



PUCP

STATA¹⁷

STATISTICS • VISUALIZATION • DATA MANIPULATION • REPORTING

DEPARTAMENTO DE ECONOMÍA
LABORATORIO DE ECONOMETRÍA: STATA
1ECO31

Sesión 3

Reporte de Tablas Estadísticas

Docente: Juan Palomino



Índice

1

Estadística Descriptiva

2

Tablas Estadísticas

3

Factor de Expansión

4

Exportar Tablas

1. Estadísticos Descriptivos

El comando summarize

El comando `summarize` provee estadísticos básicos de las variables numéricas.

```
| summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
pais	0				
dpto	90,006	13.01253	6.729924	1	25
zona	90,006	2.154767	1.003666	1	4
area	90,006	.6511788	.4765999	0	1
conglome	0				
vivienda	0				
hogar	0				
codperso	0				
mujer	90,006	.51611	.4997432	0	1
edad	90,006	41.86001	19.12005	14	98
sq_edad	90,006	2117.833	1780.555	196	9604
jefe	90,006	.3801635	.4854295	0	1
civil	90,006	2.219985	1.362159	1	4
educ	89,960	3.36285	1.505194	1	7
indigena	89,923	.8762052	.3293491	0	1
pea	89,923	.6562726	.4749541	0	1
peao	90,006	.6220696	.4848728	0	1
informal	55,990	.7909627	.4066247	0	1
ingreso	90,006	599.9798	1247.334	0	74092.16
lnwage	45,887	6.506397	1.13821	.6931472	11.21307
fac500a	90,006	276.7847	297.4005	.3713087	2820.728

```
sum ingreso, detail
```

Ingreso Mensual				
	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	90,006
25%	0	0	Sum of Wgt.	90,006
50%	47.41667		Mean	599.9798
		Largest	Std. Dev.	1247.334
75%	786	26155.33		
90%	1746.583	28198.33	Variance	1555842
95%	2559.167	53323	Skewness	8.108412
99%	5322.333	74092.16	Kurtosis	220.5458

2. Tablas Estadísticas

El comando tabulate

El comando “**tabulate**” reporta una tabla de una entrada para una variable en particular.

```
tab pea
```

PEA	Freq.	Percent	Cum.
NO PEA	30,909	34.37	34.37
PEA	59,014	65.63	100.00
Total	89,923	100.00	

Con Factor de Expansión

```
tab pea [iw=fac500a]
```

PEA	Freq.	Percent	Cum.
NO PEA	8,786,642	35.31	35.31
PEA	16094964.5	64.69	100.00
Total	24,881,606	100.00	

El comando tab1

El comando “**tab1**” permite tabular sucesivamente diferentes variables de interés.

tab1 zona area

-> tabulation of zona			
Zona Geográfica	Freq.	Percent	Cum.
Costa	27,728	30.81	30.81
Sierra	32,338	35.93	66.74
Selva	18,222	20.25	86.98
Lima Metropolitana	11,718	13.02	100.00
Total	90,006	100.00	
-> tabulation of area			
Área Geográfica	Freq.	Percent	Cum.
Rural	31,396	34.88	34.88
Urbano	58,610	65.12	100.00
Total	90,006	100.00	

El comando tab2

El comando “**tab2**” permite realizar una tabla de doble entrada de un grupo de variables.

tab2 zona area mujer

-> tabulation of zona by area			
Zona Geográfica	Área Geográfica		Total
	Rural	Urbano	
Costa	4,469	23,259	27,728
Sierra	18,762	13,576	32,338
Selva	8,165	10,057	18,222
Lima Metropolitana	0	11,718	11,718
Total	31,396	58,610	90,006

-> tabulation of zona by mujer			
Zona Geográfica	Mujer		Total
	Hombre	Mujer	
Costa	13,318	14,410	27,728
Sierra	15,552	16,786	32,338
Selva	9,061	9,161	18,222
Lima Metropolitana	5,622	6,096	11,718
Total	43,553	46,453	90,006

-> tabulation of area by mujer			
Área Geográfica	Mujer		Total
	Hombre	Mujer	
Rural	15,646	15,750	31,396
Urbano	27,907	30,703	58,610
Total	43,553	46,453	90,006

El comando table

El comando “**table**” permite producir contenido estadístico descriptivo dentro de una tabla.

```
table pais if pea==1 & ingreso>0 [iw=fac500a], c(mean ingreso)
```

País	mean(ingreso)
Perú	1258.41

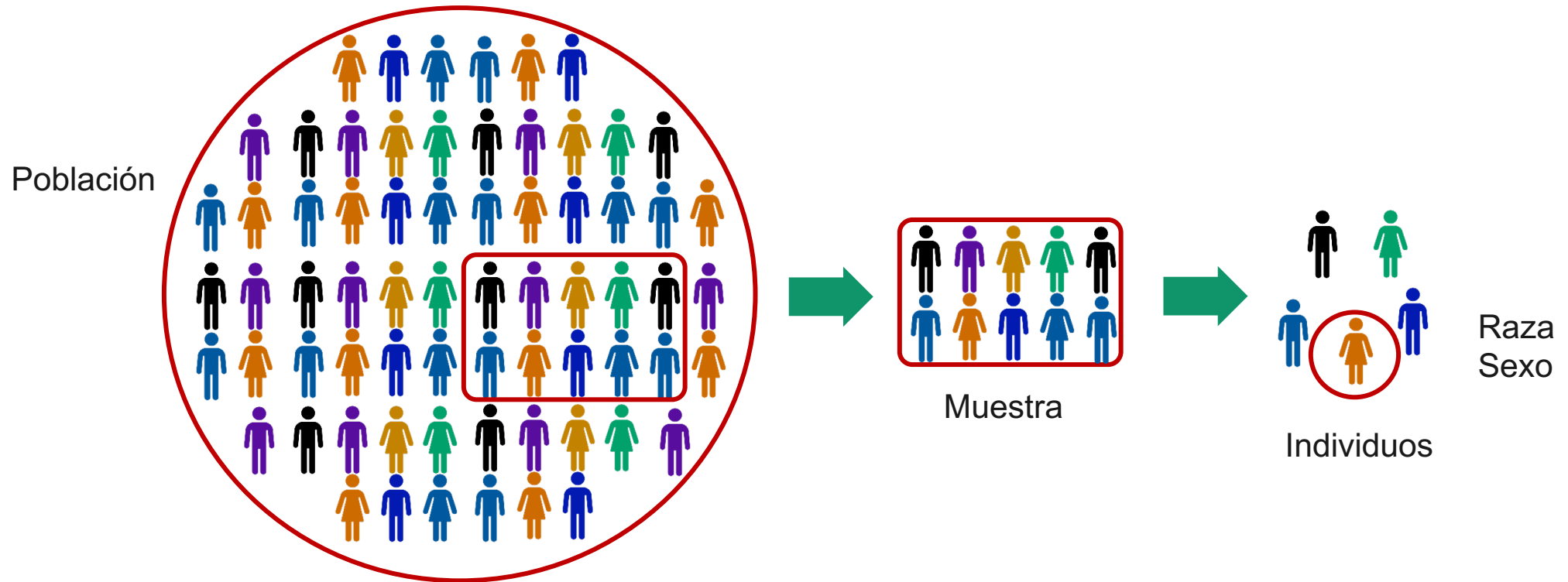
```
table zona if pea==1 & ingreso>0 [iw=fac500a], c(sum pea mean ingreso) format(%9.0fc)
```

Zona Geográfica	sum(pea)	mean(ingreso)
Costa	3,192,024	1,192
Sierra	3,845,742	980
Selva	1,743,125	1,033
Lima Metropolitana	4,060,487	1,671

3. Factor de Expansión

Factor de Expansión

- Cuando se trabaja con muestras de datos que representan una población, esta generalmente fue obtenida con un **diseño muestral** apropiado para lograr la representatividad de la población bajo estudio.
- **Factor de expansión** significa cuanta observación poblacional representa cada observación muestral.



a. Frequency Weight [fweight]

- Expande el tamaño de la encuesta al tamaño de la población.
- Indica el número de observaciones duplicadas.
- Se utiliza en **tablas** para generar frecuencias.
- Se puede usar en distribuciones de frecuencia solo cuando la **variable de peso es discreta** (sin números fraccionarios).
- Aplicar este factor a estimaciones de varianza, p-valor o calcular errores estándar sería un error, dado que infla la base de datos.

tab poverty [fweight = weight]

b. Importance Weight [iweight]

- Indica la "importancia" de la observación en un sentido vago.
- No tiene una definición estadística formal.
- Cualquier comando que admita "**iweights**" definirá exactamente cómo se tratan.
- Está destinado a ser utilizado por programadores que desean producir un cierto cálculo.
- Se puede usar en distribuciones de frecuencia incluso cuando la **variable de peso es continua** (números fraccionarios)

`tab poverty [iweight = weight]`

c. Analytical Weight [aweight]

- Este factor es apropiado de usar cuando existen **datos que representan promedios o tasas**, donde el ponderador contiene el número de observaciones con la que fue construida el promedio.
- Inversamente proporcional a la varianza de una observación
- Se supone que la varianza de la observación j es $\frac{\sigma^2}{w_j}$, donde w_j son los pesos
- Para la mayoría de los comandos de Stata, la escala registrada de aweight es irrelevante
- Stata reescala internamente las frecuencias, por lo que la suma de pesos es igual al tamaño de la muestra

tab x [aweight = weight]

c. Analytical Weight [aweight]

- Las observaciones representan promedios y los pesos son el número de elementos que dieron lugar al promedio

Grupo	x	y	n
1	3.5	26	2
2	5	20	3

- En vez de:

Grupo	x	y
1	3	22
1	4	30
2	8	25
2	2	19
2	5	16

- Por lo general, los datos de encuestas se recopilan a nivel de individuos y hogares (no como promedios).
 - Por lo tanto, usar aweight no es apropiado para la mayoría de los casos

d. Probability Weight [pweight]

- En las encuestas las muestras son extraídas mediante un proceso aleatorio, denotado por el inverso de la probabilidad de que la observación se incluya debido al diseño de muestreo
- STATA utiliza este valor como el número de sujetos en la población que representa cada observación en el compuesto de las proporciones, promedios y parámetros de regresión.
- Indicado para **regresiones estadísticas**.
- Las varianzas, los errores estándar y los intervalos de confianza se estiman con un procedimiento más preciso.
 - Estima errores estándar robustos

regress y x1 x2 [pweight = weight]

Resumen de Stata Weights

Pesos en Distribuciones de Frecuencia		
Unidad de medida de peso	Expandir al tamaño de la población	Mantener el tamaño de la muestra
Discreta	fweight	aweight
Continua	iweight	

PESOS EN REGRESIONES ESTADÍSTICAS que deben mantener el tamaño de la muestra	
Errores estándar robustos	R2 ajustado, SCR, SCT, SCE
pweight	aweight
reg y x, robust	outreg2

4. Exportar Tablas

El comando asdoc

El comando “**asdoc**” permite exportar la tabla de estadística descriptiva a Word. Hay que tener en cuenta lo siguiente:

- No funciona con los globals, ni con comillas.
- Se tiene que especificar la ruta donde se va guardar el archivo con el comando “**cd**”

```
asdoc sum, save(ejemplo.doc)
```

Descriptive Statistics					
Variable	Obs	Mean	Std. Dev.	Min	Max
dpto	90006	13.013	6.73	1	25
zona	90006	2.155	1.004	1	4
area	90006	.651	.477	0	1
mujer	90006	.516	.5	0	1
edad	90006	41.86	19.12	14	98
sq edad	90006	2117.833	1780.555	196	9604
jefe	90006	.38	.485	0	1
civil	90006	2.22	1.362	1	4
educ	89960	3.363	1.505	1	7
indigena	89923	.876	.329	0	1
pea	89923	.656	.475	0	1
peao	90006	.622	.485	0	1
informal	55990	.791	.407	0	1
ingreso	90006	599.98	1247.334	0	74092.164
lnwage	45887	6.506	1.138	.693	11.213
fac500a	90006	276.785	297.401	.371	2820.728

Exportar tabulados a excel

El comando “tab2xl” permite exportar tablas de doble entrada a Excel.

```
tab2xl dpto zona using "$works/tabla.xlsx", col(1) row(1)
```

Departamento	Zona Geográfica				Total
	Costa	Sierra	Selva	Lima Metropolitana	
Amazonas	0	1137	1862	0	2999
Ancash	1563	2307	0	0	3870
Apurímac	0	2342	0	0	2342
Arequipa	1118	3036	0	0	4154
Ayacucho	0	2578	304	0	2882
Cajamarca	123	2660	684	0	3467
Callao	0	0	0	2868	2868
Cusco	0	2834	418	0	3252
Huancavelica	0	2642	0	0	2642
Huánuco	0	2240	966	0	3206
Ica	3847	11	0	0	3858
Junín	0	2711	1052	0	3763
La Libertad	2987	1219	0	0	4206
Lambayeque	4039	131	0	0	4170
Lima	2925	690	0	8850	12465
Loreto	0	0	4331	0	4331
Madre de Dios	0	0	1467	0	1467
Moquegua	1739	540	0	0	2279
Pasco	0	1709	608	0	2317
Piura	4429	416	0	0	4845
Puno	0	2701	46	0	2747
San Martín	0	0	3346	0	3346
Tacna	2847	434	0	0	3281
Tumbes	2111	0	0	0	2111
Ucayali	0	0	3138	0	3138
Total	27728	32338	18222	11718	90006

Exportar tabulados a Word

El comando “tab2doc” permite exportar tablas de una entrada de una variable a Word.

```
putdocx begin  
tab2docx dpto  
putdocx save "$works/tabla.docx"
```

<u>Departamento</u>	Frequency	Percentage (%)	Cum. (%)
Amazonas	2999	3.33	3.33
Ancash	3870	4.30	7.63
Apurímac	2342	2.60	10.23
Arequipa	4154	4.62	14.85
Ayacucho	2882	3.20	18.05
Cajamarca	3467	3.85	21.90
Callao	2868	3.19	25.09
Cusco	3252	3.61	28.70
Huancavelica	2642	2.94	31.64
Huánuco	3206	3.56	35.20
Ica	3858	4.29	39.49
Junín	3763	4.18	43.67
La Libertad	4206	4.67	48.34
Lambayeque	4170	4.63	52.97
Lima	12465	13.85	66.82
Loreto	4331	4.81	71.63
Madre de Dios	1467	1.63	73.26
Moquegua	2279	2.53	75.80
Pasco	2317	2.57	78.37
Piura	4845	5.38	83.75
Puno	2747	3.05	86.81
San Martín	3346	3.72	90.52
Tacna	3281	3.65	94.17
Tumbes	2111	2.35	96.51
Ucayali	3138	3.49	100.00
Total:	90006	100	



PUCP