Departamento de Matemática Aplicada
Matemáticas
Licenciatura en CC. Químicas (Curso 2008-09)
Estadística unidimensional con EXCEL. Práctica 6

1. Introducción

El programa EXCEL del paquete Microsoft Office es una aplicación que trabaja con hojas de cálculo, un elemento donde se incluyen datos, gráficos, fórmulas, etc. para poder realizar cálculos de manera rápida y eficaz.

Existen diferentes paquetes informáticos que distribuyen sus propias versiones de hojas de cálculo, aunque por ser quizás la más extendida nos centraremos en esta versión distribuida por Microsoft.

En este grupo de prácticas vamos a utilizar la hoja de cálculo EXCEL para realizar cálculos estadísticos, presentarlos y analizarlos convenientemente.

En esta práctica en particular vamos a aprender a calcular parámetros de centralización y de dispersión de un conjunto de datos, tanto si los tenemos en una lista como si nos dan una tabla con los datos y su frecuencia respectiva.

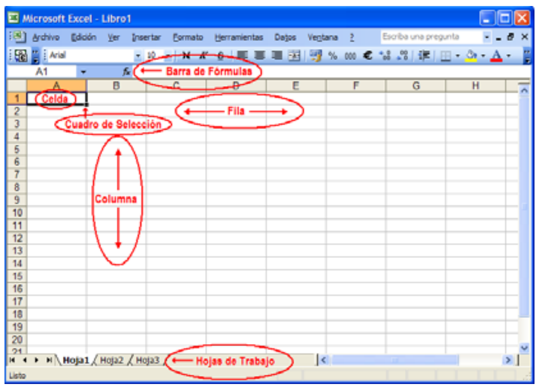


Figura 1: Aspecto de la Hoja de Cálculo EXCEL.

Al iniciar el programa EXCEL se muestra un Libro con tres Hojas en blanco. Sobre cada una de estas hojas se irán escribiendo los datos, las fórmulas y los gráficos de cada ejercicio. En la figura 1 se muestra el aspecto del programa con los elementos más importantes.

Las Hojas están divididas en Celdas que se corresponden con las intersecciones entre Filas, nombradas con números del 1 al 65536, y Columnas nombradas con letras de la A a la IV. El tamaño de una hoja puede aumentarse si se desea. Cada Celda por tanto llevará un nombre que se corresponde con la Columna y Fila a la que corresponde. Así en la figura 1 la Celda seleccionada es la A1.

Dentro de cada Celda se pueden introducir Datos o Fórmulas, los Datos podrán ser números o texto, según se requiera.

Algo importante a tener en cuenta siempre que se maneje EXCEL es que los números decimales se escribirán siempre con una coma de separación en lugar de un punto. Esto viene determinado por defecto en las versiones en castellano y aunque podría cambiarse, optaremos por dejarlo así. De esta manera, si escribimos `0.45` EXCEL entenderá que estamos introduciendo una variable tipo texto y no un número. Para introducirlo correctamente habrá que escribirlo `0,45`.

2. Trabajo con listas

Vamos a ilustrar el manejo de EXCEL con algún ejemplo sencillo:

Ejemplo 1 *Las calificaciones obtenidas por un alumno en una serie de prácticas han sido:*

7	*6*	*4*	*5*	*8*	*8*	*8*	*9*	*9*	*7*

Vamos a calcular los parámetros estadísticos más habituales, su tabla de frecuencias y su histograma.

2.1. Cálculo de los parámetros descriptivos

En primer lugar deberemos introducir los datos del problema. Por comodidad los introduciremos en la columna A. En la celda A1 escribiremos el encabezamiento "**Notas**" y pulsaremos la tecla ENTER. A continuación vamos introduciendo los datos de manera sucesiva en las celdas de la A2 hasta la A11.

EXCEL posee una librería de funciones muy amplia que permite el cálculo con rangos de datos. En esta práctica vamos a utilizar las funciones que permiten obtener los parámetros estadísticos más usuales. Para ello, en la columna A, a partir de la celda A14 por ejemplo, colocaremos los títulos de los parámetros y en la columna B, a su lado, la fórmula para su cálculo.

Título	Fórmula	Valor
Nº Datos=	`=CONTAR(A2:A11)`	10
Suma=	`=SUMA(A2:A11)`	71
Moda=	`=MODA(A2:A11)`	8
Mediana=	`=MEDIANA(A2:A11)`	7,5
Q1=	`=CUARTIL(A2:A11;1)`	6,25
Q3=	`=CUARTIL(A2:A11;3)`	8
Rango=	`=CUARTIL(A2:A11;4)-CUARTIL(A2:A11;0)`	5
Media=	`=PROMEDIO(A2:A11)`	7,1
Varianza=	`=VARP(A2:A11)`	2,49
CuasiVar.=	`=VARA(A2:A11)`	2,76667
Desv. Tip.=	`=DESVESTP(A2:A11)`	1,57798
Desv. Est.=	`=DESVESTA(A2:A11)`	1,66333
Curtosis=	`=CURTOSIS(A2:A11)`	-0,31043
Coef. Asim.=	`=COEFICIENTE.ASIMETRIA(A2:A11)`	-0,73883

2.2. Tabla de Frecuencias e Histograma

Para representar gráficamente los datos será necesario previamente agruparlos en una tabla de frecuencias donde aparezca cada nota x_i con su frecuencia absoluta f_i.

Esta tabla la situaremos en las columnas D y E de la hoja de cálculo. En la celda D1 colocaremos el título **xi** y en la celda E1 el título **fi**. A continuación, en las celdas D2 hasta la D11 colocaremos las notas desde 1 hasta 10. Para que EXCEL calcule la frecuencia correspondiente a cada nota seleccionaremos todo el rango E2:E11 donde se situarán las frecuencias, a continuación escribiremos la fórmula `=FRECUENCIA(A2:A11;D2:D11)` y presionaremos conjuntamente las teclas CONTROL+MAYÚSCULAS+INTRO para que EXCEL reconozca que se aplica a todo el rango seleccionado.

Para crear el Histograma actuaremos de la siguiente manera:

- Seleccionar el rango de frecuencias a dibujar (`E2:E11`).
- Abrir el menú **`Insertar`** y elegir la opción **`gráfico`**.
- Inmediatamente se abrirá el asistente y deberemos elegir un tipo de gráfico. Elegiremos `Columnas` y presionaremos el botón que pone **`siguiente`**.
- En la ventana siguiente del asistente elegiremos la pestaña **`Serie`** para elegir los datos a dibujar. (Ver figura 2)

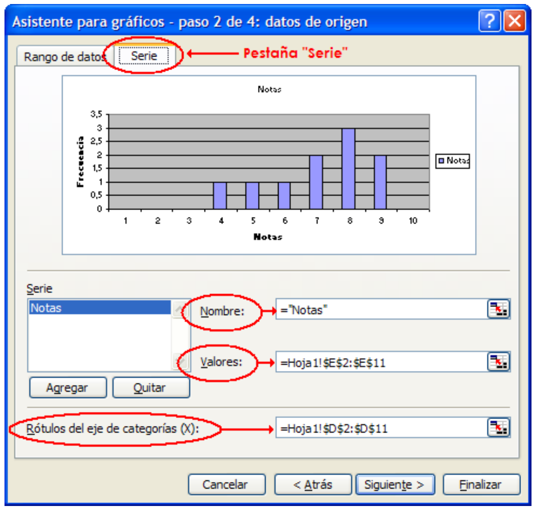


Figura 2: Ventana del asistente para gráficos de EXCEL.

- En el cuadro de texto llamado **`Rótulos`** incluiremos el rango de datos (`D2:D11`). Opcionalmente, el el cuadro de texto llamado `Nombre` podemos colocar el nombre de la serie, en este caso **Notas**.
- Presionando el botón **`Siguiente`** podremos incluir el título del gráfico, de los ejes, etc. Si se presiona el botón **`Finalizar`** se inserta directamente el Histograma en la Hoja de Cálculo.

2.3. Ejercicios

Ejercicio 1 *En una muestra de 40 bolas fabricadas por una máquina se han medido los diámetros (en cm) obteniéndose:*

0,853	0,859	0,851	0,840	0,859	0,841	0,846	0,857	0,862	0,845
0,851	0,846	0,855	0,861	0,868	0,852	0,843	0,840	0,841	0,854
0,854	0,847	0,853	0,855	0,843	0,847	0,859	0,863	0,856	0,857
0,842	0,850	0,860	0,852	0,856	0,856	0,860	0,854	0,861	0,868

1. *Calcula la media aritmética, mediana, moda, varianza y desviación típica.*
2. *Agrupa los datos en 6 clases y dibuja el Histograma resultante.*

Ejercicio 2 *Copia los datos del problema anterior en una nueva Hoja de Cálculo y realiza la siguiente transformación a todos ellos $y = 2x - 127$, donde x es la variable dada como dato en el ejercicio anterior.*

1. *Hallar la media aritmética, mediana, varianza y desviación típica de y.*

2. *Comprobar con este ejemplo que dadas dos variables x, y tales que $y = ax + b, a, b \in IR$ se cumple que*

$$\bar{y} = a\bar{x} + b$$
$$s_y^2 = a^2 s_x^2,$$

siendo $\bar{x}$ la media de x y s_x^2 su varianza.

3. Trabajo con tablas estadísticas

Ilustraremos el manejo de las tablas con el siguiente ejemplo:

Ejemplo 2 *Tras medir la concentración de un determinado reactivo en una disolución se han obtenido los siguientes resultados:*

Concentración	*[0,54–0,56)*	*[0,56–0,58)*	*[0,58–0,60)*	*[0,60–0,62)*
Nº de muestras	*3*	*12*	*22*	*6*

Vamos a calcular:

1. *La media, la desviación estándar*
2. *Representaremos el Histograma de los datos.*

3.1. Cálculo de la media y la desviación estándar

Para proceder al cálculo de éstas magnitudes empezaremos introduciendo los datos en una tabla. Para ello escribiremos en la celda A1 `datos` y en la celda B2 `frecuencia` como títulos de las columnas respectivas. A continuación iremos metiendo la marca de cada clase en una celda de la columna A, empezando por la celda A2: 0,55 0,57, ..., y en la columna B la frecuencia correspondiente a cada clase:3, 12, ...

Una vez introducidos los datos en la tabla procederemos al cálculo de la media y de la varianza:

$$\overline{x} = \frac{\sum_{n=1}^{M} x_n \cdot f_n}{N} \qquad S^2 = \frac{\sum_{n=1}^{N} x_n^2 \cdot f_n}{N} - \overline{x}^2,$$

siendo M el número total de clases y N la suma de las frecuencias.

3.1.1. Cálculos con los datos

Para poder calcular los parámetros deseados primero habrá que realizar algunos cálculos intermedios con los datos y sus frecuencias:

- En la celda C1 escribimos el título de la columna (**xi*fi**) y en la celda C2 escribimos la fórmula : `=A2*B2`. A continuación, pinchando sobre el cuadro de selección que aparece en la esquina inferior derecha de la celda (ver figura 1), con el puntero del ratón se arrastra hacia abajo hasta completar el cálculo con todos los datos.
- En la celda D1 escribimos el título de la columna (**xi^2*fi**) y en la celda D2 escribimos la fórmula : `=A2^2*B2`. A continuación, pinchando sobre el cuadro de selección con el puntero del ratón se arrastra hasta completar el cálculo con todos los datos.

3.1.2. Cálculo de los parámetros

- En la celda A8 se escribe el título (`N=`) y en la celda B8 escribimos la fórmula : `=suma(B2:B5)`.
- En la celda A9 escribimos el título (`media=`) y en la celda B9 escribimos la fórmula : `=suma(C2:C5)/B8`.
- En la celda A10 escribimos el título (`varianza=`) y en la celda B10 escribimos la fórmula : `=suma(D2:D5)/B8-B9^2`.
- En la celda A11 escribimos el título (`desviación típica=`) y en la celda B11 escribimos la fórmula : `=raiz(B10)`.

Para el histograma habremos de seguir los mismos pasos descritos anteriormente. Opcionalmente se podrá introducir una columna con las clases que servirán de rótulos en lugar de las marcas. Los resultados del Ejemplo 2 se muestran en la figura 3

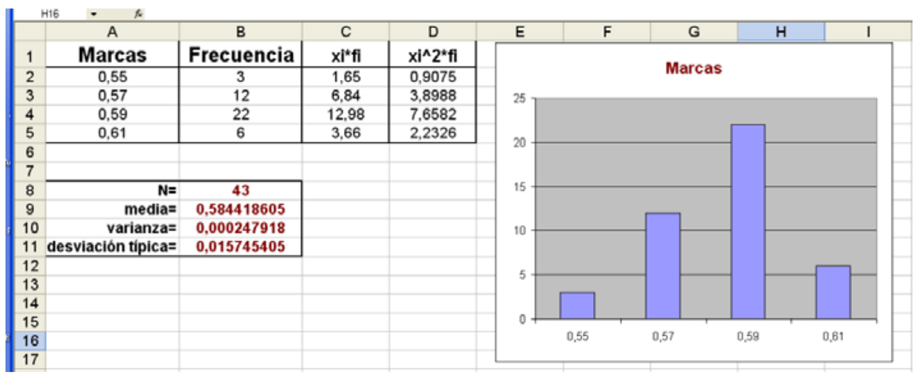


Figura 3: Resultados del Ejemplo 2.

3.2. Ejercicios

Ejercicio 3 *La siguiente tabla indica el sueldo mensual x_i (en €) de las familias de 200 alumnos de un curso, donde $h_i = fr(x_i)$ representa la frecuencia relativa simple.*

x_i	$\bar{x}_i$	h_i
500–700	600	0,06
700–900	800	0,115
900–1100	1000	0,105
1100–1300	1200	0,075
1300–1500	1400	0,085
1500–1700	1600	0,1
1700–1900	1800	0,11
1900–2100	2000	0,095
2100–2300	2200	0,07
2300–2500	2400	0,09
2500–2700	2600	0,05
2700–2900	2800	0,03
2900–3100	3000	0,015

1. *Crear un fichero con los datos del problema.*

2. *Añadir al fichero una columna con el número de familias para cada sueldo o la frecuencia absoluta de la variable.*

3. *Representar los datos mediante un histograma de frecuencias.*

4. *Calcular la media, la moda, la varianza y la desviación típica.*

Ejercicio 4 *Contabilizado el número de accidentes diarios en una gran ciudad durante un periodo de 60 días, se obtuvieron los siguientes datos:*

número de accidentes	*número de días*
0	24
1	17
2	8
3	7
4	2
5	2

1. *Calcular la media y la desviación típica.*

2. *Dibujar el histograma de la anterior distribución.*

Departamento de Matemática Aplicada

Matemáticas

Licenciatura en CC. Químicas (Curso 2008-09)

Regresión Lineal con EXCEL. Práctica 7

El Método de ajuste por Mínimos Cuadrados sirve para encontrar una función $y = f(x, \alpha_1, \alpha_2, \ldots, \alpha_m)$, en la que habrá que calcular los parámetros $\alpha_1, \alpha_2, \ldots, \alpha_m$, de tal forma que sea la función que se ajuste lo mejor posible a una tabla de valores que relaciona las dos variables x e y obtenida experimentalmente:

x	x_1	x_2	$\ldots$	x_n
y	y_1	y_2	$\ldots$	y_n

Para calcular los parámetros se impone la condición de que sea mínima la función

$$S(\alpha_1, \alpha_2, \ldots, \alpha_m) = \sum_{i=1}^{n} \left[f(x_i, \alpha_1, \alpha_2, \ldots, \alpha_m) - y_i\right]^2$$

Como $S(\alpha_1, \alpha_2, \ldots, \alpha_m)$ es una función de m variables, una condición necesaria para que tenga un valor extremo en un punto, es que sus derivadas parciales en ese punto sean todas nulas. De aquí obtenemos un sistema de m ecuaciones, con m incógnitas:

$$\frac{\partial S}{\partial \alpha_1} = 0, \quad \frac{\partial S}{\partial \alpha_2} = 0, \quad \ldots, \frac{\partial S}{\partial \alpha_m} = 0$$

cuyas soluciones son los parámetros $\alpha_1, \alpha_2, \ldots, \alpha_m$, que nos indican cómo es la función que mejor se ajusta a los datos, es decir, $y = f(x, \alpha_1, \alpha_2, \ldots, \alpha_m)$.

La función $f(x, \alpha_1, \alpha_2, \ldots, \alpha_m)$ puede ser de cualquier tipo, teóricamente. Sin embargo, en la práctica, con programas como EXCEL podremos calcular los parámetros (directamente) cuando la función es un polinomio, una exponencial, logarítmica o potencial.

Para realizar estos cálculos EXCEL dispone de la opción `Dibujar Línea de Tendencia` en el gráfico de los datos. Veamos en un ejemplo como se puede ajustar una recta con Mínimos Cuadrados:

Ejemplo 1 *Dada la tabla de valores,*

x	1	2	3	4
y	2,1	4,3	6	7,8

Encontrar la función de la forma $y = ax + b$ que mejor se ajuste a los datos.

En primer lugar procederemos a introducir los datos en dos columnas de la Hoja de Cálculo. A continuación los dibujaremos. Seleccionamos el rango con los datos, y elegimos en el menú `Insertar` la opción `Gráfico`. En el asistente elegimos el tipo de gráfico `XY (Dispesión)` y presionamos el botón `Finalizar`.

A continuación seleccionamos uno de los puntos que representa a los datos del gráfico y pulsamos el botón derecho del ratón. En el menu que se despliega elegimos la opción `Agregar línea de tendencia`.

En el asistente que aparece elegimos el tipo de regresión que nos interese (`Lineal`) y en la pestaña `Opciones` marcamos el cuadro `Presentar la ecuación en el gráfico`. Al presionar el botón aceptar se muestra el gráfico con los puntos y la línea de regresión correspondiente, con ecuación

$$y = 1{,}88x + 0{,}35$$

Para obtener el coeficiente de correlación lineal

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\displaystyle\sum_{i=1}^{4}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\displaystyle\sum_{i=1}^{4}(x_i - \overline{x})^2}\sqrt{\displaystyle\sum_{i=1}^{4}(y_i - \overline{y})^2}}$$

basta pedir `=PEARSON(A1:A4;B1:B4)` o `=COEF.DE.CORREL(A1:A4;B1:B4)` para obtener el resultado

$$r = 0,99836$$

lo que quiere decir que el ajuste lineal es muy bueno pues r es proximo a 1.

También podemos ajustar estos datos a una parábola sin más que cambiar el tipo de regresión a una polinomial de orden 2, y se obtiene:

$$y = -0,1x^2 + 2,38x - 0,15$$

cuya gráfica se ajusta mejor a los puntos.

Si se hace el ajuste a un polinomio de orden 3, se obtiene

$$y = 0,1x^3 - 0,85x^2 + 4,05x - 1,2$$

siendo el ajuste en este caso perfecto, puesto que 4 puntos determinan un único polinomio de grado 3 que pasa por ellos, de la misma forma que 2 puntos determinan una única recta que pasa por ellos. De hecho, si se pidiera un polinomio de grado 4 o mayor, se obtendría el mismo polinomio de grado 3.

EXCEL permite hacer otros tipos de ajustes. En `Agregar linea de tendencia` se pueden elegir otras opciones. El ajuste logarítmico para este ejemplo es

$$y = 3,9994\ln(x) + 1,8725$$

y el exponencial

$$y = 1{,}5592e^{0{,}427x}$$

Para determinar qué ajuste es mejor EXCEL usa el coeficiente de determinación $R2$ que se define como

$$R2 = \frac{\displaystyle\sum_{i=1}^{4}(f(x_i) - \overline{y})^2}{\displaystyle\sum_{i=1}^{4}(y_i - \overline{y})^2}$$

siendo $y = f(x)$ el ajuste obtenido sobre los puntos resultantes de linealizar el problema. En un ajuste lineal o polinómico $f(x)$ es la recta o el polinomio en cuestión, mientras que en otro tipo de ajuste $R2$ se calcula utilizando la nube de puntos (X_i, Y_i) resultante de transformar convenientemente los (x_i, y_i) para que el ajuste de los nuevos puntos sea lineal.

El coeficiente de determinación $R2$ mide cómo es la varianza de los puntos estimados por el ajuste en comparación con los de la muestra. Sus valores están entre $R2 = 0$ (ausencia de ajuste) y $R = 1$ (ajuste perfecto, la función pasa por todos los puntos).

El valor del coeficiente de determinación para cada ajuste se obtiene en EXCEL de la misma manera en la que se obtiene la ecuación pero ahora marcando el cuadro `Presentar el valor de R cuadrado en el` a la vez que mostramos la ecuación. Los valores obtenidos para los ajustes anteriores son:

$R2_{lin} = r^2 = 0,9967$ $\qquad R2_{cubic} = 1$ $\qquad R2_{exp} = 0,9416$

$R2_{cuad} = 0,999$ $\qquad R2_{log} = 0,9781$ $\qquad R2_{pot} = 0,9974$

y como se esperaba es el de la cúbica el que tiene un mejor valor.

Si queremos hallar la recta de regresión de X sobre Y debemos cambiar el orden de las columnas ya que por defecto toma siempre la primera columna como variable independiente. Esto lo podemos hacer seleccionando cada columna de datos, copiándola y pegándola en otra parte de la Hoja o en otra hoja diferente usando las opciones `Copiar` y `Pegar` del menú `Edición`. Para el ejemplo que nos ocupa la ecuación de la recta de regresión de X sobre Y es

$$x = 0,5302y - 0,1774$$

Ejercicio 1 *En las tablas siguientes se presentan tres series de datos preparadas por el estadístico Frank Anscombe. Los tres conjuntos tienen la misma recta de regresión y el mismo coeficiente de correlación:*

Serie de datos A

x	10	8	13	9	11	14	6	4	12	7	5
y	8,04	6,95	7,58	8,81	8,33	9,96	7,24	4,26	10,84	4,82	5,68

Serie de datos B

x	10	8	13	9	11	14	6	4	12	7	5
y	9,14	8,14	8,74	8,77	9,26	8,10	6,13	3,10	9,13	7,26	4,74

Serie de datos C

x	8	8	8	8	8	8	8	8	8	8	19
y	6,58	5,76	7,71	8,84	8,47	7,04	5,25	5,56	7,91	6,89	12,50

Realiza el ajuste lineal para los tres conjuntos de datos y calcula sus coeficientes de correlación r y de determinación $R2$. ¿Qué conclusión obtienes de esto?

Ejercicio 2 *Ajustar mediante el Método de Mínimos Cuadrados, los valores de x e y dados por la siguiente tabla*

x	7	8	9	10	11	12	13
y	7,4	8,4	9,1	9,4	9,5	9,5	9,4

a una función:

i) cuadrática: $f(x) = ax^2 + bx + c$. Calcular la función $f(x)$, el coeficiente $R2$ y representar la función obtenida junto a los datos de la tabla.

ii) exponencial: $f(x) = ae^{bx}$. Calcular la función $f(x)$, el coeficiente $R2$ y representar la función obtenida junto a los datos de la tabla.

iii) logarítmica: $f(x) = a\log(x) + b$. Calcular la función $f(x)$, el coeficiente $R2$ y representar la función obtenida junto a los datos de la tabla.

iv) ¿Qué ajuste es el mejor?

Ejercicio 3 *De ciertas medidas realizadas se han obtenido los siguientes valores:*

x	0	0,2	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5
y	1	0,833	0,667	0,54	0,4	0,333	0,286	0,25	0,222	0,2	0,182	0,167

a) Representar la tabla de valores gráficamente.

b) Se puede observar que se trata, aproximadamente, de una función del tipo $y = \dfrac{1}{ax+b}$. *Esta función no la podemos obtener directamente, pero la podemos transformar en*

$$\frac{1}{y} = ax + b$$

Ahora, haciendo $X = x$; $Y = \dfrac{1}{y}$, *nos queda la expresión*

$$Y = aX + b$$

Calcular esta función. (Habrá que hacer una nueva tabla de valores.)

c) Deshacer el cambio de variable y obtener la función original y representarla con los valores del apartado a).

d) Calcular el coeficiente R2.

Ejercicio 4 *Se han obtenido los siguientes datos experimentales:*

v_i(en litros)	1,65	1,03	0,74	0,61	0,53	0,45
p_i(en kg/cm^2)	0,5	1	1,5	2	2,5	3

(a) Ajustar estos datos a una curva de la forma $PV^n = m$ *(tipo de regresión* `potencial`*).*

(b) Ajustar estos datos a una curva de la forma $P = \dfrac{1}{aV+b}$.

(c) ¿Qué ajuste es mejor?

0.3. Introducción

Las tablas de probabilidad son herramientas muy útiles para el cálculo de probabilidades cuando no se tiene a mano una calculadora científica o un ordenador. Sin embargo conseguir las tablas no siempre es posible a no ser que se fotocopien de algún libro, y no siempre se tiene a mano un libro de estadística con tablas.

En la presente práctica se van a construir tablas de probabilidad para las principales distribuciones continuas y discretas. Que se podrán imprimir y así estarán disponibles en cualquier momento, siendo por tanto útiles para la realización de los problemas de la asignatura y para su estudio.

Existen diversos tipos de tablas de probabilidad, las simples en las que a partir de un valor se obtiene el valor de la probabilidad de que la variable estadística tome dicho valor, las acumuladas en las que a partir de un valor se obtiene el valor de la probabilidad de que la variable sea menor o igual que la dada, las inversas en las que a partir del valor de la probabilidad se obtiene el valor de la variable estadística a la que corresponde, etc.

0.4. Variable Discreta

Una variable discreta solo puede tomar valores enteros. En este caso no hay que hablar tanto de una tabla de probabilidad sino más bien de una lista de probabilidad en la que en una columna se sitúan los valores de la variable y a su derecha los de la probabilidad correspondiente, acumulada o no según se quiera. Sin embargo estas listas se pueden mejorar si incluimos más de una columna de probabilidades en la que se vayan variando los valores del parámetro o parámetros de la distribución.

Ejemplo 1 *Construcción de una tabla de probabilidad simple para la distribución de Poisson con valores de la media* $\lambda = 0,05 \quad 0,1 \quad 0,5 \quad 1 \quad 1,1 \quad 1,2 \quad 1,3 \quad 1,4 \quad 1,5 \quad 2 \quad 5$ *y con valores de la variable desde* $x = 0$ *hasta* $x = 10$.

En una hoja de cálculo en blanco situaremos el título de la tabla, en este caso **Distribución de Poisson**.

A continuación, en la columna A, a partir de la celda A3 y en adelante colocaremos los valores de la variable estadística, empezando por el valor `0`, `1`, ... hasta llegar a `10` como nos piden.

Una manera rápida de realizar esto es la siguiente: Se introducen los dos primeros valores en las celdas A3 y A4. A continuación se seleccionan ambos con el ratón y se pincha con el puntero del ratón en el cuadro de selección que parece en la esquina inferior derecha de las celdas seleccionadas, arrastrando hasta llegar a la celda A13 que corresponderá al valor `10` deseado.

Una operación semejante se hace entre las columnas B2 y L2 en las que metemos los valores del parámetro λ de la distribución.

Una vez introducidos los valores de la cabecera de la tabla procedemos a calcular los de su probabilidad.

- En la celda B3 introduciremos la fórmula: `=POISSON($A3;B$2;FALSO)` y automáticamente devolverá el valor de la probabilidad correspondiente al valor $k = 0$ con $\lambda = 0,05$ que es 0,95123.

- A continuación se selecciona esta celda y pinchando en el cuadro de selección se arrastra a toda la fila.

- Una vez realizada esta operación, y con toda la fila todavía seleccionada, se vuelve a pinchar en el cuadro de selección y se arrastra a toda la tabla, obteniendo todos los valores deseados.

Como podrás comprobar en este caso al referenciar a las celdas correspondientes a los valores de la variable estadística y a los de la media λ hemos utilizado el símbolo `$` delante del nombre de la fila o de la columna. Se ha usado `$A3` en lugar de `A3` y `B$2` en lugar de `B2`. Esto se hace para que al copiar la fórmula a todas las celdas de la tabla, el valor de la columna `A` (al usar `$A`) y el de la fila `2` (al usar `$2`) no cambien, usando siempre en el cálculo los valores de las cabeceras.

En la función de EXCEL utilizada `=POISSON($A3;B$2;FALSO)`, el último argumento `FALSO` o `VERDADERO` sirve para elegir el valor de la probabilidad acumulada si el valor es `VERDADERO` o el de la probabilidad simple si el valor es `FALSO`.

Ejercicio 1 *Siguiendo el procedimiento anterior construye una tabla de distribución binomial con* $n = 6$ *usando la función de EXCEL* `=DISTR.BINOM(k;n;p;FALSO)` *y valores de* $p = 0,05\ \ 0,1\ \ 0,15\ ...\ 0,5$ *y para valores de la variable* k *desde 0 hasta 6.*

Ejercicio 2 *Construye una tabla de probabilidad binomial acumulada para los valores anteriores de las variables.*

0.5. Variable Continua

En el caso de variables continuas las tablas de probabilidad que se realizan son siempre acumuladas, ya que no suele interesar conocer el valor de la densidad de probabilidad para un valor concreto de la variable estadística.

Una tabla de probabilidad acumulada para una variable continua es una tabla en la que como entrada se presenta una serie de valores de la variable en las filas y las columnas de la tabla de manera que el valor de la probabilidad que se muestra en cada celda de la tabla se corresponde el de la variable correspondiente a la suma de los valores de la fila y de la columna.

Dentro de las tablas de probabilidad directa, la que más frecuentemente se va a utilizar será la de la distribución Normal Estándar $N(0,1)$, cuya obtención vamos a describir a continuación.

Ejemplo 2 *Construcción de la tabla de probabilidad acumulada de la distribución normal estándar* $N(0,1)$.

En una hoja de cálculo en blanco situaremos el título de la tabla, en este caso **Distribución Normal Estándar Acumulada N(0,1)**.

A continuación, en la columna A, a partir de la celda A3 y en adelante colocaremos las unidades y las décimas de la variable estadística, empezando por el valor `0,0`, `0,1`, ... hasta llegar a `3,9` por ejemplo. Como ya hemos comentado para la variable discreta, esta operación se puede realizar rápidamente si introducimos los dos primeros valores, los seleccionamos y pinchando en el cuadro de selección, arrastramos hasta conseguir el último de los valores.

Una operación semejante se hace entre las columnas B2 y K2 en las que metemos los valores de las centésimas desde `0,00` hasta `0,09`.

Una vez introducidos los valores de la cabecera de la tabla que constituirán, sumados, los de la variable estadística procedemos a calcular los de su probabilidad acumulada.

- En la celda B3 introduciremos la fórmula: `=DISTR.NORM.ESTAND($A3+B$2)` y automáticamente devolverá el valor de la probabilidad acumulada correspondiente al valor 0 que es 0,5.

- A continuación se selecciona esta celda y pinchando en el cuadro de selección se arrastra a toda la fila.

- Una vez realizada esta operación, y con toda la fila todavía seleccionada, se vuelve a pinchar en el cuadro de selección y se arrastra a toda la tabla, obteniendo todos los valores deseados.

Ya disponemos de una tabla con los valores de la probabilidad acumulada para una distribución normal $N(0,1)$. Si mejoramos el aspecto con unos toques de color en el título, las columnas y las filas, podremos imprimir la tabla para usarla en el estudio de la asignatura.

Ejercicio 3 *Usando la función de EXCEL* `=DISTR.NORM(dato,media,desv,VERDADERO)` *construye una tabla de probabilidad acumulada para la distribución normal* $N(3,2)$ *(media* $\bar{x}=3$*, y desviación* $\sigma_x = 2$*) con valores de la variable desde* $x=0$ *hasta* $x=6$*.*

Ejercicio 4 *Usando la función de EXCEL* `=DISTR.EXP(dato,lambda,VERDADERO)` *construye una tabla de probabilidad acumulada para la distribución exponencial con parámetro* $\lambda = 3$*, con valores de la variable desde* $x=0$ *hasta* $x=4$*.*

Ejercicio 5 *Usando la función de EXCEL* `=DISTR.T(dato,grados,2)` *construye una tabla de probabilidad para la distribución t de Student con 9 grados de libertad y 2 colas, con valores de la variable desde* $x=0$ *hasta* $x=4$*.*

1. Tablas de Probabilidad Inversa

En ocasiones va a ser necesario utilizar tablas en las que conocido en valor de la probabilidad de una determinada variable, queramos conocer su valor correspondiente. Esto se puede hacer mediante tablas de probabilidad inversa.

Ejemplo 3 *Construcción de una tabla de probabilidad inversa para la distribución t de Student (de una cola) con grados de libertad desde 1 hasta 30.*

En una hoja de cálculo en blanco situaremos el título de la tabla, en este caso **Distribución Inversa t Student**.

A continuación, en la columna A, a partir de la celda A3 y en adelante colocaremos los grados de libertad de la variable estadística, empezando por el valor `1`, `2`, ... hasta llegar a `30`. Una operación semejante se hace entre las columnas B2 y K2 en las que metemos los valores de la probabilidad desde `0,5 0,45` hasta `0,05`.

Una vez introducidos los valores de la cabecera de la tabla procedemos a calcular los de su probabilidad inversa.

- En la celda B3 introduciremos la fórmula: `=DISTR.T.INV(2*B$2;$A3)` y automáticamente devolverá el valor de la variable cuya probabilidad acumulada es 0,5 que es 0. (El valor que aparece es el cero con la precisión simple con la que estamos trabajando).

- A continuación se selecciona esta celda y pinchando en el cuadro de selección se arrastra a toda la fila.

- Una vez realizada esta operación, y con toda la fila todavía seleccionada, se vuelve a pinchar en el cuadro de selección y se arrastra a toda la tabla, obteniendo todos los valores deseados.

Para obtener la tabla correspondiente a una distribución t de Student de una cola hemos multiplicado por 2 el valor de la probabilidad ya que por defecto la función `=DISTR.T.INV(p;n)` devuelve la variable correspondiente la la distribución de dos colas.

Ejercicio 6 *Construye una tabla de probabilidad inversa para la distribución normal* $N(0,1)$. *Usa la función* `=DISTR.NORM.ESTAND.INV(p)`.

Ejercicio 7 *Construye una tabla de probabilidad inversa para la distribución* F *de Fisher en la que en horizontal estén los grados de libertad* n_1 *de 1 a 10, y en vertical los grados de libertad* n_2 *de 1 a 10. Haz la tabla correspondiente a un valor de la probabilidad de* $0,01$. *Usa la función* `=DISTR.F.INV(p,n1,n2)`

Departamento de Matemática Aplicada
Matemáticas
Licenciatura en CC. Químicas (Curso 2008-09)
Análisis de la Varianza (ANOVA). Práctica 9

Introducción

Mediante contraste de hipótesis, utilizando una t de Student, se puede comprobar si dos medias muestrales difieren o no significativamente. En esta práctica vamos a plantearnos el mismo problema, pero tratándose de más de dos muestras. La técnica que vamos a utilizar se denomina *análisis de la varianza* o *ANOVA*.

Veamos con un ejemplo cómo se puede utilizar esta técnica con excel.

Ejemplo

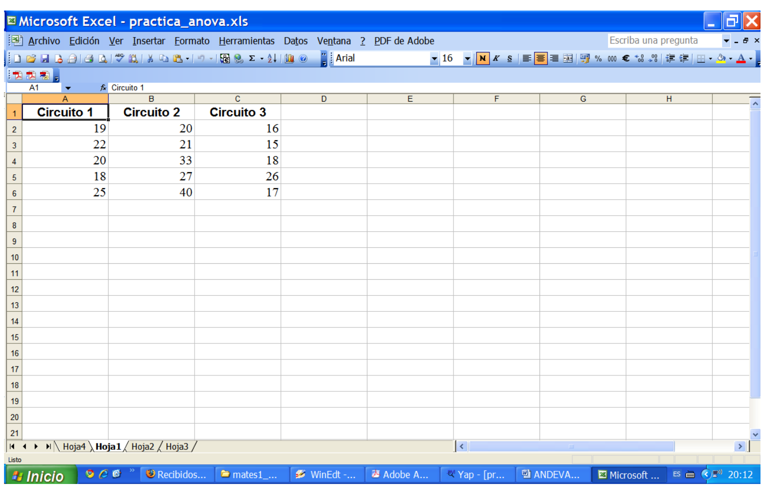
Queremos comprobar si hay alguna diferencia en los tiempos de respuesta entre los diferentes tipos de circuitos que se muestran en la siguiente tabla, en la que tenemos diferentes tiempos de respuesta para cada circuito:

Circuito 1	19	22	20	18	25
Circuito 2	20	21	33	27	40
Circuito 3	16	15	18	26	17

Vamos a hacer análisis de la varianza, para ello planteados dos posibles hipótesis, la llamada hipótesis nula, que denotaremos H_0, que supone que las medias de los tiempos de respuesta son iguales y la hipótesis alternativa, H_a, que supone que no todas son iguales.

$H_0 : \quad \mu_{circuito1} = \mu_{circuito2} = \mu_{circuito3}$
H_a : no todas las medias son iguales.

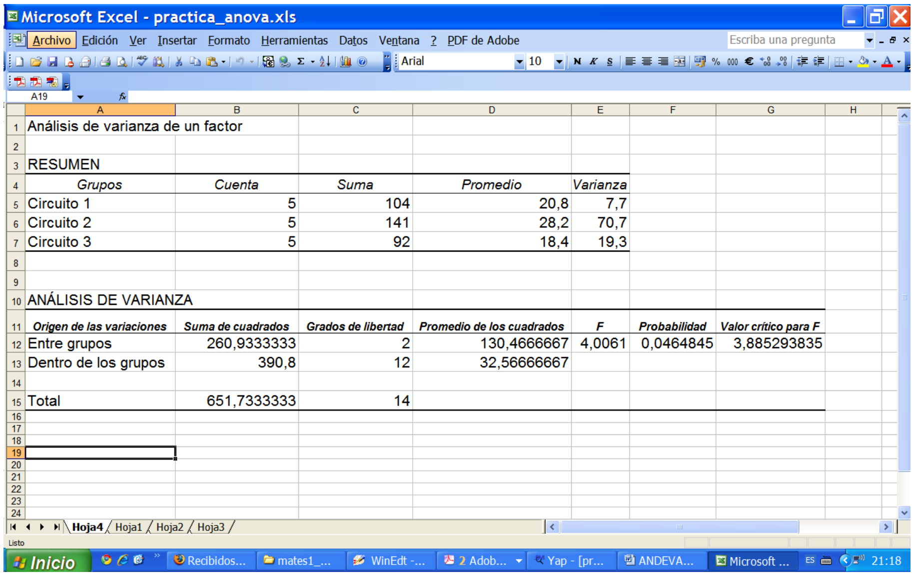
En primer lugar, introducimos los datos en una hoja excel:



En el menú principal abrimos **Herramientas** y elegimos la opción **Análisis de datos** (si esta opción no aparece en **Herramientas**, se puede activar en la opción **Complementos** buscando la opción **Herramientas para análisis**).

Se abre la ventana **Análisis de varianza de un factor** en la que se debe elegir el rango de entrada, marcándolo con el ratón. Incluimos en el rango de entrada los títulos de las columnas, en cuyo caso, para evitar que excel los confunda con los datos, también hay que marcar la opción **Rótulos en la primera fila**. Y pulsamos Aceptar

Si todo lo hemos hecho correctamente, se abrirá una hoja nueva con los siguientes datos:



Microsoft Excel - practica_anova.xls

Análisis de varianza de un factor

RESUMEN

Grupos	Cuenta	Suma	Promedio	Varianza
Circuito 1	5	104	20,8	7,7
Circuito 2	5	141	28,2	70,7
Circuito 3	5	92	18,4	19,3

ANÁLISIS DE VARIANZA

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Entre grupos	260,9333333	2	130,4666667	4,0061	0,0464845	3,885293835
Dentro de los grupos	390,8	12	32,56666667			
Total	651,7333333	14				

En la tabla RESUMEN aparecen calculadas las medias y varianzas de cada grupo.

En la tabla ANÁLISIS PARA LA VARIANZA aparecen los siguientes datos:

Entre grupos: en esta fila se mide la varianza de las medias de cada grupo respecto a la media total.

Dentro de los grupos: en esta fila se mide una varianza ponderada de las varianzas de cada grupo. Los pesos respectivos de cada varianza es el número de grados de libertad de cada grupo.

Total: mide la variación total de los datos en conjunto.

Suma de cuadrados: suma de los cuadrados de las desviaciones.

F: es el valor del estadístico. Si todas las medias fueran iguales, la varianza de las medias debería ser (aproximadamente) igual a la varianza promedio, en cuyo caso, $F \simeq 1$.

Probabilidad: nos da el menor nivel de significación para rechazar la hipótesis nula de igualdad de medias. Si la probabilidad es menor que 0,05, se rechaza la hipótesis nula.

Utilizando la función de análisis de la varianza para un factor del programa excel, obtenemos (con un nivel de significación $\alpha = 0{,}05$) el valor de F, que es el valor del estadístico de la prueba, la probabilidad y el valor crítico para F. Como la probabilidad es menos que 0,05, rechazamos la hipótesis nula. Otra forma de verlo es que el valor estadístico de F está dentro de la región de rechazo, es decir, $F >$ valor crítico para$F$. Por tanto, la conclusión es que las medias no son iguales.

Ejercicios

Ejercicio 1 *Existen tres métodos artesanales para la obtención de alcohol destilado de maíz, conocidos como: método Barleycorn, método Woodcorn y método Wincorn. Usando tecnología moderna se han aplicado los tres métodos a diferentes partidas con la misma cantidad de una determinada variedad de maíz con los siguientes resultados (obtenidos en litros):*

método Barleycorn	7	4	3	2	6		
método Woodcorn	5	4	8	7	10	4	
método Wincorn	9	4	9	2	8	10	7

Determinar si hay diferencia significativa entre las cantidades medidas destiladas mediante los 3 métodos.

Ejercicio 2 *Con el fin de comparar la resistencia de hormigones fabricados con tres tipos diferentes de áridos, se realizaron muestras experimentales utilizando el mismo cemento y se midió su resistencia. Los resultados fueron:*

Árido 1	595	610	634	619	605	620	682	
Árido 2	633	666	648	640	650	617	623	
Árido 3	480	432	507	471	516	608	686	493

¿Hay diferencias entre la resistencia media de los hormigones?

Ejercicio 3 *Se midió la humedad porcentual en muestras de arena con tres métodos diferentes, obteniéndose los siguientes resultados:*

Método 1	7,3	7,4	9,7	10,6	6,2	7,8			
Método 2	7,6	8,1	8,4	7,9	10,1	10,4	9,8	9,2	
Método 3	5,9	9,5	8,6	10,6	7,4	6,6	9,2	8,3	10,2