

Transformació de variables

A partir d'una variable existent, en crearem de noves fent servir els següents tipus de transformacions:

1. **Recodificar:** assignar codis nous als valors d'una variable.
2. **Categoritzar o classificar:** crear un conjunt de classes.
3. **Càlcul funcional:** aplicar una mateixa funció (matemàtica) a tots els valors d'una variable.

Recodificar: assignar codis nous a una variable. Per exemple,

1. Podem passar els valors d'una variable qualitativa (e.g., religió=cristiana, musulmana, budista,...) a valors numèrics (cristiana=1, musulmana=2, budista=3,...). S'anomena recodificació automàtica.
2. Podem donar nous valors a una variable numèrica (e.g. si la variable és 'persones per habitatge', podem deixar tots els valors iguals excepte i fer correspondre el valor 9 sempre que sigui més gran de 6).

Categoritzar o classificar: a partir d'una variable numèrica (e.g. edat) es crea una nova variable que s'obté agrupant els valors de l'antiga variable (e.g., joves=[0,30), adults=[30,60), grans=[60,120),...).

Càlcul funcional: a partir d'una o més variables, en creem una de nova aplicant una funció matemàtica a les anteriors. Per exemple,

1. Si X =alçada d'una persona en m, i volem definir una nova variable Y que contingui l'alçada en cm, haurem de multiplicar tots els valors de X per 100, és a dir, $Y=X \cdot 100$.
2. Si X =població d'un país i Y =superfície d'un país en Km^2 , podem crear una nova variable D que contingui la densitat de població (habitants per Km^2). Els valors d'aquesta nova variable s'obtenen dividint la població per la superfície ($D = \frac{X}{Y}$).

Què passa als descriptius centrals i de dispersió quan apliquem el càlcul funcional a una variable? Estudiarem aquesta pregunta en alguns exemples concrets de transformació:

1. Translacions: a cada observació sumem el mateix número constant C .
2. Multiplicar per constants: multipliquem cada observació pel mateix número constant K .

Translacions: a cada observació de X sumem el mateix número constant C . Si C és positiu, estem traslladant tots els valors a la dreta, i si C és negatiu, estem traslladant tots els valors cap a l'esquerra.

Per exemple, els valors màxims i mínims canviaran (els haurem de sumar la constant C). La mitjana també canviarà de la mateixa forma. La mitjana de $X + C$ serà $\bar{X} + C$.

En canvi, la forma (simetria i curtosi) i les mesures de dispersió (variància i desviació típica) són les mateixes.

Multiplicar per una constant (o canvi d'escala): multipliquem cada observació de X per un mateix número K .

És important distingir quan K és positiva i quan és negativa. Si és negativa, intercanvia els valors de costat a l'eix horitzontal. Aleshores, per exemple, els valors màxims i mínims no només es veuen multiplicats per K sinó que a més s'intercanvien.

La forma de la distribució sofreix un canvi d'escala. Si K és negativa, la simetria canvia també de signe (s'intercanvien els braços de dreta i esquerra) i si K és positiva es manté.

La mitjana es veu multiplicada per la constant K i la dispersió augmenta si $K > 1$ o $K < -1$ i disminueix si $-1 < K < 1$. Per exemple, la variància es veu multiplicada per K^2 i la desviació típica per K .

Puntuacions tipificades o stàndard: serveix per comparar la posició relativa de dos o més individus (o casos) respecte la mostra a la que pertanyen.

Si x_i és un resultat obtinguts per una variable X que té mitjana \bar{X} i desviació típica s , la seva puntuació tipificada corresponent és

$$z_i = \frac{x_i - \bar{X}}{s}$$

Observeu que si aïllem x_i de la fórmula anterior tenim $x_i = \bar{X} + z_i s$. És a dir, z_i és el número de desviacions típiques que el nostre valor s'allunya de la mitjana.

Propietats:

1. Si z_i és positiu aleshores x_i està per sobre de la mitjana i si és negatiu aleshores x_i està per sota de la mitjana.
2. Ens indica la posició d'un individu respecte el seu grup.
3. Les puntuacions tipificades no tenen unitats.
4. La nova variable obtinguda Z que s'obté a partir de X (primer restem \bar{X} i després multipliquem per $1/s$), té mitjana 0 i desviació típica 1.