UNIVERSITY OF CAMBRIDGE
Judge Business School

Updated: October 13, 2011

Lab Instructor: Michael Freeman
Contact: mef35@cam.ac.uk

Contest Quiz 1

Question Sheet

In this quiz we will be looking at salaries in a US bank.

Question 1

Load the bank data set into R from the url: http://thiloklein.de/R/bank. Inspect the data, and answer the following:

I) How many observations are there?
(a) 9 (b) 10 (c) 474 (d) 484

II) Determine the number of variables, excluding "X".
(a) 8 (b) 9 (c) 10 (d) 11

I) What is the `idnumber` for the 2nd from last entry in the data set?
(a) 88 (b) 437 (c) 474 (d) 484

Question 2

Subset the data set, choosing only the `male` employees.

I) How many `male` employees are there?
(a) 104 (b) 216 (c) 258 (d) 370

II) Now subset the `male` employees, only including those with a `salary` strictly over 30,000. How many male employees with a salary over 30,000 are there?
(a) 37 (b) 87 (c) 171 (d) 179

Question 3

Return to using the full data set for now.

I) Inspect the variable `jobcat`. Which of the following is true:

(a) `jobcat` is currently classed as a factor, but should be numeric
(b) `jobcat` is currently classed as numeric, but should be a factor
(c) `jobcat` is currently classed as a factor, and should be a factor
(d) `jobcat` is currently classed as numeric, and should be numeric

II) Are any of the observations of the variable `salary` non-integer?
(a) NO (b) YES

Question 4

Now we will produce summary statistics for the current salary, given by the variable `salary`.

I) What is the mean `salary`, to 2 decimal places?
(a) 23250.00 (b) 28725.00 (c) 33984.72 (d) 36112.50

II) Determine the interquartile range.
(a) 5259.72 (b) 7387.5 (c) 10734.72 (d) 12862.5

III) Determine the 67.5% quantile. (note: leave 'type' at the default value of 7 in the appropriate R function)
(a) 29684 (b) 30010 (c) 33300 (d) 37323

Question 5

Consider the variable `salbegin`, the starting salary of the employee.

I) What is the standard deviation of this variable?
(a) 5982.22 (b) 7933.55 (c) 35786956 (d) 62941229

II) What is the skewness? (rounded to 2 decimal places)
(a) 2.78 (b) 4.19 (c) 8.92 (d) 14.85

III) What is the kurtosis? (rounded to 2 decimal places)
(a) 2.78 (b) 4.19 (c) 8.92 (d) 14.85

Question 6

We will now explore the relationship between starting salary and current salary.

I) What is the covariance between `salbegin` and `salary`?
(a) 62941229 (b) 120877311 (c) 174920501 (d) 418305184

II) What is the correlation between `salbegin` and `salary`? (rounded to 2 decimal places)
(a) -1.23 (b) 0.14 (c) 0.88 (d) 1.102

III) Which of the following statemenets is true?

a) Starting and current salary are strongly negatively correlated, if starting salary is high, current salary is highly likely to be low

b) Starting and current salary are weakly negatively correlated, if starting salary is high, current salary is somewhat likely to be low

c) Starting and current salary are weakly positively correlated, if starting salary is high, current salary is highly likely to be high also

d) Starting and current salary are strongly positively correlated, if starting salary is high, current salary is highly likely to be high also

Question 7

Create a new variable `smartrichmale`, with `smartrichmale` = 1 if the employee is male, with a salary strictly over 30,000, and education strictly over 16.

I) What is the proportion, to the nearest 3 decimal places, of smart rich men in the bank?
(a) 0.091 (b) 0.101 (c) 0.136 (d) 0.171

II) What proportion of smart rich men working in the bank come from a minority?
(a) 0.083 (b) 0.243 (c) 0.341 (d) 0.472

Question 8

We are now interested in how `salary` depends on certain traits of the employees.

I) How much higher is the average `salary` of non-minorities as compared to minorities? (rounded to 2 decimal places)
(a) 1752.21 (b) 2489.28 (c) 4719.91 (d) 8582.04

II) Returning to the `smartrichmale` employees we specified earlier, in our sample, which group exhibits a higher percentage increase in wage from starting (`salbegin`) to current (`salary`) salary?
(a) not a `smartrichmale` (b) is a `smartrichmale`

Question 9

Create two new variables, `logsalary` equal to the natural logarithm of `salary`, and `logsalbegin` equal to the natural logarithm of `salbegin`.

I) Plot `logsalbegin` against `logsalary` (i.e. `logsalbegin` on y-axis, `logsalary` on x-axis). How are the data points distributed?

a) from bottom left to top right of the plot

b) from top left to bottom right of the plot

II) Insert a straight line of best fit through the plot. To the nearest integer, what is the y-intercept of the line?
(a) 1 (b) 6 (c) 9 (d) 12

Question 10

Produce a histogram of `salary`.

I) Is the histogram skewed to the left or right?
(a) left (b) right

II) How many employees (i.e. what frequency) have a `salary` between 20000 and 30000? Select an appropriate band.
(a) 50-100 (b) 100-150 (c) 150-200 (d) 200-250