QUEEN MARY, UNIVERSITY OF LONDON

MTH 4106 Introduction to Statistics

Practical 8 6 March 2012

Today we will

- see how Minitab simulates data from standard distributions,
- become more familiar with covariance and the correlation coefficient,
- verify some theorems from lectures in particular cases.

1 (Simulating data from a Poisson distribution)

Suppose that $X_1 \sim \text{Poisson}(4)$. Use

`Calc` → `Random Data` → `Poisson...`

to simulate 200 rows of data from this distribution, and store them in a column named `X1`.

Before you go any further, write down $\mathbb{E}(X_1)$ and $\text{Var}(X_1)$.

Now use

`Stat` → `Basic Statistics` ▶ `Display Descriptive Statistics...`

to find out the sample mean of `X1` and the sample variance of `X1`. Are they close to the theoretical values that you wrote down?

2 (Drawing a histogram) A histogram gives a good visual impression of a probability mass function. You can get the simple histogram for `X1` by using

`Graph` → `Histogram...`

Does this histogram have the shape that you expect for a Poisson distribution?

3 (Fitting a theoretical distribution) Right-click in the graph, and choose

`Add` ▶ `Distribution Fit...`

Look at the list of distributions to choose from. "Poisson" is not there, so choose `Normal`. What happens? What does the new graph show?

This particular feature is not very useful for what we are doing this week, but we will come back to it in later weeks.

4 (More Poisson distributions) Let X_2 and X_3 be two further Poisson random variables with $X_2 \sim \text{Poisson}(2)$ and $X_3 \sim \text{Poisson}(6)$. Simulate 200 rows of data for each of these, storing them in columns `X2` and `X3`. Then check that the sample mean, sample variance and histogram for each of them is close to what you expect.

5 (Plotting the simulated data on a scatterplot) Draw a scatterplot with `X2` as the Y variable and `X1` as the X variable. Since `X1` and `X2` were generated independently, there should be no relationship between the two variables. Is that what the scatterplot shows?

If X_1 and X_2 are independent random variables then $\text{corr}(X_1, X_2) = 0$. Use

`Stat` → `Basic Statistics` ▶ `Correlation`

to find the sample correlation coefficient of `X1` and `X2`. Is it close to zero?

6 (Sums of Poisson random variables) Create a column called `T` containing the sum `X1` + `X2` + `X3`. What should the distribution of the random variable $X_1 + X_2 + X_3$ be? (Hint: look at the theorems in the section about probability generating functions.)

Verify that the histogram, sample mean and sample variance of `T` agree with this.

What is the relationship between the sample mean for `T` and those for `X1`, `X2` and `X3`?

7 (Two variables which are not independent) Calculate $\text{Cov}(T, X_1)$ and $\text{corr}(T, X_1)$ by hand, where $T = X_1 + X_2 + X_3$. What sort of relationship do you expect between T and X_1? Verify this by plotting `T` against `X1` and finding the value of the sample correlation coefficient.

8 (Binomial random variables) Let X and Y be independent random variables with $X \sim \text{Bin}(50, 0.7)$ and $Y \sim \text{Bin}(48, 0.7)$. Put $Z = X + Y$.

State the distribution of Z. Calculate $\text{Cov}(X, Z)$ and $\text{corr}(X, Z)$ by hand.

Simulate 200 values of X, Y and Z in Minitab and produce a scatterplot of Z against X. Find the sample correlation coefficient of the simulated values.

9 (Saving and erasing) Before continuing, open a report pad and copy to it those graphs and those parts of the output that will help you to remember what you have learnt in this practical. You may find it helpful to keep a printed version of this report in your lecture notes.

When you have done that, you need to make some space in Minitab, so close down all the graphs and delete the columns of data by using

`Data` $\rightarrow$ `Erase`.

10 (Assignment question) You need to use Minitab to answer part of the Feedback question from Assignment 7, which is copied below.

Let X and Y be independent random variables with $X \sim N(3, 1)$ and $Y \sim N(6, 9)$. Put $Z = Y - 4X + 7$.

1. State the distribution of Z.
2. Calculate $\text{Cov}(X, Z)$.
3. Calculate $\text{corr}(X, Z)$.
4. Simulate 200 values of X, Y and Z in Minitab and produce a scatterplot of Z against X.