Load the data set Miami.mtw from the data directory Student1. This contains various information collected on n=83 female college students attending a statistics class. From the data description this is clearly not a rigorous random sample from the population of female college students, but we will work under the assumption that it is. We will concentrate on a few of the variables for which information was collected.

**A. Dichotomous variables and sample proportions:**

Consider two categorical (dichotomous) variables in the data set, for which a numerical coding is provided:

**Smoke:** smoke regularly Yes=1, No=2  
**Diet:** currently dieting Yes=1, No=2

1. Find the sample proportions and corresponding standard errors for smoking and for dieting. One way to find the sample proportions is to use Stat > Table > Cross-tabulation, and “Display column percentages”. The relevant percentages can then be turned into proportions in (0,1) dividing by 100. To find the standard errors, you need to use the formula based on the sample proportion.

Rows: Smoke

<table>
<thead>
<tr>
<th>% of Col</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>All</td>
</tr>
</tbody>
</table>

\[ p\text{-}hat = 0.1446, \text{ se}(p\text{-}hat) = \sqrt{\frac{0.1446(1-0.1446)}{83}} = 0.0386 \]

Rows: Diet

<table>
<thead>
<tr>
<th>% of Col</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>All</td>
</tr>
</tbody>
</table>

\[ p\text{-}hat = 0.3614, \text{ se}(p\text{-}hat) = \sqrt{\frac{0.3614(1-0.3614)}{83}} = 0.0527 \]

2. Now, to observe how the data relative to Smoke or Diet can be thought of as binomial experiment, let us change the numeric coding that was provided for these categorical variables. Instead of Yes=1, No=2, we want to code as Yes=1, No=0. This can be done using Manip > Code > Numeric to Numeric. In the dialog box, give the following specifications:

- Code data from columns: Smoke  Diet
- Into Columns: Smoke01  Diet01
- Original value 1, New 1
- Original value 2, New 0

With this new coding, Stat > Column Statistics > Sum, for the column Smoke01 (Diet01) gives the corresponding binomial count for smoking (dieting). Also, Stat > Column
Statistics > Mean, for the column Smoke01 (Diet01) divides this binomial count by \( n=83 \), providing the sample proportion for smoking (dieting). Verify that you obtain the same numbers as in (1) above.

\textbf{Mean of Smoke01}  
\[ \text{Mean of Smoke01} = 0.14458 \]

\textbf{Mean of Diet01}  
\[ \text{Mean of Diet01} = 0.36145 \]

3. Next, verify whether we can use a normal approximation for the sampling distribution of the sample proportion for smoking and dieting. The general rule is: \( np > 10 \) and \( n(1-p) > 10 \). Of course we don’t know the population proportions, so to get an idea of whether the sample is large enough to use normal approximations, replace the sample proportions you computed above in this rule (i.e. use your \( p \)-hat’s instead of \( p \))

- for smoking: \( 0.1446 \times 83 = 12 \), \( (1-0.1446) \times 83 = 70.99 \)
- for dieting: \( 0.3614 \times 83 = 29.99 \), \( (1-0.3614) \times 83 = 53 \)

We can use a normal approximation for the sampling distribution of \( p \)-hat in both cases.

4. Based on the approximately normal sampling distributions of the sample proportions for smoking and dieting, construct 95% and 98% confidence intervals for both proportions (4 intervals in all). To do this, you can use the sample proportions and standard errors you computed in (1) above, finding the appropriate multipliers with Calc > Probability distributions > Normal. Specify mean 0 and standard deviation 1, and use “Inverse cumulative probability” giving the appropriate input constant.

Input constant 0.025 (half of 5% under the symmetric normal curve)

\textbf{Inverse Cumulative Distribution Function}  
\[
\begin{array}{ll}
\text{P( X <= x )} & \text{x} \\
0.0250 & -1.9600
\end{array}
\]

Input constant 0.01 (half of 2% under the symmetric normal curve)

\textbf{Inverse Cumulative Distribution Function}  
\[
\begin{array}{ll}
\text{P( X <= x )} & \text{x} \\
0.0100 & -2.3263
\end{array}
\]

for smoking:
- confidence level 95% \( 0.1446 +/- 1.96 \times 0.0386 \) 14.46% +/- 7.56%
- confidence level 98% \( 0.1446 +/- 2.33 \times 0.0386 \) 14.46% +/- 8.99%

for dieting:
- confidence level 95% \( 0.3614 +/- 1.96 \times 0.0527 \) 36.14% +/- 10.33%
- confidence level 98% \( 0.3614 +/- 2.33 \times 0.0527 \) 36.14% +/- 12.28%
Comment on these intervals, their confidence level and the margins of error they embody, in a fashion that would be understandable to someone not knowing statistics.

5. An easier way to go here is Stat > Basic Statistics >1 Proportion. This menu computes confidence intervals (and performs tests, which we have not studied yet) for one proportion. For instance, specify

   Samples in columns: Smoke01

then, in “Options” give the confidence level, say 95% (ignore “Test proportion” and “Alternative” which have to do with testing). Last, check the box for using the normal approximation. Ignore the parts of the output that have to do with testing, and verify that the resulting 95% confidence level interval coincides with the one you computed in (4) above for smoking.

Test and CI for One Proportion: Smoke01
Test of p = 0.5 vs p not = 0.5
Success = 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>X</th>
<th>N</th>
<th>Sample p</th>
<th>95.0% CI</th>
<th>Z-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoke01</td>
<td>12</td>
<td>83</td>
<td>0.144578</td>
<td>(0.068921, 0.220236)</td>
<td>-6.48</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The interval coincides with the one computed above.

6. If you don’t check the box for using the normal approximation, Minitab performs exact calculations with a binomial distribution to compute the confidence interval (and perform testing). Try this out and compare the resulting 95% confidence level interval with the one based on the normal approximation.

Test and CI for One Proportion: Smoke01
Test of p = 0.5 vs p not = 0.5
Success = 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>X</th>
<th>N</th>
<th>Sample p</th>
<th>95.0% CI</th>
<th>Exact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoke01</td>
<td>12</td>
<td>83</td>
<td>0.144578</td>
<td>(0.076999, 0.238934)</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Performing exact calculations with a binomial distribution, for the same confidence level of 95%, produces a slightly broader interval that is not exactly centered at \( p\text{-hat}=0.1446 \).

B. Quantitative variables and sample means

Consider two quantitative variables in the data set, namely

**Hgt:** height in inches

**Wgt:** weight in pounds

1. Find the sample means and corresponding standard errors for Hgt and Wgt. You can use Stat > Basic Statistics > Display predictive statistics. To get the standard error of a sample mean you can divide the sample standard deviation by the square root of n (=83),
or simply read the standard error off the output column “SE Mean” (verify that indeed dividing StDev by the square root of N in the output gives SE Mean)

### Descriptive Statistics: Hgt, Wgt

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>TrMean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hgt</td>
<td>83</td>
<td>65.446</td>
<td>66.000</td>
<td>65.493</td>
<td>2.586</td>
<td>0.284</td>
</tr>
<tr>
<td>Wgt</td>
<td>83</td>
<td>126.64</td>
<td>125.00</td>
<td>126.09</td>
<td>14.72</td>
<td>1.62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hgt</td>
<td>58.000</td>
<td>71.000</td>
<td>64.000</td>
<td>67.000</td>
</tr>
<tr>
<td>Wgt</td>
<td>98.00</td>
<td>175.00</td>
<td>120.00</td>
<td>135.00</td>
</tr>
</tbody>
</table>