y) = np = 100(.3) = 30$$

$$Var(Y) = np(1-p) = 100(.3)(.7) = 21$$

(d) [6] Use the Binomial distribution to write an expression for $P[26 \leq Y < 42]$.

$$\sum_{y=26}^{41} \binom{100}{y} (.3)^y (.7)^{100-y}$$

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

$$np(1-p) = 100(.3)(.7) = 21 > 10$$
(f) Use the normal distribution to approximate $P[26 \leq Y < 42]$.

$$P[26 \leq Y < 42] = P[25.5 \leq Y \leq 41.5]$$

$$= P\left[\frac{25.5 - 30}{\sqrt{21}} \leq \frac{Y - 30}{\sqrt{21}} \leq \frac{41.5 - 30}{\sqrt{21}}\right]$$

$$= P[-0.98 \leq Z \leq 2.51]$$

$$= .9940 - .1635$$

$$= .8305$$

3. A certain jogger runs one hour per day and burns an average of 500 calories with a standard deviation of $\sigma = 20$. (Assume that calories burned follows a normal distribution.) Let $X$ denote the calories burned by this jogger in a daily run.


$$P\left[X > 505\right] = P\left[\frac{X - 500}{20} > \frac{505 - 500}{20}\right]$$

$$= P[Z > .25]$$

$$= 1 - .5987 = .4013$$


$$P\left[-10 < \frac{X - 500}{20} < 10\right] = P\left[-\frac{10}{20} < \frac{X - 500}{20} < \frac{10}{20}\right]$$

$$= P[-.5 < Z < .5]$$

$$= .6915 - .3085 = .3830$$

(c) [6] Let $\bar{X}$ denote the average number of calories burned per daily run over a 7 day week. Find $P[\bar{X} > 505]$.

$$M_X = M = 500 \quad \sigma_X = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{7}} = 7.56$$

$$P[\bar{X} > 505] = P\left[\frac{\bar{X} - 500}{\frac{20}{\sqrt{7}}} > \frac{505 - 500}{\frac{20}{\sqrt{7}}}\right]$$

$$= P[Z > .66] = 1 - .7446$$

$$= .2554$$
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, \( y \)?

\[ 3.55502 \]

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

\[ + \sqrt{R^2} = + \sqrt{.493599} = .703 \]

(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.

\[ \hat{y} = .295 + .00248(\text{quant}) + .4014(\text{under GPA}) \]

(\( R^2 \) is reasonable and the equation is simpler)

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

\[ y_i - \hat{y}_i = 3.68 - ( .295 + .00248(800) + .4014(3.65) ) \]

\[ = 3.68 - 3.74 \]

\[ = - .06 \]