**Bachelor Degree in Computer Engineering**

**Statistics** **group E (English)**

# FIRST PARTIAL EXAM

March 30th 2015

| | |
|---|---|
| Surname, name | |
| Signature | |

## Instructions

1. Write your name and sign in this page.
2. Answer each question in the corresponding page.
3. All answers must be justified.
4. Personal notes in the formula tables will not be allowed.
5. Mobile phones are not permitted over the table. It is only permitted to have the DNI (identification document), calculator, pen, and the formula tables. Mobile phones cannot be used as calculators.
6. Do not unstaple any page of the exam (do not remove the staple).
7. All questions score the same (over 10).
8. At the end, it is compulsory to sign in the list on the professor's table in order to justify that the exam has been handed in.
9. Time available: **2 hours**.

**1.** Answer the following questions.

**a)** If two data sets are considered (A and B), is it possible that the kurtosis coefficient of A is higher than B but the standard deviation of A is lower than B? Justify your reply. *(2 points)*

**b)** Indicate what is the population under study, the random variable, the variable's dimension and the type of variable of each one of these examples:

1. Number of enquiries attended daily and number of pending requests in certain computer company. *(2 points)*

2. Sales volume of a sample of 100 Spanish companies in the computer business during the year 2014. *(2 points)*

**c)** One company measures certain quality parameter that follows a Normal distribution when the process operates under control. Explain with detail the different patterns of non-normality (i.e. patterns that deviate from the Normal model) that could be detected by means of a frequency histogram of the data, indicating how such histogram would highlight the atypical patterns.

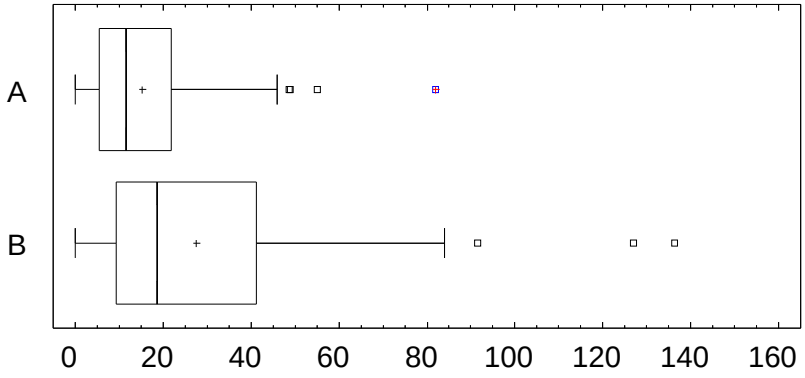
*(4 points)*

**2.** A company that provides certain service by internet works with 3 different servers (A, B and C). It turns out that 25% of requests (accesses) of this service are executed in server A, 35% on B and the rest on C. The company has verified that 1% of requests running on server A present an error while accessing the server, while this percentage is 2% in case of server B. Furthermore, if a request is randomly chosen, the probability of having an error and being executed in server C is 1%.

**a)** If a request is chosen at random, what is the probability of occurring an error while accessing the server? *(5 points)*

**b)** If a request is randomly taken and it turns out to be correct, what is the probability of having been executed on server A? *(5 points)*

**3.** Certain components of two different types (A and B) are used for the manufacturing of an electronic device. After obtaining experimentally the time of operation until failure (in months) of 100 component of each type, the following box-whisker plot is obtained. It is assumed that the values follow an exponential distribution and that the average time of B is 27.5.




**a)** If a circuit is arranged with 4 components in series, two of which are of type A and two of type B (that is: -A-A-B-B-), what is the probability to fail before one year? It is assumed that the failure is independent of the type of component.

*(5 points)*

**b)** What is the minimum number of components of type B that should be connected in parallel to have a reliability > 95% after one year of operation?

*(5 points)*

**4.** In certain communication network, the system that checks errors has to guarantee that the percentage of incorrect packages circulating through the network has to be less than **one** per thousand. For this purpose, in every session, the system randomly selects *n* packages at constant periods of time and verifies if they are correct or not. If the verification system finds 3 or more incorrect packages, the session is not accepted as appropriate and the communication is temporarily stopped to perform a correction procedure.

Determine the minimum value *n* of packages to verify in order to get a probability less than 5 per thousand to accept as appropriate a session not satisfying the requirement. *(10 points)*

**5.** One company manufactures a type of electronic component to be used in computer equipment. Certain quality parameter is controlled which follows a Normal distribution with average m=90. A given component is considered defective (not correct) when this parameter is greater than 95.

**a)** The company has found that 2% of the components produced are defective. Calculate the standard deviation of this quality parameter. *(5 points)*

**b)** If a sample of 2000 components is selected at random, what is approximately the probability to find less than 60 defective components? *(5 points)*

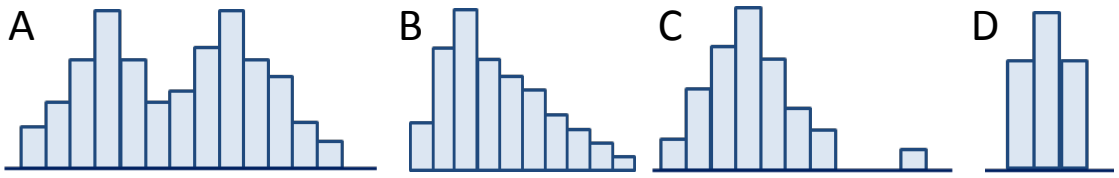
# SOLUTION

**1a)** Yes. The kurtosis coefficient (KC) provides very different information and independent with respect to the dispersion. KC indicates leptokurtic data (if KC>0, taking 0 as reference value), or platykurtic data (KC<0 that might be caused by truncated values). By contrast, the data dispersion, measured for example with the standard deviation, can be high or low regardless the presence of leptokurtic or platykurtic data.

**1b1)** Population: days of the year. Random variable (registered for every day): number of enquiries attended and number of pending requests. Variable's dimension = 2 (two-dimensional). Type: both components of the two-dimensional variable are quantitative and discrete.

**1b2)** Population: all Spanish companies in the computer business. Random variable: sales volume during 2014 of each company (one-dimensional variable). The sales volume is usually expressed in euros and, hence, the type is quantitative and continuous.

**1c)** A histogram obtained from a set of values that presumably follow a Normal distribution can reflect different types of problems:




- Fig. A: Mixture of two populations (for example, quality parameter corresponding to two different products).

- Fig. B: positively skewed distribution.

- Fig. C: Existence of outliers (one or a few values extremely high compared with the rest).

- Fig. D: Truncated values. Data above or below certain value are disregarded and they do not appear in the histogram. This could indicate that values outside the interval considered as "good quality" are discarded.

- Other patterns of non-normality would be: histogram with a platykurtic shape; abnormal frequency of certain value (one bar higher or lower than expected) due to data manipulation (e.g. values rounded up), etc.

**2a)** Events:
A: the request is executed in server A → P(A) = 0.25
B: the request is executed in server B → P(B) = 0.35
C: the request is executed in server C → P(C) =1-P(A)-P(B) = 0.4
E: the request presents an error while accessing the server.
P(E/A)=0.01; P(E/B)=0.02; P(E∩C)=0.01

According to the Total Probability Theorem:
P(E)=P(E∩A)+ P(E∩B)+ P(E∩C)=P(A)·P(E/A)+P(B)·P(E/B)+ P(E∩C)=
= 0.25·0.01+0.35·0.02+0.01 = **0.0195**

**2b)** According to the Bayes' theorem:

$$P\left(A/\overline{E}\right)=\frac{P\left(A\cap\overline{E}\right)}{P\left(\overline{E}\right)}=\frac{P(A)\cdot P\left(\overline{E}/A\right)}{1-P(E)}=\frac{0.25\cdot(1-0.01)}{1-0.0195}=\mathbf{0.2524}$$

**3a)** Event A: component of type A is operative more than one year.
Event B: component of type B is operative more than one year.
Event C: the circuit in series -A-A-B-B- is operative more than one year.

$T_B \approx \exp(\alpha=1/27.5)$ ; $P(B)=P(T_B>12)=e^{-\alpha\cdot t}=e^{-12/27.5}=0.646$

The average value of the data from A is represented as a small cross (+) inside the box, which is approximately 15, so that: $\alpha = 1/15$.

$P(A)=P(T_A>12)=e^{-\alpha\cdot t}=e^{-12/15}=0.449$

$P(C)=P(A_1\cap A_2\cap B_1\cap B_2)=P(A_1)\cdot P(A_2)\cdot P(B_1)\cdot P(B_2)=0.449^2\cdot 0.646^2=0.084$

The probability to fail before one year is the complementary: 1-0.084 = **0.916**

**3b)** Event F: component of type B fails before one year: $F=\overline{B}$

Event D: the circuit with *n* components of type B in parallel <u>fails</u> before 1 year.

Reliability (probability of the circuit to be operative more than 1 year) > 95%; P(D)< 0.05

$P(F)=P(T_B<12)=1-e^{-\alpha\cdot t}=1-e^{-12/27.5}=0.354$

$P(D)=P(F_1\cap\ldots\cap F_n)=P\left(F_1\right)\cdot\ldots\cdot P\left(F_n\right)=\left[P\left(F\right)^n\right]<0.05;\quad 0.354^n<0.05$

n > (log 0.05) / log (0.354) ; n > 2.88 ; $\mathbf{n_{min} = 3}$

**4)** Random var. X = number of incorrect packages in the sample

X ~ B(n, p<0.001)

If X ≤ 2 the session is accepted as appropriate; If X ≥ 3 it is not accepted.

¿n? / P(accept a session / p ≤ 0.001) = P(X ≤ 2 / p ≤ 0.001) ≤ 0.005

P(X ≤ 2 / p ≤ 0.001) = P[B(n, p≤0.001) ≤ 2] = P[Ps(**λ**=n·0.001) ≤ **2**] ≤ 0.005

Looking at the Poisson abacus (reading on curve 2), it is obtained that **λ ≥** 9.25

n·0.001 **≥** 9.25 → n ≥ 9.25/0.001 → **n ≥ 9250**

**5a)** $P\left[N(90,\sigma)>95\right]=0.02$; $P\left[N(0;\,1)>(95-90)/\sigma\right]=0.02$ ; $(95-90)/\sigma=2.05$;

**σ = 2.44**

**5b)** Y= number of defective components in the sample of 2000 components.

Y ≈ B (n=2000, p=0.02). Given that n·p·(1-p)=39.2 > 9 it can be approximated by means of the Normal distribution, with average n·p=40 and variance 39.2 (**σ** = 6.26).

P(Y<60) ≈ P[N(40; 6.26)<59.5] = P[N(0; 1)<3.11] = 1-0.00094 = **0.99906**