Generating the Gaussian or Normal Probability Density Function using EXCEL

The instructions below will take you through the generation and plotting of the normal probability density function:

\[ p(x) = \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-m)^2}{2s^2}} \]

using EXCEL. The probability density function provides a graphical display of the probability of occurrence of individual values in the sample of pebble mass assuming it is normally distributed with mean \( \hat{m} \) and standard deviation \( \hat{s} \).

Begin by OPENing EXCEL and, in column I, add the values 100 and 110 in cells I2 and I3. Click on Cell I2 and while holding down the SHIFT key click on the next cell - cell I3. Both cells will be highlighted. Drag the lower right corner of the selected cells down to row 52 or until the scrolling text box reads 600. Let up and the column fills with values from 100 to 600 at 10gram intervals.

Next go over to Column A and in cell number 5 type in mean pebble mass. In cell A6 type in standard deviation of pebble masses.

In cell B5 enter the value 350.18 and in cell B7 enter the value 48 (i.e. the mean and standard deviation of the pebble masses).
The values in column B and rows 5 and 6 can be referred to later in our computation of $p(m)$.

Cell references in EXCEL can be either **absolute or relative**. The absolute reference is useful, because it allows us to refer back to a specific cell and avoids having that value change when a formula cell into which it is entered is copied. The absolute reference is fixed reference. Whether a reference is fixed or absolute depends only on how you refer to it. The following table illustrates various absolute references.

<table>
<thead>
<tr>
<th>Formula</th>
<th>Result when formula is copied</th>
</tr>
</thead>
<tbody>
<tr>
<td>=2*$A1</td>
<td>The column remains constant – fixed</td>
</tr>
<tr>
<td>=2*A$1</td>
<td>The row remains constant – fixed</td>
</tr>
<tr>
<td>=2*$A$1</td>
<td>The row and column remain constant – fixed</td>
</tr>
</tbody>
</table>

Now in our example, the mean mass is stored in cell B5 and the standard deviation in cell B6.

$$p(m) = \frac{1}{\sqrt{2\pi s^2}} e^{-\left(\frac{(m-\hat{m})^2}{2s^2}\right)}$$

When we go to compute the normal probability density of a mass distribution having average value 350.18 and standard deviation 48 then we can make absolute reference to cells B5 and B6.

**Simplifying Complex Calculations:**

Here are some suggestions that might be useful when undertaking complex computations of the sort that you might be asked to do on the job. This advice applies whether you use PSIPlot or EXCEL.

We can break any complicated mathematical expression down into parts that can be solved for individually and then combined later in a complicated mathematical expression. The computation of $p(m)$ is actually not that complex and you might prefer not to take this approach in the future, but the suggestions are made for you to consider. When complicated expressions are broken down into parts, individual parts can be checked for accuracy, and there is less likelihood of error showing up in the final solution.

In the present application, we could compute the factor $\frac{1}{\sqrt{2\pi s^2}}$ separately and store it in a cell for later use. To do this, go to cell B7 and type in $=1/(2*3.141593*$B$6^2)^0.5$. In cell A7 enter - **Inverse of square root of $2\pi s^2$**. This will identify the value 0.0083 that appears in cell B7.
We could also calculate the variable \( z = \frac{m - \hat{m}}{s} \).

Go to cell J2 and enter \((I2-\$B$5)/\$B$6\). This is the \( z \) value for a mass of 100 grams. You should get \( z = -5.21 \) returned to cell J1. You may be saying to yourself that you could just as easily have entered \((I2-350.18)/48\). But again, the purpose of using the absolute cell reference in this case is to give you a simple example of its use. Absolute references can help you organize more complicated calculations than the one we just dealt with. The present example is offered as an illustration of the possible use. Copy this cell into cells J3 through J52.

**Computing the probability distribution \( p(m) \)**

Go to column K and enter the following formula in cell K2:

\[ =\$B$7*EXP(-0.5*J2^2) \]

\( \$B$7 \) is the constant \( \frac{1}{\sqrt{2\pi s^2}} \).

\( J2 \) is \( z \) for a pebble having mass 200 grams. The \( J \)'s equal \( \frac{m - \hat{m}}{s} \).

\( EXP \) is the natural base \( e \).

Copy cell K2 into cells K3 through K52.

Your window should look like the one below.

**Plotting the Probability Distribution \( p(m) \) vs. \( m \)**

Before generating a plot of the probability versus mass select these two columns. We’ll review the procedures for selecting non-adjacent columns. *Take Notes!*
Click on the chart wizard icon on the EXCEL menu bar
(see illustration at right)

This will bring up the Chart Wizard window (below). Select the illustrated options, then click on Next.

The next window to appear requests information about the series of cells that you want to plot. Click on the Series folder (see below).
If you wanted to, you could click on the series tab and replace the I’s with J’s to get the plot of probability versus Z than Mass. However, we have a specific interest in evaluating the probability of occurrence of various pebble masses.

Click OK
Change the plot title and axis labels to suit the problem (see below).

![Chart Wizard - Step 3 of 4 - Chart Options](chart)  

The Chart Location window appears. Take the default, which will place the plot in your worksheet.

![Chart Wizard - Step 4 of 4 - Chart Location](chart)
Your worksheet should look something like that shown below.

![Excel screenshot]

**Changing Absolute Cell Values**
One very powerful aspect of the absolute cell referencing features in EXCEL is that this will allow you to reproduce the entire series of computations for a different mean and different standard deviation simply by changing the values you entered in cells B5 and B6.

Go ahead and change the mean value to 400. Note what happens to your plot. The change is automatically made to your plot. Now change the standard deviation to 100.

**Note these various changes - Experiment**
Plot Esthetics:

For a professional presentation, you may wish to enhance the appearance of your plot with different backgrounds, text colors, etc. Several cosmetic features will be illustrated in the class presentation, so please take notes!

You will find it useful to add the drawing tools. To do this, right click on blank areas in the upper menu panel. The menu shown below will drop down; **Click on Drawing**.

The following menu bar will appear at the bottom of your Excel page.

Click here to create a text box.
NOTES

- Add Text Box
- Format text in text box
- Format background in text box
- Format Cells
- Format Chart
Your completed display might end up looking like that below.

**Statement of problem:** Compute and plot the normal probability densities of pebble masses tabulated in Table 7.1 of Waltham’s text.
Generating Histograms with Excel

Up to this point we have concentrated on the use of PsiPlot to do spreadsheet calculations and basic plotting and analysis of data. This is largely because PsiPlot is easier to use in some respects than Excel, particularly for plotting, plot formatting and regression computation and plotting. The following exercise is provided to increase your familiarity with Excel functionality. Excel will generate data histograms similar to those we developed using PsiPlot.

First click on tools and examine the drop down window (see below).

Look for Data Analysis. If you can’t find it in the list, you will have to install it. To install the Data Analysis Tool Pak click on Add-Ins and then check the Analysis ToolPak option box (see below). Click OK and the add in should take place.
Now return to the Tools menu item and you should now see a Data Analysis option in the list. Select the Data Analysis option.

When you click on Data Analysis in the above list, the following window will open up. Highlight the Histogram Analysis Tool and click OK.
In the following menu, place your cursor in the Input Range box and then with your mouse left-click on the pebble mass data values and slide down selecting the data in that column. When you finish, your screen should look like that shown below.

Click on the BinRange box and select the bin cells provided for you in the worksheet. Finally, click on the Output Range box and then with the mouse sweep through a few rows and columns to give it some cells to output to. Your window should look something like the following.

Also check the Chart Output option box. Then click OK. Your worksheet should look like the following.
Note that the columns of Bin and Frequency appear along with the histogram to the right.

You can edit your histogram plot to increase plot size, change label names, etc.
Left-clicking on the histogram bars will bring up the following window. From there, you can select the options folder and adjust the Gap width in the histogram to eliminate the gaps altogether (set Gap width to 0).

Today’s assignment:

Hand in a plot of the probability density function generated in the first part of this exercise for the pebble mass sample.

Make sure it is correctly labeled and that your name is on it.