



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

INFORME

TAREA II

**“ANALISIS UNIVARIADO,
COMPARACION DE
PROMEDIOS Y
VARIANZAS,
CHI CUADRADO,
CORRELACION”**

ECONOMETRIA

PROFESOR: RODRIGO ORTEGA

INTEGRANTE:

BARBARA LIZAMA

18 DE ABRIL DE 2021

PROBLEMA

Se adjunta un set de datos con la muestra de salario y educación de dos grupos raciales en EEUU, generando diversos test de hipótesis en relación a la población.

OBJETIVO

Realizar diversas pruebas de hipótesis con diversos análisis como estadística descriptiva, pruebas de t, pruebas de Chi Cuadrado y análisis de correlación para establecer diversas conclusiones en relación a estas hipótesis planteadas.

1) Realice un análisis univariado de las dos variables de interés considerando el total de la muestra y por cada grupo racial. Incluya un histograma y estadística descriptiva en cada caso. Realice además una prueba de normalidad de cada variable en la muestra completa. A partir de los resultados obtenidos pruebe la hipótesis de que ambos grupos tienen niveles de educación y salario similares. Especifique como realizó dicha prueba y establezca sus conclusiones. En la población general, ¿cuál de las dos variables tiene mayor variación?

El análisis univariado se realizó de manera total de la muestra y por la variable cualitativa raza (afroamericano y caucásico).

POBLACION		
	Salario (dólares/mes)	educación (años)
Media	957,945	13,468
Error típico	13,224	0,072
Mediana	905	12
Moda	1000	12
Desviación estándar	404,361	2,197
Varianza de la muestra	163507,675	4,825
Curtosis	2,718	-0,735
Coefficiente de asimetría	1,201	0,549
Rango	2963	9
Mínimo	115	9
Máximo	3078	18
Suma	895679	12593
Cuenta	935	935
Mayor (1)	3078	18
Menor(1)	115	9
Nivel de confianza(95,0%)	25,952	0,141
Límite Inferior	931,993	13,327
Límite Superior	983,898	13,609
Coefficiente de variación	42,211	16,310

Ilustración 1 Estadística Descriptiva Población

AFROAMERICANO		
	Salario (dólares/mes)	educación (años)
Media	735,842	12,442
Error típico	27,015	0,159
Mediana	683,50	12,00
Moda	577,00	12,00
Desviación estándar	295,931	1,738
Varianza de la muestra	87575,092	3,022
Curtosis	1,295	0,825
Coefficiente de asimetría	1,002	0,768
Rango	1614	9
Mínimo	260	9
Máximo	1874	18
Suma	88301	1493
Cuenta	120	120
Mayor (1)	1874	18
Menor(1)	260	9
Nivel de confianza(95,0%)	53,492	0,314
Límite Inferior	682,350	12,127
Límite Superior	789,333	12,756
Coefficiente de variación	40,217	13,972

Ilustración 2 Estadística Descriptiva Variable Afroamericano

CAUCASICO		
	Salario (dólares/mes)	educación (años)
Media	990,648	13,620
Error típico	14,292	0,078
Mediana	938	13
Moda	1000	12
Desviación estándar	408,003	2,217
Varianza de la muestra	166466,243	4,917
Curtosis	2,711	-0,888
Coefficiente de asimetría	1,195	0,493
Rango	2963	9
Mínimo	115	9
Máximo	3078	18
Suma	807378	11100
Cuenta	815	815
Mayor (1)	3078	18
Menor(1)	115	9
Nivel de confianza(95,0%)	28,053	0,152
Límite Inferior	962,595	13,467
Límite Superior	1018,701	13,772
Coefficiente de variación	41,185	16,280

Ilustración 3 Estadística Descriptiva Variable Caucásico

Clase	Frecuencia	% acumulado
115,00	1	0,11%
213,77	1	0,21%
312,53	6	0,86%
411,30	32	4,28%
510,07	63	11,02%
608,83	80	19,57%
707,60	89	29,09%
806,37	93	39,04%
905,13	103	50,05%
1003,90	118	62,67%
1102,67	71	70,27%
1201,43	70	77,75%
1300,20	55	83,64%
1398,97	37	87,59%
1497,73	31	90,91%
1596,50	21	93,16%
1695,27	17	94,97%
1794,03	14	96,47%
1892,80	4	96,90%
1991,57	9	97,86%
2090,33	2	98,07%
2189,10	8	98,93%
2287,87	0	98,93%
2386,63	2	99,14%
2485,40	2	99,36%
2584,17	2	99,57%
2682,93	1	99,68%
2781,70	1	99,79%
2880,47	0	99,79%
2979,23	0	99,79%
y mayor...	2	100,00%

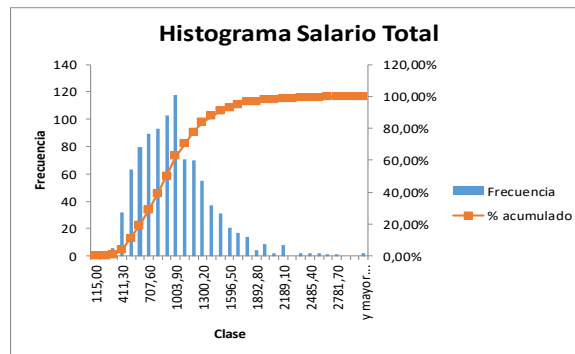


Ilustración 4 Histograma Salario Población

Clase	Frecuencia	% acumulado
9	10	1,07%
9,3	0	1,07%
9,6	0	1,07%
9,9	0	1,07%
10,2	35	4,81%
10,5	0	4,81%
10,8	0	4,81%
11,1	43	9,41%
11,4	0	9,41%
11,7	0	9,41%
12	393	51,44%
12,3	0	51,44%
12,6	0	51,44%
12,9	0	51,44%
13,2	85	60,53%
13,5	0	60,53%
13,8	0	60,53%
14,1	77	68,77%
14,4	0	68,77%
14,7	0	68,77%
15	45	73,58%
15,3	0	73,58%
15,6	0	73,58%
15,9	0	73,58%
16,2	150	89,63%
16,5	0	89,63%
16,8	0	89,63%
17,1	40	93,90%
17,4	0	93,90%
17,7	0	93,90%
y mayor...	57	100,00%

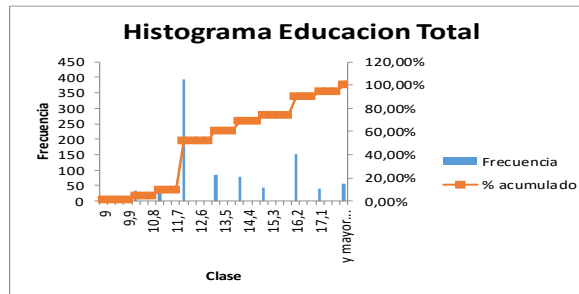


Ilustración 5 Histograma Educación Población

Clase	Frecuencia	% acumulado
260	1	0,83%
421,4	14	12,50%
582,8	33	40,00%
744,2	20	56,67%
905,6	19	72,50%
1067	20	89,17%
1228,4	7	95,00%
1389,8	0	95,00%
1551,2	4	98,33%
1712,6	1	99,17%
y mayor...	1	100,00%

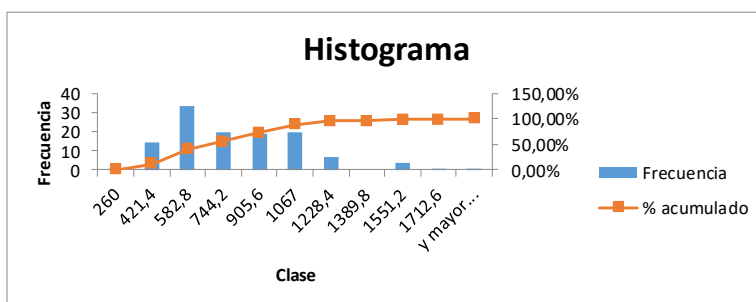


Ilustración 6 Histograma Salario Afroamericano

Clase	Frecuencia	% acumulado
9	4	3,36%
9,9	0	3,36%
10,8	9	10,92%
11,7	9	18,49%
12,6	62	70,59%
13,5	8	77,31%
14,4	11	86,55%
15,3	5	90,76%
16,2	9	98,32%
17,1	1	99,16%
y mayor...	1	100,00%

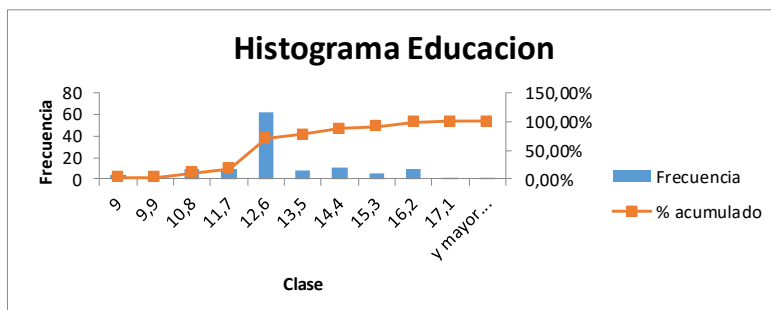


Ilustración 7 Histograma Educación Afroamericano

Clase	Frecuencia	% acumulado
115,000	1	0,12%
220,821	1	0,25%
326,643	5	0,86%
432,464	23	3,68%
538,286	57	10,67%
644,107	71	19,39%
749,929	69	27,85%
855,750	106	40,86%
961,571	92	52,15%
1067,393	100	64,42%
1173,214	71	73,13%
1279,036	65	81,10%
1384,857	40	86,01%
1490,679	29	89,57%
1596,500	22	92,27%
1702,321	19	94,60%
1808,143	12	96,07%
1913,964	5	96,69%
2019,786	8	97,67%
2125,607	2	97,91%
2231,429	7	98,77%
2337,250	2	99,02%
2443,071	2	99,26%
2548,893	2	99,51%
2654,714	0	99,51%
2760,536	1	99,63%
2866,357	1	99,75%
2972,179	0	99,75%
y mayor...	2	100,00%

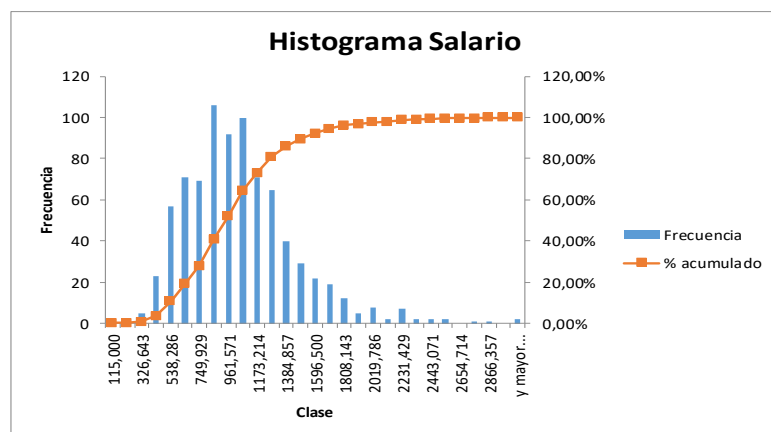


Ilustración 8 Histograma Salario Caucásico

Clase	Frecuencia	% acumulado
9,000	6	0,74%
9,321	0	0,74%
9,643	0	0,74%
9,964	0	0,74%
10,286	26	3,93%
10,607	0	3,93%
10,929	0	3,93%
11,250	34	8,10%
11,571	0	8,10%
11,893	0	8,10%
12,214	330	48,59%
12,536	0	48,59%
12,857	0	48,59%
13,179	77	58,04%
13,500	0	58,04%
13,821	0	58,04%
14,143	66	66,13%
14,464	0	66,13%
14,786	0	66,13%
15,107	40	71,04%
15,429	0	71,04%
15,750	0	71,04%
16,071	141	88,34%
16,393	0	88,34%
16,714	0	88,34%
17,036	39	93,13%
17,357	0	93,13%
17,679	0	93,13%
y mayor...	56	100,00%

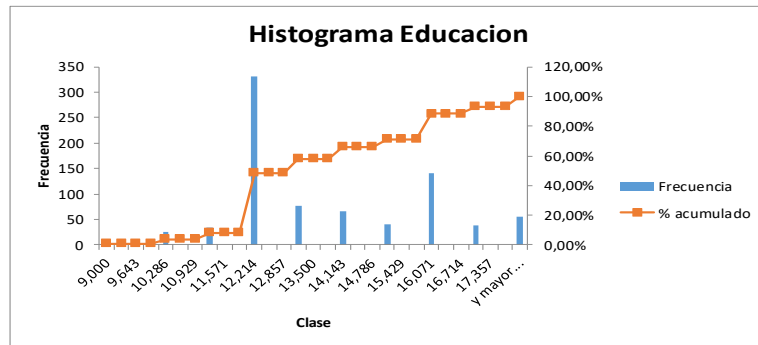


Ilustración 9 Histograma Educación Caucásico

Prueba de Normalidad (Se ocupa el Test de Shapiro-Wilk y D' Agostino- Pearson)

Shapiro-Wilk Test

	Salario (dólares/mes)	educación (años)
W-stat	0,938	0,869
p-value	3,31424E-05	7,0342E-09
alpha	0,05	0,05
normal	no	no

d'Agostino-Pearson

DA-stat	21,186	13,302
p-value	2,50856E-05	0,001292779
alpha	0,05	0,05
normal	no	no

Ilustración 10 Variable Afroamericano

Shapiro-Wilk Test

	Salario (dólares/mes)	educación (años)
W-stat	0,936	0,882
p-value	0,00	0,00
alpha	0,05	0,05
normal	no	no

d'Agostino-Pearson

DA-stat	184,081	130,190
p-value	0,000	0,000
alpha	0,05	0,05
normal	no	no

Ilustración 11 Variable Caucásico

A partir del output de la prueba de Shapiro, se puede observar que los p-values para ambas variables son menores a 0,05 por lo cual se deduce que las variables, no se distribuyen de manera normal.

Test de Hipótesis

μ_1 : Variable Afroamericano
 μ_2 : Variable Caucásico

$H_0: \mu_1 = \mu_2$
 $H_a: \mu_1 \neq \mu_2$

Prueba z para medias de dos muestras

	Salario Afroamericano	Salario Caucasico
Media	735,842	990,648
Varianza (conocida)	87575,092	166466,243
Desviación Estandar	295,931	408,003
Observaciones	120,000	815,000
Diferencia hipotética de las medias	0	
z	-8,337	
P(Z<=z) una cola	0,000	
Valor crítico de z (una cola)	1,645	
Valor crítico de z (dos colas)	0,000	
Valor crítico de z (dos colas)	1,960	

Ilustración 12 Test Hipótesis Salario

Se rechaza la Hipótesis nula de que ambas variables poseen el mismo rango salarial, ya que el p-value es menor al α de 0,05 y además el z calculado en valor absoluto es mayor al valor crítico de z.

μ_1 : Variable Afroamericano
 μ_2 : Variable Caucásico

$H_0: \mu_1 = \mu_2$
 $H_a: \mu_1 \neq \mu_2$

Prueba z para medias de dos muestras

	educación Afroamericano	educación Caucasico
Media	12,442	13,620
Varianza (conocida)	3,022	4,917
Desviación Estandar	1,738	2,217
Observaciones	120,000	815,000
Diferencia hipotética de las medias	0	
z	-6,667	
P(Z<=z) una cola	0,000	
Valor crítico de z (una cola)	1,645	
Valor crítico de z (dos colas)	0,000	
Valor crítico de z (dos colas)	1,960	

Ilustración 13 Test Hipótesis Educación

Se rechaza la H_0 de que ambas variables poseen el mismo rango salarial, ya que el p-value es menor al α de 0,05 y además el z calculado en valor absoluto es mayor al valor crítico de z.

POBLACION		
	Salario (dólares/mes)	educación (años)
Media	957,945	13,468
Desviación estándar	404,361	2,197
Varianza de la muestra	163507,675	4,825
Coeficiente de variación	42,211	16,310

Ilustración 14 Coeficiente Variación Población

Según el calculo realizado la variable con mayor variación es el salario, ya que presenta un 42,21 contra un 16,31 de la variable educación.

2) Utilizando la prueba de t apropiada pruebe que ambos grupos raciales tienen igual nivel salarial y educacional. Para definir la prueba apropiada, y utilizando la prueba de F, compare las varianzas de cada variable en cada grupo. Establezca sus conclusiones.

	Salario Afroamericano	Salario Caucasico	educación Afroamericano	educación Caucasico
Media	735,842	990,648	12,442	13,620
Varianza	87575,092	166466,243	3,022	4,917
Observaciones	120	815	120	815
Diferencia hipotética de las medias	0		0	
Grados de libertad	193		181	
Estadístico t	-8,337		-6,667	
P(T<=t) una cola	0,000		0,000	
Valor crítico de t (una cola)	1,653		1,653	
P(T<=t) dos colas	0,000		0,000	
Valor crítico de t (dos colas)	1,972		1,973	

Ilustración 15 Prueba de t varianzas desiguales

	Salario Afroamericano	Salario Caucasico	educación Afroamericano	educación Caucasico
Media	735,842	990,648	12,442	13,620
Varianza	87575,092	166466,243	3,022	4,917
Observaciones	120	815	120	815
Grados de libertad	119	814	119	814
F	0,526		0,615	
P(F<=f) una cola	0,000		0,001	
Valor crítico para F (una cola)	0,786		0,786	

Ilustración 16 Prueba de F para varianza de dos muestras

Como conclusión se puede indicar que, en ambas pruebas tanto de F para dos muestras como la prueba de t para varianzas desiguales, se ve que el p value es menor al α del 0,05, por lo cual se rechaza la Hipótesis nula de que el nivel salarial y educacional es igual entre los grupos raciales.

También al comparar las varianzas de cada variable en cada grupo, se puede concluir que la variable caucásica es la que muestra una mayor varianza tanto para el nivel educacional como salarial.

3) Utilizando una prueba de Chi-Cuadrado determine si el número de afroamericanos y caucásicos encuestados, por cada nivel de educación, fue similar al esperado.

Para realizar este análisis, los encuestados se agruparon por raza y por nivel educacional como sigue:

Media: Hasta los 12 años de estudio.
Superior: Mayor a 12 años de estudio.

	Media	Superior	Total
Afroamericanos	85	35	120
Caucasicos	396	419	815
Total	481	454	935

Ilustración 17 Valores Observados

	Media	Superior	Total
Afroamericanos	62	58	120
Caucasicos	419	396	815
Total	481	454	935

Ilustración 18 Valores Esperados

Analizando el calculo realizado para los valores observados se puede identificar que en el caso de los afroamericanos que se encuestaron en el nivel educacional media son mayores a los observados y en el caso del nivel educacional superior los encuestados son menores a los esperados. En el caso de los caucásicos los encuestados en el nivel de educación media fue menor a lo esperado y en el nivel educacional superior fue mayor al esperado.

SUMMARY		Alpha		0,05
Count	Rows	Cols	df	
935	2	2	1	

CHI-SQUARE						
	chi-sq	p-value	x-crit	sig	Cramer V	Odds Ratio
Pearson's	20,720012	5,3158E-06	3,84145882	yes	0,14886383	2,56962482
Max likelihood	21,3517614	3,8227E-06	3,84145882	yes	0,15111621	2,56962482

Ilustración 19 Prueba Chi-cuadrado

Al revisar la prueba de chi-cuadrado se puede deducir que el chi calculado es mayor al chi critico por lo cual se rechaza la hipótesis nula de que los valores observados son iguales a los esperados.

4) Realice un análisis de correlación general entre las variables salario y educación. Realice el mismo ejercicio por grupo racial. Establezca sus conclusiones en relación al grado, tipo y significancia de las relaciones obtenidas.

$H_0: \rho = 0$

$H_a: \rho \neq 0$

	Salario (dólares/mes)	educación (años)
Salario (dólares/mes)		1
educación (años)	0,32710869	1

Ilustración 20 Correlación Población

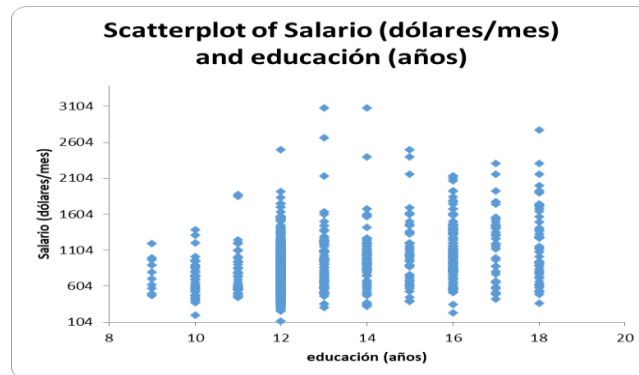


Ilustración 21 Grafico de Dispersión Población

Correlation Coefficients

Pearson	0,32710869
Spearman	0,313764

Pearson's coeff (t test)

Alpha	0,05
Tails	2
corr	0,327
std err	0,031
t	10,573
p-value	0,000
lower	0,266
upper	0,388

Pearson's coeff (Fisher)

Rho	0
Alpha	0,05
Tails	2
corr	0,32710869
std err	0,032721
z	10,367143
p-value	3,4983E-25
lower	0,26862965
upper	0,38318529

Ilustración 22 Coeficiente de Correlación Población

Viendo el grafico y el calculo del coeficiente de correlación, se puede concluir que hay una baja correlación positiva entre las variables educación y salario en la población, por lo tanto, hay una baja significancia.

Según el calculo de la prueba t, nos indica que se rechaza la Hipótesis nula de que no existe correlación entre las variables, dado que el p value es menor al α de 0,05.

	Salario (dólares/mes)	educación (años)
Salario (dólares/mes)		1
educación (años)	0,103703908	1

Ilustración 23 Correlación Afroamericano

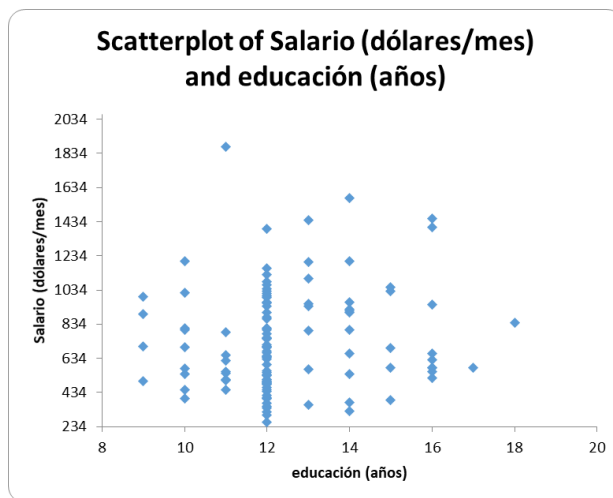


Ilustración 24 Grafico de Dispersión Afroamericano

Al realizar el cálculo del coeficiente de correlación, se puede concluir que en el grupo afroamericano es aún más baja pero positiva la correlación entre las variables educación y salario, por lo cual tiene una baja significancia.

	Salario (dólares/mes)	educación (años)
Salario (dólares/mes)	1	
educación (años)	0,317257331	1

Ilustración 25 Correlación Caucásico

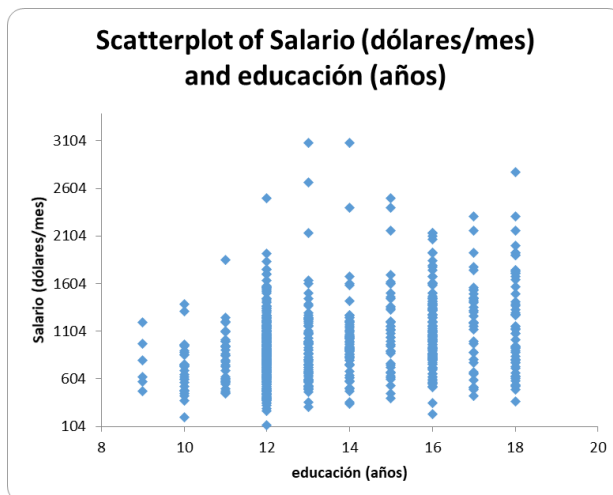


Ilustración 26 Grafico de Dispersión Caucásico

Aquí también se puede ver que la correlación entre las variables educación y salario en el grupo caucásico es baja pero positiva, con una baja significancia.

CONCLUSION

Por medio de las diferentes pruebas que se realizó durante el informe se puede concluir que las variables salario y educación no son iguales dentro de los grupos raciales afroamericanos y caucásicos. Por consiguiente, se puede indicar que existe un sesgo por grupos raciales, ya que los salarios son mayores para los caucásicos que para los afroamericanos y además esta variable salario es la que tiene una mayor variación en la población.

Por otra parte, las variables salario y educación tienen una correlación positiva pero baja.