

## TEMA 4

### ANÁLISIS DE LOS DATOS

El proceso de la investigación. El proceso de investigación ya fue analizado, acá ponemos en perspectiva algunos puntos.

El proceso de los datos. Los datos se entran al Excel, se analizan con el SAS o SPSS y se hace un informe escrito, con el Word y oral con el PowerPoint.

El análisis estadístico de la información.

El informe de investigación. Es todo un arte. Algunos puntos merecen comentarios.

Estudio de casos. *Tomaremos como referencia los casos de procesamiento de los datos que hemos trabajado en Salto 1) del Centro de Salud 2) que estamos llevando a cabo.*

#### **4.1. CONCEPTOS ESTADÍSTICOS DE PROCESAMIENTO DE LOS DATOS.**

Discutiremos algunas perspectivas estadísticas y algunos problemas usuales del procesamiento de datos.

**PASOS DEL ANÁLISIS.** El procesamiento de los datos tiene las siguientes etapas:

**1. Análisis univariado.** Siempre comenzamos con un análisis univariado, algunas veces la investigación no pasa de acá.

**2. Análisis bivariado.** En el caso de variables cualitativas se habla de "cruce de variables". El resultado se expresa en una tabla de contingencia. Las tablas resultantes también se conocen como "de doble entrada". Puede haber de triple entrada o más, pero son difíciles de interpretar.

**3. Modelación.** La modelación estadística es la culminación del proceso de investigación. El estado actual de la investigación que nosotros frecuentamos es hasta el paso 2. El primer ejemplo ilustrativo que hemos incluido es la predicción del colesterol a partir del peso corporal y la edad.

**EL SOFTWARE ESTADÍSTICO.** A nivel mundial los mejores programas son: SAS, SPSS, BMDP, en ese orden. Hay infinidad de programas buenos de menor nivel y hay programas de aún mayor nivel como el GLIM, S-PLUS y otros. Estos últimos solo pueden tener interés para estadísticos. Hay programas como el EpiInfo que son para el área de enfermería y tienen valor. El otro día dijimos: el Excel para un individuo, el EpiInfo para un hospital y el SPSS para COCEMI.

**Identificación de las variable.** El primer paso entonces es la definición adecuada de las variables. Este paso es muy simple y muy importante. Creemos que se aprende con la experiencia y no con una explicación. En la siguiente página repasamos algunos ejemplos, especialmente los errores nos enseñan mucho.

Algunos disparate(citos):

-Un autor(cito) dice que va a estudiar "*como afecta el rendimiento el número de hijos*".

-Otro habla de la variable *calles*, la variable *iluminación* y dice que lo *normal* es 3 focos por cuadra

-Otro dice: "*frecuencia diaria=2 veces por semana*".

-La forma de decirlo también debe ser pulida: "*ninguna conectada a la red cloacal*". Creo que es más exacto decir que no hay red cloacal. Es como cuando me liquidaron mal el sueldo y me descuentan por *cobro indebido*, cuando deben decir *pago indebido*.

Claro Uds. me pueden decir: Sus ejemplos son demasiado obvios. Entonces le doy uno menos obvio.

Una investigadora define "*pirosis*" con los siguientes valores: 1=no hubo síntomas en el último año; 2= menos de una vez por mes; 3= una vez por mes; 4=una vez por semana; 5=varias veces por semana y 6=diariamente. Mas adelante la investigadora se pregunta: *¿Existe relación entre "pirosis" y "frecuencia de los síntomas"?* ;Si son la misma cosa!

**Ejemplo de un encuadre para definir variables**

Definición conceptual de variable	Definición operacional, es decir, indicador	Escala de medición
Edad	Edad el día del último cumpleaños	Continua, por mes
Nivel de hemoglobina	Concentración de hemoglobina en la sangre capilar, medida por un hemoglobinómetro	Continua, por ejemplo, gramos por cada 100 ml, redondeados al gramo más próximo.
Estado de nutrición	Peso comparado con la edad mediante una curva de crecimiento normal.	Ordinal, por ejemplo, bien nutridos $\geq 80\%$ de la norma; moderadamente desnutridos = de 60%-80% de la norma; seriamente desnutridos $\leq 60\%$ de la norma.
Satisfacción del paciente	Respuesta a una pregunta concreta dirigida a los pacientes.	Ordinal, por ejemplo, 1. muy satisfecho 2. satisfecho en general 3. algo descontento 4. muy descontento 5. lo ignoro 6. ninguna respuesta
Cobertura de inmunización	Porcentaje de niños inmunizados en un determinado grupo de edades	Continua, por ejemplo, porcentajes o número ordinal. alta $\geq 80\%$ media = 60 - 80% baja $< 60\%$
Fuente principal de carbohidratos en la dieta	Tipo principal de alimento primario	Nominal, por ejemplo, maíz, mijo, arroz, mandioca.

### VARIABLES E INDICADORES. Puesta a punto de un lenguaje común.

**1. Escalas.** Otra clasificación que usamos le llamamos escalas:

- 1) Nominal.
- 2) Ordinal.
- 3) De intervalos.
- 4) Racional.

No existe una equivalencia total entre esta clasificación y la otra pero algunos autores dicen que las variables cualitativas se expresan en escala nominal y las cuantitativas en escalas de intervalos o racionales. Otros ven la situación más compleja.

**2. Variables dependientes e independientes.** Acá la conversación se pone así: en ciertas condiciones, por ejemplo cuando usamos modelos estadísticos, tenemos variables que se pueden suponer que varían en función de otras, que dependen de otras. Por ejemplo la estatura de los niños depende de la edad. Entonces las llamamos *dependientes* e *independientes*. En otros casos no es clara la situación y podemos dudar en como clasificarlas. Si medimos la estatura y el peso de los niños, no es claro cual debe ser considerada la variable dependiente y cual independiente. De tener una situación en la cual una variable depende de otra pasamos fácil a una situación en la que una variable *causa* a la otra, pero ahí nos metemos en otras complicaciones. El artículo de Austin Bradford Hill enfoca este tema con maestría. Por otro lado tenemos la palabra *factores*, que generalmente se reserva para las variables independientes. Algunos autores solo usan esa expresión para las variables independientes cualitativas, pero esto no es general.

**3. Parámetros y variables.** Entre enfermeras y médicos no es tan frecuente como entre veterinarios confundir parámetros con variables. Los parámetros son fijos y las variables varían (que novedad). Si decimos los parámetros medidos son: el peso, la altura de los animales o personas, nos quedamos sin palabras para describir a la media o la correlación entre esas variables.

**Descripción univariada de cada variable.** Luego de definida las variables que queremos estudiar (que es como decir luego que sabemos lo que queremos hacer) el primer paso entonces es la descripción de cada variable. Los datos que varían los llamamos *variables*, a las constantes les llamamos *parámetros*. Polit y otros autores no hablan así de variables y datos. Escuchenlos a ellos y formen su propia visión. Las variables pueden ser *cualitativas o cuantitativas*, estas a su vez pueden ser *continuas o discretas* (discontinuas).

***Variables cualitativas.*** Las variables cualitativas se cuentan, el procedimiento se llama análisis de conteo y como resultado obtenemos tablas de frecuencia.

***Variables cuantitativas.*** Las variables cuantitativas se miden. Las medidas más usadas son de posición o localización o tendencia central (media, mediana, moda, etc.) de dispersión (varianza, desviación estándar o típica, etc.)

**Asociación entre variables (cruce de variables).** El siguiente paso del análisis es el llamado (especialmente por los sociólogos) cruce de variables donde estudiamos la posible asociación entre variables. Según que variables estemos intentando asociar tenemos diferentes situaciones

***Cuali x cuali.*** El cruce de variables cuali da origen en general a tablas o cuadros de contingencia. A partir de esas tablas se elaboran muchos otros cálculos como los riesgos relativos de Epidemiología.

***Cuanti x cuanti.*** Para estudiar la asociación entre variables cuantitativas las herramientas estadísticas son la correlación y la regresión.

***Cuali x cuanti.*** Para analizar la relación entre variables de las cuales una es cuali y la otra cuanti tenemos básicamente que ser criteriosos. Probablemente lo más frecuente sea la comparación de medias.

Luego de este paso tenemos diferentes situaciones. Una es la modelación más avanzada del punto de vista estadístico, otra es la interpretación criteriosa de los resultados.

**Interpretación de las asociaciones entre variables.** No hay recetas sobre como se interpretan resultados, pero hay herramientas. Las herramientas estadísticas (por ejemplo, como distinguir entre asociaciones reales (significativas) y asociaciones casuales) serán tratadas en Bioestadística II. Otros aspectos como la interpretación biológica de la asociación y la causalidad deben ser tratados de otro modo, por ejemplo recomendamos los criterios de Bradford Hill sobre causalidad.

**Variable cuantitativa.** Tomaremos como ejemplo el peso de los niños al nacer en Salto.

**Univariate Procedure**  
**Variable=PESO**

Momentos			
N	1670	Sum Wgts	1670
Mean	3160.608	Sum	5278216
Std Dev	638.4719	Variance	407646.4
Skewness	-1.05707	Kurtosis	2.636382
USS	1.736E10	CSS	6.8036E8
CV	20.20092	Std Mean	15.62369
T:Mean=0	202.2959	Pr> T	0.0001
Num ^= 0	1670	Num > 0	1670
M(Sign)	835	Pr>= M	0.0001
Sgn Rank	697642.5	Pr>= S	0.0001
Cuantiles(Def=5)			
100% Max	5200	99%	4500
75% Q3	3600	95%	4000
50% Med	3200	90%	3850
25% Q1	2900	10%	2435
0% Min	300	5%	2000
		1%	840
Range	4900		
Q3-Q1	700		
Mode	3100		
Extremos			
	Lowest	Obs	Highest
	300(	655)	4600(
	400(	909)	4800(
	400(	817)	5000(
	600(	1328)	5100(
	680(	348)	5200(
			1186)
			1625)
			1684)
			1021)
			1697)
	Missing Value		
	Count	52	
	% Count/Nobs	3.02	

Analysis Variable : PESO

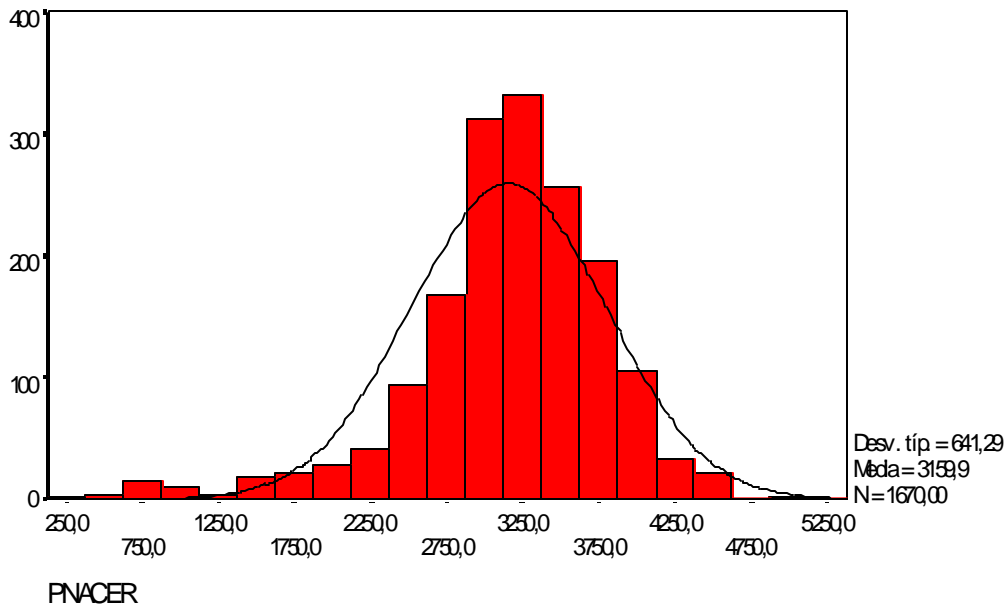
N	Mean	Std Dev	Minimum	Maximum
			GESTAS=.	
59	3149.83	581.0112811	1300.00	4200.00
			GESTAS=1	
400	3064.70	564.4922841	800.0000000	4600.00
			GESTAS=2	
971	3174.20	661.3340049	300.0000000	5100.00
			GESTAS=3	
240	3268.13	655.7082990	600.0000000	5200.00

```
----- SEXO=' ' GESTAS=, -----
N      Mean    Std Dev   Minimum  Maximum
-----
3      2433.33  981.4954576  1300.00  3000.00
```

```

-----SEXO='' GESTAS=1 -----
1  2750.00      .  2750.00  2750.00
-----SEXO='' GESTAS=2 -----
10  2885.00 290.6410539  2150.00  3100.00
-----SEXO='' GESTAS=3 -----
3  3000.00      0  3000.00  3000.00
-----SEXO=F GESTAS=1 -----
190  3033.37 520.4223164  1400.00  4600.00
-----SEXO=F GESTAS=2 -----
526  3125.86 609.2652328  300.0000000  4500.00
-----SEXO=F GESTAS=3 -----
114  3175.53 706.1663823  600.0000000  5000.00
-----SEXO=M GESTAS=1 -----
209  3094.69 602.3982710  800.0000000  4300.00
-----SEXO=M GESTAS=2 -----
435  3239.30 719.5269240  400.0000000  5100.00
-----SEXO=M GESTAS=3 -----
123  3360.49 601.6818691  1400.00  5200.00

```



### 3.2. ESTUDIO DE CASOS.

Los casos que tenemos presentes para estudiar son los siguientes:

- 1) Investigación sobre el **Cáncer de Mama en Salto**. Esta investigación fue hecha un tiempo atrás y ya no recuerdo bien, pero la hemos usado como ejemplo de algunos procesamientos.
- 2) A la autora se le proporcionó un banco de datos con información sobre internos en un **Hospital Pereyra del Campo**. Los datos tenían los siguientes problemas: algunas variables estaban repetidas y contradictorias como el país de origen de la persona. Otras variables como la ocupación estaban codificadas con errores: Peón, jornalero etc.
- 3) **Centro de Salud de Salto**. El Centro de Salud de Salto tiene en un cuaderno, amarillo y viejo pero bien guardado, los datos de los nacimientos. Cada vez que se necesita una cosa hay que hacer un rastreo terrible. Falta velocidad de recuperación de la información. Con ayuda de los estudiantes estamos entrando los datos y proponiendo hipótesis de estudio.
- 4) **Incidencia de Reflujo Gastro Esofágico**. Entre otros objetivos, la autora desea estudiar diferentes factores de riesgo, entonces se plantea uno de esos, el mate por ejemplo y cuenta si las personas que toman mate tienen mayor porcentaje de incidencia de RGE. Hace pruebas de  $\chi^2$  y examina los valores p. El grado de "representatividad" de la muestra es aceptable. El tamaño de la muestra fue calculado con precisión, pero para una variable. Cuando intentamos hacer análisis más complejos aparece como que el tamaño de la muestra no es suficiente.

Del punto de vista del diseño tenemos:

Investigación	Tipo de investigación	
Cáncer de Mama	Banco de datos incompleto	Entrevista
Hospital Pereyra del Campo	Banco de datos	
Centro de Salud de Salto	Banco de datos	
Reflujo Gastro Esofágico	Muestreo	

**ALGUNOS PROBLEMAS IMPORTANTES.** Los problemas que hemos encontrado son de dos tipos: básicos y complicados.

Problema	Solución
Los enfermeros no saben entrar los datos.	Aprendan
Hay errores en la entrada de datos	Revisemos, van a volver
Codificaciones inadecuadas	Recodifiquemos juntos
Hay dificultades en la interpretación	Consulten, discutamos.

#### 4.2.1. ANALISIS DE DATOS DEL HOSPITAL PIÑEYRO DEL CAMPO

VARIABLES. Las variables que se proporcionaron fueron:

F_ING	Fecha de ingreso
SEXO	1=hombre 2=mujer
EDAD	Edad al ingresar
P_NAC	País de nacimiento
C_NAC	Ciudad de nacimiento
PROCEDE	Procedencia
E_CIVIL	Estado civil 1=soltero 2=casado 3=divorciado 4=viudo
PROFESIO	Ocupación al ingresar son muchas
F_EGR	Fecha de egreso
T_EG	Tipo de egreso 1=fallecimiento 2-alta
DIAGN	Diagnostico al egresar 1=enf infecciosas y parasitarias 2=Enf respiratorias 3=Enf apar. circulatorio 4= cáncer 5=accidentes 6=otros
D	Diagnostico recodificada
N_EDUC	Nivel de educación 0=analfabeto 1=sabe leer y/o escribir
INGRE	Ingresos que percibe 0=no percibe 1=pension 9=otros
ASCEN	Ascendencia 0=no 1=si
DESC	Descendencia 0=no 1=si
FICHA	Numero de ficha
RESIDENC	Años de residencia en el país
T_A	Tiempo de internación en años y días
T_D	Tiempo de internación en días
T_T	Tiempo de internación en años
T_R	
P_NAC	País de nacimiento
NACION	Nacionalidad
NACION1	Nacionalidad por región
COD_DIAG	Diagnostico recodificado por Didier agrega códigos 7=ets 8=neurológicos 9=mentales 10=reumatismo 11=resp no infeccioso

1. Positivo: la investigadora ya trajo los datos en Excel.
2. Le asignó a cada variable un nombre en código (nombre corto que identifica a la variable).
3. Negativo: algunas variables estaban redundantes, repetidas en diferente forma. Por ejemplo: PAIS DE NACIMIENTO, CIUDAD DE NACIMIENTO, PROCEDENCIA y NACIONALIDAD. Para comenzar nos desinteresamos de la ciudad donde nació. Luego vimos se vio que había solo un paciente de Suiza y de otros países, entonces es mejor crear una categoría "otros" o "resto de Europa" y dejar en menos clases las variables. Con decisiones de ese tipo se simplifica la base de datos, aunque se pueden perder algunos datos. Al cruzar las variables las tablas se vuelven mucho mas manejables.
4. Variables de fecha, son muchas veces difíciles (o al menos peculiares) de trabajar.
5. Codificación de las variables. Sexo y estado civil estaban codificados con un número, esto no es malo, pero era una exigencia de los computadores de hace unos años, hoy en día es mejor codificar *Mujer* como *M* y *Hombre* como *H* que usar 1 o 0 que no se sabe cual es cual. Lo mismo sucede con el estado civil, una letra (la inicial) es mejor representante que un número. Tipo de egreso, nivel de alfabetización, si recibe o no ingresos, eran variables con el mismo problema. La recodificación del lugar de origen (la variable llamada PROCED) es un ejemplo. La variable es evidentemente una codificación del lugar de origen, pero no nos dijeron que es cada categoría y no lo sabemos. No sirve para nada. En cambio recodificar como se hizo en la otra variable es autoexplicativo.
6. La recodificación de las enfermedades. Al hacer el análisis vimos que había una enorme cantidad de enfermedades. La codificación de las enfermedades en categorías no se había completado, entonces lo tuvimos que hacer nosotros. Esta es una tarea que la tiene que hacer una enfermera o un médico, no la puede hacer el estadístico (por que no sé lo suficiente de eso). Visto de otro punto de vista: es responsabilidad de Uds. En los bancos de datos que andan en la vuelta no se tiene la información al

día (actualizada) ni completa, ahí hay un rubro en el que Uds. tienen que actuar. En el presente caso, la investigadora agregó códigos que no existían para reflejar una problemática sanitaria detallada.

7. Un problema semejante fue la profesión, había muchos oficios y era imposible manejarlos así. La solución puede venir por el lado de codificar los oficios como lo hace el INE - Instituto Nacional de Estadística. Ellos tienen una lista de profesiones.
8. Tiempo de internación. El tiempo de internación se tomó solamente en años, no vale la pena mayor detalle.
9. Número de ficha, es útil para identificar al paciente, pero en una investigación no sirve para nada, además el problema de la confidencialidad de la información debe estar siempre presente.

Luego de seleccionadas las variables que quedaron para analizar (borramos las otras) fueron:

F_ING	Fecha de ingreso
SEXO	1=hombre 2=mujer
PROCEDE	Procedencia
E_CIVIL	Estado civil 1=soltero 2=casado 3=divorciado 4=viudo
PROFESIO	Ocupación al ingresar son muchas
F_EGR	Fecha de egreso
T_EG	Tipo de egreso 1=fallecimiento 2=alta
DIAGN	Diagnostico al egresar 1=enf infecciosas y parasitarias 2=Enf respiratorias 3=Enf apar. circulatorio 4= cáncer 5=accidentes 6=otros
D	Diagnostico recodificada
N_EDUC	Nivel de educación 0=analfabeto 1=sabe leer y/o escribir
INGRE	Ingresos que percibe 0=no percibe 1=pension 9=otros
ASCEN	Ascendencia 0=no 1=si
DESC	Descendencia 0=no 1=si
RESIDENC	Años de residencia en el país
T_T	Tiempo de internación en años
P_NAC	País de nacimiento
NACION1	Nacionalidad por región
COD_DIAG	Diagnostico recodificado por Didier agrega códigos 7=ets 8=neurológicos 9=mentales 10=reumatismo 11=resp no infeccioso

## Resultados.

En mes de internación, hay una distribución razonable.

En año vemos que los datos abarcan un período muy extendido. Vamos a agruparlos en períodos, veremos como. Lo mismo pasa con edad, hay demasiados valores de la variable.

Sexo. Llama la atención que el 63% son hombres.

Estado civil Un solo divorciado. En esa época no existía el divorcio!

Profesión. Ya hablamos de eso, son demasiadas categorías. Noten el problema de las mayúsculas, "Jornalero" y "jornalero" son dos valores diferentes para el computador, por tanto dos profesiones diferentes!

Cuando vimos causa de egresos, vimos que había dos: fallecimiento y alta. Los fallecimientos eran el 99,7 %! Era un hospital terminal.

En años de internación vemos un fenómeno de interés estadístico: la sobrevivencia sigue una función exponencial negativa que es típica de los procesos de sobrevivencia.

PROFESIO	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Agricult	15	1.0	15	1.0
Ajusmeca	1	0.1	16	1.1
Albañil	3	0.2	19	1.3
Artista	1	0.1	20	1.3
CARBONER	1	0.1	21	1.4
CARPINTE	1	0.1	22	1.5
CARRERO	1	0.1	23	1.5
Carboner	1	0.1	24	1.6
Carpinte	10	0.7	34	2.2
Cigarrer	1	0.1	35	2.3
Cochero	1	0.1	36	2.4
Cocinera	6	0.4	42	2.8
Cocinero	4	0.3	46	3.0
Jardiner	3	0.2	67	4.4
Jornaler	652	43.0	719	47.5
LABRADOR	1	0.1	720	47.5
Laboores	1	0.1	721	47.6
Labores	399	26.3	1120	73.9
Labrador	3	0.2	1123	74.1
Lavander	9	0.6	1132	74.7
Limpiado	1	0.1	1133	74.8
Maestro	1	0.1	1134	74.9
Marinero	2	0.1	1136	75.0
Marino	1	0.1	1137	75.0
Mecanico	1	0.1	1138	75.1
Mendiga	1	0.1	1139	75.2
Mercachi	1	0.1	1140	75.2
Militar	1	0.1	1141	75.3
MozoHote	1	0.1	1142	75.4
Mucama	1	0.1	1143	75.4
Mucamo	1	0.1	1144	75.5
NOTRABAJ	1	0.1	1145	75.6
Ninguna	1	0.1	1146	75.6
Ninguno	1	0.1	1147	75.7
No	1	0.1	1148	75.8
Notiene	3	0.2	1151	76.0
jardiner	3	0.2	1332	87.9
jorneler	1	0.1	1333	88.0
jornlero	1	0.1	1334	88.1
labander	1	0.1	1335	88.1
labrador	10	0.7	1345	88.8

lavander	17	1.1	1362	89.9
lechero	1	0.1	1363	90.0
lotero	4	0.3	1367	90.2
lustmueb	1	0.1	1368	90.3
maestro	2	0.1	1370	90.4
marinero	13	0.9	1383	91.3
marino	3	0.2	1386	91.5
mecanico	1	0.1	1387	91.6
mendiga	1	0.1	1388	91.6
mercachi	1	0.1	1389	91.7
militar	1	0.1	1390	91.7
modista	1	0.1	1391	91.8
mucama	1	0.1	1392	91.9
mueblera	1	0.1	1393	91.9
no	3	0.2	1396	92.1
notiene	13	0.9	1409	93.0
panadero	4	0.3	1413	93.3
partera	1	0.1	1414	93.3
peluquer	7	0.5	1421	93.8
peon	3	0.2	1424	94.0
peoncoci	1	0.1	1425	94.1
perfumis	1	0.1	1426	94.1
pescador	2	0.1	1428	94.3
picapedr	2	0.1	1430	94.4
pintor	7	0.5	1437	94.9
planchad	2	0.1	1439	95.0
portero	1	0.1	1440	95.0
puestero	1	0.1	1441	95.1
quimico	1	0.1	1442	95.2
quintera	1	0.1	1443	95.2
quintero	18	1.2	1461	96.4
repartid	1	0.1	1462	96.5
sastre	1	0.1	1463	96.6
servient	1	0.1	1464	96.6
sinprof	1	0.1	1465	96.7
sinprofe	1	0.1	1466	96.8
sirvient	26	1.7	1492	98.5
soldado	2	0.1	1494	98.6
sprofesi	1	0.1	1495	98.7
talabart	1	0.1	1496	98.7
telegraf	1	0.1	1497	98.8
tropero	2	0.1	1499	98.9
vendedor	5	0.3	1504	99.3
zapatero	11	0.7	1515	100.0

Frequency Missing = 242

NACION	Frequency	Percent	Frequency	Percent
Argentina	42	2.4	42	2.4
Brasil	45	2.6	87	5.1
Espana	363	21.2	450	26.2
Francia	93	5.4	543	31.6
Italia	433	25.2	976	56.9
Otros	7	0.4	983	57.3
Portugal	10	0.6	993	57.9
RestoAL	12	0.7	1005	58.6
RestoEur	35	2.0	1040	60.6
Uruguay	676	39.4	1716	100.0

Frequency Missing = 41

Código y Nacionalidad son dos variables que merecen comentarios especiales.

CRUCE DE VARIABLES. Cruzamos sexo con estado civil, y sexo con el código de enfermedades completado por Didier.

TABLE OF SEX BY E\_CIVI

SEX	E_CIVI				
Frequency	C	D	S	V	Total
Percent					
Row Pct					
Col Pct					
F	63	0	253	309	625
	3.72	0.00	14.94	18.24	36.89
	10.08	0.00	40.48	49.44	
	28.38	0.00	27.77	55.18	
H	159	1	658	251	1069
	9.39	0.06	38.84	14.82	63.11
	14.87	0.09	61.55	23.48	
	71.62	100.00	72.23	44.82	
Total	222	1	911	560	1694
	13.11	0.06	53.78	33.06	100.00

Frequency Missing = 63

Otro de los problemas que encontramos es cuando queríamos cruzar variables con muchas categorías que los cuadros quedaban muy grandes.

#### 4.2.2.PREVALENCIA DE SINTOMAS DE REFLUJO GASTRO ESOFÁGICO EN UN BARRIO DE MONTEVIDEO

Procedimiento
La investigadora me contactó. Nos vimos el 30/12/99 a la hora 11. Me introdujo al problema. Tenía una serie de interrogantes. Le explique lo que era un IdeC PdeH y pvalue Asociación entre variables cualitativas. $\chi^2$ p<0,05 También le expliqué las tres etapas usuales del análisis que decimos: uni, bi y modelación. Quedamos en: i) Traer proyecto escrito para leer ii) datos en diskette iii) los cruces que le interesaba decididos.
Todo esto en una hora.

DEFINICIÓN DE LAS VARIABLES.
El reflujo gastro esofágico (RGE) es el mas común de los trastornos esofágicos. Hipotonía primaria del esfínter esofágico inferior, relajaciones transitorias de dicho esfínter, trastornos del aclaramiento esofágico y retardo en el vaciamiento gástrico se han vinculado a la fisiología del RGE. La verdadera prevalencia del RGE en la comunidad es desconocida ya que muchos pacientes nunca consultaron por sus síntomas y además esta solo podría ser establecida por peachimetría de 24 horas de una muestra seleccionada al azar de la población. En el renglón de arriba dice en que consiste el RGE ahora dice que solo se establece por pH. No todos los pacientes con reflujo tienen esofagitis o pirosis y no todos los que tienen pirosis tienen peachimetrías anormales. Entonces la medición a través del pH no es perfecta? Me llevo dos reuniones entender (obligarla a definir lo que quería): pirosis es RGE Pirosis y regurgitaciones ácidas son consideradas sintomas con alto grado de especificidad para el diagnóstico de enfermedad por RGE. A pesar de esto muchos estudios de prevalencia de RGE han usado el síntoma pirosis como marcador de enfermedad por RGE, y se ha visto que cuando la pirosis es el síntoma predominante, el diagnóstico puede hacerse con una especificidad del 92%.

VARIABLE	VALORES POSIBLES
<b>PREGUNTAS BÁSICAS</b>	
No:	No del cuestionario
1)Sexo:	fem: 1, mas: 2
2)Edad:	1:de18a29, 2:30a39, 3:40a49, 4:50a59, 5:60a69, 6:mayores de70Años
3)Dolencia	1: si, 2: no
3 <sup>a</sup> )enf gasin	1:reflujo, 2:gastritis, 3:hernia hiatal, 0:no corresponde
3b)enf otra	1: asma, 2: coronariopatías, 3:otras, 0: no corresponde
4)medicamentos	1: si, 2: no
4 <sup>a</sup> )aines	1: diaria, 2: irregular, 0: no corresponde
4b)corticoides	1: diaria, 2: irregular, 0: no corresponde
4c)nifedipina	1: diaria, 2: irregular, 0: no corresponde
4d)teofilina	1: diaria, 2: irregular, 0: no corresponde
4e)benzodiacepina	1: diaria, 2: irregular, 0: no corresponde
<b>II. SINTOMAS TÍPICOS.</b>	
5) <b>pirosis</b>	1:no en último año, 2:- lvez/mes, 3:alrededorlvez/mes, 4: alrededorlvez/semana, 5:varias/semana, 6:diariamente
6)intesidad	1:leve, 2:moderado, 3:severo, 4:muy severo, 0:no corresponde
7)momentaparición	1:día, 2:noche, 3:las dos anteriores, 0:no corresponde
8)ingiere/alivio	1: si, 2: no, 0: no corresponde
8 <sup>a</sup> )que ingiere	1: antiacidos, 2: otros
9) <b>regurgitaciones</b>	1:no en ultimo año, 2:-lvez/mes, 3:alrededorlvez/mes, 4: alrededorlvez/semana, 5:varias/semana, 6:diariamente
10)aparic/regurgi	1:dia, 2:noche, 3:las dos anteriores, 0:no corresponde
<b>III. SINTOMAS ATÍPICOS.</b>	
11) <b>globus</b>	1: si, 2: no
12) <b>disfagia</b>	1: si, 2: no
13) odinofagia	1: si, 2: no
14) <b>dolorprecordial</b>	1: si, 2: no
15) <b>bronquitis</b>	1: si, 2: no
16) <b>tos nocturna</b>	1: si, 2: no
17)levantar/respaldo	1: si, 2: no
17a) porque?	1: pirosis, 2: disnea, 3: tos, 0: no corresponde
18) <b>disfonía</b>	1: si, 2: no
<b>IV. FACTORES DE RIESGO</b>	
19a) mate	1: si, 2: no
19b) té	1: si, 2: no
19c) café	1: si, 2: no
20) cigarrillos	1: si, 2: no
20a)cantid cigarrillos	1:menosde5, 2:6a15, 3:mas de 15, 0:no corresponde
21)alcohol/diario	1: si, 2: no
22) Visita al Dr.	1: si, 2: no
22a)frecuencia Dr.	1: no corresponde, 2: 1 a 2 veces, 3: 3 a 5 veces, 4: 6 a 10, 5: m s de 10
22b)motivo Dr.	1: pirosis, 2: disnea, 3: tos, 0: no corresponde
23)dijoenfermedad	1: si, 2: no, 0: no corresponde
23a)gastritis	1: si, 2: no, 0: no corresponde
23b)reflujo	1: si, 2: no, 0: no corresponde
23c)ulcera	1: si, 2: no, 0: no corresponde
23d)hernia hiatal	1: si, 2: no, 0: no corresponde
23e) otras	1: si, 2: no, 0: no corresponde

**Objetivo.** El propósito de este estudio es determinar la prevalencia de pirosis como síntoma de RGE en una población de Montevideo. Como objetivo secundario, evaluar la severidad del mismo, la necesidad de uso de medicación y el requerimiento de asistencia médica. Además se evaluará en que porcentaje de personas se asocia pirosis y regurgitaciones con síntomas atípicos y la vinculación entre el RGE y factores de riesgo.

**Materiales & Métodos.** Las siguientes definiciones se usaron para identificar los síntomas en el cuestionario:

1) Pirois - acidez, ardor o quemazón en el pecho, detrás del esternón.

2) Regurgitaciones - fluido ácido o amargo que viene a la garganta o a la boca.

La severidad de los síntomas se valora a través de la medición de la frecuencia de aparición de los mismos y de la percepción subjetiva del síntoma por el encuestado.

La frecuencia se medirá en la siguiente escala: 1) nada en el año previo 2) menos de una vez por mes, 3) alrededor de una vez al mes, 4) alrededor de una vez a la semana, 5) varias veces a la semana y 6) diariamente

La percepción subjetiva se clasificará en: 1) leve, puede ser ignorado 2) moderado, no puede ser ignorado pero no afecta sus actividades diarias 3) severo, afecta las actividades diarias y 4) muy severo, afecta marcadamente su estilo de vida.

Entre las preguntas que el investigador quiere contestar:

¿Hay diferencias (significativas) por sexo y edad?

¿La severidad del síntoma esta asociada proporcionalmente con la frecuencia del síntoma?

Vamos a ponernos de acuerdo, se definió pirosis a partir de la frecuencia ¿Y ahora me pregunta si está relacionado con la frecuencia?

Lo que quiere es cruzar pirosis con intensidad.

con los siguientes valores:

**COMENTARIOS.** Haremos una crítica (más o menos minuciosa) de los planteos de investigación.

Primera objecion: calificar como población a un barrio, parece prematuro.

\* Se toma una población al azar, querra decir una muestra al azar...

\*Criterios de muestreo Se procuró lograr un márgen de error ("absoluto")  $d=\pm 0,05$ .

\* Algunas expresiones no muy convenientes: "sesgar indebidamente los resultados" y "enfermedades significativas".

La investigación quería "medir los niveles de prevalencia específicos por edad y sexo de pirosis y regurgitaciones experimentados en cualquier momento en el pasado año y por los que reportaron experimentarlos por lo menos semanalmente". Luego se intenta lo mismo para regurgitaciones.

Para medir la prevalencia total de pirosis lo que hay que hace es simplemente contar el porcentaje de gente que presenta pirosis (tables pirosis).

Para medir la prevalencia por edad y sexo hay que hacer lo mismo por edad y sexo, o sea procedemos como sigue:

tables sexo\*(pirosis regurg)

edad\*(pirosis regurg)

notemos que se hace la prevalencia de ambas situaciones por edad y sexo casi a la vez.

¿Hay diferencias significativas por s exo y edad? El  $\chi^2$  contesta eso.

### **FACTORES DE RIESGO**

De los 95 sujetos que reportaron pirosis 24% consumian drogas favorecedoras de reflujo:

Analicemos al reves, de los que consumen drogas la incidencia es mayor que entre los que no? Es sign?

Un 86% son consumidores de yerba mate, un 34% tiene sintomas de reflujo, es diferente entre los que toman y los que no?

Idem para el cafe y el te. EL CC contesta eso

idem para fumadores, consumen alcoholicas

Ojo con los factores de confusión, lo dejamos para despues.

#### 4.2..1. UNA SALIDA DE SAS.

ECIVIL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
C	11	47.8	11	47.8
D	2	8.7	13	56.5
S	2	8.7	15	65.2
U	1	4.3	16	69.6
V	7	30.4	23	100.0

EDADM	Frequency	Percent	Cumulative Frequency	Cumulative Percent
10	1	5.0	1	5.0
11	6	30.0	7	35.0
12	6	30.0	13	65.0
13	4	20.0	17	85.0
14	3	15.0	20	100.0

EDADC	Frequency	Percent	Cumulative Frequency	Cumulative Percent
42	1	4.0	1	4.0
48	2	8.0	3	12.0
49	1	4.0	4	16.0
52	2	8.0	6	24.0
56	2	8.0	8	32.0
58	1	4.0	9	36.0
59	1	4.0	10	40.0
61	1	4.0	11	44.0
65	3	12.0	14	56.0
70	1	4.0	15	60.0
71	1	4.0	16	64.0
72	2	8.0	18	72.0
74	1	4.0	19	76.0
75	1	4.0	20	80.0
76	2	8.0	22	88.0
77	1	4.0	23	92.0
80	1	4.0	24	96.0
82	1	4.0	25	100.0

EDAD	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	4	16.0	4	16.0
2	6	24.0	10	40.0
3	5	20.0	15	60.0
4	9	36.0	24	96.0
5	1	4.0	25	100.0

Este es un ejemplo de una investigación hecha por alumnos de Enfermería en 1997.

Tabla de edades

EDADES	MUJERES	HOMBRES
65 - 69	8	1
70 - 74	15	3
75 - 79	3	10
80 - 84	2	7
85 - más	3	6
Totales	31 53,4%	27 46,6%
Media de edades	71,2	78,3

Media total de edades: 74,7 años