

Análisis de muestras complejas con R. La Encuesta de Calidad de Vida en el Trabajo.

III Jornadas de Usuarios de R

Tania Iglesias, Patricia Díaz, Alexandra González

Unidad de Consultoría Estadística
Servicios Científico-Técnicos
Universidad de Oviedo

17 de noviembre de 2011

- 1 Introducción
- 2 Software disponible
- 3 El paquete survey de R
- 4 Ejemplo: ECVT 2010

- Representar adecuadamente la población en estudio.
- Utilización de ponderaciones, estratos o efectos de diseño.
- Diseño empleado en multitud de encuestas oficiales:
 - Encuesta de Calidad de Vida en el Trabajo (ECVT)
 - Encuesta sobre la participación de la población adulta en las actividades de aprendizaje (EADA)
 - Encuesta sobre el gasto de los hogares en educación (módulo piloto asociado a EPF)

Ejemplo

- Objetivo: estimar los ingresos familiares en una población
- Dos zonas: zona rica y zona pobre
- Muestra de 200 viviendas: 100 en zona rica y 100 en zona pobre
- Distribución población: 2.500 hogares zona pobre y 1.200 zona rica

Uso de ponderaciones

	Frec. muestra	Frec. población	Prob. selección	Ponderaciones
Zona Rica	100	1.200	100/1.200	1.200/100=12
Zona Pobre	100	2.500	100/2.500	2.500/100=25

Ingreso medio en cada zona

Ingreso familiar medio en zonas pobres: 12.000 €

Ingreso familiar medio en zonas ricas: 25.000 €

Estimación ingreso medio

Media sin ponderar:

$$(12000 * 100 + 25000 * 100) / (100 + 100) = 18.500 \text{ €}$$

Media ponderada:

$$(12000 * 100 * 25 + 25000 * 100 * 12) / (100 * 25 + 100 * 12) = 16.216 \text{ €}$$

- Stata
- SPSS
- SAS
- Epi Info
- Sudaan

Nuestra elección



El paquete **survey** de R

- Definición diseño: upe, estratos, ponderaciones, deff,...
- Técnicas post-estratificación, calibración, muestro multietápico
- Descriptivos: medias, totales, ratios,...
- Tablas de contingencia
- Modelos de regresión lineal y logística
- Análisis de supervivencia
- Análisis de componentes principales

ECVT

- Objetivos:
 - Investigar la calidad de vida que tiene el ocupado en su trabajo
- Periodicidad anual
- Elaborada por el Ministerio de Trabajo e Inmigración
- Resultados:
 - Información acerca de la situación laboral del ocupado y de su entorno familiar
 - Características del puesto de trabajo, movilidad laboral, satisfacción en el trabajo, conciliación de la vida laboral y familiar..

Ámbito encuesta

- Geográfico: territorio nacional, con excepción de Ceuta y Melilla
- Poblacional: población ocupada mayor de 16 años que reside en viviendas familiares

Cuestionario

- Datos sociodemográficos
- Situación laboral
- Calidad de vida en el trabajo:
 - Satisfacción en el trabajo
 - Entorno laboral
 - Tiempo de trabajo
 - Formación académica y profesional
 - Compensación por el trabajo
 - Asociacionismo

- Encuesta dirigida a ocupados residentes en viviendas familiares
- Relación de viviendas y habitantes: padrón

Tipo de muestreo

- Muestreo trietápico estratificado
- Estratos: cruce Comunidad Autónoma y tramo de tamaño de municipio

- 1 Primera etapa: muestra de secciones censales con probabilidades proporcionales al tamaño de cada sección
- 2 Segunda etapa: muestra de viviendas con probabilidad igual para cada vivienda de la sección
- 3 Tercera etapa: selección de un ocupado residente en cada vivienda

- 1 Primera etapa: muestra de secciones censales con probabilidades proporcionales al tamaño de cada sección
- 2 Segunda etapa: muestra de viviendas con probabilidad igual para cada vivienda de la sección
- 3 Tercera etapa: selección de un ocupado residente en cada vivienda

- 1 Primera etapa: muestra de secciones censales con probabilidades proporcionales al tamaño de cada sección
- 2 Segunda etapa: muestra de viviendas con probabilidad igual para cada vivienda de la sección
- 3 Tercera etapa: selección de un ocupado residente en cada vivienda

Perfil de la Afiliación Sindical

- Variables ECVT:
 - CCAA y tramo municipio
 - Antigüedad
 - Amistades en el trabajo
 - Ingresos mensuales
 - Frecuencia con la que trabaja horas de más
 - Conocimiento del convenio o de las actividades realizadas por el sindicato
 - Participación de los beneficios de la empresa
 - Trabajo a turnos o de noche
 - Nivel riesgo
 - Satisfacción trabajo, empleados, directivos
 - Sector

Etapas

- 1 Definición diseño muestral
- 2 Análisis descriptivo
- 3 Análisis bivariante
- 4 Construcción de un modelo de regresión

Posibilidades *svydesign*

- unidad de muestreo
- estrato
- tipo de ponderación
- fpc
- deff,...

Ejemplo

```
svydesign(id=~1, strata=~estrato,  
weights=~Factorelevacion, fpc=~fpc, data= datos)
```


Funciones

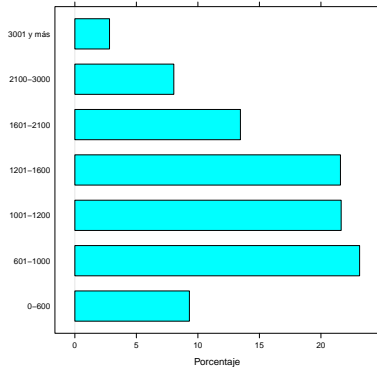
- svytotal
- svymean

Ingresos mensuales

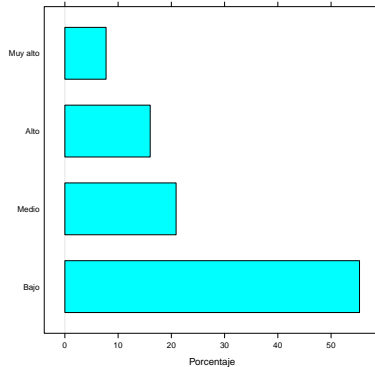
	Frec.	%
0-600	1753844	0.09
601-1000	4359882	0.23
1001-1200	4077647	0.22
1201-1600	4066564	0.22
1601-2100	2534963	0.13
2100-3000	1515205	0.08
3001 y más	530807	0.03

Análisis descriptivo

Ingresos mensuales



Nivel riesgo



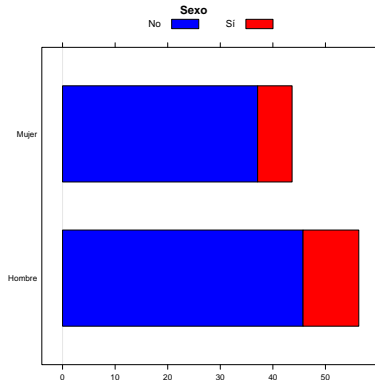
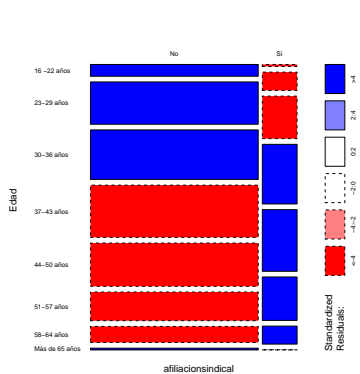
Análisis bivariente: *svytable*

	No	%	Sí	%
16-22 años	695729	3.78	30875	0.17
23-29 años	2343458	12.73	250338	1.36
30-36 años	3046445	16.55	518098	2.81
37-43 años	3420021	18.58	699115	3.80
44-50 años	2862149	15.55	712610	3.87
51-57 años	1890968	10.27	575594	3.13
58-64 años	1041460	5.66	232689	1.26
Más de 65 años	82615	0.45	7460	0.04

Relación entre edad y afiliación sindical

Se ha realizado el análisis para estudiar la relación entre *Edad* y *afiliacionsindical*, obteniéndose que se rechaza la hipótesis de independencia (test de Pearson, p-valor=0.00).

Análisis bivariate

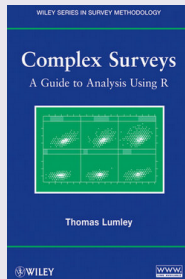


Modelo de regresión: `svyglm` family binomial

	Coeficientes	Odds	Error estándar	Estadístico t	p valor
(Intercept)	-1.63	0.20	0.72	-2.26	0.02
AntigüedadEntre3y10años	-0.32	0.72	0.13	-2.45	0.01
AntigüedadMenosde3años	-0.79	0.45	0.15	-5.35	0.00
EdadIntervalos3036años	1.09	2.98	0.51	2.15	0.03
EdadIntervalos3743años	1.18	3.25	0.51	2.30	0.02
EdadIntervalos4450años	1.38	3.97	0.51	2.69	0.01
EdadIntervalos5157años	1.33	3.79	0.52	2.57	0.01
FrecuenciatrabajadeMAS	-0.27	0.76	0.13	-2.09	0.04
Ingresos.Bajo	-0.65	0.52	0.22	-2.96	0.00
NivelRiesgo.Alto	0.54	1.71	0.11	4.79	0.00
ProvinciaNacimientoAsturias	0.65	1.92	0.27	2.38	0.02
ProvinciaNacimientoBarcelona	-0.31	0.73	0.15	-2.08	0.04
ProvinciaNacimientoCádiz	-0.82	0.44	0.33	-2.46	0.01
ProvinciaNacimientoSta.CruzdeTenerife	0.87	2.40	0.36	2.40	0.02
RealizacionJornadaPartida	-0.25	0.78	0.11	-2.28	0.02
SectoresAgricult.ganad.silvicult.ypesca	-1.39	0.25	0.54	-2.55	0.01
SectoresConstrucción	-0.76	0.47	0.20	-3.75	0.00
SectoresIndustriamanufacturera	-0.53	0.59	0.16	-3.30	0.00
TipoPuestoMandoIntermedio	0.96	2.61	0.47	2.02	0.04
TrabajaTurnos.Sí	0.30	1.35	0.12	2.49	0.01

Paquete *survey*

- Thomas Lumley
- <http://faculty.washington.edu/tlumley/survey>
- Bibliografía: *Complex Surveys: a guide to analysis using R*



¡Gracias por vuestra atención!

