

IN3401- Estadística para Economía y Gestión

Profesores: Marcelo Henríquez, Sebastián Maldonado

Auxiliares: José Miguel Carrasco, Gabriel Espejo

Fecha: 12 de Mayo del 2010

P1 Estadística Básica.

Considere la base construida a partir de los datos que se adjuntan con esta tarea en formato Excel. Impórtelos desde SPSS.

i) ¿Cuál es la escala de las variables?

Al analizar las variables en estudio, se tiene que éstas pueden ser clasificadas como:

a) Nominale

b) Ordinale

c) Escala. Incluye a “intervalos” y “razón”

Se ha usado esta separación dado que es la misma que utiliza SPSS:

Código	Descripción	Escala
Id	Identificación	Nominal
GSE	Grupo Socio Económico	Ordinal
PS0	Nombre	Nominal
PS1	Edad	Razón
PS2	Sexo	Nominal
PS3	Posición dentro de la Familia	Nominal
PS4	Actividad jefe de Familia	Ordinal
PS5	Estudios del jefe de Familia	Ordinal
PS6	Si entrevistada no es jefe de familia, actividad	Ordinal
PS7	Si entrevistada no es jefe de familia, estudios	Ordinal
PC1	¿Qué marcas de ropa de mujer conoce? (primera mención)	Nominal
PC2	¿Cuál otra? (segunda mención)	Nominal
PC3	¿Y de las siguientes marcas, conoce alguna? (primera mención guiada)	Nominal
PC4	¿Cuál usa habitualmente? (primera mención)	Nominal
PC5	¿Cuál es su marca favorita?	Nominal
PC6	Pensando en su próxima compra ¿qué marca consideraría comprar?	Nominal
PC7	¿Qué marcas recomendaría? (primera mención)	Nominal
PC8	¿Por qué es su marca favorita (PC5)? (primera mención)	Nominal
PC9	¿Dónde compra habitualmente su ropa femenina? (primera mención)	Nominal
PO1	Distribución de Zara en una multitienda	Ordinal

¹ Al importar los datos desde SPSS no se obtienen los mismos datos, esto se produce debido a que en la base de datos a cada mención se le asigna un número, esto hace que el programa interprete de manera distinta los datos (como simples números), asignando escalas que no siempre son las correctas.

- ii) ¿Qué distribución es de esperar tengan estas variables?
 Para ver que distribuciones son las esperadas, es necesario analizar los tipos de respuestas de las variables de estudio; si se permiten muchas respuestas o un número reducido, o si el tipo de medición se ajusta a alguna ley que permita estimar la distribución.

Código	Descripción	Escala
GSE	Grupo Socio Económico	Multinomial
PS0	Nombre	Multinomial
PS1	Edad	Normal
PS2	Sexo	Binomial
PS3	Posición dentro de la Familia	Multinomial
PS4	Actividad jefe de Familia	Multinomial
PS5	Estudios del jefe de Familia	Multinomial
PS6	Si entrevistada no es jefe de familia, actividad	Multinomial
PS7	Si entrevistada no es jefe de familia, estudios	Multinomial
PC1	¿Qué marcas de ropa de mujer conoce? (primera mención)	Multinomial
PC2	¿Cuál otra? (segunda mención)	Multinomial
PC3	¿Y de las siguientes marcas, conoce alguna? (primera mención guiada)	Multinomial
PC4	¿Cuál usa habitualmente? (primera mención)	Multinomial
PC5	¿Cuál es su marca favorita?	Multinomial
PC6	Pensando en su próxima compra ¿qué marca consideraría comprar?	Multinomial
PC7	¿Qué marcas recomendaría? (primera mención)	Multinomial
PC8	¿Por qué es su marca favorita (PC5)? (primera mención)	Multinomial
PC9	¿Dónde compra habitualmente su ropa femenina? (primera mención)	Multinomial
PO1	Distribución de Zara en una multitienda	Multinomial
PO2	Tienda de distribución	Multinomial
PO3	Tiempo de última compra de producto Zara	Normal
PO4	Calificación para Zara	Normal

- iii) Calcule estadísticos descriptivos e histogramas de todas ellas.
SPSS calcula dentro de las posibilidades los estadísticos descriptivos.

	N	Rango	Mínimo	Máximo	Media	Desv. típ.	Varianza
PS1	301	20	25	45	35,16	5,914	34,981
PS2	301	1	1	2	1,00	,058	,003
PS3	301	3	1	4	1,94	,577	,333
PS4	301	4	1	5	2,86	,864	,747
PS5	301	5	1	6	2,85	,995	,990
PS6	245	96	3	99	33,17	17,890	320,052
PS7	245	98	1	99	5,70	14,929	222,866
PC1	301	139	1	140	30,84	30,552	933,399
PC2	228	108	1	109	24,41	23,555	554,851
PC3	124	118	2	120	28,10	22,360	499,981
PC4	301	96	1	97	5,49	12,230	149,584
PC5	301	98	1	99	24,15	34,417	1184,530
PC6	301	98	1	99	36,91	38,627	1492,038
PC7	301	118	1	119	31,48	39,236	1539,430
PC8	223	36	1	37	25,04	11,831	139,980
PC9	258	98	1	99	5,18	12,578	158,194
PO1	135	8	1	9	1,81	2,286	5,227
PO2	113	98	1	99	6,47	21,787	474,662
PO3	38	998	1	999	30,32	161,435	26061,411
PO4	52	5	2	7	6,15	1,036	1,074
N válido (según lista)	16						

- iv) Desarrolle un ranking de el TOP (top of mind, correspondiente a la PC1).

La PC1 corresponde a “¿Qué marcas de ropa de mujer conoce? (solo primera mención)”. El desarrollo del ranking consiste en ver la frecuencia de las distintas menciones, para luego confeccionar el ranking. Este proceso se realiza mediante SPSS. El “Top 5 of Mind” es:

Marca	Frecuencia	Porcentaje	Porcentaje acumulado
Wados	33	10,9634551	10,9634551
Zara	27	8,97009967	19,9335548
Basement	18	5,98006645	25,9136213
Wrangler	16	5,31561462	31,2292359
Ellus	10	3,32225914	34,551495

P2 SPSS.

v) Cambie todas las etiquetas de variables a máximo dos palabras. Use la sintaxis:

var label

variable ‘etiqueta’.

Las etiquetas asignadas a cada categoría son:

Código	Etiqueta
GSE	Grupo Socioeconómico
PS0	Nombre
PS1	Edad
PS2	Sexo
PS3	Posición familia
PS4	Actividad jefe
PS5	Estudios jefe
PS6	Actividad Otro
PS7	Estudios Otro
PC1	Marcas conoce
PC2	Otra
PC3	Conoce alguna
PC4	Usa habitualmente
PC5	Marca preferida
PC6	Consideraría comprar
PC7	Marca recomendaría
PC8	Causa preferencia
PC9	Donde compra
PO1	Zara Multitienda
PO2	Tienda distribución
PO3	Tiempo compró
PO4	Nota Zara

vi) Realice tests de diferencias de medias de al menos 3 variables que considere relevantes, comparando las muestras ¿Se verifican los supuestos para aplicar esta metodología? Fundamente su respuesta e interprete los resultados.

Podemos realizar test de diferencias mediante Analyze -> Compare means -> Independent Samples T-test

Group Statistics				
	PS3	N	Mean	Std. Deviation
PS6	1,00	1	3,0000	.
	3,00	36	29,8889	22,41400
PC1	1,00	57	24,6140	24,25663
	3,00	36	25,0278	25,67432

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
PS6	Equal variances assumed			-1,183	35	,245	-26,88889	22,72317	-73,01938	19,24160
	Equal variances not assumed						-26,88889			
PC1	Equal variances assumed	,197	,658	-,078	91	,938	-,41374	5,28209	-10,90596	10,07847
	Equal variances not assumed			-,077	71,403	,939	-,41374	5,35096	-11,08222	10,25473

vii) Defina la variable “etareo”, como un resumen de “edad” según 4 tramos. Use la sintaxis (por ejemplo):

Compute educacion=1*(edad>20)*(edad<=30)+2*...

Execute.

viii) Ponga etiquetas a esta variable etareo que acaba de crear, con la sintaxis: val label etareo caso ‘etiqueta’.

Las etiquetas asignadas fueron:

Tramo	Etiqueta
Tramo 1	Entre 25 y 30
Tramo 2	Entre 31 y 35
Tramo 3	Entre 36 y 40
Tramo 4	Entre 41 y 45

ix) ¿Qué escala tiene esta nueva variable? Calcule su frecuencia.

La escala de la variable ‘Etareo’ es “intervalo”, ya que sus valores representan magnitudes y la distancia entre los números de su escala es igual. Como no hay cero absoluto, corresponde a Intervalo y no a Razón.

En cuanto a la frecuencia, ésta se obtiene mediante SPSS:

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Entre 36 y 40	87	28,9	28,9	28,9
Entre 25 y 30	86	28,6	28,6	57,5
Entre 31 y 35	65	21,6	21,6	79,1
Entre 41 y 45	63	20,9	20,9	100,0
Total	301	100,0	100,0	

x) Calcule el cruce de esta variable con la variable TOP (en SPSS: Analyze → Descriptive Statistics → Crosstabs). Interprete. ¿Cómo cambia su

interpretación al repetir el cálculo con la segunda mención (PC2) y la mención guiada (PC3)?

Al observar las tablas resultantes del cruce de las variables *etéreo* y *TOP*, se puede apreciar que la marca más conocida en general, también es la más conocida en casi todos los rangos de edad, en base a esto, es posible inferir que su popularidad no depende de la edad de las mujeres, por lo tanto, no habría una relación significativa entre las variables.

Las únicas marcas en que se observa algo distinto, son la marca 24, que claramente es preferida por las mujeres más jóvenes, y la marca 40, que es preferida por los rangos de mayor edad.

En el caso del cruce de *etéreo* con la segunda mención de marca de ropa de mujer conocida, tampoco se aprecia una relación entre la edad y la marca. Lo que sí llama la atención es la falta de respuesta en los rangos de mayor edad, no así en los rangos de menor edad donde la ausencia de respuesta es menor. Se podría decir que la gente joven retiene mucho más las marcas de moda.

Cabe señalar que en ambas menciones espontáneas, la marca 4 fue la más nombrada.

En el caso del cruce con la mención guiada, tampoco se observa una relación, el hecho que haya menos respuestas hace que sea aun más complicada poder interpretar algo. Se repite el patrón anterior donde, las marcas más conocidas en general, son marcas muy conocidas en cada grupo *etéreo*.

xi) ¿Existe alguna relación significativa? (dependencia significativa bajo test chi-2) (en el Crosstabs de SPSS, antes de apretar OK para calcular, seleccione Statistics → Chi-square).

a) *Variables TOP - etéreo:*

Para iniciar el test chi-cuadrado, se plantean las hipótesis con las cuáles se trabajará:

H_0 = *Etéreo* y *Top* son independientes (no asociación)

H_1 = *Etéreo* y *Top* son dependientes (asociación).

Luego, con la ayuda de la tabla de contingencia se calculan las frecuencias esperadas suponiendo que estamos bajo la hipótesis de no asociación. Para dicho fin, se utiliza la expresión:

$$e_{ij} = (n_{i.} \cdot n_{.j}) / n$$

Donde

- e_{ij} es la frecuencia esperada para la marca i en el tramo etéreo j .
 $n_{i.}$ es el total de personas que señalaron a la marca i .
 $n_{.j}$ es el total de personas que se encuentran en el tramo etéreo j .
 n es el tamaño total de la población.

Con estos datos, se construye el siguiente estadístico, que bajo H_0 sigue una distribución chi-cuadrado:

$$\chi^2 = \sum_{i=1}^{140} \sum_{j=1}^4 \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

Siendo

- n_{ij} la frecuencia observada para la marca i en el tramo etéreo j .

Pruebas de chi-cuadrado:

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	226,144 (a)	222	,410
Razón de verosimilitudes	247,321	222	,117
Asociación lineal por lineal	1,203	1	,273
N de casos válidos	301		

(a) 290 casillas (96,7%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es ,21.

Viendo la tabla de resultados del test, se concluye que el valor de la suma de las “distancias” entre las frecuencias esperadas y observadas es igual a 226.144. Con esto, y sabiendo que se tienen 222 grados de libertad se puede calcular el p-value.

$$P\text{-value} = p(\chi_{222}^2 > 226.144) = 0.41 > 0.05$$

Finalmente como el p-value es mayor que el 5% de significancia, entonces no podemos rechazar la hipótesis de independencia. No existe relación significativa entre el Top of mind y el rango de edad.

b) Variables Segunda mención - Etéreo:

Razonando de forma similar, y teniendo la tabla de resultados, se aprecia que el p-value en esta oportunidad vale 0.205, que pese a ser menor que la vez anterior, continúa siendo mayor que 0.05 (5%). Por ende, se concluye que tampoco hay una relación significativa entre la segunda mención (PC2) y la variable etéreo.

Pruebas de chi-cuadrado:

		Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	de	204,804 (a)	189	,205
Razón de verosimilitudes	de	209,071	189	,151
Asociación lineal por lineal	por	,488	1	,485
N de casos válidos		228		

(a) 253 casillas (98,8%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es ,19.

c) Variables mención guiada (PC3) - Etéreo:

Pruebas de chi-cuadrado:

		Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	de	132,607 (a)	132	,469
Razón de verosimilitudes	de	133,136	132	,456
Asociación lineal por lineal	por	1,784	1	,182
N de casos válidos		124		

(a) 180 casillas (100,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es ,18.

Para el test chi-cuadrado de estas dos variables se obtuvo un p-value igual a 0.469, que es mucho mayor que el 5% con el cual se rechazaría la hipótesis de independencia, por lo cual se concluye que no hay relación significativa entre la mención guiada y la edad de las personas (con los rangos autodefinidos).

Nota: para todos los cálculos entregue outputs y sintaxis. Para obtener las sintaxis de los cálculos, apretar “Paste” en vez de “Ok”. Es posible correr desde spss o desde las sintaxis.

xii) Defina una variable que integre el TOP, segunda mención y mención guiada y llámela conoc. Calcule su frecuencia.

Como en SPSS es imposible establecer un sistema de vectores que contengan y extraigan los tres valores de mención por individuo y por esto aplicamos un proceso de cuenta simple de cada mención vía Excel, para luego determinar el total de menciones que existió dentro de la muestra y así establecer la fracción de menciones por marca dentro del total de menciones y así fijar la frecuencia de menciones por marca dentro del rango.

Junto con esto establecimos un nivel de marcas para el estudio que son preponderantes a nivel de menciones y que a su vez representan un 68,61% del total de las mismas. A continuación se presenta el listado de dichas marcas:

Marca	Porcentaje
4	9,34%
27	7,35%
14	7,20%
2	6,13%
22	4,13%
28	3,68%
3	3,22%
24	3,06%
5	2,45%
18	2,45%
1	2,30%
30	2,30%
33	2,30%
17	2,14%
7	1,99%
21	1,99%
40	1,84%
9	1,68%
32	1,53%
61	1,53%
Acumulado	68,61

xiii) Seleccione (filtre) la submuestra correspondiente a la edad ≤ 35 años (en SPSS vaya a Data → Select Cases → If condition is satisfied y escriba el criterio de selección; seleccione Unselected cases are “deleted” en vez de “Filtered”; no olvide guardar esta operación en la sintaxis apretando Paste en vez de Ok). Guarde este nuevo archivo con otro nombre.

El código de sintaxis de este filtro es:

```
FILTER OFF.  
USE ALL.  
SELECT IF(PS1 <= 35).  
EXECUTE .
```

Para realizar esta instrucción, en seleccionar datos, se debe clicar la opción “Si satisface la condición” y luego imponer que $PS1 \leq 35$, es decir que la edad sea menor que dicho valor. Para guardar esta selección en otro archivo, se elige: Copiar casos seleccionados a un nuevo conjunto de datos y posteriormente se escoge el nombre del nuevo archivo.