



**PUCP**

**STATA**<sup>17</sup>

STATISTICS • VISUALIZATION • DATA MANIPULATION • REPORTING

DEPARTAMENTO DE ECONOMÍA  
LABORATORIO DE ECONOMETRÍA: STATA  
1ECO31

# Sesión 11

## Test Paramétricos y Correlaciones

Docente: Juan Palomino



# Índice

1

Test de Proporciones

2

Test de Varianza

3

Test de Medias

4

Correlaciones

# 1. Test de Proporciones

---

# Test de proporciones para dos muestras

- El objetivo de esta prueba es evaluar si la diferencia de dos proporciones (o porcentajes) de dos grupos son estadísticamente iguales a cero.
- En este caso la hipótesis nula se expresa como :

$$H_0: p_1 - p_2 = 0$$

# Test de proporciones para dos muestras

Evaluemos la hipótesis de que la proporción de individuos con ingresos mayores al promedio es idéntica entre individuos ubicados en áreas urbanas y rurales.

```
tab2 area ing_sup, row
```

Área Geográfica	ing_sup		Total
	Inferior	Superior	
Rural	18,251 85.66	3,056 14.34	21,307 100.00
Urbano	14,794 50.52	14,487 49.48	29,281 100.00
Total	33,045 65.32	17,543 34.68	50,588 100.00

# Test de proporciones para dos muestras

¿La proporción de individuos con ingresos mayores al promedio es la misma independientemente si están ubicados en áreas urbanas o rurales?

```
prtest ing_sup, by(area) level(95)
```

Two-sample test of proportions				Rural: Number of obs =	21307
				Urbano: Number of obs =	29281
Group	Mean	Std. Err.	z	P> z	[95% Conf. Interval]
Rural	.143427	.0024012			.1387207 .1481334
Urbano	.4947577	.0029218			.489031 .5004844
diff	-.3513306	.0037819			-.3587431 -.3439182
	under Ho:	.0042858	-81.98	0.000	
diff = prop(Rural) - prop(Urbano)				z = -81.9764	
Ho: diff = 0					
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0	
Pr(Z < z) = 0.0000		Pr( Z  >  z ) = 0.0000		Pr(Z > z) = 1.0000	

## 2. Test de Varianza

---



# Test de varianza para dos muestras

- Se utiliza para decidir si las varianzas de dos o más grupos de observaciones son iguales.
- Evaluamos si el ratio entre las varianzas son iguales a 1.

$$H_0: \frac{\sigma_1}{\sigma_2} = 1$$

# Test de varianza para dos muestras

Evaluaremos la hipótesis que la varianza de los ingresos de los individuos ubicados en áreas urbanas es igual a la varianza de aquellos ubicados en áreas rurales.

```
bys area: sum ingreso
```

area = Rural					
Variable	Obs	Mean	Std. Dev.	Min	Max
ingreso	21,307	470.252	820.8899	0	28198.33
area = Urbano					
Variable	Obs	Mean	Std. Dev.	Min	Max
ingreso	29,281	1365.343	1719.986	0	74092.16

# Test de varianza para dos muestras

¿La dispersión de los ingresos de las áreas urbanas es diferente a la dispersión de los que están en áreas rurales?

```
sdtest ingreso, by(area)
```

Variance ratio test						
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Rural	21,307	470.252	5.623721	820.8899	459.2291	481.2749
Urbano	29,281	1365.343	10.05152	1719.986	1345.642	1385.045
combined	50,588	988.3427	6.581754	1480.353	975.4424	1001.243
ratio = sd(Rural) / sd(Urbano)				f = 0.2278		
Ho: ratio = 1				degrees of freedom = 21306, 29280		
Ha: ratio < 1		Ha: ratio != 1		Ha: ratio > 1		
Pr(F < f) = 0.0000		2*Pr(F < f) = 0.0000		Pr(F > f) = 1.0000		

# 3. Test de Medias

---

# Test de medias para dos muestras

Queremos contrastar si el promedio de dos grupos son iguales.

La hipótesis nula:

$$H_0: \mu_1 = \mu_2$$

Equivale a:

$$H_0: \mu_1 - \mu_2 = 0$$

# Test de medias para dos muestras

Evaluaremos la hipótesis nula que el promedio de ingresos es idéntico entre aquellos ubicados en áreas rurales y urbanas.

```
mean ingreso, over(area) level(90)
```

Mean estimation		Number of obs = 50,588		
		Mean	Std. Err.	[90% Conf. Interval]
c.ingreso@area				
	Rural	470.252	5.623721	461.0016 479.5024
	Urbano	1365.343	10.05152	1348.81 1381.877

# Test de medias para dos muestras

Para nuestro caso:

```
ttest ingreso, by(area) unequal level(95)
```

Two-sample t test with unequal variances						
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Rural	21,307	470.252	5.623721	820.8899	459.2291	481.2749
Urbano	29,281	1365.343	10.05152	1719.986	1345.642	1385.045
combined	50,588	988.3427	6.581754	1480.353	975.4424	1001.243
diff		-895.0915	11.51779		-917.6665	-872.5164
diff = mean(Rural) - mean(Urbano) t = -77.7139						
Ho: diff = 0 Satterthwaite's degrees of freedom = 44489.1						
Ha: diff < 0 Ha: diff != 0 Ha: diff > 0						
Pr(T < t) = 0.0000 Pr( T  >  t ) = 0.0000 Pr(T > t) = 1.0000						

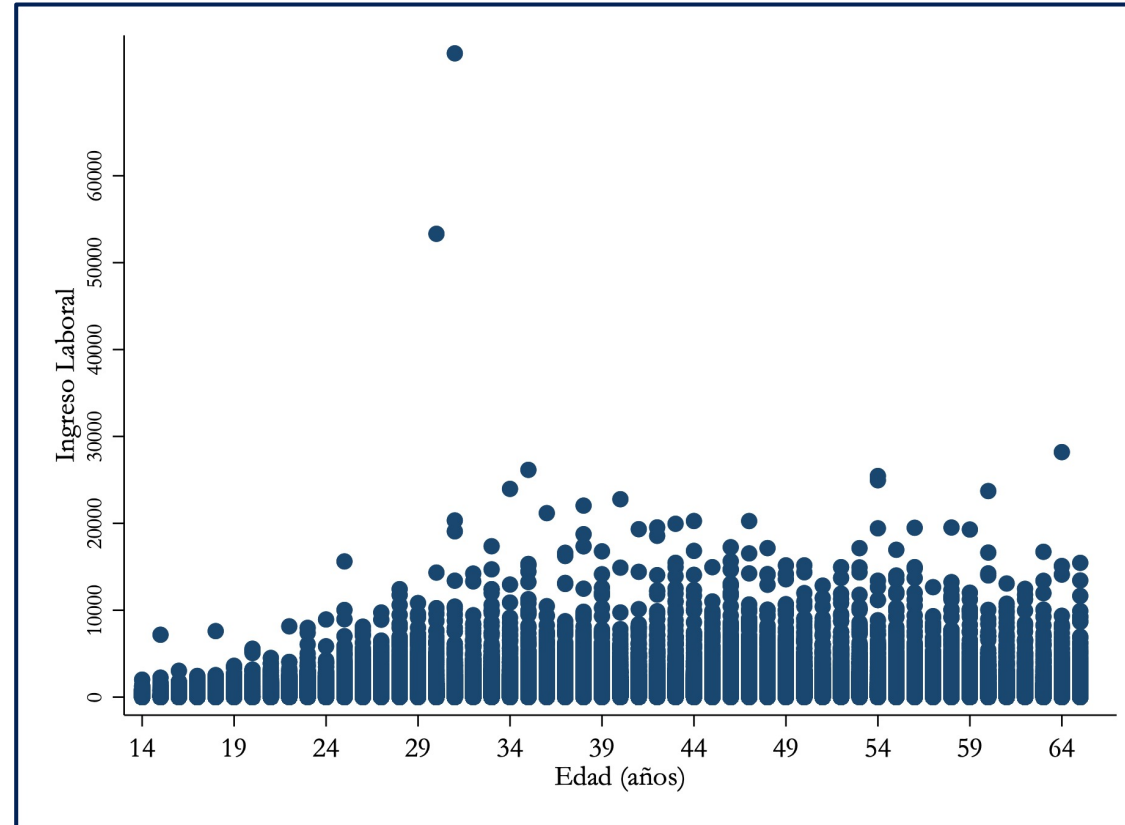
# 4. Correlaciones

---



# Relación entre variables

Gráfico de dispersión: relación entre ingreso laboral y edad:



# Coeficiente de Correlación de Pearson

- Una medida de asociación **lineal** entre dos variables aleatorias continuas.
- Dadas dos variables “x” e “y” con “n” observaciones ( $i = 1, \dots, n$ ), su estimador muestral, el cual es un estimador consistente, es:

$$\hat{\rho} = \frac{\sum_N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_N (x_i - \bar{x})^2} \sqrt{\sum_N (y_i - \bar{y})^2}} = \frac{S_{XY}}{S_X S_Y}$$

Este está acotado en el intervalo  $\hat{\rho} \in [-1, 1]$ .

# Coeficiente de Correlación de Pearson

¿Cómo generamos este coeficiente en Stata?

¿La correlación es estadísticamente diferente de cero?

$$H_0: \rho = 0$$

Stata tiene un comando que estima las correlaciones pero, además, asigna niveles de significancia:

```
. corr ingreso edad  
(obs=50,588)
```

	ingreso	edad
ingreso	1.0000	
edad	0.1326	1.0000

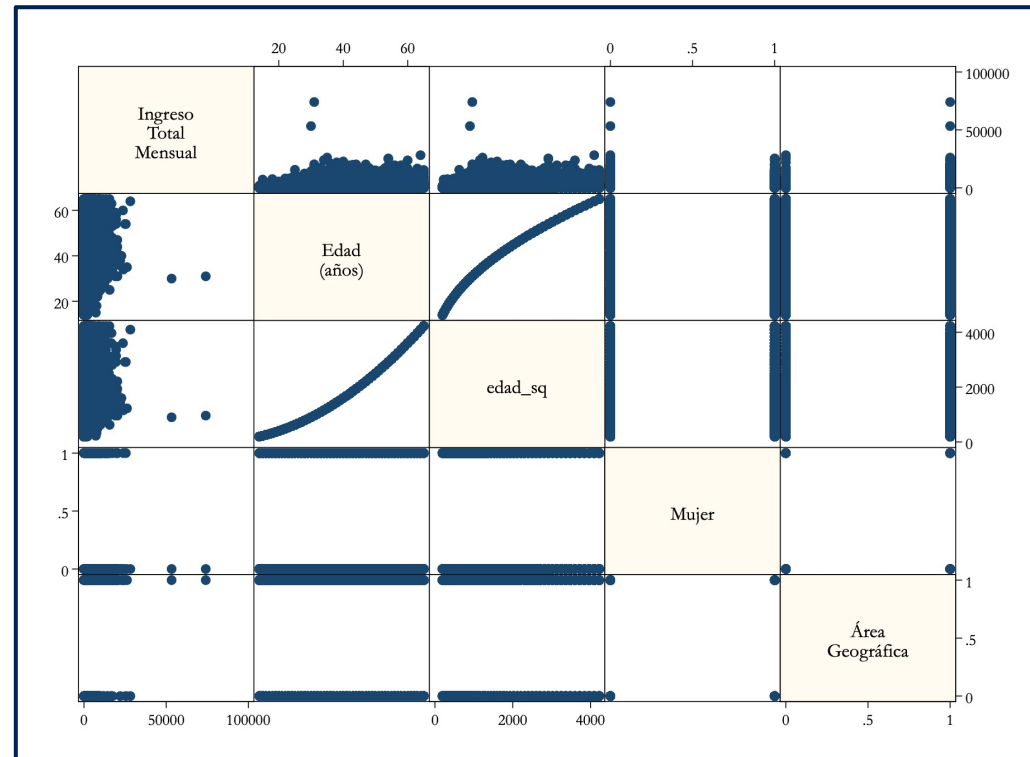
```
. pwcorr ingreso edad, obs sig star(0.01)
```

	ingreso	edad
ingreso	1.0000	
	50588	
edad	0.1326* 1.0000	
	0.0000	
	50588	50588

# Gráfico Matriz de Correlaciones

El gráfico matriz muestra la relación entre todas las variables.

```
graph matrix ingreso edad edad_sq mujer area, ///  
diagonal(,bfcolor(eggshell)) graphregion(color(white))
```





**PUCP**