Answers Assignment 3

Exercise 1

The joint probability distribution of the random variables *X* and *Y* is shown in the table below, where *X* is the number of tennis racquets and *Y* is the number of golf clubs sold daily in a small sports store.

		X	
Y	1	2	3
1	0.30	0.18	0.12
2	0.15	0.09	0.06
3	0.05	0.03	0.02

1. *X* and *Y* are independent.

True: For independence: P(X,Y)=P(X)P(Y). This must hold for all elements:

P(X=1,Y=1) = 0.30 = 0.60 * 0.5.
P(X=2,Y=1) = 0.18 = 0.60*0.30.
P(X=3,Y=1) = 0.12 = 0.60 * 0.20.
P(X=1,Y=2) = 0.15 = 0.5 *0.3.
P(X=2,Y=2) = 0.3*0.3 =0.09.
P(X=3,Y=2) = 0.3*0.2=0.06.
P(X=1,Y=3) = 0.5*0.1 =0.05.
P(X=2,Y=3) = 0.1*0.3 =0.03.
P(X=3,Y=3) = 0.1*0.2 =0.02.

2. The standard deviation of the number of sold tennis racquets is, rounded to two decimals accuracy, 0.78.

True: E(X2) = 0.5 +0.3*4 + 0.2* 9 = 3.5. V(X) = 3.5 – 1.7^2 = 0.61. Sqrt(0.61)= 0.78.

3. As the sales for tennis racquets and golf clubs are positively correlated, the covariance between the number of sold tennis racquets and golf clubs is positive. That is, COV(X,Y) > 0.

False: X and Y independent -> Cov =0.

4. The expected value of *X* + *Y* is 3.2.

True: E(X+Y) = E(X) +E(Y) = 1.7+1.5 = 3.2

Exercise 2
In a company a large number of PCs are in use. From time to time there are failures of the hard disk, and from time to time there are failures of the screen. Define the following random variables:

H: the total number of failures in a week of the hard disk of all PCs.

S: the total number of failures in a week of the screen of all PCs.

Assume that the number of failures in a week is independent from the number of failures in weeks prior to this particular week, as well as independent from the number of failures in weeks after this particular week.

Based on experience the technical service has established the following joint probability distribution.

S H	0	1
0	0.44	0.16
1	0.21	0.07
2	0.12	0.00

5. Find the expected values of H and S.

First: Complete the table:

B S	0	1	Marginal distribution S
0	0.44	0.16	0.60
1	0.21	0.07	0.28
2	0.12	0.00	0.12
Marginal distribution B	0.77	0.23	1.00

For calculations the following table is useful:

s	P(S=s)	s.P(S=s)	s^2 P(S=s)
0	0.60	0	0
1	0.28	0.28	0.28
2	0.12	0.24	0.48
Som		0.52	0.76

We can calculate:

$\mu_S = E(S) = 0.52 \qquad \sigma_S^2 = 0.76 - (0.52)^2 = 0.4896$

And:

$\mu_B = E(B) = 0.23 \qquad \sigma_B^2 = 0.76 - (0.52)^2 = 0.1771$

6. Find the expected value of the total number of failures in 52 weeks.

We have: $E(X+Y) = E(X) + E(Y)$ (always). Thus, the expected number of failures in a week is: 0.52 + 0.23 = 0.75

Now, this is the expected number for each week so that the expected number of failures in 5 weeks is the sum of 52 of these: 52 x 0.75 = 39

7. Find the covariance between H and S (round off at 4 decimal places).

Use the following table to see that the covariance is -0.0496.

B	S	P(b,s)	$(b-\mu_b)$	$(s-\mu_s)$	$(b-\mu_b)\times(s-\mu_s)$	$(b-\mu_b)\times(s-\mu_s)\times$ P(b,s)
0	0	0.44	-0.23	-0.52	0.1196	0.052624
0	1	0.21	-0.23	0.48	-0.1104	-0.023184
0	2	0.12	-0.23	1.48	-0.3404	-0.040848
1	0	0.16	0.77	-0.52	-0.4004	-0.064064
1	1	0.07	0.77	0.48	0.3696	0.025872
1	2	0.00				0
Som						-0.049600

8. Find the coefficient of correlation (round off at 4 decimal places).

$$\rho = \frac{cov_{BS}}{\sigma_B \sigma_S} = \frac{-0.0496}{\sqrt{0.1771 \times 0.4896}} = -0.1684$$

9. Find the variance of the total number of failures in that week (round off at 4 decimal places).

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y$$

$$cov_{XY} = \rho\sigma_X\sigma_Y$$

Thus: $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\,cov_{XY}$

$$\sigma^2_{B+S} = \sigma^2_B + \sigma^2_S + 2\,cov(B,S) = 0.1771 + 0.4896 - 2 \times 0.0496 = 0.5675$$

Exercise 3

For an elevator, constructed to transport at most 4 adults, the average weight per person should not exceed 80 kg. If it does, the elevator is unhappy, refuses to depart, and makes an awful noise!

Assume that the weight of adults is approximately normally distributed with a mean of 72 kg and a standard deviation of 8 kg.

10. If four adults, at random, enter the elevator, find the probability that it does not depart.

$\overline{X}$ is normal distributed with $\mu = 72$

And standard error

$$\frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{4}} = 4$$

Thus:

$$P(\overline{X} > 80) = P(Z > \frac{80-72}{4} = 2) = 0.0228$$

11. If the probability that the elevator does not depart, is not allowed to be larger than 0.0062, find the average weight per person of 4 adults that should not be exceeded.

Let g denote the value not to be exceeded. Then,

$$P(\overline{X} > g) = P(Z > \frac{g-72}{4}) = 0.0062 .$$

Hence:

$$\frac{g-72}{4} = 2.5 \text{ zodat } g = 82$$

Exercise 4

The Education Service Center of the Erasmus School of Economics evaluates the Applied Statistics 1 course through a questionnaire on SIN-Online.

12. Students participating in this questionnaire form a “simple random sample” (SRS) from the population of students registered for the course.

False: Not a SRS but a voluntary response sample.

Exercise 5

In the appendix of a report we find the following table:

Absence due to illness				
	Department 1		Department 2	
Age category	Number of employees	Absence (%)	Number of employees	Absence(%)
Old	80	10	20	20
Young	20	2	80	4
Total	100	8.4	100	7.2

13. These data present an example of Simpson’s paradox.

True: Dep 1 absence is lower for both groups. But, as there are more old people in Dep 1 (and old people are more often absent here) the total absence is higher for Dep 1.