1. The table below is a contingency table showing the cross-classification of fatal accidents at speeds greater than 55 mph involving compact spots utility vehicles between 1995 and 1999 in the U.S. Accidents are cross-classified by make of car (Ford or other) and cause of the accident (tire or other).\(^1\)

<table>
<thead>
<tr>
<th>Cause of accident</th>
<th>Other</th>
<th>Tire</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ford</td>
<td>206</td>
<td>24</td>
<td>230</td>
</tr>
<tr>
<td>Other</td>
<td>598</td>
<td>11</td>
<td>609</td>
</tr>
<tr>
<td>Total</td>
<td>804</td>
<td>35</td>
<td>839</td>
</tr>
</tbody>
</table>

Let \(A\) denote the event that a randomly selected fatal accident involved a Ford, and let \(B\) denote the event that a randomly selected fatal accident was caused by a tire.

(a) Compute \(P(A)\).

\[
P(A) = \frac{230}{839} = .274
\]

(b) Compute \(P(A \text{ and } B)\).

\[
P(A \text{ and } B) = \frac{24}{839} = .0286
\]

(c) Compute the probability that a randomly selected fatal accident was caused by a tire given that the accident involved a Ford.

\[
P(B|A) = \frac{P(A \text{ and } B)}{P(B)} = \frac{24}{230} = .104
\]

(d) What is the probability that the accident was caused by a tire given that the car make was Other?

\[
P(B|A^c) = \frac{11}{609} = .0181,
\]

where \(A^c = \text{Other}\). The easiest computation counts the number of accidents caused by tires among the Other row (11) and divides by the number of accidents involving other cars (609).

---

\(^1\)From Ramsey, F.L. and Schafer, D.W. 2002. *The Statistical Sleuth*. These data explain by the *Ford Explorer* had been popularly called the *Ford Exploder*. 
2. Roger Fouts taught four chimpanzees 10 signs of the American sign language with the intent of determining whether some signs are easier to learn, and whether some chimps tended to learn more quickly than others.2 The results of his study are shown in the table below.

Table 1: Length of time (minutes) to learn a word.

<table>
<thead>
<tr>
<th>Word</th>
<th>Booee</th>
<th>Cindy</th>
<th>Bruno</th>
<th>Thelma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listen</td>
<td>12</td>
<td>10</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>Drink</td>
<td>15</td>
<td>25</td>
<td>36</td>
<td>18</td>
</tr>
<tr>
<td>Shoe</td>
<td>14</td>
<td>18</td>
<td>60</td>
<td>20</td>
</tr>
<tr>
<td>Key</td>
<td>10</td>
<td>25</td>
<td>25</td>
<td>40</td>
</tr>
<tr>
<td>More</td>
<td>10</td>
<td>15</td>
<td>225</td>
<td>24</td>
</tr>
<tr>
<td>Food</td>
<td>80</td>
<td>55</td>
<td>14</td>
<td>190</td>
</tr>
<tr>
<td>Fruit</td>
<td>80</td>
<td>20</td>
<td>177</td>
<td>297</td>
</tr>
<tr>
<td>Hat</td>
<td>78</td>
<td>99</td>
<td>178</td>
<td>297</td>
</tr>
<tr>
<td>Look</td>
<td>115</td>
<td>54</td>
<td>345</td>
<td>420</td>
</tr>
<tr>
<td>String</td>
<td>129</td>
<td>476</td>
<td>287</td>
<td>372</td>
</tr>
</tbody>
</table>

(a) Name the factor(s) in this study.

Word (sign) and chimp

(b) The number of treatments in this study is \(40 = 10 \times 4\).

(c) Name the response variable in this study.

Time to learn a sign.

(d) List the three essential elements of a controlled experiment.

(i) randomization
(ii) replication
(iii) comparison

(e) Two of the essential elements are absent from this experiment, or at least inadequate. Identify the only one of these elements that is unambiguously present.

Comparison. There’s no replication since the number of treatments is the same as the number of observations (40), and there’s no randomization since the experimental units were no randomly assigned to groups. It’s hard even to define an experimental unit; I suppose that it might be called a teaching episode.

3. Network servers count the number of attempts to gain access (log on) made by an individual. Let \(X\) denote a random variable that counts the number of attempts. The distribution of \(X\) is shown below.

\[
\begin{array}{c|cccc}
  x & 1 & 2 & 3 & 4 \\
  P(X = x) & .70 & .25 & .03 & .02 \\
\end{array}
\]

---

(6 pts)(a) What is the probability that an individual attempting to gain access attempts two or more times?

\[ P(X \geq 2) = .25 + .03 + .02 = .3 \]

(6 pts)(b) What is the expected number of attempts made by an individual?

\[ E(X) = 1 \times .7 + 2 \times .25 + 3 \times .3 + 4 \times .02 = 1.37. \]

(3 pts)(c) What is the variance of the number of attempts? You are encouraged to use the following table.

<table>
<thead>
<tr>
<th>((x - \mu)^2)</th>
<th>.137</th>
<th>.397</th>
<th>2.66</th>
<th>6.92</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P(X = x))</td>
<td>.70</td>
<td>.25</td>
<td>.03</td>
<td>.02</td>
</tr>
</tbody>
</table>

\[ \sigma^2(X) = .137 \times .7 + .397 \times .25 + \cdots + 6.92 \times .02 = .4133. \]

(3 pts)(d) What is the standard deviation of the number of attempts?

\[ \sigma(X) = \sqrt{.4133} = .6429. \]

4. Dogs are sometimes used to determine whether fresh grizzly bear scat is on or near a hiking trail. Suppose that a particular dog correctly identifies grizzly bear scat with probability \( p = .8 \). For simplicity, if \( n \) scat samples are presented to the dog, suppose that the outcomes (correct or not) can be viewed as independent trials, all with the same probability of resulting in a correct identification.

(6 pts)(a) Suppose that the dog is on a trail and encounters 5 grizzly scats (each in a different location). What is the probability that none are correctly identified?

\[ P(X = 0) = \binom{5}{0} .8^0 .2^5 = 1 \times 1 \times .2^5 = .00032. \]

(3 pts)(b) What is the probability that one or more scats is correctly identified?

\[ P(X \geq 1) = 1 - P(X = 0) = 1 - .00032 = .99968. \]

(6 pts)(c) What is the expected number and standard deviation of the number of correct identifications?

\[ E(X) = np = 5 \times .8 = 4 \]
\[ \sigma(X) = \sqrt{npq} = \sqrt{5 \times .8 \times .2} = .8945. \]

5. A veterinary clinic wants to estimate the mean annual veterinary cost of the animals that are regularly treated by the clinic. The clinic knows that 850 dogs, 900 cats, and 100 horses are regularly treated (these animals are the clients). Three random samples are to be drawn from each of the species (50 dogs, 50 cats and 10 horses) and the records of these clients are to be selected. From these records, the annual cost for each selected client is to be recorded and summarized by species.
What is the population of interest?
The 1850 clients.

Is this an observational study or controlled experiment, or neither? (Circle one.)
Observational study.

What type of sampling design is being proposed? Stratified random sampling.

The veterinary clinic staff realized that the client records are filed by client (animal
owner), so for example, if an owner has a cat and two dogs, the records for all three
animals are filed under the owners name. The sampling design outlined in the previous
problem is too inefficient. Instead, a random sample of owners was drawn and the
records on every client animal owned by the selected owner were collected. What type
of sampling design was used to collect the data?
Cluster sampling.

6. In 1982, The Centers for Disease Control and Prevention (CDC) reported 11,413 individuals
contracted HIV/AIDS via heterosexual contact. Of these, 4,301 were males and 7,112 were
females.³

Compute the sample proportion of males to have contracted HIV/AIDS via heterosexual
contact in 1982, assuming that the 4,301+7,112 = 11,413 individuals represent a random
sample from an infinitely large population of individuals that contract HIV/AIDS via
heterosexual contact. Compute the sample proportion of females to have contracted
HIV/AIDS.

\[ \hat{p}_F = \frac{4301}{11413} = .377. \]

Compute the probability of observing 4,301 males or fewer among a sample of \( n = \)
11,413 individuals if the probability of a male contracting HIV/AIDS via heterosexual
contact is \( p = .5. \)

Since \( X \sim \text{Binom}(11,413,.5), E(X) = np = 11413 \times .5 = 5706.5 \) and \( \sigma(X) = \sqrt{npq} = \sqrt{11413 \times .5 \times .5} = 53.4. \) The success/failure conditions are easily satisfied and the
normal approximation can be used to compute the probability.

\[
P(X \leq 4301) = P(Z \leq \frac{4301 - 5706.5}{53.4}) = P(Z \leq -26.3).\]

\( P(Z \leq -26.3) \) is not in the table; on the other hand, the probability is smaller than the
smallest tabled value (.0002), so the answer is \( P(X \leq 4301) < .0002. \)

Suppose that .0005 is the probability of observing 4,301 males or fewer (part b). Do the
data support the conjecture that males and females are equally susceptible to contracting
HIV/AIDS via heterosexual contact? Why or why not?

No, because it’s extremely unlikely to observe 4,301 males, or fewer, if \( p = .5. \) My
judgement is that \( p = .5 \) is wrong, and that \( p \) is less than .5, perhaps substantially less.

³http://www.cdc.gov/hiv/topics/surveillance/basic.htm#hivaidsexposure