Prueba de evaluación continua 1 - PEC1

RESOLUCIÓN

- Presentación y objetivos
- Enunciados: descripción teórica de la práctica a realizar
- Materiales
- Criterios de evaluación
- Formato presentación
- Fecha presentación

Presentación y objetivos

Objetivos

El objetivo principal de esta **PEC1** es llegar a dominar los conceptos básicos asociados a la estadística descriptiva (incluyendo datos agrupados) y la distribución normal.

Presentación de la prueba

La prueba consta de **tres ejercicios** más **dos prácticas** de Minitab.

En cada **ejercicio**, el alumno tendrá que resolver y responder razonadamente las preguntas que se formulan sin recurrir al Minitab.

Las **prácticas**, sin embargo, se resolverán con la ayuda del programa Minitab.

Tenéis que enviar **un único fichero DOC, RTF ó PDF** con vuestras respuestas a los ejercicios y las prácticas. Cuando el fichero tenga un tamaño considerable -debido a las capturas de pantalla que incluyáis en la respuesta a las prácticas Minitab- debéis enviarlo comprimido.

Parte 1 de la PEC: Preguntas (a resolver sin Minitab)

EJERCICIO 1

La siguiente tabla recoge la distribución del tiempo (en minutos) de conexiones a la red por los usuarios de una biblioteca pública, calculada a partir de una muestra de 200 consultas:

Tiempo	0-10	10-20	20-30	30-40
Frecuencia	60	90	40	10

a) Representa gráficamente la distribución.
b) Calcula mediana y media, y compara los 2 valores.
c) Calcula la varianza y el coeficiente de variación. ¿Qué mide este segundo?
d) Si el tiempo se expresa en horas, ¿cuáles serían los valores de la media y la varianza de la nueva distribución?
e) ¿Por encima de qué tiempo se encuentran el 25% de las conexiones más largas?

RESOLUCIÓN

a) Histograma o box-plot (1Q=8.33; Mediana=19.44; 3Q=25)
Efectivamente:

$$Q1 = 0 + (\frac{50-0}{60} * 10) = 8.33 \qquad Mediana = 10 + (\frac{100-60}{90} * 10) = 14.44$$

b) $$\overline{X} = \frac{(5*60)+(15*90)+(25*40)+(35*10)}{200} = \frac{3000}{200} = 15$$

La proximidad de los estadísticos media y mediana indica que la distribución es bastante simétrica.

c)

$$S^2 = \frac{\sum (X_i - \overline{X})^2 n_i}{n-1} = \frac{(5-15)^2 * 60 + (15-15)^2 * 90 + (25-15)^2 * 40 + (35-15)^2 * 10}{200-1} =$$

$$= \frac{6000+0+4000+4000}{199} = 70.35$$

$$CV = \frac{S}{\overline{X}} = \frac{\sqrt{70.35}}{15} = 0.56$$

El coeficiente de variación es una medida relativa de la dispersión de una distribución.

d) La media quedaria dividida por 60 y la varianza dividida por el cuadrado de esta constante (3600).

e) Dado que nos piden el 25% más alto, eso se corresponderia con el valor del tercer cuartil.

$$Q_3 = L_{i-1} + \frac{3.\frac{N}{4} - N_{i-1}}{n_i}.c_1 = 10 + \frac{3.\frac{200}{4} - 60}{90}.10 = \mathbf{20}$$

Resultado que ya se podia deducir de la simple observación de los datos

EJERCICIO 2

Si Z es una variable aleatoria del tipo N(0, 1), encontrar k en cada uno de los casos siguientes:

a) $P(Z \leq k) = 0'6772$

b) $P(Z \leq k) = 0'2004$

c) $P(-3 \leq Z \leq k) = 0'9893$

d) $P(|Z| \leq k) = 0'8444$

Resolución

a) Buscamos en tablas el valor 0'6672 o el que más se aproxime, y en este caso encontramos: 0'46. Entonces, k = 0'46

b) Si realizamos el mismo proceso, tendremos k = 0'84

c) En este caso el resultado no es inmediato, primero hemos de ver que la probabilidad de que el intervalo esté entre z = -3 y z = k, es 0'9893, y para encontrar la k le hemos de añadir el valor de la probabilidad a la izquierda de –3 que es 0'0013.

   Entonces: 0'9893 + 0'0013 = 0'9906 que equivale a una k = 2'35.

O bien:

$$P(-3 \leq Z \leq k) = 0'9893$$
$$P(-3 \leq Z \leq k) = P(Z \leq k) - P(-3 \leq k) = 0'9893$$
$$P(Z \leq k) = 0'9893 + P(-3 \leq k) = 0'9893 + 0'00135 = 0'99065 \Rightarrow \mathbf{k = 2.35}$$

d) En este caso, sabemos que la probabilidad en el intervalo: -k, k vale 0'8444 y sabemos también que la probabilidad que queda a la derecha de k es la misma que queda a la izquierda de –k.

   Entonces, si hacemos: (1-0'8444)/2 = 0'0778 (para repartir el área a la derecha y a la izquierda de k) y buscamos la k correspondiente a este valor encontramos: k = 1'42.

   O bien:

$$P(|Z| \leq k) = 0'8444$$

   P(- k < Z < k) = 0.8444

$$P(|Z| \leq k) = P(-k \leq Z \leq k) = P(Z \leq k) - P(Z \leq -k) = 1 - P(Z \leq -k) - P(Z \leq -k) =$$
$$1 - 2 \times P(Z \leq -k) = 0'8444$$
$$\frac{1 - 0'8444}{2} = 0'0778 = P(Z \leq -k) \Rightarrow \text{k = 1-0.0778=0.9222}$$

k = 1.42

EJERCICIO 3

Una empresa lleva a término una prueba para seleccionar nuevos empleados. Por la experiencia de pruebas anteriores, se sabe que las puntuaciones siguen una distribución normal de $\mu = 80$ i $\sigma = 25$

¿Qué porcentaje de candidatos obtendran entre 75 y 100 puntos?

Resolución

Nos piden encontrar unas probabilidades relativas a una variable X que sigue una distribución N(80, 25). Entonces, primero hemos de estandarizar la variable para poder utilizar las tablas de la distribución N(0, 1) y después calcularemos la probabilidad de que la variable X tome valores entre 75 y 100.

Para estandarizar, haremos: $z = \frac{x - \mu}{\sigma}$

Como en nuestro caso se trata de encontrar: $P(75 \leq X \leq 100)$

Para X = 75 tendremos: (75-80)/25 = -0’2
Para X = 100 tendremos (100-80)/25 = 0’8

Entonces, hemos de buscar la probabilidad para el intervalo: $P(-0'2 \leq z \leq 0'8)$

Y si ahora vamos a las tablas (pg. 244 del manual) vemos los resultados correspondientes para estos valores:

z = -0’2 = 0’4207
z = 0’8 = 0’7881

Entonces, ya podremos decir que el porcentage de candidatos que obtendran entre 75 y 100 puntos será: 0’7881 – 0’4207 = 0’3674 => 36’74%.

Parte 2 de la PEC: Prácticas Minitab

(se recomienda consultar la web www.uoc.edu/in3/e-math , sección materiales)

PRÁCTICA 1

Una determinada Universidad privada española que imparte la licenciatura en Documentación, oferta asignaturas semestrales que pueden ser cursadas presencialmente o bien, telemáticamente. Tras 7 años en funcionamiento de este sistema, se realiza un estudio sobre una determinada asignatura y se quiere conocer el porcentaje de estudiantes insatisfechos de una y otra modalidad. Los datos que aparecen a continuación, corresponden a los últimos 12 semestres:

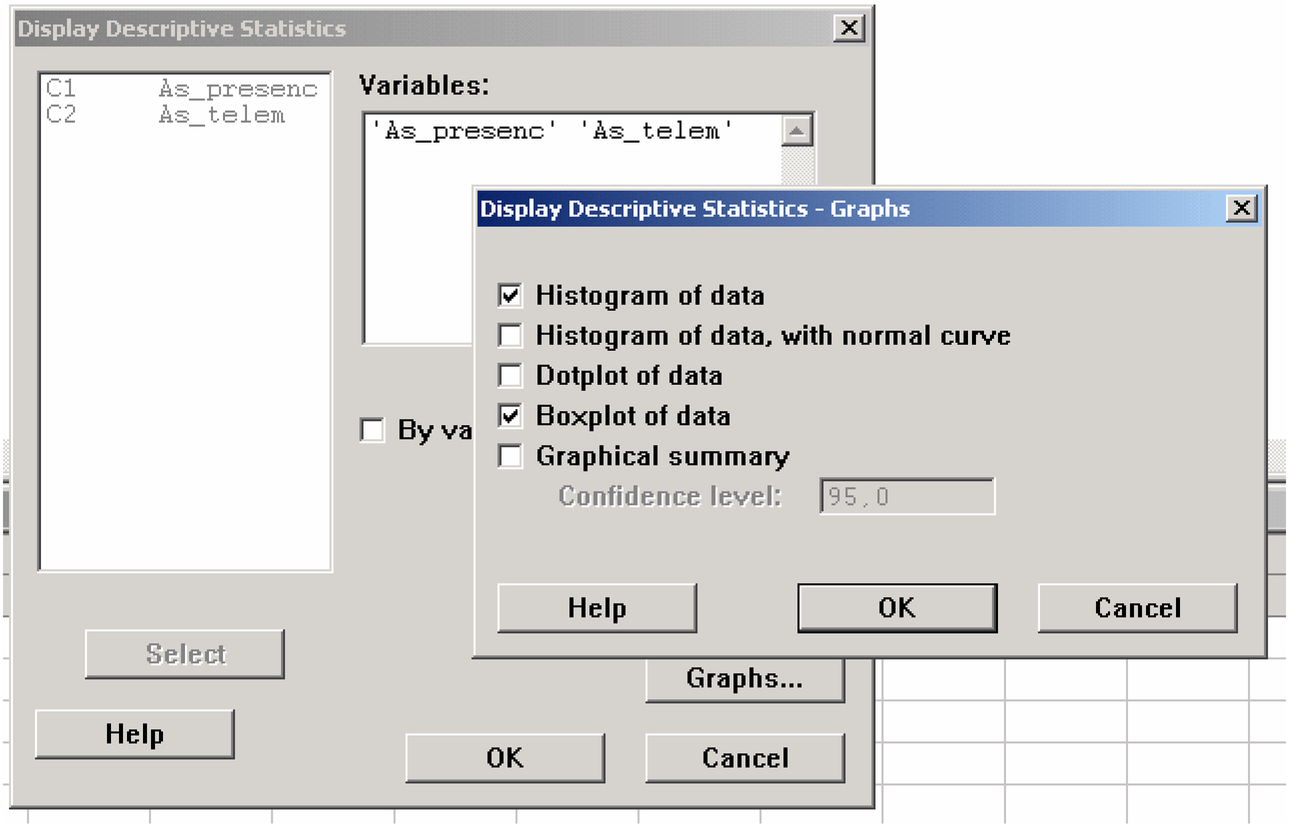
Estudiantes insatisfechos (Asignatura presencial)	Estudiantes insatisfechos (Asignatura telemática)
18	22
12	14
17	17
13	15
11	15
10	12
17	20
16	21
9	12
17	23
14	18
13	19

a) Calcula los estadísticos descriptivos de ambas variables y representa los datos en un “boxplot” y en un histograma. Comenta los resultados (valores “outliers”, comparación de parámetros (p.e., medias de ambas variables),...)

b) Razona si ambas distribuciones son o no simétricas y comenta si, a partir de los datos anteriores, existen diferencias significativas en cuanto a porcentaje de insatisfacción respecto a cursar dicha asignatura presencial o telemáticamente.

Resolución

a) Para calcular los estadísticos descriptivos, insertamos los datos de ambas variables en Minitab y seleccionamos la opción ***Stat > Basic Statistics > Display Descriptive Statistics***, y dentro de la opción ***Graphs***, seleccionamos las opciones "Histogram of data" y "Boxplot of data":



El resultado que obtenemos es el siguiente:

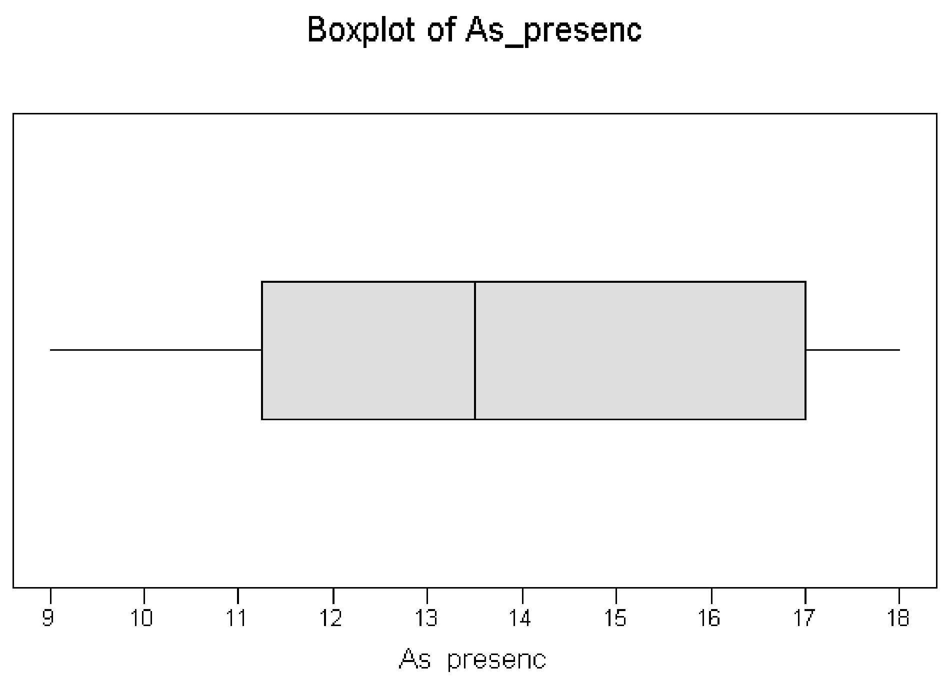
Descriptive Statistics: As_presenc; As_telem

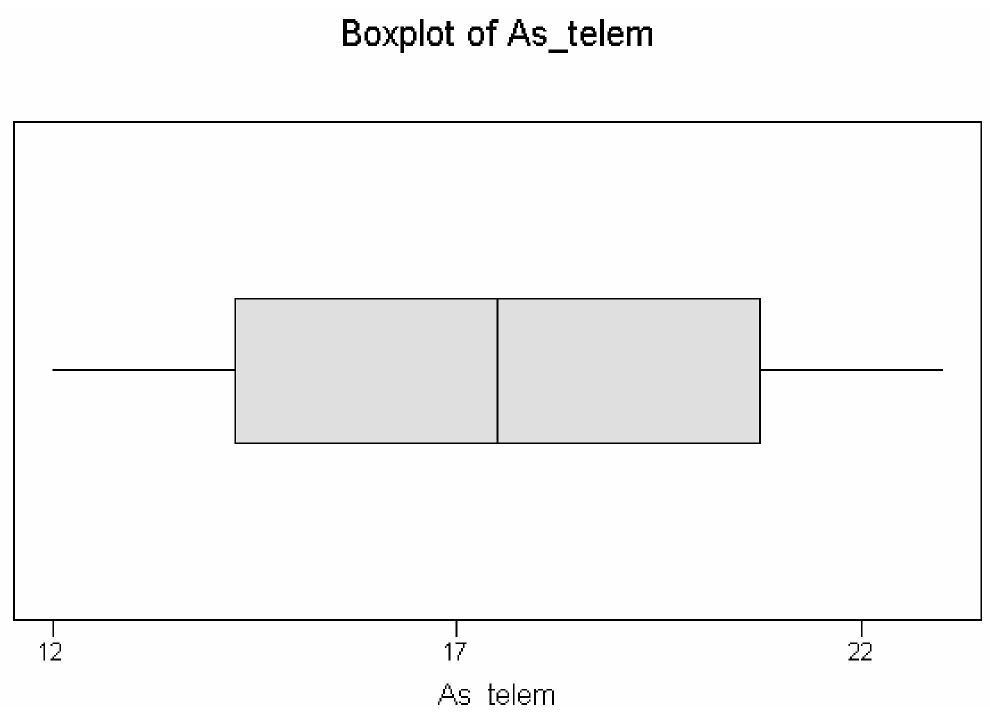
```
Variable           N     Mean   Median   TrMean    StDev  SE Mean
As_prese          12   13,917   13,500   14,000    3,059    0,883
As_telem          12    17,33    17,50    17,30     3,77     1,09

Variable     Minimum  Maximum       Q1       Q3
As_prese       9,000   18,000   11,250   17,000
As_telem       12,00    23,00    14,25    20,75
```

De estos resultados podemos deducir que el grado de insatisfacción de los estudiantes que han cursado la asignatura telemáticamente es ligeramente superior a aquellos que la han cursado presencialmente.

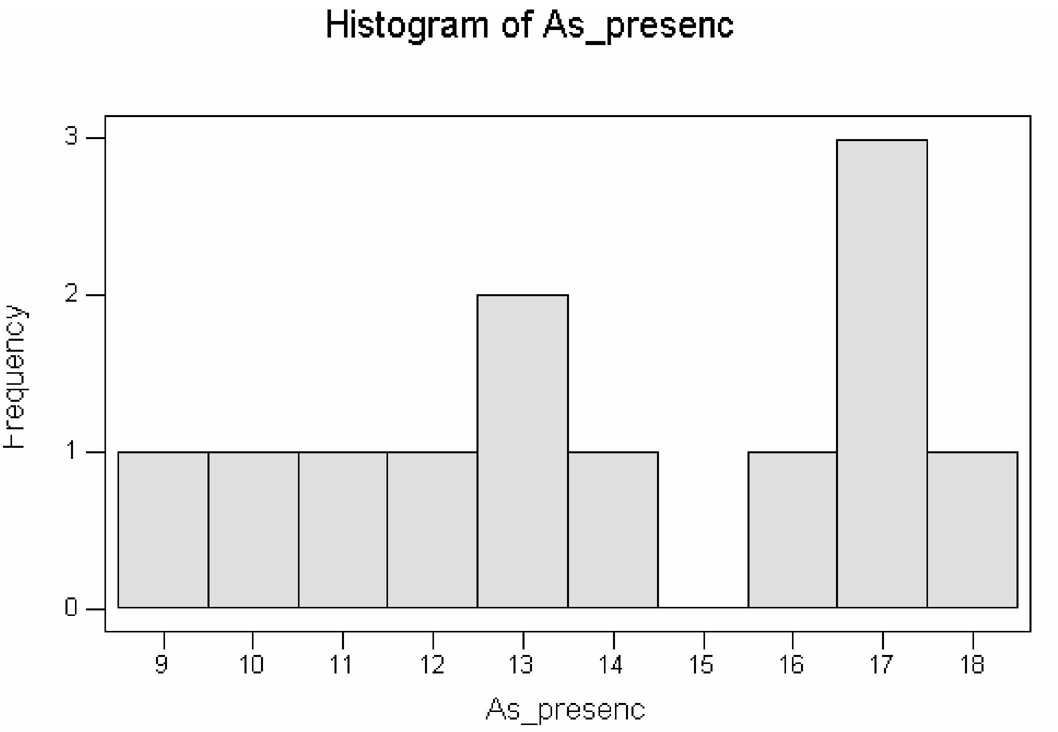
Los gráficos "boxplot" obtenidos son:

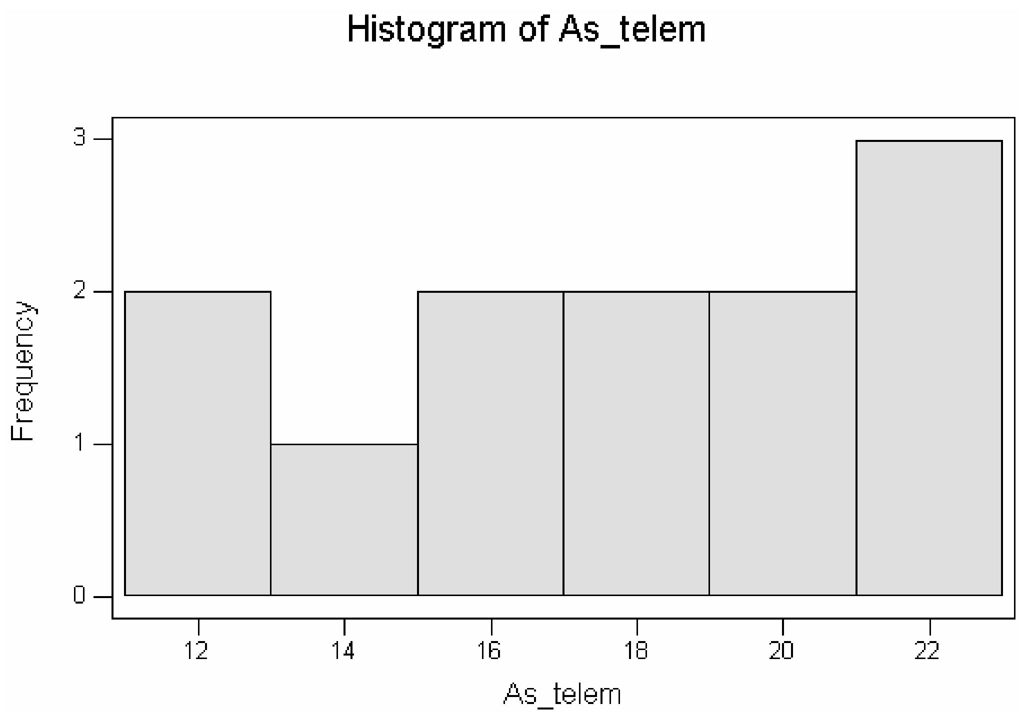




De ambos gráficos deducimos que no hay ningún valor "outlier" y observamos, entre otras características, que los valores máximos son superiores en la variable "Asignatura telemática"

Los histogramas resultantes son:





De ambos histogramas podemos deducir qué valores se repiten con mayor frecuencia, alrededor de 17 en el caso de "Asignatura presencial" y alrededor de 22 en el caso de "Asignatura telemática".

b) De los datos anteriores podemos deducir lo siguiente:

a. En el caso de "Asignatura presencial", la distribución es prácticamente simétrica, ya que la mediana y la media estan muy próximas, pero como la mediana es ligeramente inferior a la media, diremos que la distrubución es ligeramente sesgada a la derecha.
b. En cuanto a "Asignatura telemática", ocurre lo mismo que en el caso anterior, la distribución es prácticamente simétrica. No obstante, como la mediana es ligeramente superior a la media, diremos que la distribución es ligeramente sesgada a la izquierda.

Gráficamente, podemos observar este fenómeno en los "boxplot" obtenidos en el apartado anterior.

Por otra parte, a partir de los estadísticos descriptivos obtenidos podríamos concluir que el porcentaje de estudiantes insatisfechos al realizar la asignatura telemáticamente es ligeramente superior a los que la han realizado presencialmente. Sin embargo, estos datos no son suficientes para afirmar que existen diferencias significativas ya que pueden haber otro tipo de factores no contemplados y, por ello, necesitaríamos realizar un estudio de mayor profundidad.

PRÀCTICA 2

Siguiendo con el caso anterior, sabemos que los estudiantes que estan matriculados de dicha asignatura telemáticamente, acceden vía web a consultar los materiales, prácticas, mensajes del consultor, etc. con cierta regularidad. Se sabe que el número de estudiantes que, diariamente, leen los mensajes que el consultor envia para describir las tareas del dia, sigue una distribución normal de media 82, con una desviación estándar de 65.

a) Si tomamos como muestra aleatoria los 64 últimos dias, ¿cuál es la probabilidad de que la media de accesos sea inferior a 80?

b) A partir de la muestra anterior, ¿qué probabilidad existe de que la media de accesos esté entre 80 y 100?

c) Si quisieramos obtener una probabilidad del 90%, ¿cuántos accesos diarios se tendrían que producir?

d) ¿En qué situaciones se utiliza el Teorema Central del Límite?

Resolución

a) Si denotamos por X = "número de accesos al aula de la asignatura", sabemos que X sigue una distribución normal de media 82 y desviación estándar 65, esto es, X: N(82,65)

En el caso de una muestra aleatoria de tamaño n>30, la distribución de la media muestral, $\bar{X}$, seguirá una normal, en concreto:

$$\bar{X} : N(82,65/\sqrt{64})=N(82,8.125)$$

Tenemos que calcular P($\overline{X}$ < 80). Usemos la opción *Calc* > *Probability Distributions* > *Normal*:

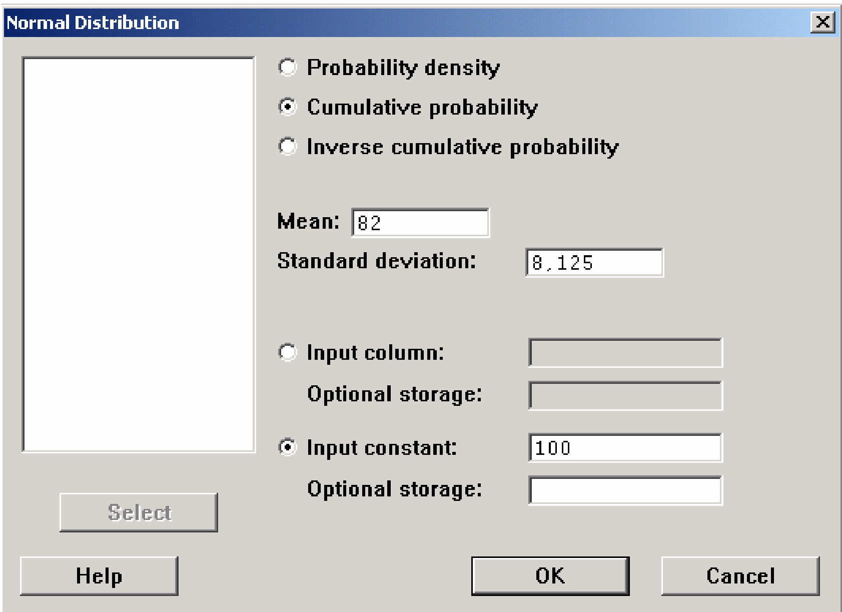
```
Cumulative Distribution Function

Normal with mean = 82,0000 and standard deviation = 8,12500

         x     P( X <= x )
   80,0000        0,4028
```

Por tanto, la probabilidad de que el número de accesos sea inferior a 80 es de aproximadamente, el 40%

b) Se pide calcular, P(80<$\overline{X}$ <100). Procedemos de forma análoga al caso anterior y obtenemos:




```
Cumulative Distribution Function

Normal with mean = 82,0000 and standard deviation = 8,12500

         x     P( X <= x )
  100,0000        0,9866
```

Por tanto, P(80<$\overline{X}$ <100)=P($\overline{X}$ <100)-P($\overline{X}$ <80) = 0.9866 – 0.4028 = 0.5838

Es decir, el porcentaje de que el número de accesos esté entre 80 y 100 es de aprox. el 58%

c) Se pide calcular c, tal que $P(\overline{X} < c) = 0{,}90$

Seleccionamos *Calc* > *Probability Distributions* > *Normal*, y seleccionamos la opción "*Inverse Cumulative Probability*":

```
Inverse Cumulative Distribution Function

Normal with mean = 82,0000 and standard deviation = 8,12500

P( X <= x )          x
    0,9000       92,4126
```

Es decir, hay una probabilidad del 90% de que el número medio de visitas diarias al aula sea inferior a aprox. 92.

d) Se utiliza cuando la distribución de la población es desconocida o no es normal de forma que la distribución muestral de la media se aproximará a la normal conforme aumenta el tamaño de la muestra, en la práctica suele valer para n>30. Para un tamaño d la muestra pequeño diremos que la distribución muestral de la media es aproximadamente normal pero para un tamaño grande de la muestra aleatoria, la distribución muestral de la media seguirá una normal, es decir a medida que aumenta el tamaño de la muestra la distribución de la media tiende a ser casi normal.

$$\bar{X} \cong N\left(\mu_{\bar{X}} = \mu; \sigma_{\bar{X}} = \sigma / \sqrt{n}\right)$$

La importancia de este teorema es que nos permite utilizar los estadísticos de la muestra para poder analizar lo que pasa en la población sin conocer la distribución que sigue la población o si no sigue una distribución normal.

Materiales

Para trabajar esta **PEC1**, se necesita haber trabajado:

- Los módulos 1 al 11.
- Las GES y videos correspondientes
- La documentación asociada al software Minitab (aunque no es imprescindible, se recomienda consultar la web www.uoc.edu/in3/e-math, sección materiales)

Criterios de evaluación

La práctica se evaluará atendiendo a los siguientes criterios:

- Cada uno de los **ejercicios** se puntuará con un máximo de **2 puntos**.
- Cada **práctica** de Minitab se puntuará con un máximo de **2 puntos**.
- La puntuación máxima se consigue cuando **la respuesta y su justificación** son correctas.

Formado de entrega

La **PEC** se entregará en un documento con formato **DOC, RTF ó PDF** en el buzón de: Ev.Continua que tenéis en el aula.

El nombre del fichero/documento tendrá el siguiente formado:

"**apellido1_apellido2_PEC1.doc**"

Los apellidos se escribirán sin acentos. Por ejemplo, el archivo de un estudiante nombrado Alfredo García Melgar seguiría la siguiente denominación: “garcia_melgar_PEC1.doc”.

Fecha de entrega

La **fecha límite de entrega son las 24 horas del día XX/XX/2006**. Recordad que las fechas límite tienen que respetarse estrictamente.

Asimismo, el día XX/XX/2006, se publicará la resolución de la **PEC1**.