In this lab session you will use MINITAB to:

- investigate the probability distribution of simple discrete random variables;
- investigate the difference between $X+X$ and $X+Y$ using simulation;
- evaluate the pmd, cdf and quantiles for the Binomial distribution.

MINITAB is intended to deal with data, i.e. samples, and not populations. However, we can fool it into effectively dealing with a population by using a very big sample.

For example, for the first tutorial problem (4A.4):

the random variable $X$ has pmf given by

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(x)$</td>
<td>0.021</td>
<td>0.085</td>
<td>0.163</td>
<td>0.206</td>
<td>0.195</td>
<td>0.149</td>
<td>0.094</td>
<td>0.051</td>
<td>0.024</td>
<td>0.009</td>
<td>0.003</td>
</tr>
</tbody>
</table>

we imagine we have a “sample” of 210 0s, 850 1s, 1630 2s, and so on. To do this, name C1 ‘$x$’ and enter these data using the `set` command:

```
MTB > set x
DATA> 210(0) 850(1) 1630(2) 2060(3) 1950(4) 1490(5) 940(6) 510(7) 240(8) 90(9) 30(10)
DATA> end
```

We now have a “sample” of 10000, which closely resembles the population specified by the pmf.

Use **Descriptive Statistics** to obtain the mean and quantiles for this “sample”. Compare with your answers for Problem 4A.4.

**exercise 1**: **Descriptive Statistics** also gives the standard deviation, which will approximate $\sigma$.

Evaluate $\mu-2\sigma$ and $\mu+2\sigma$ and hence find $\Pr(\mu-2\sigma < X < \mu+2\sigma)$ in this case.

$[\mu = 3.78, \sigma = 1.90; \Pr(\mu-2\sigma < X < \mu+2\sigma) = \Pr(-0.02 < X < 7.58) = \Pr(0 \leq X \leq 7) = 0.964.]$

**exercise 2**: Use this approach to check the answers for Problem 4A.7(a).

Suppose that $X$ and $Y$ are independent and identically distributed. For example: rolling fair dice, with $X$ the result of one die and $Y$ the result of the other. We use simulation to demonstrate the difference between $X+X$ and $X+Y$.

Open a new worksheet:

File > New . . . and choose “Minitab Worksheet” and OK.

In the new worksheet, name columns C1 and C2 ‘$X$’ and ‘$Y$’. To generate the results of the dice rolling:

```
Calc > Random Data ▶ Integer . . .
Generate 1000 rows of data
Store in column(s):
X  Y
Minimum value: 1
Maximum value: 6
OK
```

This generates 1000 observations on each of $X$ and $Y$ (saving us from rolling two dice 1000 times).

Now, name C3 and C4 as XpX and XpY. Put $X+X$ into XpX, and $X+Y$ into XpY using let as follows:

```
MTB > let XpX = X+X
MTB > let XpY = X+Y
```

Compare the distributions — using Dotplots (multiple Ys) and Descriptive Statistics.

Note that, if $E(X) = E(Y) = \mu$ and $\text{var}(X) = \text{var}(Y) = \sigma^2$:

$E(X+X) = 2\mu$ and $E(X+Y) = 2\mu$; so the means are the same.

$\text{var}(X+X) = 4\sigma^2$ and $\text{var}(X+Y) = 2\sigma^2$; so the variances are different.

And, the dotplots indicate that the distributions are quite different.
Simulate the lecture example of tossing 20 dice and adding up the scores as follows. We do this 1000 times (you can use more if you want!)

Recall that we found \( T = X_1 + X_2 + \cdots + X_{20} \) has mean \( E(T) = 70 \) and \( sd(T) = 7.64 \), so that an approximate 95% confidence interval for \( T \) is \( 55 \leq T \leq 85 \).

Open another worksheet, and generate die-rolling results as follows:

**Calc > Random Data ▶ Integer ...**
- Generate 1000 rows of data
- Store in column(s): C1-C20
- Minimum value: 1
- Maximum value: 6
- OK

This generates 1000 observations on rolling 20 fair dice. To find the sum in each case, we use **Row Statistics**:

**Calc > Row Statistics ...**
- Statistic ⊙ Sum
- Input variables: C1-C20
- Store result in: C21
- OK

This puts the sum \( X_1 + X_2 + \cdots + X_{20} \) into C21 for each row. Name C21 ‘t’, and produce a dotplot and descriptive statistics for \( T \). Compare this to our predictions for \( \mu, \sigma \) and the 95% probability interval.

Note: to check the 95% interval you could plot the sample cdf, using:

**Graph > Empirical CDF ... (Single)**
- Graph variable T
- OK

Compare the distribution of \( X_1 + X_2 + \cdots + X_{20} \) to the distribution of \( 20X_1 \).

MINITAB is useful for evaluating probabilities, cumulative probabilities and quantiles for standard distributions. We consider the Binomial distribution. The routine is the same for any of the specified distributions.

As an example we evaluate \( \Pr(X = 4) \) for \( X \overset{d}{=} Bi(n = 20, p = 0.4) \).

**Calc > Probability Distributions ▶ Binomial ...**
- There are three choices (of which we must choose one):
  - Probability (pmf)
  - Cumulative probability (cdf)
  - Inverse cumulative probability (quantile)
- Selecting Probability gives the pmf. Now we must enter \( n \) and \( p \):
  - Number of trials: 10 (enter \( n \))
  - Probability of success: 0.4 (enter \( p \)).
- There are now two choices to specify the value for which we want to evaluate the pmf. For now we use:
  - Input constant 4 (enter \( x \)).
- The output is \( \Pr(X = 4) \), for \( X \overset{d}{=} Bi(10, 0.4) \); or, in general, \( \Pr(X = x) \), for \( X \overset{d}{=} Bi(n, p) \).

If we select ⊙ Cumulative probability, the output is \( \Pr(X \leq x) \), for \( X \overset{d}{=} Bi(n, p) \).

**exercise 3**: Use MINITAB to evaluate \( \Pr(X = 4), \Pr(X \leq 2) \) and \( \Pr(2 \leq X \leq 4) \), where \( X \overset{d}{=} Bi(20, 0.2) \); \[0.2182, 0.2061, 0.5605\]

\( \Pr(X = 12), \Pr(X \geq 15) \) and \( \Pr(12 \leq X \leq 14) \), where \( X \overset{d}{=} Bi(16, 0.9) \); \[0.0514, 0.5147, 0.4683\]

and check your answers using the Statistical Tables.

Tables of values for the pmf or the cdf can be produced using ⊙ Input column.

Start another worksheet. Name three columns \( x, pmf \) and \( cdf \). We produce a table of values for the \( Bi(10, 0.4) \) pmf. Put 0, 1, 2, ..., 10 into column \( x \):

**MTB > set x**
DATA> 0:10
DATA> end
then

Calc > Probability Distributions ▶ Binomial . . .; select ○ Probability;

enter $n = 10$ and $p = 0.4$; and choose Input column:

○ Input column x (enter the column containing the $x$ values)

Optional storage pmf (enter where you want the Probabilities to be put).

This puts the pmf $p(x)$, $x = 0, 1, \ldots, 10$, into the column we called pmf. If we omitted the optional storage option, these results would be outputted to the Session window.

This can then be used as a table, or graphed. To graph the pmf use Scatterplot with Data View > Data display ⧫ Project lines. Change scale so that the $y$-scale starts at 0 and the $x$-scale is 0:10.

This can be done for the cdf using ○ Cumulative probability, putting the output into column cdf.

Selecting ○ Inverse cumulative probability gives the quantiles.

Again $q$ can be entered as a constant; or, if several quantiles are required, a column of $q$ values can be entered—for example \{0.25, 0.5, 0.75\} for the quartiles.

For a discrete (integer-valued) distribution, in response to a request for the inverse cdf, MINITAB gives the consecutive values $a$ and $b$ for which $F(a) < q$ and $F(b) \geq q$, in the following form:

| Binomial with $n = 25$ and $p = 0.4$ | 9 | 0.424617 | 10 | 0.585775 |

[This is the response to the median of $\text{Bi}(25, 0.4)$.]

We take $c_q = b$ if $F(b) > q$; and $c_q = b + 0.5$ if $F(b) = 0.5$. MINITAB, if forced to make a choice, uses $c_q = b$ in both cases.

**exercise 4**: Use MINITAB to

i. find the quartiles of $X \overset{d}{=} \text{Bi}(100, 0.5)$. \[47, 50, 53\]

ii. find $c_{0.025}$ and $c_{0.975}$ for $Y \overset{d}{=} \text{Bi}(160, 0.6)$. \[84, 108\]

Exactly the same approach applies to evaluating probabilities, cumulative probabilities and quantiles for Hypergeometric and Poisson distributions.

**exercise 5*:**

Compare $\text{Hg}(10, 50, 100)$ and $\text{Bi}(10, 0.5)$.

Compare $\text{Bi}(100, 0.05)$ and $\text{Pn}(5)$.