THE UNIVERSITY OF SYDNEY
SCHOOL OF MATHEMATICS AND STATISTICS

Tutorial Week 3

MATH1905: Statistics (Advanced) Semester 2, 2017

Web Page: `http://sydney.edu.au/science/maths/MATH1905`
Lecturer: Michael Stewart

Please ask your tutor about any difficulties from week 2.

1. Evaluate the correlation coefficient for the following data set:

x_i:	0.3	0.6	0.9	1.2	1.5	1.8	2.1	2.4
y_i:	10	15	30	35	25	30	50	45

Using your calculator, the value of r^2 is (2dp):

(a) 0.88 (b) 0.99 (c) 0.77 (d) 0.23 (e) none of the above

2. Evaluate the correlation coefficient for the following data set, both by hand and then check it with R:

x_i:	5	3	10	1
y_i:	2	1	5	0

3. In R Type

- `data(swiss)` to obtain the `swiss` data set;
- `attach(swiss)` to obtain the 6 variables from the data frame;
- `help(swiss)` and read information about this data set.

(a) Type `cor(swiss)` to obtain the matrix of pairwise correlations. What are the 3 most correlated pairs?

(b) Produce (separate) scatter plots: `plot(Education,Examination)`, `plot(Education,Fertility)`, `plot(Agriculture,Examination)`, `plot(Catholic,Fertility)`. Do you see any pattern? If yes does it agree with the corresponding correlation. What do we learn from this data analysis?

(c) Type `pairs(swiss)` to obtain all the paired scatter plots. Comment on the plots as well as on the pairwise correlations.

4. Suppose $a = \bar{y} - b\bar{x}$ and $b = S_{xy}/S_{xx}$ (using the usual notation) are the least-squares intercept and slope associated with the points $(x_1, y_1), \ldots, (x_n, y_n)$. Writing $\hat{y}_i = a + bx_i$ for the i-th fitted value and $\hat{\varepsilon}_i = y_i - \hat{y}_i$ for the i-th residual, show that $\sum_{i=1}^n (\hat{y}_i - \bar{y})\hat{\varepsilon}_i = 0$ and hence that $\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n \hat{\varepsilon}_i^2$.

5. Using the fact that for events A, B and C, $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, apply the general addition rule

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{1}$$

repeatedly to prove that

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)\,.$$

6. Two six-sided dice (of different colours) are rolled in such a way that all possible sequences of pairs of values are equally likely to show facing up when the dice come to rest. Let A be the event that a total of strictly less than 4 occurs. Its probability, $P(A)$ is:

(a) $1/6$ (b) $3/6$ (c) $9/36$ (d) $1/36$ (e) $3/36$

7. In the setting of the previous question let B be the event that the total showing is divisible by 3.

(a) Write down the event B.

(b) Determine the conditional probability $P(A|B)$.

8. Suppose that for a group of n students it is known that none of them were born in a leap year. The students line up alphabetically and write their birth date (ignoring the year) in order on a whiteboard. Assuming each possible sequence of n birth dates is equally likely, write an expression (as a function of n) giving the probability that all birth dates are different. Plot this function using R for $n = 2, 3, \ldots, 30$ (**hint**: use the functions `choose()` and `factorial()`). What is the smallest n so that this probability is less than 0.5?