

PROPORCIONES, PORCENTAJES Y RAZONES (Guía de clase)

FICHA N° 11

A. Los **estadísticos** que se calculan para las variables nominales y ordinales conforman lo que clásicamente fuera denominado como “estadística no paramétrica”(por ejemplo, ver Siegel 2003). Actualmente, el análisis de estas variables recibe el nombre de análisis de datos categóricos o cualitativos (Ver, por ejemplo, Agresti 1996).



En la estadística no paramétrica las operaciones aritmética que pueden hacerse con las variables son muy reducidas.

- i) En estos dos tipos de variables o escalas de medida, los números que eventualmente se hayan podido asignar a cada categoría o valor de la variable sólo dan cuenta de “cualidades” o de “ordenaciones”.
- ii) No se pueden realizar ninguna operación aritmética con las categorías de una variable nominal. Carece de sentido restar asiáticos a europeos, católicos a protestantes, etc se le ha asignado a los primeros el valor 2 y a los segundos el valor 1.
- iii) Un caso particular lo constituyen las variables dicotómicas: aquellas que adquieren sólo dos valores que por convención son 0 y 1. Esto se tratará más abajo.
- iv) Una modificación de los números que se han asignado a las categorías de una variable nominal, no transforma la escala de medida de la variable.
- v) Las versiones más recientes del SPSS (c) por ejemplo, admiten que en lugar de números se asigne cualquier combinación de letras a las categorías de las variables.
- vi) La distinción entre los tipos de variables es responsabilidad del investigador: los paquetes estadísticos de hecho realizan cualquier cálculo mientras que haya números identificando las categorías, sin importar si esto tiene o no algún sentido.



Los estadísticos calculables para variables nominales y ordinales son las proporciones y los porcentajes. Adicionalmente se pueden calcular razones.

- i) Nuestro pasaje por la educación básica y media nos ha proporcionado una noción elemental y operativa de qué significan cada uno de estos términos.
- ii) La mayor dificultad de exposición es hacer explícita esta noción y reflexionar sobre ella.

PROPORCIONES, PORCENTAJES Y RAZONES (Guía de clase)

A. Una proporción informa de tamaño que tiene una parte en relación a la totalidad.



Una variable permite analizar cómo se distribuyen las unidades en cada una de sus categorías o valores.

- i) Si la variable cumple con las propiedades requeridas, la totalidad de las unidades, N , ha sido “particionado” en tantas partes como categorías o valores hayan empíricamente, y cada unidad se clasificará en una sola categoría.
- ii) La suma de todas las unidades así clasificadas dará como resultado N .
- iii) En consecuencia, cada categoría tendrá un tamaño que lógicamente será menor a N . Podría ocurrir el caso extremo de que todas las unidades se hayan clasificado en una categoría; sin embargo, se trataría de una constante y como tal no tiene interés para el análisis.
- iv) Si denominados con la letra j a una categoría cualquiera de la variable, su tamaño en relación al todo o proporción (H_j) se calculará dividiendo el número de casos (N_j) que se han contado en esa categoría entre el total de unidades (N).
- v) Expresado en términos de una función:

$$[I.1] \quad H_j = N_j / N \quad (\text{para las categorías } J = 1, 2, 3, \dots, J)$$

- vi) Lógicamente se deriva de la anterior función tres conclusiones:
 - se calcularán tantas proporciones como categorías tenga una variable.
 - la proporción de una categoría cualquiera será menor que 1. (Excepto en el caso extremo de una constante).
 - la suma de todas las proporciones de las categorías de una variable sumará 1. Expresado en una función:

$$[I.2] \quad H_1 + H_2 + H_3 + \dots + H_j = 1$$

$$[I.3] \quad \sum_{J=1}^J H_J = 1$$

PROPORCIONES, PORCENTAJES Y RAZONES (Guía de clase)



En términos más abstractos se puede concluir diciendo que una proporción es un estadístico que informa cómo se distribuye una variable en términos *normalizados*.

- i) Dado que las distribuciones están normalizadas pueden compararse dos conjuntos de unidades que difieren en el tamaño (N).
- ii) Si el tamaño de la matriz es reducido, por ejemplo, menor de 30, las magnitudes de las proporciones (en términos absolutos, en tanto números) pueden prestarse a errores para la interpretación.
- iii) En algunos campos disciplinarios, la proporción puede ser denominada como incidencia: tal es el caso de los estudios de pobreza por ejemplo.

B. Un porcentaje es una proporción que ha sido multiplicada por 100. Esta transformación lineal de los valores no altera sustantivamente la distribución.



En términos formales, el porcentaje de unidades (que podemos simbolizar como P) que se han clasificado en la *j-ésima* categoría de la variable de interés se puede escribir como:

$$[I.4] \quad P_J = N_J / N * 100 \quad (\text{para toda categoría } J = 1,2,3,\dots,J)$$

- i) De la anterior ecuación se derivan algunas conclusiones formales:

- [I.5] $P_J \leq 100$

- [I.6] $P_1 + P_2 + P_3 + \dots + P_J = 100$

- ii) Si se recodifica una variable para agrupar dos categorías, por ejemplo las categorías identificadas como 1 y 2, el porcentaje de la nueva categoría recodificada (llamémosle P_r) será igual a la suma de las dos anteriores. La misma generalización se puede realizar para las proporciones. Formalmente:

- [I.7] $P_r = (N_1 + N_2) / N * 100 \quad (\text{Para } r = 1 + 2)$

- [I.8] $P_r = (N_1 / N * 100) + (N_2 / N * 100)$

- [I.9] $P_r = P_1 + P_2$

PROPORCIONES, PORCENTAJES Y RAZONES (Guía de clase)

iii) Expresado en una tabla de frecuencias:

Variable	Frecuencia absoluta	Proporción	Porcentaje
Categoría 1	N_1	H_1	P_1
Categoría 2	N_2	H_2	P_2
Categoría 3	N_3	H_3	P_3
.....
Categoría J	N_J	H_J	P_J
Total	N	1,0	100,0



Claramente la interpretación de una distribución se hace más sencilla y elocuente cuando se presenta mediante porcentajes, aunque numéricamente sean equivalentes.

- i) Intuitivamente, el porcentaje permite entender más rápidamente cuál es el tamaño relativo de las unidades que se han clasificado en una categoría de la variable. Sencillamente se puede entender que de cada 100 unidades, P se encuentran en la categoría J.
- ii) La normalización se ha llevado a cabo suponiendo que el total de unidades es igual a 100.



Es de advertir que la opción por multiplicar por 100 una proporción es el resultado de una convención general, pero no excluye otros multiplicadores, como podría ser por 1000. Las tasas de natalidad y mortalidad generalmente se expresan por millares.

C. Dadas las definiciones realizadas anteriormente respecto de las proporciones y de las razones, se sigue que estos estadísticos pueden calcularse perfectamente también para las variables intervalales y de razón, siguiendo los algoritmos presentados.

- i) Se sigue que habrán tantos estadísticos (proporciones o razones) como valores tenga una variable interval (o de razón)
- ii) Puede darse el caso extremo de una distribución donde los valores de la variable no se repitan entre las unidades; o lo que es lo mismo:

PROPORCIONES, PORCENTAJES Y RAZONES (Guía de clase)

[I.10] $N_1 = N_2 = N_3 = \dots = N_J = 1$

[I.11] $(N_1 / N) = (N_2 / N) = (N_3 / N) = \dots = (N_J / N)$

[I.12] $H_1 = H_2 = H_3 = \dots = H_J$

[I.13] $P_1 = P_2 = P_3 = \dots = P_J$

- ii) Resulta evidente de lo anterior que si para una variable se han observado 30 valores por ejemplo, existirán 30 estadísticos potenciales que podrían ser utilizados para describir la distribución empírica.
- iii) En estos casos, el uso de proporciones y porcentajes para variables intervalos o de razón resulta poco práctico. De hecho, se pierde el objetivo principal de la estadística: resumir las características de una distribución empírica.

D. Una **razón** es una relación que se establece entre dos números. Cuando estos números son proporciones o porcentajes, se puede afirmar que una razón es un “estadístico de estadísticos”.



Al menos podrían definirse tres grandes tipos de razones según cuál sean los estadísticos que se utilizan para calcularlas.

- i) una razón puede calcularse relacionando las frecuencias absolutas de dos valores de una misma variable. Si las categorías o valores de interés son las identificadas con los números 1 y 2, la razón R entre ambas se expresará formalmente:

- [I.14] $R = N_1 / N_2$

- Por ejemplo, se podría decir que según las estimaciones de la incidencia de la pobreza de ingresos en el Uruguay para el 2001 (0.270) y según la población estimada, existían 864.000 uruguayos pobres sobre una población total estimada de 3.200.000. Lo cual supone una razón de 2,70 personas no pobres por cada pobre en el país:

- [I.15] $R = N_{np} / N_p$

- [I.16] $R = (2.336.000) / (864.000)$

- [I.17] $R = 2,70$

PROPORCIONES, PORCENTAJES Y RAZONES (Guía de clase)

- ii) el segundo tipo de razones puede calcularse relacionando dos proporciones calculadas para una misma variable. En este caso, primero se calcularán las proporciones para todas las categorías de la variable y luego se calcularán las razones de acuerdo al interés. Formalmente:

- **[I.18]** $R = H_1 / H_2$

- Sustituyendo en la anterior ecuación por las funciones respectivas, se observa que el resultado es el mismo que en el caso anterior:

- **[I.19]** $R = H_{np} / H_p$

- **[I.20]** $R = (N_{np} / N) / (N_p / N)$

- **[I.21]** $R = N_{np} / N_p$

- Siguiendo con el mismo ejemplo anterior, la razón se calcularía:

- **[I.22]** $R = 0.730 / 0.270$

- **[I.23]** $R = 2.70$

- iii) Se podría calcular una razón que combine información proveniente de categorías de distintas variables. En este caso, la razón constituye una relación entre variables o mejor entre categorías o valores de dos variables distintas. Si se designa como J la categoría de interés de una variable y M la categoría de interés de otra variable, la razón se calculará:

- **[I.24]** $R = P_J / P_m$

- Considerése el siguiente ejemplo tomado del informe de Cortés, F.; Banegas, I.; Fernández, T. & Mora, M. (2003) *Perfiles de la Pobreza en Chiapas*. El Colegio de México.

“Cuando estas carencias se relativizan con referencia al estrato menos “desfavorecido”, se observan cuáles carencias son debidas a falta de servicios en general y cuáles a la falta de servicios en los territorios habitados por los más pobres. En particular, un hogar rural y pobre según LPI tiene 19 veces más chance de carecer de servicio sanitario; 16 veces más de carecer de drenaje; 15 veces más de carecer de electricidad; casi 9 de carecer de piso firme y 3 veces más chance de carecer de agua entubada, frente al hogar de referencia (urbano y no pobre de nivel 1).

PROPORCIONES, PORCENTAJES Y RAZONES (Guía de clase)

Porcentajes y Razones de carencias en la vivienda según el contexto rural / urbano para el nivel I de pobreza

	Sin agua entubada	Con piso de tierra	Sin drenaje	Sin servicio sanitario	Sin electricidad
PORCENTAJE DE HOGARES					
Pobres rurales	85,6%	58,4%	59,7%	30,9%	19,7%
Pobres urbanos	47,8%	20,8%	12,4%	5,9%	4,9%
No pobres rurales	68,4%	30,5%	29,5%	15,4%	9,9%
No pobres urbanos	26,2%	6,7%	3,7%	1,6%	1,3%
RAZONES DE CARENCIA					
Pobres rurales	3,3	8,7	16,1	19,3	15,2
Pobres urbanos	1,8	3,1	3,4	3,7	3,8
No pobres rurales	2,6	4,6	8,0	9,6	7,6

Fuente: COLMEX (2003). Cálculos hechos con base en la muestra del XII Censo de Población del año 2000 según el método propuesto por el Comité para la Medición de la Pobreza en México.



Debe señalarse de que no existe ninguna convención a los efectos de elegir cuál es la categoría que se utilizará en el numerador y cuál en el denominador. Depende básicamente del interés del investigador.

- i) En el ejemplo de la pobreza sobre Uruguay, se siguió el criterio de los tamaños absolutos de ambas categorías. Puesto que los no pobres son más personas, la información que proporcionaría la razón sería más elocuente.
- ii) En el ejemplo sobre Chiapas, el criterio fue sustantivo: brechas de bienestar entre grupos sociales caracterizados por la presencia o ausencia de los atributos de pobreza y de ruralidad.
- iii) Como sugerencia general, se podría afirmar que conviene utilizar en el numerador la categoría con mayor cantidad de unidades, para facilitar la lectura.



Tal como se ha podido apreciar, los tipos (i) y (ii) conducen a resultados exactamente iguales; el tipo (iii) en cambio entrega una información sustantivamente distinta.