TUTORIAL EXERCISE PACKAGE - 2013

Math1015-Biostatistics

Semester 1 **Tutorial Week 12** 2013

Given $L_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$, $L_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$ and $L_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2$.

The correlation coefficient, which measures the strength of a linear relationship between x and y, is $r = \frac{L_{xy}}{\sqrt{L_{xx}L_{yy}}}$ and r^2 gives the proportion of variability in y explained by the linear regression model.

The linear regression line of y on x is given by $y = a + bx$ where $b = \frac{L_{xy}}{L_{xx}}$ and $a = \bar{y} - b\bar{x}$.

Tutorial discussion: Q1, Q2 Q7 and Q8 marked with *

1. *In an industry, 200 workers, employed at a specific job, were classified according to their performance and training to test independence of the two variables. The data are summarised below:

			Total
	Good	Not good	
Trained	100	50	150
Untrained	20	30	50
Total	120	80	200

Use the χ^2-test of independence at 5% level of significance to draw a conclusion.

2. *Do people have preference for movie type? A random sample of 100 people revealed that 35 of them prefer comedy, 30 horror, 20 drama and 15 sci-fi movies. Use an appropriate test to determine if there is a difference in the proportions of the preferred movie type.

3. (**Multiple choice**) For a set of 12 pairs of observations on (x, y) from an experiment, the scatter plot suggests that it is possible to fit a linear regression model for the data. The following summary for x and y is obtained: $\sum x_i = 25, \sum y_i = 432, \sum x_i^2 = 59, \sum y_i^2 = 15648, \sum x_i y_i = 880$. The sample correlation coefficient between x and y is (2dp):

(a) 7.57 (b) -7.57 (c) -0.08 (d) -0.78 (e) 1.00.

4. (**Multiple choice**) Using the summary statistics from Q3 above, the values of L_{xx}, L_{yy} and L_{xy} are (respectively):

(a) 96, -20, 6.917 (b) 59, 15648, 880.5 (c) 6.917, 96, -20

(d) 56.917, 15612, -20 (e) can not calculate.

5. (**Multiple choice**) The least squares estimate of β in Q3 is:

(a) 2.891 (b) -19.5 (c) 6.917 (d) -2.891 (e) -0.760

6. (**Multiple choice**) The estimated value of y at $x = 5$ for the data in Q3 (from the regression line in Q5) is (2dp) :

(a) 27.57 (b) 47.77 (c) 41.87 (d) 55.97 (e) can not calculate.

7. *A researcher is responsible for showing how carbon dioxide levels in blood influence breathing rates by affecting the acidity of the blood. In one experiment he administered varying doses of sodium bicarbonate with the following results:

Dose (in grams):	x	30	40	50	60	70	80	90	100
Breathing rate (in breath/min):	y	16	14	13	13	11	12	9	9

(a) Calculate the coefficient of correlation.

(b) Find the least squares regression line of breathing rate on dose.

(c) What breathing rate would you predict for a dose of 85g?

(d) Draw a scatter plot of the data and mark the regression line on the diagram.

(e) Draw a diagnostic plot and comment on it.

Use R to answer Q8 to Q9

8. *

(a) Consider the built-in `cars` data set, which gives the speed of cars (in mph) and distance taken to stop (in ft). Note that the data are from the 1920s. Read the data by executing `attach(cars)`.

(b) Find the correlation coefficient between `speed` and `dist` by using the built-in function `cor` and specifying the correct arguments. **Hint:** `cor(dist,speed)`

(c) Fit a simple linear regression line for the above data with `dist` as response variable and `speed` explanatory variable, by using the built-in function `lm(dist~speed)`

(d) Write down the estimated regression line in (c).

9. Do Q7 in `R`.

Semester 1 Problem Sheet Week 12 2013

1. Use the χ^2 tables to answer the following.

(a) If $P(\chi^2_{10} > a) = 0.10$, find a.

(b) If $P(\chi^2_{15} \geq a) = 0.05$, find a.

(c) Bound $P(\chi^2_{22} \geq 36.5)$.

(d) Bound $P(\chi^2_{40} \geq 52.1)$.

2. A market researcher wishes to assess consumers' preference among four different colours available on a name-brand household washing machine. The following frequencies were observed from a random sample of 198 recent sales:

Colour	Avocado	Tan	White	Blue	Total
Observed frequency, O_i	61	55	41	41	198

What frequencies are expected (ie. find E_i's) under the hypothesis that

(a) the four colours are equally likely.

(b) the four colours are in the ratios 6:5:4:3.

3. Calculate X^2 in each case of Q2. Examine each model for goodness of fit using appropriate χ^2 tests.

4. Check the answers for Q2 using `R`.

5. Book P.185, Q10.37-10.38.

6. Book P.219, Q11.1, 11.3 and 11.4.