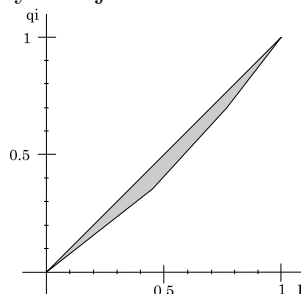


1. La curva de Lorenz se utiliza para representar gráficamente el grado de igualdad que hay en el reparto de total de una variable entre los individuos de una muestra. La curva de Lorenz se realiza marcando sobre unos ejes de coordenadas los pares (F_i, q_i) , uniéndolos mediante líneas y cerrando los extremos con los puntos $(0, 0)$ y $(1, 1)$ respectivamente. Es decir se van marcando los porcentajes del total de la variable q_i que le corresponden a cada fracción de la población F_i acumulando siempre de menor a mayor. A modo de ejemplo consideramos el experimento consistente en observar el sueldo (variable X) de cada persona de una oficina. La población la constituirían todos empleados y tendríamos como muestra toda la población. La variable es cardinal, de razón y continua porque los posibles valores son números cualesquiera y 0 significa ausencia de sueldo. Supongamos que en esa oficina hay 4 empleados que cobran 1500 euros, otros 3 que cobran 2000 y 2 que cobran 2500. Construimos la tabla de la desigualdad que contiene las columnas (F_i, q_i) y dibujamos la curva:

x_i	n_i	F_i	$x_i n_i$	q_i
1500	4	0.4444	6000	0.3529
2000	3	0.7778	6000	0.7059
2500	2	1	5000	1
Total	9	—	17000	—



lo que interpretaríamos diciendo que al 44.44 % de los empleados que menos cobran les corresponde un 35.29 % del sueldo conjunto, mientras que al 77.78 % les corresponde el 70.59 % (para que no hubiese desigualdad al 44.44 % le debería corresponder el 44.44 % y al 77.78 % también el 77.78 %). Vemos que la curva se parece bastante a la recta, lo que significa que no hay mucha desigualdad.

2. Planteamiento: tenemos 4 bienes, que son los destinos de los distintos vuelos (Madrid, Barcelona, París y Londres). Las magnitudes son los precios y el número (cantidad) de vuelos a cada destino. Los períodos son noviembre y diciembre del año pasado. En el apartado a) nos piden el índice de precios de Laspeyres. Tomaremos como base noviembre y como período actual diciembre y aplicamos una de las posibles fórmulas:

$$L_{pnov}^{dic} = \frac{\sum_{i=1}^N p_{it} \times q_{i0}}{\sum_{i=1}^N p_{i0} \times q_{i0}} = \frac{150 \times 220 + 200 \times 130 + 350 \times 25 + 560 \times 30}{150 \times 220 + 180 \times 130 + 410 \times 25 + 470 \times 30} = \frac{84550}{80750} = 1,0471$$

Podemos decir entonces que el precio medio de los vuelos de esa compañía subió un 4.71 % de noviembre a diciembre teniendo en cuenta el valor (o la facturación) de dichos vuelos en el mes de noviembre.

En el apartado b) nos preguntan por la parte de la variación del índice que se debe a los vuelos domésticos. Podríamos contestar tanto en términos absolutos (repercusión absoluta) como en términos relativos (repercusión en porcentaje o participación). Como lo más habitual es medir dicha contribución con la repercusión en porcentaje, será lo que haremos en este caso. Sabemos que la variación del índice de noviembre a diciembre fue del 4.71 %, ya que $\% \Delta_{nov \rightarrow dic}^{I_0^{t2}} = \frac{I_0^{t2}}{I_0^{t1}} = \frac{1,0471}{1} = 1,0471$. Se trata de determinar qué parte de ese 4.71 % se debe a los vuelos de Madrid y Barcelona (repercusión en porcentaje). Para hallar la repercusión en porcentaje hay que hallar primero la absoluta:

$$R_{nov}^{nov \rightarrow dic}(\text{Madrid}) = \frac{(I_0^{t2}(i) - I_0^{t1}(i))w_i}{\sum_{i=1}^N w_i} = \frac{(\frac{150}{150} - \frac{150}{150})150 \times 220}{150 \times 220 + 180 \times 130 + 410 \times 25 + 470 \times 30} = 0 \text{ puntos}$$

$$R_{nov}^{nov \rightarrow dic}(\text{Barc.}) = \frac{(I_0^{t2}(i) - I_0^{t1}(i))w_i}{\sum_{i=1}^N w_i} = \frac{(\frac{200}{180} - \frac{180}{180})180 \times 130}{150 \times 220 + 180 \times 130 + 410 \times 25 + 470 \times 30} = 0,0322 \text{ puntos}$$

es decir, el índice no subió nada (0 puntos) a causa de los vuelos a Madrid y subió 3.22 puntos porcentuales a causa de los vuelos a Barcelona. Para hallar la repercusión en porcentaje tenemos que dividir esos valores por el valor inicial del índice $I_{nov}^{nov} = 1$, por lo que el resultado numérico es igual, aunque cambia la interpretación: diríamos que el índice de precios no subió nada (un 0 %) a causa de los vuelos a Madrid y subió un 3.22 % a causa de los vuelos a Barcelona. En definitiva, del 4.71 % que subió el índice, el 3.22 % se debe a los vuelos domésticos.

3. En el apartado a) nos piden explicar el modelo exponencial con detalle. Planteamiento: el experimento consiste en observar el precio de las patatas (variable 1) y la cantidad (variable 2) que se vende cada día (individuo). La población la constituirán entonces todos los días y tenemos una muestra de tamaño desconocido. Según vemos en la salida del SPSS la variable independiente X es el precio (es la que controlaremos) y la variable dependiente Y es la cantidad (es la que podremos predecir). La fórmula que nos indica el SPSS es $Y = b0 * (e * (b1 * t))$, así que sustituyendo tendríamos que la fórmula de la regresión en este caso sería $Y = 2077,50e^{-0,0529X}$. Según vemos, el 95.9 % de la variabilidad de la cantidad de patatas vendidas se explica por la regresión exponencial con el precio, por lo que la fiabilidad de esta regresión será del 95.9 %. Esto significa que los valores de nuestra muestra se ajustan bastante bien a ese modelo exponencial, por lo que las predicciones serán fiables. Para realizar nuevas predicciones de la cantidad a partir de un precio dado mediante esta fórmula, deberíamos asegurarnos primero que el precio dado está entre 15 y 70 céntimos (mínimo y máximo) y luego sustituir en esa fórmula. Por ejemplo, si el precio fuera $x = 50$ céntimos, nuestra predicción de ventas sería de $Y = 2077,50e^{-0,0529 \times 50} = 147,5136$ kilos.

Para hacer las predicciones que nos piden en el apartado b) por regresión, en primer lugar debemos asegurarnos que los precios que nos dan están entre el mínimo y el máximo de la muestra. Vemos que 35 céntimos sí está entre 15 y 70, pero 90 no, luego podremos hacer una predicción en el primer caso pero no en el segundo. Para hacer la predicción debemos utilizar el mejor modelo. Miramos en la columna de Rsq y vemos que el mejor modelo es el inverso (INV), que explica el 99 % de la variabilidad de la cantidad vendida. La fórmula del modelo inverso es $y = b0 + (b1/t)$, por lo que vemos que nos quedaría $y = -256,24 + 20389,5/X$. Para hacer la predicción, sustituimos x por el 35 y nos queda $y = -256,24 + 20389,5/35 = 326,3171$ kilos. Es decir, si el precio de las patatas se fija en 35 céntimos, pronosticamos unas ventas de 326,3771 kilos. La predicción no es exacta (nunca podemos asegurar que va a ser una cantidad exacta, siempre puede haber algún error) aunque sí es bastante fiable, ya que el $R^2 = 0,990$ indica que los datos de la muestra se ajustan muy bien a ese modelo.

3. Planteamiento: el experimento consiste en observar la edad de los hombres y mujeres de Asturias, luego la variable será la edad y los individuos los hombres y las mujeres. Disponemos de muestras separadas para la población de hombres y mujeres. El tamaño de la muestra de la población de hombres es $109306 + 163192 + \dots + 14841 = 522980$, mientras que el tamaño de la muestra de la población de mujeres es $103931 + 162982 + \dots + 32676 = 565193$. La variable edad es cardinal (porque los valores que toma son números), continua (porque se pueden tener 23 años o 23.54, depende de la precisión que se emplee) y de razón (porque 0 años representa ausencia de edad). Nos dan los datos agrupados en clase-

En el apartado a) nos piden comparar la edad de un hombre de 45 años y una mujer de la misma edad relativamente, es decir, en comparación con el resto de hombre y de mujeres de sus respectivas muestras. Para ello podemos hacer 2 cosas: bien calcular el rango percentil y mirar en qué posición se sitúan dentro de sus muestras esas personas (% de personas con la misma edad o menos) o bien tipificar, lo que significa comparar la edad de esas personas con las medias de sus grupos en términos relativos. Vamos a optar por esta segunda forma. Para que no pueda haber lugar a confusión, vamos a llamar X a la edad de los hombres e Y a la edad de las mujeres. Para tipificar el valor 45 debemos calcular $z_{45}^x = \frac{45 - \bar{x}}{S_x}$ y $z_{45}^y = \frac{45 - \bar{y}}{S_y}$, lo que podemos hacer a partir de las tablas:

Distribución de edades de los hombres

Clase	x_i	n_i	$x_i n_i$	$x_i^2 n_i$
de 0 a 19 años	9.5	109306	$9,5 \times 109306$	$9,5^2 \times 109306$
de 20 a 40 años	30	163192	30×163192	$30^2 \times 163192$
de 40 a 60 años	50	130919	50×130919	$50^2 \times 130919$
de 60 a 80 años	70	104722	70×104722	$70^2 \times 104722$
de 80 a 100 años	90	14841	90×14841	$90^2 \times 14841$
Total	—	522980	21146347	1117385066.5

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^n x_i n_i}{N} = \frac{21146347}{522980} = 40.4343 \text{ años.} \\ \bar{x}^2 &= \frac{\sum_{i=1}^n x_i^2 n_i}{N} = \frac{1117385066,5}{522980} = \\ &2136.5732 \text{ años}^2 \\ S_x^2 &= \bar{x}^2 - \bar{x}^2 = 2136,5732 - 40,4343^2 = \\ &501.6406 \text{ años}^2 \\ S_x &= \sqrt{S_x^2} = \sqrt{501,6406} = 22,3973 \text{ años.}\end{aligned}$$

Podemos decir entonces que la edad de los hombres varía alrededor de 40.4343 años con una dispersión de 22.3973 años.

Distribución de edades de las mujeres

Clase	x_i	n_i	$x_i n_i$	$x_i^2 n_i$
de 0 a 19 años	9.5	103931	$9,5 \times 103931$	$9,5^2 \times 103931$
de 20 a 40 años	30	162982	30×162982	$30^2 \times 162982$
de 40 a 60 años	50	136036	50×136036	$50^2 \times 136036$
de 60 a 80 años	70	129568	70×129568	$70^2 \times 129568$
de 80 a 100 años	90	32676	90×32676	$90^2 \times 32676$
Total	—	565193	24689204.5	1395712350

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^n x_i n_i}{N} = \frac{24689204,5}{565193} = 43.6828 \text{ años.} \\ \bar{x}^2 &= \frac{\sum_{i=1}^n x_i^2 n_i}{N} = \frac{1395712350}{565193} = \\ &2469.4438 \text{ años}^2 \\ S_x^2 &= \bar{x}^2 - \bar{x}^2 = 2469,4438 - 43,6828^2 = \\ &561.2569 \text{ años}^2 \\ S_x &= \sqrt{S_x^2} = \sqrt{561,2569} = 23,6909 \text{ años.}\end{aligned}$$

Por lo que la edad de los mujeres varía alrededor de 43.6828 años con una dispersión de 23.6909 años.

Podemos aplicar entonces las fórmulas de tipificación anteriores para las personas de 45 años y obtendríamos:

$$z_{45}^x = \frac{45 - 40,4343}{22,3973} = 0,2039 \quad z_{45}^y = \frac{45 - 43,6828}{23,6909} = 0,0556,$$

Vemos que el hombre de 45 años está por encima de la media 0.2039 puntos en términos relativos, mientras que la mujer de 45 años sólo está 0.0566 puntos por encima en términos relativos. Luego, en comparación con su grupo, es mayor un hombre de 45 años que una mujer de esa edad.

En el apartado b) nos piden comparar la variabilidad de la edad en las muestras de hombres y de mujeres. La variabilidad se puede medir en términos absolutos mediante la desviación típica, que ya la tenemos calculada y vemos que es superior para el grupo de las mujeres. Sin embargo, sabemos que no se pueden comparar medidas absolutas porque dependen de las magnitudes, por lo que debemos utilizar el coeficiente de variación, que mide la variabilidad en términos relativos:

$$C.V.(x) = \frac{S_x}{\bar{x}} = \frac{22,3973}{40,4343} = 0,5539 \quad C.V.(y) = \frac{S_y}{\bar{y}} = \frac{23,6909}{43,6828} = 0,5423$$

Es decir, la edad de los hombres presenta una dispersión del 55.39 % en relación a su media, mientras que la variabilidad de la edad de las mujeres es del 54.23 % en relación a su media. Vemos que la variabilidad es similar en ambos grupos, aunque ligeramente superior en el de los hombres. Eso significa que la edad de los hombres está más dispersa alrededor de su media y, por lo tanto, dicha media es menos representativa que la del grupo de las mujeres.