Bachelor Degree in Computer Engineering

Statistics

FIRST PARTIAL EXAM

April 2nd 2012

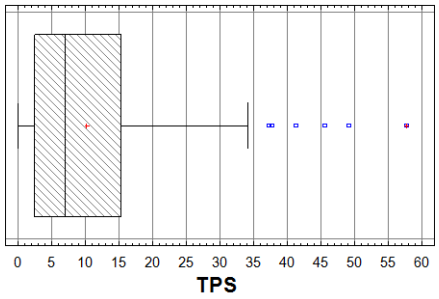
Surname, name	
Signature	

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. 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.
5. Do not unstaple any page of the exam (do not remove the staple).
6. All questions score the same (over 10).
7. 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.
8. Time available: 2 hours.

1. In order to study the performance of an operative system, the time required for processing certain service (TPS) (expressed in ms) has been registered during one week.

The following plot shows the results obtained:



Answer the following questions:

a) What is the variable under study in this case? What type is it? **(2 points)**

b) What is the name of this type of plot? For what types of variables is its use recommended? **(2 points)**

According to the information provided by the plot shown above:

c) What can we say about the distribution of the variable analyzed? **(2 points)**

d) What parameter would provide reliable information about the position of the data analyzed? Justify your answer and calculate, if possible, the approximate value of that parameter. **(2 points)**

e) What parameter would provide reliable information about the dispersion of the data analyzed? Justify your answer and calculate, if possible, the approximate value of this parameter. **(2 points)**

2. In a group of 30 students of Computer Engineering, each one has a Tablet PC that are distributed according to their brands as follows: half of them are Sungsam, 10 are CERA and the rest are PH.

We know that 30% of CERA Tablets were purchased on-line, which also applies to 20% of Sungsam and 40% of PH Tablets.

Answer the following questions justifying conveniently the reply:

a) If one Tablet is randomly chosen, what is the probability of not having been purchased on-line? **(5 points)**

b) If we choose a Tablet that was purchased on-line, what is the probability to be Sungsam? **(5 points)**

3. One company that produces pieces for airplanes is interested in the quality study of the production of one of the pieces manufactured. The quality of this type of piece is determined by the length of one of its parts.
In order to carry out the quality study, the company has decided to conduct a statistical process control for that piece by measuring the key parameter.
The company sets a maximum and minimum tolerance limit for the parameter to measure. It is known that, under normal operative conditions, 99% of pieces manufactured remain within the tolerance limits, and therefore they are regarded as correct.

a) If one sample of 5 pieces is randomly taken, what is the probability of finding 2 defective pieces? **(3 points)**

b) If one sample of 50 pieces is randomly taken, what is the probability of finding at least 3 defective pieces? (use the approximation to the Poisson distribution) **(3.5 points)**

c) If one sample of 10,000 pieces is randomly taken, calculate approximately the probability of finding less than 120 defective pieces. **(3.5 points)**

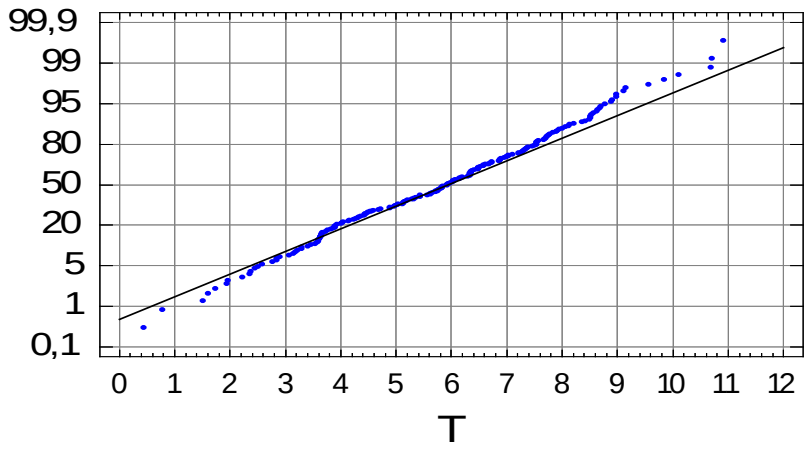
4. In certain corporate computer network, the time between access of users (that is, requests for access or *login*) varies according to an exponential distribution with median 1.04 minutes.

a) Define completely the distribution followed by the random variable under study by specifying which are the fundamental parameter(s). **(4 points)**

b) Calculate the probability of having one request for access in the 30 seconds after setting up the computer network. **(3 points)**

c) Assuming that the network has been running during 3 minutes and no *login* request has occurred, what is the probability of occurring one *login* request in the next 30 seconds? **(3 points)**

5. In order to check the performance of a mass data storage unit, the access time to a set of data stored in this unit have been measured (in milliseconds). This access time fluctuates randomly depending on the position of data within the device. Results have been represented in the chart below:



a) What is the name of this type of chart? **(1 point)**

b) Based on the information provided by the chart, estimate approximately, if possible, the average and the standard deviation of the time for accessing one datum stored in the unit. **(4 points)**

c) What is the probability of taking more than 1750 miliseconds in total for reading a vector of 300 data? **(5 points)**

SOLUTION:

1a) The variable under study is the time required for processing certain service (TPS), expressed in ms. It is a one-dimensional quantitative continuous variable, because the values of time can be measured with decimals.

1b) Box-and-whisker plot. It is recommended for continuous variables or for discrete variables with a high number of different values.

1c) The distribution is characterized by: position, dispersion and shape parameters. TPS ranges from 0 to 57 with an average of 10 and a median of 7. Fifty percent of the values are comprised between 7.5 and 15.5. The distribution is positively skewed (because the right whisker is much longer than the left one) and there is no clear evidence about outliers. The data might follow an exponential distribution because most values are close to zero.

1d) As the distribution is skewed the median is a better parameter of position than the average because it is not affected by extreme values. Median = 7 (line inside the box).

1e) As the distribution is skewed, the interquartile range (IQR) is a better parameter of dispersion than the variance or the standard deviation, because it is not affected by extreme values. IQR = Q3 - Q1 = 15.5 - 2.5 = **13**

2a) The group has 30 tablets: 15 are Sungsmam, 10 are Cera, 5 are PH
Tablets purchased on-line of Sungsam = 0.2 · 13 = 3;
Tablets purchased on-line of Cera = 0.3 · 10 = 3 ; of PH = 04 · 5 = 2.

Event S: the tablet is Sungsam; C: the table is Cera; PH: the tablet is PH
O: the tablet was purchased on-line.

P(S)=15/30=0.5 ; P(C)=10/30 = 0.33 ; P(PH) = 5/30 = 0.167.
P(O/S)=0.2 ; P(O/C)=0.3 ; P(O/PH)=0.4

P(O)=P(O·S)+P(O·C)+P(O·PH) = 3/30 + 3/30 + 2/30 = 8/30
$P(\overline{O}) = 1 - P(O) = 1 - 8/30 = 22/30 = \mathbf{0.73}$

2b) P(S/O) = 3 / (3+3+2) = 3/8
Another way: $P(S/O) = \frac{P(S \cap O)}{P(O)} = \frac{3/30}{8/30} = \mathbf{3/8}$

3a) X: number of defective pieces in the sample; X ~ Bi(n=5, p=0.01)

$$P(X=2) = \binom{5}{2} \cdot 0.01^2 \cdot 0.99^3 = \frac{5!}{2!\cdot 3!} 0.01^2 \cdot 0.99^3 = 10 \cdot 0.01^2 \cdot 0.99^3 = \mathbf{0.00097}$$

3b) X ~ Bi(n=50, p=0.01) Approximation: X ~ Ps (λ=n·p=0.5)
P(X≥3)=1 - P(X≤2) = 1-0.984 = **0.016** (the approximate value 0.984 is obtained from the Poisson abacus, reading on the curve "2" with λ=0.5).

Exact value assuming a Poisson distribution:

$$P(X \geq 3) = 1 - e^{-0.5}\left(\frac{0.5^0}{0!} + \frac{0.5^1}{1!} + \frac{0.5^2}{2!}\right) = 1 - e^{-0.5}(1 + 0.5 + 0.125) = \mathbf{0.014}$$

3c) X ~ Bi(n=10000, p=0.01) Approximation:

$$X \approx N\left(m = n{\cdot}p;\ \sigma^2 = n\cdot p\cdot(1-p)\right) \approx N\left(100;\ \sigma^2 = 99\right)$$

$$P(X<120) \cong P\left[N(100;\sqrt{99})<119.5\right] = P\left[N(0;1) < \frac{119.5-100}{\sqrt{99}}\right] = P\left[N(0;1)<1.96\right] = \mathbf{0.975}$$

4a) $P(X>x) = e^{-\alpha\cdot x}$; $P(X>1.04) = 0.5 = e^{-\alpha\cdot 1.04}$; $-1.04\alpha = \ln 0.5$; $\alpha = 0.6665$

In an exponential distribution, the fundamental parameter is α:

X ~ exp (α=0.6665).

4b) P(T < 0.5 min)= $1-e^{-0.6665\cdot 0.5} = \mathbf{0.2834}$

4c)

$$P(T<3.5/T>3) = \frac{P\left[(T<3.5)\cap(T>3)\right]}{P(T>3)} = \frac{P(3<T<3.5)}{P(T>3)} = \frac{P(T>3)-P(T>3.5)}{P(T>3)} =$$

$$= 1-\frac{P(T>3.5)}{P(T>3)} = 1-\frac{e^{-0.6665\cdot 3.5}}{e^{-0.6665\cdot 3}} = \mathbf{0.2834}$$

Same result as in 4b) due to the "lack of memory" property of the exponential distribution:

$$= 1-\frac{P(T>3.5)}{P(T>3)} = 1-P\left[T>(3.5-3)\right] = 1-P(T>0.5) = P(T<0.5)$$ [question 4b]

5a) The name of the chart is Normal Probability Plot

5b) Median = 6 (value of T that corresponds to a 50% in the vertical scale). As all points fit to a straight line, the distribution is approximately Normal, which implies that median ≈ **mean=6**.

Percentile 2.5% ≈ 2 (value of T that corresponds approximately to 2.5% in the vertical scale).

In a Normal distribution: percentile 2.5% = $m-2\sigma = 2$; $6-2\sigma = 2$; **σ=2**

5c) X~N(6; 2) ; $X_1 + ... + X_{300} \approx N\left(m = 300\cdot 6;\ \sigma^2 = 300\cdot 2^2 = 1200\right)$

$$X \approx N\left(m = 1800; \sigma = \sqrt{1200}\right);$$

$$P\left[N\left(1800;\sqrt{1200}\right) > 1750\right] = P\left[N(0;1) > \frac{1750-1800}{\sqrt{1200}}\right] = P\left[N(0;1) > -1.44\right] = \mathbf{0.925}$$