Chapter II: Bivariate Descriptive Statistics - Exercises

Bernardo D'Auria

Statistics Department

Universidad Carlos III de Madrid

GROUP 89 - COMPUTER ENGINEERING

September 14th, 2010

Exercise Exam I.I. Sep'06 - C4

In a Statistics group of a first-year Computer Engineering, we know that 60% of the students passed the Continuous Evaluation (C.E.) that consists of practical tests made during all the semester. To get full approval of that course they require to be evaluated only on the theoretical part by passing a Final Exam (F.E.). 70% of the ones that did not pass the C.E. did not pass F.E. either that in this case consists of both a theoretical and practical part. Finally we know that the 18% of all students passed both the C.E. and the F.E.

a) What is the percentage of students that **did not pass** both the C.E. and the F.E.?

b) What percentage of students **did not pass** the F.E.?

c) Among the students who **do not passed** the F.E. what percentage **did pass** the C.E.?

Solution

Bivariate Relative joint-Frequency Table

	C.E.		
F.E.	Passed	Not Passes	
Passed	42%	12%	54%
Not Passed	18%	28%	46%
	60%	40%	

a) 28%

b) 46%

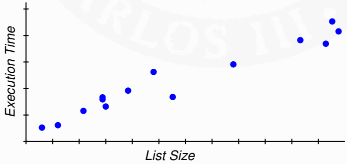
c) 39.13%

Exercise

n	T
65	208
118	327
116	352
66	162
114	296
30	91
60	225
106	305
25	85
38	127
45	138
44	161
85	244
52	178
44	156

We have a computer program that executes a set of tasks with a list containing n objects. We execute the program 15 times using each time a different number of objects and we record its execution times T (seconds). As the CPU has to do other operations during the execution of the program, if we execute two times the software with the same values for the number of objects n we would get every time different values of its execution times. The following table shows the numbers of objects, n, in the list and the associated executions times T . The picture shows the scatterplot of these data.

$\bar{n}$	67.2
$\bar{T}$	203.67
corr(n, T)	0.972
s_n^2	990.03
s_T^2	6770.76



a) How much time should we wait to execute the program with a list made of $n = 90$ elements?

b) If we want that the execution time were less than $T = 100$ seconds, what should the maximum size of the list we could use be?

SOLUTION

a) How much time should we wait to execute the program with a list made of $n = 90$ elements?
We can compute the parameters of the regression line

$$\hat{T}(n) = b_T \cdot n + a_T \quad : \qquad b_T = 2.54 \quad \text{and} \quad a_T = 32.87$$

and find that

$$\hat{T}(90) = 261.62 \text{ sec;}$$

b) If we want that the execution time were less than $T = 100$ seconds, what should the maximum size of the list we could use be?
We can compute the parameters of the regression line

$$\hat{n}(T) = b_n \cdot T + a_n \quad : \qquad b_n = 0.3716 \quad \text{and} \quad a_n = -8.488$$

and find that

$$\lfloor \hat{n}(100) \rfloor = \lfloor 28.67 \rfloor = 28 \text{ elements.}$$

Conditional Absolute and Relative Frequencies

Bivariate Absolute joint-Frequency Table

	y_1	y_2	
x_1	$f_A[1,1]$	$f_A[1,2]$	$f_A[1,\cdot]$
x_2	$f_A[2,1]$	$f_A[2,2]$	$f_A[2,\cdot]$
	$f_A[\cdot,1]$	$f_A[\cdot,2]$	$f_A[\cdot,\cdot]$

f_A stays for Absolute Frequency.

$f_A[\cdot,\cdot]$ is the sample size.

f_R stays for Relative Frequency, e.g.

$$f_R[2,1] = f_A[2,1]/f_A[\cdot,\cdot].$$

Example of Conditional Frequency

absolute

$$f_A[2|y_2] = f_A[2,2] \qquad f_A[2|x_1] = f_A[1,2]$$

relative

$$f_R[2|y_2] = \frac{f_A[2|y_2]}{f_A[\cdot|y_2]} = \frac{f_A[2,2]}{f_A[\cdot,2]} = \frac{f_A[2,2]}{f_A[\cdot,\cdot]}\frac{f_A[\cdot,\cdot]}{f_A[\cdot,2]} = \frac{f_R[2,2]}{f_R[\cdot,2]}$$

$$f_R[2|x_1] = \frac{f_A[2|x_1]}{f_A[\cdot|x_1]} = \frac{f_A[1,2]}{f_A[1,\cdot]} = \frac{f_A[1,2]}{f_A[\cdot,\cdot]}\frac{f_A[\cdot,\cdot]}{f_A[1,\cdot]} = \frac{f_R[1,2]}{f_R[1,\cdot]}$$