Practice: Introduction to Statistical Inference

Bernardo D'Auria

Statistics Department

Universidad Carlos III de Madrid

GROUP 89 - COMPUTER ENGINEERING

November 4th, 2010

Practice Exam - June 2007

The file *AlumnosIndustriales.sf3* contains information over a sample of 95 students. The variable "Dinero" denotes the amount of money (pesetas) that they had with them.

a) Fit a Normal distribution to this variable, (try to transform data if necessary). What is the best fit? [Explain your answer]

Practice Exam - June 2007

The file *AlumnosIndustriales.sf3* contains information over a sample of 95 students. The variable "Dinero" denotes the amount of money (pesetas) that they had with them.

a) Fit a Normal distribution to this variable, (try to transform data if necessary). What is the best fit? [Explain your answer]

The distribution does not look Normal: the histogram is asymmetric and the chi-squared test has a very low p-value. Since data are positive asymmetric we look for a transformation of type x^c with $c < 1$. Selecting $c = 0.40$ the histogram looks like a normal and the chi-squared test gives a p-value$= 0.539421$ (All transformations with $0.25 < c < 0.55$ are valid). We use a Normal fitting for $X^{0.40}$.

Practice Exam - June 2007

The file *AlumnosIndustriales.sf3* contains information over a sample of 95 students. The variable "Dinero" denotes the amount of money (pesetas) that they had with them.

a) Fit a Normal distribution to this variable, (try to transform data if necessary). What is the best fit? [Explain your answer]

b) Using the above fitting compute the probability that a student has more than 500 pesetas with him/her.

Practice Exam - June 2007

The file *AlumnosIndustriales.sf3* contains information over a sample of 95 students. The variable "Dinero" denotes the amount of money (pesetas) that they had with them.

a) Fit a Normal distribution to this variable, (try to transform data if necessary). What is the best fit? [Explain your answer]

b) Using the above fitting compute the probability that a student has more than 500 pesetas with him/her.

Assuming $X^{0.40} \sim N(13.593, 7.21556^2)$ we have

$$\Pr(X > 500) = \Pr(X^{0.40} > 500^{0.40}) = \Pr(X^{0.40} > 12.011) = 0.586774$$

Practice Exam - June 2007

A web server receives accesses from independent clients at a constant average speed of 60 access/minute. The server can handle at full speed a maximum of 80 accesses in one minute. After that amount it works under saturation and the users start to be rejected. Compute by using STATGRAPHICS:

a) What proportion of time the server is under saturation?

Practice Exam - June 2007

A web server receives accesses from independent clients at a constant average speed of 60 access/minute. The server can handle at full speed a maximum of 80 accesses in one minute. After that amount it works under saturation and the users start to be rejected. Compute by using STATGRAPHICS:

a) What proportion of time the server is under saturation?

$X =$ accesses in one minute $\sim \text{Po}(\lambda = 60)$.

$$\Pr(X > 80) = 0.0056$$

Practice Exam - June 2007

A web server receives accesses from independent clients at a constant average speed of 60 access/minute. The server can handle at full speed a maximum of 80 accesses in one minute. After that amount it works under saturation and the users start to be rejected. Compute by using STATGRAPHICS:

a) What proportion of time the server is under saturation?

b) What is the minimum capacity the server should have to reduce the saturation probability to less than 0.001?

Practice Exam - June 2007

A web server receives accesses from independent clients at a constant average speed of 60 access/minute. The server can handle at full speed a maximum of 80 accesses in one minute. After that amount it works under saturation and the users start to be rejected. Compute by using STATGRAPHICS:

a) What proportion of time the server is under saturation?

b) What is the minimum capacity the server should have to reduce the saturation probability to less than 0.001?

We look in ICDF for the value C such that $\Pr(X > C) = 0.001$, and we get $C = 85$.

Practice Exam - June 2007

A web server receives accesses from independent clients at a constant average speed of 60 access/minute. The server can handle at full speed a maximum of 80 accesses in one minute. After that amount it works under saturation and the users start to be rejected. Compute by using STATGRAPHICS:

a) What proportion of time the server is under saturation?

b) What is the minimum capacity the server should have to reduce the saturation probability to less than 0.001?

c) What is the probability to have an interval of time of more than 5 seconds with no accesses?

Practice Exam - June 2007

A web server receives accesses from independent clients at a constant average speed of 60 access/minute. The server can handle at full speed a maximum of 80 accesses in one minute. After that amount it works under saturation and the users start to be rejected. Compute by using STATGRAPHICS:

a) What proportion of time the server is under saturation?

b) What is the minimum capacity the server should have to reduce the saturation probability to less than 0.001?

c) What is the probability to have an interval of time of more than 5 seconds with no accesses?

$T =$ time in seconds between two accesses $\sim \text{Exp}(\lambda = 1 \text{ acc/sec})$.

$$\Pr(T > 5) = 0.0067$$

Practice Exam - September 2007

The file *VelViento730.sf3* contains information about wind speeds (in meter per second) measured each hour into two wind parks (*Park 1* and *Park 2*)

a) Fit a Normal distribution to the variable *Parque1*, (try to transform data if necessary). What is the best fit? [Explain your answer]

Practice Exam - September 2007

The file *VelViento730.sf3* contains information about wind speeds (in meter per second) measured each hour into two wind parks (*Park 1* and *Park 2*)

a) Fit a Normal distribution to the variable *Parque1*, (try to transform data if necessary). What is the best fit? [Explain your answer]

The distribution does not look Normal: the histogram is asymmetric and the chi-squared test has a very low p-value. Since data are positive asymmetric we look for a transformation of type x^c with $c < 1$. Selecting $c \in [0.40, 0.50]$ the histogram looks more symmetric like a normal however the chi-squared test gives a p-value< 0.05 (not good). We decide to use a Normal fitting for $X^{0.45}$.

Practice Exam - September 2007

The file *VelViento730.sf3* contains information about wind speeds (in meter per second) measured each hour into two wind parks (*Park 1* and *Park 2*)

a) Fit a Normal distribution to the variable *Parque1*, (try to transform data if necessary). What is the best fit? [Explain your answer]

b) Using the above fitting compute the probability that in a hour the wind speed in Park 1 is greater than $5\,m/s$.

Practice Exam - September 2007

The file *VelViento730.sf3* contains information about wind speeds (in meter per second) measured each hour into two wind parks (*Park 1* and *Park 2*)

a) Fit a Normal distribution to the variable *Parque1*, (try to transform data if necessary). What is the best fit? [Explain your answer]

b) Using the above fitting compute the probability that in a hour the wind speed in Park 1 is greater than $5\,m/s$.

Assuming $X^{0.45} \sim N(9.62, 0.57^2)$ we have

$$\Pr(X > 5) = \Pr(X^{0.45} > 5^{0.45}) = \Pr(X^{0.45} > 2.063) = 1 - 0.46 = 0.54$$

Practice Exam - June 2008

The clients that access to web server of a travel agency do it in an independent way and with an average rate of 5 accesses per minute. What is the probability that in one given minute 10 clients simultaneously access?

Practice Exam - June 2008

The clients that access to web server of a travel agency do it in an independent way and with an average rate of 5 accesses per minute. What is the probability that in one given minute 10 clients simultaneously access?

$X =$ accesses in one minute $\sim \text{Po}(\lambda = 5)$.

$$\Pr(X > 10) = 0.0137$$

Practice Exam - June 2008

We have a batch made of 200 items produced y a mechanical process that on average makes 1% of defective items. The production of a defective item is independent of the state of other articles, and the rate of the production of defective items is constant. We are interested in the random variable

$$X = \text{number of defective items in a batch.}$$

What is the probability to produce 2 or less defective items?

Practice Exam - June 2008

We have a batch made of 200 items produced y a mechanical process that on average makes 1% of defective items. The production of a defective item is independent of the state of other articles, and the rate of the production of defective items is constant. We are interested in the random variable

$$X = \text{number of defective items in a batch.}$$

What is the probability to produce 2 or less defective items?

$X \sim \text{Bin}(200, 0.01)$.

$$\Pr(X < 3) = 0.6767$$

Practice Exam - June 2008

The file *FrenoITV.sf3* contains information about the kilometers traveled by a sample of vehicles that go to a ITV service station.

a) Fit a Normal distribution to the variable *KM*, (try to transform data if necessary). What is the best fit? [Explain your answer]

Practice Exam - June 2008

The file *FrenoITV.sf3* contains information about the kilometers traveled by a sample of vehicles that go to a ITV service station.

a) Fit a Normal distribution to the variable *KM*, (try to transform data if necessary). What is the best fit? [Explain your answer]

The distribution does not look Normal: the histogram is asymmetric and the chi-squared test has a very low p-value (0.002). Since data are positive asymmetric we look for a transformation of type x^c with $c < 1$. Selecting $c = 0.5$ the histogram looks more symmetric and the chi-squared test gives a p-value$= 0.386$. Therefore we use a Normal fitting for $X^{0.5}$.

Practice Exam - June 2008

The file *FrenoITV.sf3* contains information about the kilometers traveled by a sample of vehicles that go to a ITV service station.

a) Fit a Normal distribution to the variable *KM*, (try to transform data if necessary). What is the best fit? [Explain your answer]

b) Using the above fitting compute the percentage of vehicles that go to the ITV service station with more than 150.000 traveled kilometers.

Practice Exam - June 2008

The file *FrenoITV.sf3* contains information about the kilometers traveled by a sample of vehicles that go to a ITV service station.

a) Fit a Normal distribution to the variable *KM*, (try to transform data if necessary). What is the best fit? [Explain your answer]

b) Using the above fitting compute the percentage of vehicles that go to the ITV service station with more than 150.000 traveled kilometers.

Assuming $X^{0.5} \sim N(340.5, 95.98)$ we have

$$\Pr(X > 150000) = \Pr(X^{0.5} > 387.2983) = 0.3131$$

Practice Exam - September 2008

A productive process produces defective items with a proportion of 3 each 1000 items. We have a batch of 200 items made by this productive process.
What is the probability that the batch contains more than 1 defective items?

Practice Exam - September 2008

A productive process produces defective items with a proportion of 3 each 1000 items. We have a batch of 200 items made by this productive process.
What is the probability that the batch contains more than 1 defective items?

Assuming
$X =$ defective items in a batch of 200 items $\sim \text{Bin}(200, 0.003)$.

$$\Pr(X > 1) = 0.1217$$

Practice Exam - September 2008

The file *radiacestacion.sf3* contains information about the intensities of solar radiation in a given geographical region.

a) Fit a Normal distribution to the variable *Radiac_Loc2*, (try to transform data if necessary). What is the best fit? [Explain your answer]

Practice Exam - September 2008

The file *radiacestacion.sf3* contains information about the intensities of solar radiation in a given geographical region.

a) Fit a Normal distribution to the variable *Radiac_Loc2*, (try to transform data if necessary). What is the best fit? [Explain your answer]

The distribution does not look Normal: the histogram is asymmetric even if the chi-squared test gives a p-value$= 0.14$ that could be considered acceptable. Since data are negative asymmetric we look for a transformation of type x^c with $c > 1$. Selecting $c \in [1.5, 2]$ the histogram looks more symmetric and the chi-squared test gives an higher p-value. We use a Normal fitting for $X^{1.75}$ that gives a p-value$= 0.726$.

Practice Exam - September 2008

The file *radiacestacion.sf3* contains information about the intensities of solar radiation in a given geographical region.

a) Fit a Normal distribution to the variable *Radiac_Loc2*, (try to transform data if necessary). What is the best fit? [Explain your answer]

b) Using the above fitting compute the probability that one day the solar radiation will be more than 1000 units.

Practice Exam - September 2008

The file *radiacestacion.sf3* contains information about the intensities of solar radiation in a given geographical region.

a) Fit a Normal distribution to the variable *Radiac_Loc2*, (try to transform data if necessary). What is the best fit? [Explain your answer]

b) Using the above fitting compute the probability that one day the solar radiation will be more than 1000 units.

Assuming $X^{1.75} \sim N(658833, 349774^2)$ we have

$$\Pr(X > 1000) = \Pr(X^{1.75} > 17783) = 0.9666$$

Practice Exam - June 2009

A machine produces on average 8% of defective items with independence and at constant speed. We produce a batch of 100 items.
Compute the probability to have 7 or more defective items in a batch.

Practice Exam - June 2009

A machine produces on average 8% of defective items with independence and at constant speed. We produce a batch of 100 items.
Compute the probability to have 7 or more defective items in a batch.

Assuming
$X =$ defective items in a batch of 100 items $\sim \text{Bin}(100, 0.08)$.

$$\Pr(X > 7) = 0.55289$$

Practice Exam - June 2009

A web server receives on average 10 accesses per minute, we can assume that these accesses arrive with independence and at constant speed.
What is the probability that there are no accesses during 2 minutes?

Practice Exam - June 2009

A web server receives on average 10 accesses per minute, we can assume that these accesses arrive with independence and at constant speed.
What is the probability that there are no accesses during 2 minutes?

$X =$ number of accesses in 1 minute $\sim \text{Po}(\lambda = 10)$.
$Y =$ number of accesses in 2 minute $\sim \text{Po}(\lambda = 20)$.

$$\Pr(Y = 0) = 2.06 \times 10^{-9}$$

Practice Exam - June 2009

The file *IndiceMC.sf3* contains information about the Body Mass Index (kg/m^2) measured in a sample of graduate students.

a) Fit a Normal distribution to the variable *Indice de Masa Corporal*, (try to transform data if necessary). What is the best fit? [Explain your answer]

Practice Exam - June 2009

The file *IndiceMC.sf3* contains information about the Body Mass Index (kg/m^2) measured in a sample of graduate students.

a) Fit a Normal distribution to the variable *Indice de Masa Corporal*, (try to transform data if necessary). What is the best fit? [Explain your answer]

The variable is unimodal and positive asymmetric (Skewness=0.886). We can improve its symmetry by using a transformation of the kind X^c with $c < 1$. Without transforming data the p-value of the chi-squared test is $p = 0.10$ that could be considered already acceptable. However if we transform data we can improve the fitting. For example using $c = -1.6$ we get a skewness of only -0.01 and a p-value equal to 0.51. The histogram of the transformed variable looks very similar to a Normal distribution, therefore our model is

$$Y = X^{-1.6} \sim N(0.00725; 0.00117^2)$$

Practice Exam - June 2009

The file *IndiceMC.sf3* contains information about the Body Mass Index (kg/m^2) measured in a sample of graduate students.

a) Fit a Normal distribution to the variable *Indice de Masa Corporal*, (try to transform data if necessary). What is the best fit? [Explain your answer]

b) Using the above fitting compute the percentage of students with Body Mass Index greater than 25.

Practice Exam - June 2009

The file *IndiceMC.sf3* contains information about the Body Mass Index (kg/m^2) measured in a sample of graduate students.

a) Fit a Normal distribution to the variable *Indice de Masa Corporal*, (try to transform data if necessary). What is the best fit? [Explain your answer]

b) Using the above fitting compute the percentage of students with Body Mass Index greater than 25.

Assuming $Y = X^{-1.6} \sim N(0.00725; 0.00117^2)$ we have

$$\Pr(X > 25) = \Pr(X^{-1.6} < 25^{-1.6}) = \Pr(Y < 0.0058) = 0.10$$