eman ta zabal zazu
euskal herriko unibertsitatea
universidad del país vasco

Departamento de Matemática Aplicada
Escuela Universitaria de Ingeniería Técnica Industrial
Universidad del País Vasco
Plaza de la Casilla, 3
48012 Bilbao

MÉTODOS ESTADÍSTICOS DE LA INGENIERÍA
CONVOCATORIA DE JUNIO (2011)

EJERCICIO 1

Históricamente, la empresa EUBAT ha fabricado baterías para coches con una vida media máxima estimada en 4 años. Se implanta un nuevo proceso de producción y, debido a esta reestructuración, el Departamento de Calidad ha establecido la media aritmética del nuevo proceso de producción en 3.53 años pero se desea homologar. Para ello, una agencia externa ha tomado una muestra aleatoria significativa dada por el siguiente diagrama de tallo y hojas:

		Frecuencia
1	6 9	2
2	2 5 6 6 9	5
3	0 0 1 $\boxed{1}$ 1 1 2 2 2 3 3 3 4 $\boxed{4}$ 4 5 5 6 7 7 7 8 8 $\boxed{8}$ 8 9	26
4	1 1 2 3 4 5 7 7	8

En la empresa EUBAT se sospecha que alguna de las observaciones es atípica.

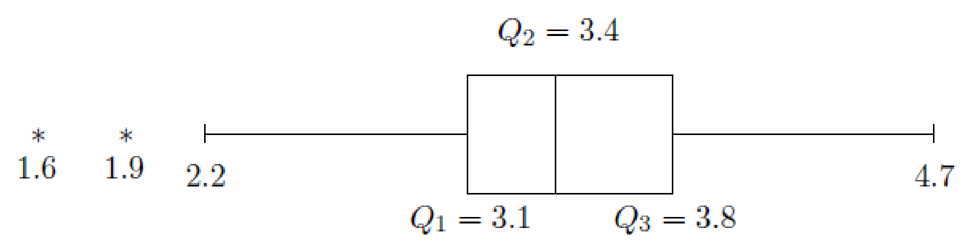
(A) Construye un diagrama de caja (*boxplot*) de Tukey para destacar dichos casos, si es que en realidad existen, justificando de forma detallada el proceso seguido en el cálculo de los estadísticos necesarios (**8 PUNTOS**).

Se trabaja con datos no agrupados (son pocos y se gana en exactitud en los resultados que se obtengan). Los valores intermedios que intervienen son: $\sum_{i=1}^{40} x_i = 136.50\ años; \sum_{i=1}^{40} x_i^2 = 485.07\ años$, que llevan a los estadísticos: $\overline{x} = 3.41951\ años; \hat{s}_X = 0.695061\ años$. A partir de ella se han de calcular los cuartiles correspondientes para hacer el análisis de las observaciones atípicas (**2 PUNTOS**).

CUARTIL	CÁLCULOS NECESARIOS	VALOR
Q_1	$x_{n/4} = x_{11}$	3.1 años
Me = Q_2	$x_{n/2} = x_{21}$	3.4 años
Q_3	$x_{3n/4} = x_{24}$	3.8 años
RIC	$Q_3 - Q_1$	0.7 años
D_1	$x_{n/10} = x_{4.1}$	2.6 años
D_9	$X_{9n/10} = x_{36.9}$	4.3 años

(**2 PUNTOS**)

Se construye el diagrama de Tukey "resumido":



(**1 PUNTO**)

Ahora, se calculan los puntos:

$$a_1 = Q_1 - 3RIC = 1\ año$$

$$a_2 = Q_1 - 1.5RIC = 2.05\ años$$

eman ta zabal zazu
euskal herriko unibertsitatea
universidad del país vasco

Departamento de Matemática Aplicada
Escuela Universitaria de Ingeniería Técnica Industrial
Universidad del País Vasco
Plaza de la Casilla, 3
48012 Bilbao

$$a_3 = Q_3 + 1.5RIC = 4.85 \text{ años}$$
$$a_4 = Q_3 + 3RIC = 5.9 \text{ años}$$

(**2 PUNTOS**)

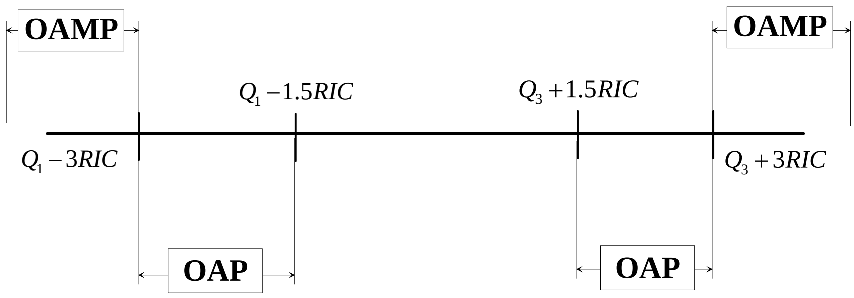
que establecerán los intervalos en los que pueden darse observaciones atípicas moderadas o posibles OAP (a saber,

$$I_1 = (-\infty, a1) \cup (a4, \infty)$$

y/o observaciones atípicas muy probables OAMP (es decir, si está en

$$I_2 = (a1, a2) \cup (a3, a4)$$

según la distribución de la figura



(**1 PUNTO**)

Concluyendo, se observan dos OAP en $x_1 = 1.6$ y $x_2 = 1.9$, que se mantendrán en los estudios que se hagan.

(B) Razona qué medida de centralización consideras la más adecuada para describir el centro de la muestra (**1 PUNTO**).

La distribución es bastante simétrica. Por lo tanto, la media y la mediana van a ser dos medidas de centralización bastante parecidas. Resultará más cómodo elegir la media como medida de centralización.

(C) Razona en cuánto podría aumentar la observación más grande de la serie estadística sin que el rango intercuartílico se viera afectado (**1 PUNTO**).

Se puede aumentar todo lo que se quiera. Los cuartiles no se van a ver alterados por ese posible dato atípico, y por tanto, tampoco el rango intercuartílico.

EJERCICIO 2

(A) Con un nivel de significación del 1 %, calcula el intervalo de confianza para la precisión del proceso de medida correspondiente a la población, supuesta normal, de la que se ha extraído la muestra dada (**4 PUNTOS**).

Al hablar de precisión el ejercicio consiste en una estimación confidencial (es decir, intervalar) de desviaciones típicas (así se usará la distribución chi cuadrado como modelo de probabilidad). Para este caso los valores que intervienen en la estimación son:

$$\hat{s}_X \sqrt{n-1} = 0.70\sqrt{40} = 4.427189 \text{ años}$$

$$\text{Valores críticos exactos: } \begin{cases} \chi_1^2 = \chi^2_{\alpha=0.5\%,\nu=40\ gdl} = 20.7065 \\ \chi_2^2 = \chi^2_{\alpha=99.5\%,\nu=40\ gdl} = 66.766 \end{cases}$$

Finalmente, el intervalo de confianza buscado viene dado por

$$\sigma \in [\mathsf{I}, \mathsf{L}] = \left[\frac{\hat{s}_X \sqrt{n-1}}{\sqrt{\chi_2^2}}, \frac{\hat{s}_X \sqrt{n-1}}{\sqrt{\chi_2^2}} \right] = \left[\frac{4.427189}{\sqrt{66.766}}, \frac{4.427189}{\sqrt{20.7065}} \right] =$$
$$\sigma \in [\mathsf{I}, \mathsf{L}] = [0.5418 \text{ años}, 0.9729 \text{ años}]$$

(B) ¿A qué conclusión llegará la agencia externa de calidad cuando tenga que decidir sobre la exactitud del nuevo proceso de medida de esta empresa con un nivel de significación del 2.5 %? Justifica la respuesta (**4 PUNTOS**).

Como se trata de trabajar con exactitudes el parámetro a considerar es la media aritmética: una población, grandes muestras (en consecuencia, se usará la distribución normal como modelo de probabilidad). Así para este caso, el valor del error probable es $\sigma_{\hat{\mu}} \equiv \frac{\hat{s}_X}{\sqrt{n}} = \frac{0.695061}{\sqrt{41}} = 0.1086 \text{ años}$, teniendo el contraste propuesto las características:

$$\begin{cases} H_0 : \mu_0 = 3.53 \text{ años} \\ H_a : \mu < 3.53 \text{ años} \end{cases}$$

El estadístico del contraste es

$$z = \frac{\overline{x} - \mu_0}{\sigma_{\hat{\mu}}} = \frac{3.41951 - 3.53}{0.1086} = -1.017403$$

y el valor crítico que separa la región de aceptración de la zona de rechazo es

$$z_\alpha = z_{H_a} = z_{2.5\,\%} = -1.959961082$$

Conclusión: con la muestra trabajada no existe evidencia estadística que exista un problema de exactitud en el histórico de esta empresa al nivel de significación $\alpha = 2.5\,\%$. Y no se puede rechazar la hipótesis nula.

Para encontrar el valor crítico se ha de buscar cuál es el nivel de significación correspondiente al estadístico del contraste obtenido; a saber, $z_\alpha = -1.959961082$. Entonces, por ensayo y error (aproximación) se deduce que $p - valor = \alpha_c = \text{Ï}\left(Z < -1.01785\right) = 15.4375\,\%$.

(C) ¿Cuál debe ser el tamaño mínimo de la muestra si el error en la estimación de la media tiene que ser como mucho de 0.01 años? (**2 PUNTOS**)

Como debemos actualizar el error muestral a $\sigma_{\hat{\mu}} = 0.01\,años$, se deduce la siguiente desigualdad

$$z_\alpha \frac{\hat{s}_X}{\sqrt{n}} = 1.959961082\frac{0.695061}{\sqrt{n}} \leq 0.01\,años \Rightarrow$$

$$n = \left(\frac{0.695061 \times 1.959961082}{0.01}\right)^2 = 13.6232^2 = 185.591$$

con lo que n = 186 unidades en la muestra.

EJERCICIO 3

La producción anual en la fábrica ACERBASK es una variable aleatoria X con distribución normal. Se sabe que el 90 % de los años la producción es inferior a 1300 miles de Tm y el 40 % de los años es superior a 1100 miles de Tm. La producción se puede considerar independiente de unos años a otros. Se pide:

(A) Calcula la media aritmética y la varianza de dicha distribución (**4 PUNTOS**).

Solución.– Vamos a denotar por X la variable aleatoria que mide la producción anual en millones de Tm.

$$X \sim N(\mu;\sigma)$$

$$\left.\begin{array}{l} P(X<1.3)=0.900 \implies P\left(Z<\dfrac{1.3-\mu}{\sigma}\right)=0.900 \implies \dfrac{1.3-\mu}{\sigma}=1.28155 \\ P(X>1.1)=0.400 \implies P\left(Z>\dfrac{1.1-\mu}{\sigma}\right)=0.400 \implies \dfrac{1.1-\mu}{\sigma}=0.253347 \end{array}\right\} \implies$$

$$\implies \left\{\begin{array}{lcl} \mu &=& 1.05072 \\ \sigma &=& 0.194514 \end{array}\right.$$

(B) Calcula la probabilidad de que la producción anual sea superior a 1000 miles de Tm (**3 PUNTOS**)

Solución.– Sabiendo que $X \sim (1.05072; 0.194514)$,

$$\begin{array}{rcl} P(X>1) & \overset{\text{tipificar}}{\underset{\downarrow}{=}} & P\left(Z>\dfrac{1-1.05072}{0.194514}\right)=P(Z>-0.260755) \overset{\text{simetría}}{\underset{\downarrow}{=}} \\ & = & P(Z<0.260755) \overset{\text{tablas}}{\underset{\downarrow}{=}} \boxed{0.602859=P(X>1)} \end{array}$$

(C) Calcula la probabilidad de que en 10 años la producción total sea inferior a 10000 miles de Tm (**3 PUNTOS**).

Solución.– Como las producciones anuales son normales e independientes, si denotamos por X_i la producción anual en millones de Tm, se tiene:

$$X_i \sim N(1.05072; 0.194514) \qquad i=1,2,\ldots,10$$

Utilizando la propiedad reproductiva de la ley normal, la producción total en 10 años en millones de euros, sigue una ley normal, cuya media es la suma de las medias y la varianza es la suma de las varianzas (por ser independientes). Es decir:

$$Y \sim N\left(10.5072; \sqrt{0.378356}=0.615107\right)$$

Por lo tanto:

$$\begin{array}{rcl} P(Y>10) & \overset{\text{tipificar}}{\underset{\downarrow}{=}} & P\left(Z>\dfrac{10-10.5072}{0.1945140.615107}\right)=P(Z>-0.82458) \overset{\text{simetría}}{\underset{\downarrow}{=}} \\ & = & P(Z<0.82458) \overset{\text{tablas}}{\underset{\downarrow}{=}} \boxed{0.795195=P(Y>10)} \end{array}$$

Departamento de Matemática Aplicada
Escuela Universitaria de Ingeniería Técnica Industrial
Universidad del País Vasco
Plaza de la Casilla, 3
48012 Bilbao

EJERCICIO 4

La compañía aérea BILBAIR oferta, a las 07:45 a.m., tres vuelos diarios desde Bilbao a Barcelona, Madrid y Sevilla. Un ejecutivo debe visitar las tres ciudades a lo largo de la semana y se presenta en el aeropuerto, sin plaza en ningún vuelo, con intención de comprar un billete para uno de ellos sin importarle el destino. Las probabilidades de que estén completos los aviones de Madrid, Barcelona y Sevilla son, respectivamente, del 60%, del 50% y del 40%. Se supone que las ocupaciones de los diferentes vuelos son sucesos independientes.

(**A**) ¿Cuál es la probabilidad de que no consiga billete para ningún vuelo?, ¿y la probabilidad de que consiga volar a alguno de los tres destinos? (**4 PUNTOS**).

<u>Solución.–</u>

M :	"plaza en vuelo a Madrid"	$P(M) = 0.40$
B :	"plaza en vuelo a Barcelona"	$P(B) = 0.50$
S :	"plaza en vuelo a Sevilla"	$P(S) = 0.60$

Sabemos que los sucesos M, B y S son independientes, luego también lo son los sucesos contrarios.

- $P(\overline{M} \cap \overline{B} \cap \overline{S}) = P(\overline{M}) \cdot P(\overline{B}) \cdot P(\overline{S}) = 0.60 \cdot 0.50 \cdot 0.40 = 0.12$
- $P(M \cup B \cup S) = 1 - P(\overline{M \cup B \cup S}) = 1 - P(\overline{M} \cap \overline{B} \cap \overline{S}) = 1 - 0.12 = 0.88$

(**B**) ¿Cuál es la probabilidad de que pueda volar a Barcelona o a Madrid? (**3 PUNTOS**).

<u>Solución.–</u>

$$P(B \cup M) = 1 - P(\overline{B \cup M}) = 1 - P(\overline{B} \cap \overline{M}) = 1 - P(\overline{B}) \cdot P(\overline{M}) = 1 - 0.50 \cdot 0.60 = 1 - 0.30 = 0.70$$

(**C**) Sabiendo que ha volado a Barcelona, ¿cuál es la probabilidad de que no hubiese plaza en ninguno de los otros dos vuelos? (**3 PUNTOS**).

<u>Solución.–</u> Como la probabilidad condicionada es una probabilidad podemos usar la independencia de los sucesos:

$$P(\overline{M} \cap \overline{S} | B) = P(\overline{M} | B) \cdot P(\overline{S} | B) = P(\overline{M}) \cdot P(\overline{S}) = 0.60 \cdot 0.40 = 0.24$$

Departamento de Matemática Aplicada
Escuela Universitaria de Ingeniería Técnica Industrial
Universidad del País Vasco
Plaza de la Casilla, 3
48012 Bilbao

eman ta zabal zazu

euskal herriko unibertsitatea — universidad del país vasco

MÉTODOS ESTADÍSTICOS DE LA INGENIERÍA

CONVOCATORIA DE JUNIO (2011)

EJERCICIO 5. La dirección de un Centro de Secundaria realiza un estudio de los resultados obtenidos por el alumnado en tres tipos de materias diferentes, contando el número de suspensos obtenidos en ellas. Los resultados han sido los de la tabla:

	Materias		
Alumnado	**Matemáticas**	**Lengua**	**Inglés**
Bachillerato Tecnológico	12	10	17
Bachillerato de la Naturaleza y S.	15	11	13
Bachillerato de Sociales y H.	26	14	12

(A) ¿Se puede afirmar que hay relación entre cursar un tipo de bachillerato y suspender las Matemáticas? (**2 PUNTOS**)

Este problema es un problema de homogeneidad para lo que se ha de construir la tabla unidireccional siguiente

Materia de Matemáticas

Alumnado	n_i^o	n_i^t	χ_i^2
Bachillerato Tecnológico	12	15,9	0,956603774
Bachillerato de la Naturaleza y S.	15	15,9	0,050943396
Bachillerato de Sociales y H.	26	21,2	1,086792453
	53		**2,094339623**

donde las frecuencias teóricas, n_i^t, se han calculado tendiendo en cuenta la influencia de todas las materias consideradas, así como los tipos de bachillerato considerado (es decir, se han trabajado las probabilidades marginales).

Es un contraste no paramétrico con estadístico $\chi^2 = \sum_{i=1}^{3} \chi_i^2 = \sum_{i=1}^{3} \frac{\left[n_i^t - n_i^o\right]^2}{n_i^t} = 2.0943$ y siendo la distribución de probabilidad chi cuadrado χ^2 el modelo a seguir con un contraste unidireccional de cola superior $\chi_1^2 = \chi^2_{\alpha=99\%,\nu=c-1=2\,gdl} = 9.21 > \chi^2$.

En resumen, de la muestra proporcionada no existe evidencia estadística para suponer que los alumnos suspendidos en Matemáticas dependa del tipo de bachillerato que se cursa con un nivel del significación $\alpha = 1\%$.

(B) ¿Y entre ser el tipo de bachillerato que se cursa y suspender Inglés? (**2 PUNTOS**)

Análogamente a como se ha procedido en el apartado anterior

Materia de Inglés

Alumnado	n_i^o	n_i^t	χ_i^2
Bachillerato Tecnológico	17	12,6	1,536507937
Bachillerato de la Naturaleza y S.	13	12,6	0,012698413

Bachillerato de Sociales y H.	12	16,8	1,37142857 1
	42		**2,92063492** **1**

$$\chi^2 = \sum_{i=1}^{3} \chi_i^2 = \sum_{i=1}^{3} \frac{\left[n_i^t - n_i^o\right]^2}{n_i^t} = 2.9206$$

$$\chi_1^2 = \chi^2_{\alpha=99\,\%,\nu=c-1=2\,gdl} = 9.21 > \chi^2$$

En resumen, de la muestra proporcionada no existe evidencia estadística para suponer que los alumnos suspendidos en Inglés dependa del tipo de bachillerato que se cursa con un nivel del significación $\alpha = 1\%$.

(C) En general, ¿se puede decir que hay relación entre el tipo de bachillerato cursado y el tipo de materias suspendidas? (**2 PUNTOS**)

Es un problema de independencia de criterios y se tendrá que construir la tabla de contingencia de frecuencias teóricas (esperadas), n_i^t, a partir de las frecuencias observadas, n_i^o, que se dan en la tabla del ejercicio.

	Materias			
Alumnado	**Matemáticas**	**Lengua**	**Inglés**	
Bachillerato Tecnológico	15,9000	10,5000	12,6000	**39**
Bachillerato de la Naturaleza y S.	15,9000	10,5000	12,6000	**39**
Bachillerato de Sociales y H.	21,2000	14,0000	16,8000	**52**
	53	**35**	**42**	**130**

El contraste es

$$\begin{cases} H_0: \text{No existe relación entre el tipo de bachillerato cursado y las materias suspendidas} \\ H_a: \text{Sí existe relación entre el tipo de bachillerato cursado y las materias suspendidas} \end{cases}$$

que siempre es unilateral (de cola superior). El estadístico del contraste es la prueba chi cuadrado de Pearson χ^2:

$$\chi^2 = \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{\left[n_{ij}^t - n_{ij}^o\right]^2}{n_{ij}^t} = 2.3134$$

cuyo cálculo lleva a

	Asignatura			
Bachillerato	**Matemáticas**	**Lengua**	**Inglés**	
Tecnológico	0,9566	0,0238	1,5365	2,5169
Naturaleza	0,0509	0,0238	0,0127	0,0875
Sociales	1,0868	0,0000	1,3714	2,4582

2,0943	0,0476	2,9206	**5,0626**

De las tablas de la distribución χ^2 (que es el modelo de probabilidad que rige en este tipo de problemas) se deduce que:

$$\chi_1^2 = \chi^2_{\alpha,\nu=(c-1)(r-1)} \equiv \chi^2_{99\%,4\,gdl} = 13.28 > \chi^2$$

Resumiendo, a partir de la muestra suministrada no existe evidencia estadística para rechazar la hipótesis nula; y en consecuencia, no existe relación alguna entre el tipo de bachillerato cursado y las materias suspendidas con un nivel de significación $\alpha = 0.01$.

(D) ¿Se puede afirmar que hay relación entre cursar bachillerato de Sociales y suspender las Matemáticas? (**2 PUNTOS**)

Del análisis de la tabla del apartado (1º) se deduce que la relación de atracción entre ser del bachillerato de Sociales y suspender Matemáticas es

$$n^t_{Sociales,\,Matematicas} = 26 > n^o_{Sociales,\,Matematicas} = \frac{N_{Sociales} N_{Matematicas}}{N} = \frac{52 \times 53}{130} = 21.2$$

Es decir, hay relación de atracción a pesar del resultado en el contraste general realizado en aquel apartado.

(E) ¿Y entre ser del bachillerato Tecnológico y suspender Inglés? Utiliza para hacer los cálculos un nivel de significación $\alpha = 1\%$ (**2 PUNTOS**).

Análogamente, del análisis de la tabla del apartado (2º) la relación de atracción entre ser del bachillerato tecnológico y suspender Inglés es

$$n^t_{Tecnologico,\,Ingles} = 17 > n^o_{Tecnologico,\,Ingles} = \frac{N_{Tecnologico} N_{Ingles}}{N} = \frac{42 \times 39}{130} = 12.6$$

Es decir, también existe relación de atracción a pesar del resultado en el contraste general realizado en el apartado (2º).

MÉTODOS ESTADÍSTICOS DE LA INGENIERÍA

CONVOCATORIA DE JUNIO (2011)

EJERCICIO 6.