Prueba de evaluación continua 1 - PEC1

Solución

- Presentación y objetivos
- Enunciados: descripción teórica de las prácticas a realizar
- Materiales
- Criterios de evaluación
- Formato de entrega
- Fecha de entrega

Presentación y objetivos

Objetivos

El objetivo principal de esta **PEC1** es llegar a dominar los conceptos básicos asociados a la estadística descriptiva (medidas de centralización y de dispersión, muestreo, …), así como ser capaz de calcular probabilidades a partir de la distribución normal.

Presentación de la prueba

La prueba consta de **dos 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, sólo con tablas y calculadora.

Las **prácticas**, se resolverán con la ayuda del programa Minitab. Por ello no solos tenéis que enviar el fichero Word con los ejercicios desarrollados, sino también el ficheros de Minitab con extensión ***.MPJ**.

Enunciados de la PEC1

Parte 1 de la Práctica: Preguntas

EJERCICIO 1

Hemos realizado un experimento que consiste en medir las velocidades instantáneas (en km/h) de 92 vehículos que han pasado por un determinado punto de una autopista. Los resultados son:

145	110	190	155	150	190	155	116	120
135	160	160	130	175	153	138	121	108
150	150	116	150	130	102	123	112	180
215	130	165	140	150	95	120	150	155
145	180	125	122	155	157	190	115	175
133	170	115	155	150	140	135	140	150
138	148	155	150	142	164	155	155	170
135	170	150	125	118	170	180	108	130
145	190	185	125	160	110	130	155	136
140	118	131	125	145	145	195	125	160
120	155							

Se pide:

E1.1. Identifica la variable. ¿De qué tipo de datos estamos hablando?

La variable es la velocidad en que un coche pasa por delante de la autopista. Se trata de datos de tipo cuantitativo o numéricas; además, dentro de las cuantitativas son continuas.

E1.2. Haz un gráfico de tallos y hojas, una tabla de frecuencias absolutas y un histograma.

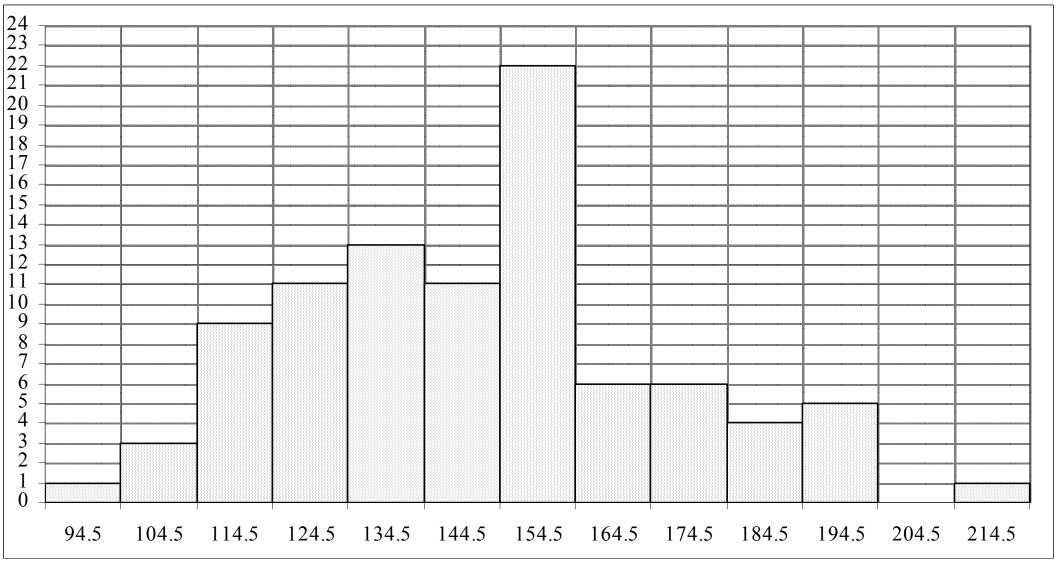
Según cómo se distribuyan los datos, podemos hacer diferentes gráficos de tallos y hojas:

9	5
10	288
11	002556688
12	00012355555
13	0000013555688
14	00002555558
15	0000000000355555555557
16	000045
17	000055
18	0005
19	00005
20	
21	5

A partir de este gráfico, obtenemos la tabla de frecuencias:

Intervalo	Marca de clase	Frecuencia
90-99	94.5	1
100-109	104.5	3
110-119	114.5	9
120-129	124.5	11
130-139	134.5	13
140-149	144.5	11
150-159	154.5	22
160-169	164.5	6
170-179	174.5	6
180-189	184.5	4
190-199	194.5	5
200-209	204.5	0
210-219	214.5	1

Que nos sirve para hacer el histograma:



E1.3. Comenta brevemente la distribución que habéis obtenido (utilizando el vocabulario del Módulo didáctico).

La distribución es asimétrica por la derecha, es decir, tiene una cola por la derecha correspondiente a los coches que pasan más rápidamente por delante del colegio. Tiene un valor alejado o insólito dentro del intervalo 210-219, concretamente el valor máximo (215).

E1.4. Calcula la media y la desviación típica.

El valor de la media será la suma de los valores que tenemos partido por el número de valores, es decir,

$$\bar{x} = \frac{\sum_{i=1}^{92} x_i}{92} = \frac{13354}{92} = 145.15$$

y la desviación típica será la raíz cuadrada de la varianza, es decir,

$$s^2 = \frac{\sum (x_i - 145.15)^2}{92 - 1} = \frac{51283.87}{91} = 563.56$$

$$s = \sqrt{563.56} = 23.74$$

E1.5. Calcula los cinco números resumen.

El mínimo es 95 y el máximo 215. Como tenemos 92 datos (número par), la mediana se situará entre las dos puntuaciones que quedan en medio, en que son ambas 145. La mediana será pues 145. Así nos quedan dos mitades de 46 cifras cada uno. La primera mitad se puede dividir en dos trozos de 23 valores cada uno y los dos números que tenemos en medio también son iguales: el 125. El primer cuartil será, entonces, el 125. En contraste, en la segunda mitad quedan en medio el 155 y el 157, es decir, en que el tercer cuartil será el 156.

E1.6. ¿Cuáles de los indicadores calculados en los apartados 4 y 5 crees que son más correctos para medir la dispersión de esta distribución? Razona la respuesta.

Como la distribución es asimétrica es mejor utilizar los 5 números resumen en la media y la desviación, que se utilizan más bien con distribuciones simétricas.

E1.7. Elabora una tabla de frecuencias relativas (en porcentajes) y de frecuencias relativas acumuladas.

La tabla de frecuencias relativas será

Intervalo	Frecuencia relativa (%)	Frecuencia relativa acumulada(%)
90-99	1.09	1.09
100-109	3.26	4.35
110-119	9.78	14.13
120-129	11.96	26.09
130-139	14.13	40.22
140-149	11.96	52.17
150-159	23.91	76.09
160-169	6.52	82.61
170-179	6.52	89.13
180-189	4.35	93.48
190-199	5.43	98.91
200-209	0.00	98.91
210-219	1.09	100.00
Total	100.00	

E1.8. ¿Cuál es el porcentaje de coches que pasaron por delante de una escuela a 160 km/hora o a más velocidad?

Tenemos en la tabla anterior que un 76.09% de los coches pasaron a menos de 160 Km/hora. Así pues, si restamos este porcentaje de 100 encontraremos los que pasaron a 160 Km/hora o más, que serán: 100 - 76.09 = 23.91%.

E1.9. Al hacer muchas más observaciones, vemos que la distribución de la velocidad con la que los coches pasan por delante de la escuela sigue una distribución normal con una media de 140 Km/hora y una desviación típica de 25 Km/hora. A partir de éstos parámetros, responde a las siguientes cuestiones:

a) ¿Cuál será el intervalo de velocidades con la que pasarán el 68% de los coches por delante del colegio?

Un 68% de los coches pasarán a una velocidad entre la marcada por la media más una desviación estándar y la media menos una desviación estándar, es decir:

$$(\mu \pm \sigma) = (140 \pm 25) = (115 : 165)$$

Es decir, un 68% de los coches pasarán entre 115 y 165 Km/hora.

b) ¿Cuál será la proporción de coches que pasarán a una velocidad superior a los 200 Km/hora?

Primero estandarizamos esta puntuación de 200 Km/hora.

$$z = \frac{x-\mu}{\sigma} = \frac{200-140}{25} = 2.4$$

Si miramos en la tabla el área bajo la curva normal estándar para la puntuación 2.4 nos saldrá 0.9918 (un 99.18%). Sin embargo, esta área corresponde a los coches que pasan a MENOS de 200 Km/h. Para saber la proporción de aquéllos que pasarán a MÁS de 200 Km/h. encontramos la proporción complementaria, es decir,

100 - 99.18 = 0.82; en conclusión, 0.82% de los coches pasaran por delante de la escuela a más de 200 km/hora.

c) ¿Cuál será la proporción de coches que pasarán entre 80 y 100 Km/h.?

Primero tenemos que ver cuántos pasarán con 100 o menos kilómetros por hora.

Estandarizamos $z = \frac{x-\mu}{\sigma} = \frac{100-140}{25} = -1.6$

Miramos a la tabla que una Z de -1. 6 corresponde a una área de 0.0548 (es decir un 5.48%). En definitiva, un 5.48% de los coches pasan a menos de 100 Km/h.

A continuación miramos cuántos pasan a 80 Km/hora o menos.

Estandarizamos $z = \frac{x-\mu}{\sigma} = \frac{80-140}{25} = -2.4$

Miramos a la tabla que una Z de -2. 4 corresponde a una área de 0.0082 (es decir un 0.82%). En definitiva, un 0.82% de los coches pasan a menos de 80 Km/h.

El resultado de la pregunta lo encontraremos restando a la opción de menos de 100 Km/h. lo que hemos encontrado con menos de 80 Km/hora. Así, 5.48 - 0.82 = 4.66; un 4.66% de los coches pasan delante del colegio entre 80 y 100 Km/h.

EJERCICIO 2

Para aprobar unas oposiciones se necesita obtener un mínimo de 100 puntos en una prueba. Por experiencias anteriores se sabe que la distribución de los puntos obtenidos por los opositores es una normal de media 110 puntos, y una desviación estándar de 15 puntos.

Se pide:

E2.1. ¿Cuál es la probabilidad de que un opositor apruebe?

Dado que los datos siguen una distribución normal con $\overline{X}$ =110, y s = 15, y como para aprobar se necesitan 100 puntos, si estandarizamos y buscamos el resultado en las tablas, encontramos el valor: 0'7486 que es la probabilidad de que un opositor apruebe. Eso se podría dar como un porcentaje y sería el 74 '86% de posibilidades de aprobar.

E2.2. ¿Si sabemos que hay 1.000 opositores y sólo 300 plazas, cuántos puntos se tendrán que exigir para ajustar el número de plazas al número de opositores aprobados?

Teniendo en cuenta que tenemos 1000 opositores y 300 plazas con la misma distribución normal del apartado anterior, tenemos que encontrar la probabilidad de aprobar condicionado al número de plazas.

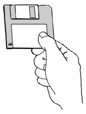
1000*p = 300 => 300/1000 = 0'3, que es la probabilidad de aprobar.

Si ahora buscamos en las tablas de la normal el valor: 1 - 0'3 = 0'7 encontramos que eso corresponde a una z = 0'52, entonces, si x es el número de puntos necesarios para aprobar, haríamos:

$$(x - 110)/15 = 0'52 \Rightarrow x = 117'8$$

y si exigimos 117 '8 puntos para aprobar, ya estamos ajustando el número de aprobados al número de plazas vacantes.

Parte 2 de la Práctica: Prácticas Minitab



PRÁCTICA 1

Una compañía aérea quiere hacer un estudio para averiguar cuál es el porcentaje de viajeros insatisfechos después de haber hecho una reserva de vuelo a través del Portal de la compañía en Internet, y cuál es el porcentaje de viajeros insatisfechos que han hecho la reserva a través de una agencia de viajes.

En la siguiente tabla aparecen los datos correspondientes a los últimos veinte meses.

Insatisfechos (Billete electrónico)	Insatisfechos (Billete por agencia)
18	22
12	14
17	17
13	15
11	15
10	12
17	20
16	21
9	12
17	23
14	18
13	19
15	22
18	24
14	21
8	10
19	29
18	25

10 15	14 15

Se pide:

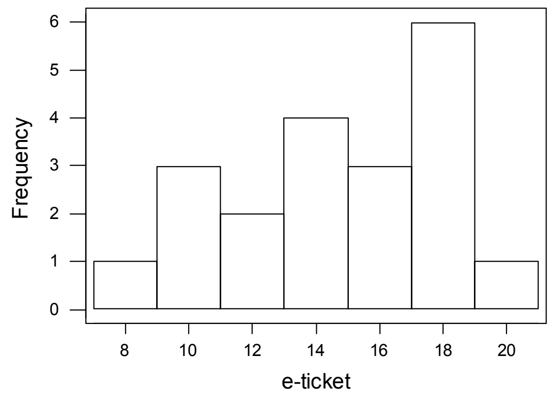
P1.1. Realizar un diagrama de tallos y hojas para cada grupo de datos, así como un histograma y un diagrama de cajas.

A continuación se muestran los diagramas de tallos y hojas para las dos series de datos anteriores.

Seleccionamos ***Graph > Stem-and-Leaf***

```
Stem-and-leaf of e-ticket N = 20
Leaf Unit = 0,10

    1    8 0
    2    9 0
    4   10 00
    5   11 0
    6   12 0
    8   13 00
   10   14 00
   10   15 00
    8   16 0
    7   17 000
    4   18 000
    1   19 0


Stem-and-leaf of Otros N = 20

Leaf Unit = 1,0


    1    1 0
    3    1 22
    8    1 44555
    9    1 7
   (2)   1 89
    9    2 011
    6    2 223
    3    2 45
    1    2
    1    2 9
```

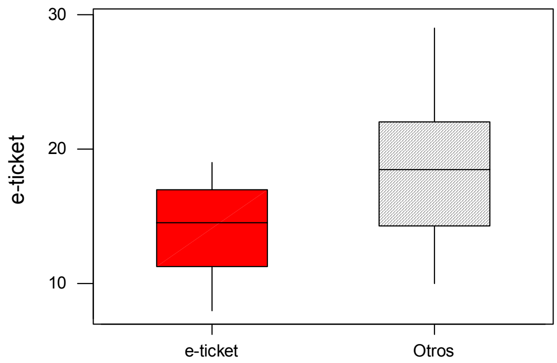
Seleccionamos ***Graph > Histogram***



Seleccionamos ***Graph > Boxplot***

Seleccionamos ***Graph > Boxplot***

Seleccionamos en la opción ***Frame*** la opción ***Multiple graphs***.



P1.2. Calcular los estadísticos descriptivos más destacables de las dos muestras de datos. Comenta los resultados.

Seleccionamos ***Stat > Basic statistics > Display Descriptive Statistics***:

```
Descriptive Statistics


Variable              N       Mean     Median     TrMean
StDev     SE Mean
e-ticket             20     14,200     14,500     14,278
3,350       0,749
Otros                20      18,40      18,50      18,28
5,02        1,12

Variable        Minimum    Maximum     Q1         Q3
e-ticket          8,000     19,000     11,250     17,000
Otros             10,00      29,00      14,25      22,00
```

Podemos observar cómo las medias en los dos grupos difieren ligeramente, así como sus distribuciones. Los rangos de variación de los datos en los dos casos también son muy diferentes. Destacamos el valor máximo de 19% para los usuarios de reserva de billetes vía Internet frente al 29% d'altres sistemas de reserva.

P1.3. Razona si la distribución correspondiente a los datos de "Viajeros insatisfechos (billete electrónico)" es simétrica.

A partir de los datos anteriores, podemos concluir que la distribución es muy simétrica ya que los datos de la media aritmética y la mediana son prácticamente el mismo valor. Como la mediana es ligeramente superior a la media aritmética, podríamos afirmar que la distribución es ligeramente sesgada hacia la izquierda.

PRÁCTICA 2

El portal de reservas de vuelo por Internet del apartado anterior ha observado que el número de reservas semanales que al final no se formalizan sigue una distribución de media 92, y de desviación estándar 17,2.

Se pide:

P2.1. Si se coge como muestra los datos de las 40 últimas semanas, ¿cuál es la probabilidad de que la media semanal de reservas de vuelo que al final se cancelan esté entre 86 y 100?

Tenemos que la variable X es la distribución del número semanal de reservas que finalmente no se formalizan y que no sabemos si es normal o no. Si se coge una muestra aleatoria de tamaño muestral 40, para el TCL, la distribución de la media se aproximará a una normal independientemente de la variable X.

Es decir, como n=40 > 30, podemos utilizar el TCL para afirmar que la distribución de medias muestrales $\bar{X}$ se podrá aproximar por una normal ($\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$) con media 92 y desviación estándar 2,72

Por lo tanto, seleccionamos ***Calc > Probability Distributions > Normal*** y activamos la opción ***Cumulative Probability***:

Cumulative Distribution Function

```
Normal with mean = 92,0000 and standard deviation = 2,72000

         x P( X <= x)
  100,0000        0,9984
```

Cumulative Distribution Function

```
Normal with mean = 92,0000 and standard deviation = 2,72000

         x P( X <= x)
   86,0000        0,0137
```

$P(86<\bar{X}<100)=P(\bar{X}<100)-P(\bar{X}<86)=0,9984 -0,0137=0,9847$

En conclusión, podemos decir que existe una probabilidad del 98,48% de que el número medio semanal de reservas canceladas de vuelos de este portal de Internet esté entre 86 y 100.

P2.2. ¿Si queremos conseguir sólo una probabilidad del 50%, qué número de vuelos exactamente tendrían que ser cancelados?

Queremos saber en cuanto a tiene que valer "c" para que $P(\bar{X}<c)=0.50$

Seleccionamos ***Calc > Probability Distributions > Normal*** y activamos la opción "**Inverse Cumulative Probability**".

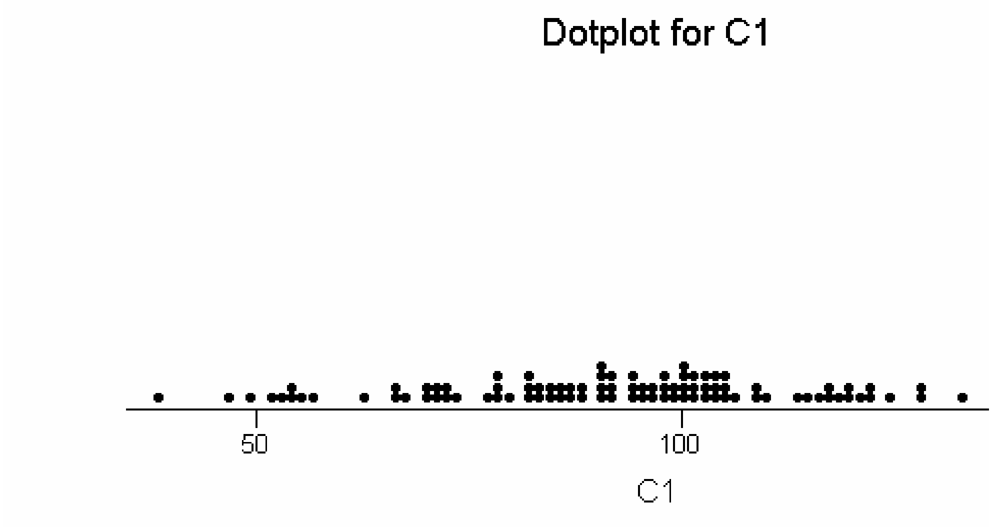
Inverse Cumulative Distribution Function

```
Normal with mean = 92,0000 and standard deviation = 2,72000

      P( X <= x)    x
    0,5000        92,0000
```

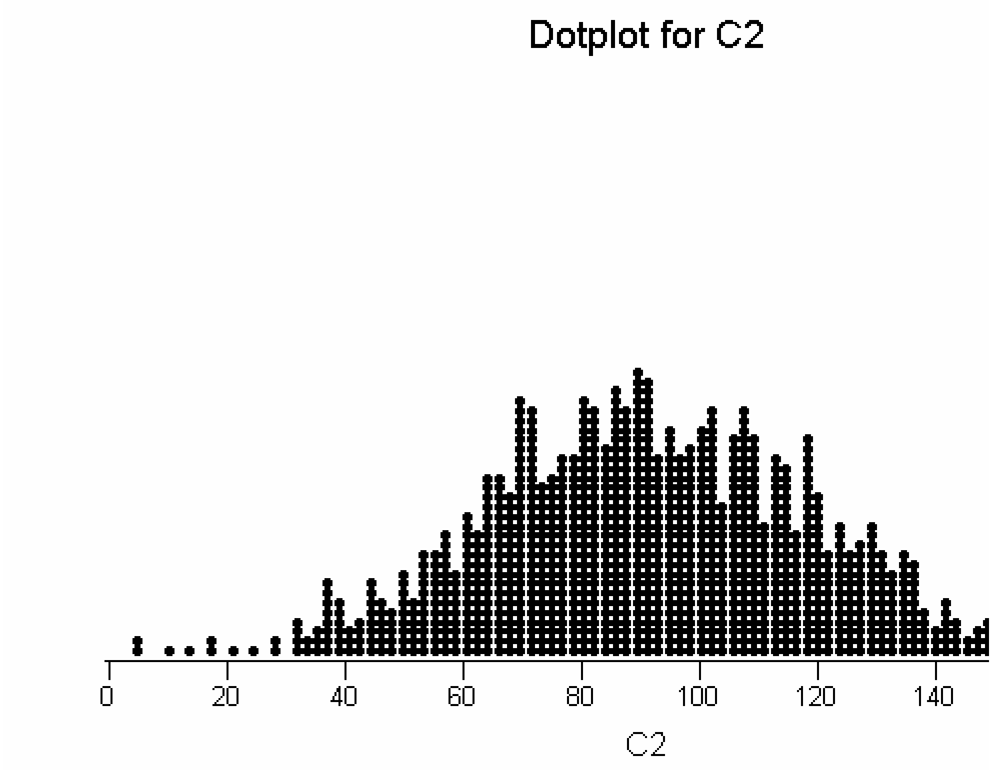
Por lo tanto, el número de vuelos cancelados tendría que ser 92.

P2.3. A partir de la media y de la desviación estándar del apartado anterior, genera 100 valores aleatorios y representa gráficamente el resultado mediante un *Dotplot*. Repite el proceso con 1000 valores. Comenta los resultados.

Seleccionamos ***Calc > Random Fecha > Normal*** y guardamos los datos en la columna C1. Posteriormente seleccionamos ***Graph > Dotplot***:



Hacemos el mismo proceso para 1000 datos y obtenemos:



Se puede observar que a medida que aumenta el tamaño de n, la distribución se va aproximando más a una distribución normal.

Materiales

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

- Los módulos 1 al 11
- Las **GES1**, **GES2**, **GES3 y GES4**
- Los vídeos correspondientes
- Los materiales complementarios enviados al Tablon
- El software Minitab: especialmente los documentos de actividades resueltas

Criterios de evaluación

La práctica se evaluará atendiendo a los siguientes criterios:

- Los e**jercicios** se puntuará con un máximo de 3'5 y 2'5 puntos respectivamente.
- 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.

Formato de entrega

La **PEC1** se entregará en un documento **Word** para Windows en el buzón de: "Ev.Continua" que tenéis en el aula junto con el Tablón y Foro.

El nombre del fichero/documento tendrá el siguiente formato: "apellido1_apellido2_PEC1.doc". Los apellidos se escribirán sin acentos. Por ejemplo, el archivo de un estudiante nombrado Alfredo García Melgar llevaría la siguiente denominación: "garcia_melgar_PEC1.doc".

Cada alumno tendrá que enviar dos archivos:

- **Garcia_melgar_PEC1.doc**
- **Garcia-melgar_PEC1.MPJ**

* * * MUY IMPORTANTE: Para poder entender las fórmulas, nos tendríamos que acostumbrar a hacerlas con el Editor de ecuaciones del programa Word.

Fecha de entrega

La **fecha límite de entrega** son las **24 horas** del día **24 de octubre del 2004**. Recordad que las fechas límite tienen que respetarse estrictamente.

Asimismo, el día 25/10/2004, se publicará la resolución de la PEC1.

Las calificaciones se publicarán el día 04/11/2004.