

Análisis Desempeño Académico

Estadística Multivariada | Roxana Pérez Medina

Variables

Cualitativas

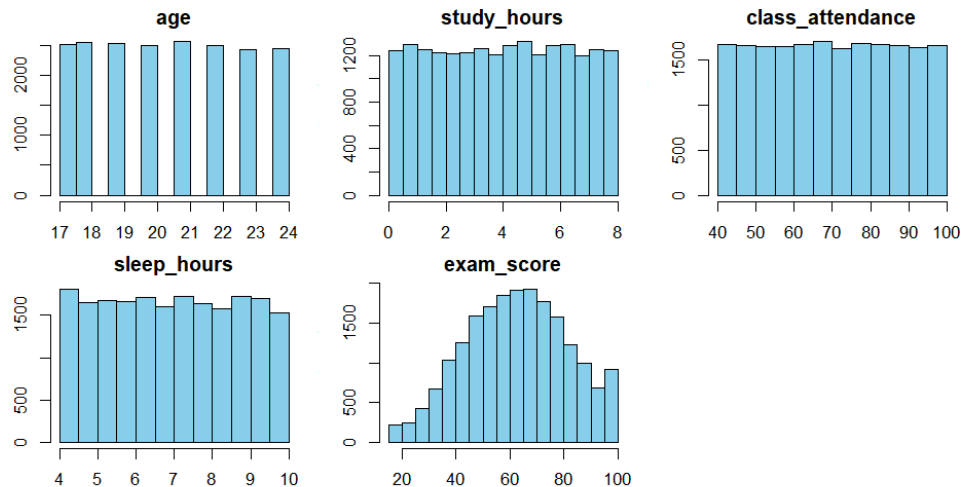
Age
Study hours
Class attendance
Sleep hours
Exam score

Cuantitativas

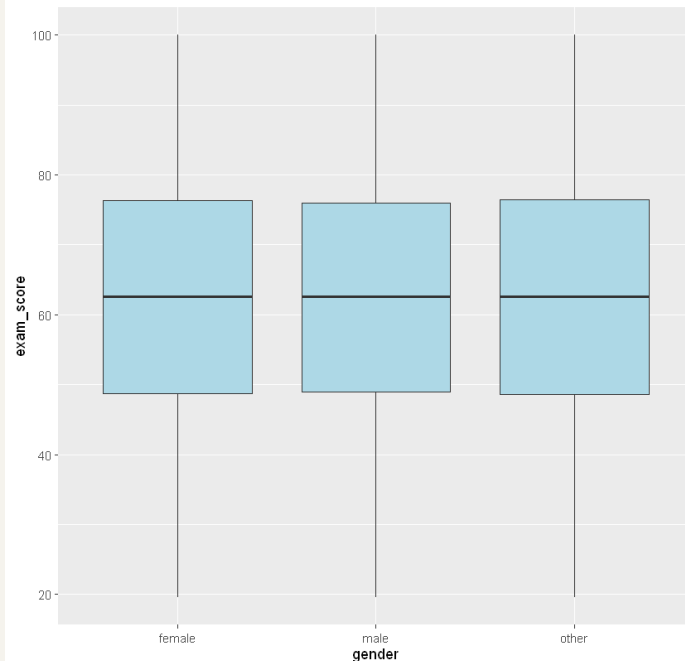
Gender
Course
Internet Access
Sleep quality
Study method
Facility rating
Exam difficulty

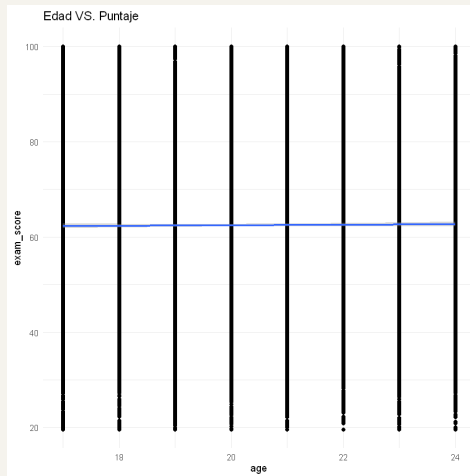
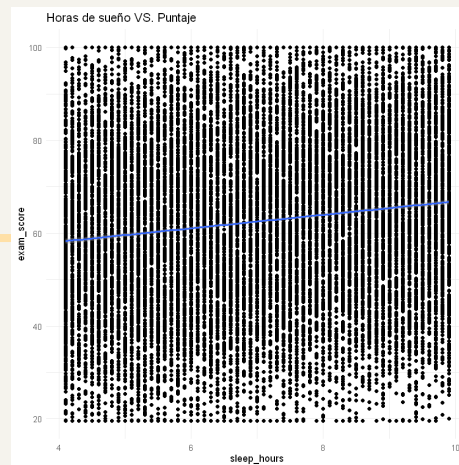
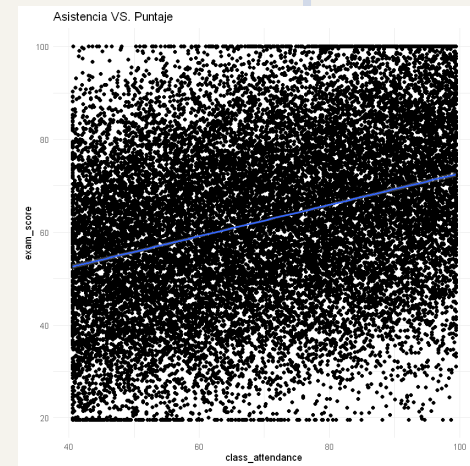
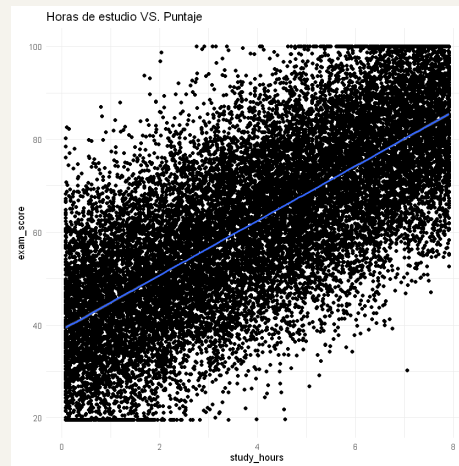
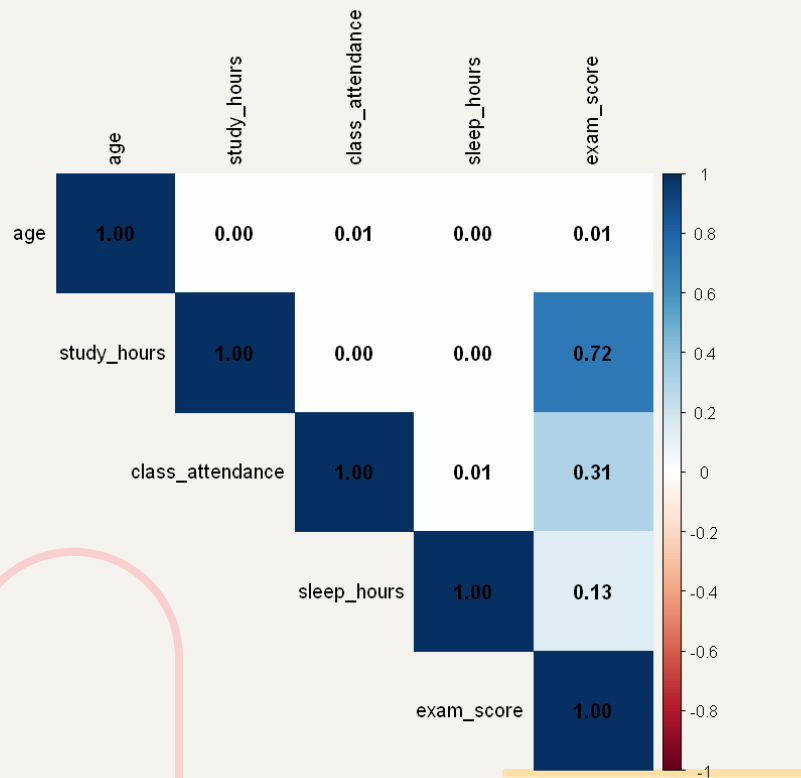
Analisis Exploratorio

Distribución de Variables Cuantitativas



Distribución de puntajes por género





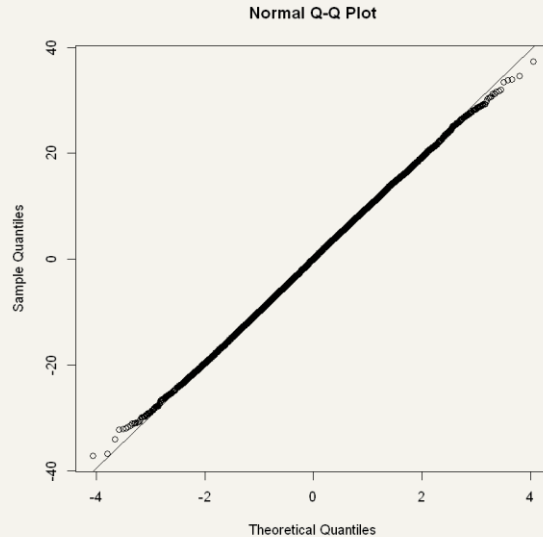
Regresión Lineal Múltiple

Multicolinealidad

	Horas Estudio	Asistencia Clases	Horas Sueño	Edad
Horas Estudio	1.000000	-0.001645	-0.004533	0.002955
Asistencia Clases	-0.001645	1.000000	0.007187	0.008449
Horas Sueño	-0.004533	0.007187	1.000000	-0.000385
Edad	0.002955	0.008449	-0.000385	1.000000

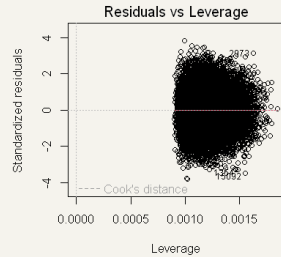
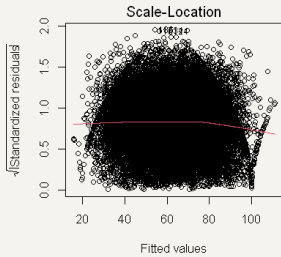
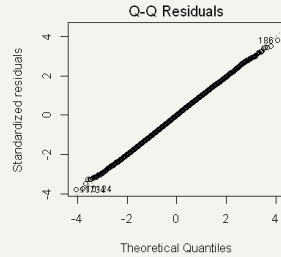
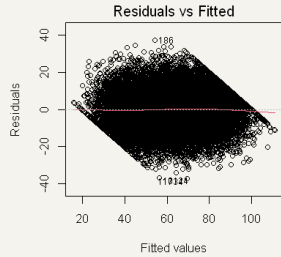
Regresión Lineal Múltiple

Normalidad en los Residuales



Regresión Lineal Múltiple

Homocedasticidad



Parámetro	Valor
Estadístico BP	32.29
Grados de libertad (df)	23
Valor p	0.0943

Regresión Lineal Múltiple

No Autocorrelación

Parámetro	Valor
Estadístico DW	2.017
Valor p	0.8853

Pruebas de Hipótesis

Hipótesis Nula: $\beta_0 = \beta_1 = \dots = \beta_p = 0$

Hipótesis Alternativa: Al menos un coeficiente es distinto de cero

Métricas del modelo

Residual Standard Error	9.79 (19976 gl)
Multiple R^2	0.7322
Adjusted R^2	0.7319
F-statistic	2375 (23, 19976 gl), $p < 0,001$
Número de observaciones	20000

Hipótesis Individuales

Hipótesis Nula: $\beta_i = 0$

Hipótesis Alterna: $\beta_i \neq 0$

Variables	Coefficiente	Error Est.	Valor t	p
Intercepto	15,113 972	0,819 229	18,449	<.001***
Género (ref: femenino)				
Masculino	0,081 963	0,170 079	0,482	0,630
Otro	0,177 948	0,169 861	1,048	0,295
Curso (ref: no especificado)				
B.Sc	-0,181 423	0,258 515	-0,702	0,483
B.Tech	0,032 088	0,260 375	0,123	0,902
BA	0,053 079	0,258 061	0,206	0,837
BBA	0,080 964	0,259 440	0,312	0,755
BCA	0,096 910	0,257 936	0,376	0,707
Diploma	-0,038 715	0,259 651	-0,149	0,881
Horas de sueño	1,447 385	0,039 985	36,198	<.001***
Método de estudio (ref: coaching)				
Grupo	-7,696 113	0,219 573	-35,050	<.001***
Mixto	-4,850 444	0,219 974	-22,050	<.001***
Videos online	-8,825 206	0,217 562	-40,564	<.001***
Auto-estudio	-9,473 660	0,217 438	-43,570	<.001***
Horas de estudio	5,896 578	0,030 009	196,496	<.001***
Asistencia a clases	0,339 531	0,004 007	84,729	<.001***
Calidad del sueño (ref: regular)				
Buena	4,587 914	0,169 760	27,026	<.001***
Mala	-4,816 629	0,169 324	-28,446	<.001***
Acceso a internet (ref: no)				
Sí	0,085 751	0,193 659	0,443	0,658
Calificación instalaciones (ref: alta)				
Baja	-7,821 262	0,170 203	-45,953	<.001***
Media	-3,767 067	0,169 454	-22,231	<.001***
Dificultad examen (ref: fácil)				
Difícil	0,044 889	0,199 271	0,225	0,822
Moderada	0,062 434	0,159 158	0,392	0,695
Edad	-0,010 385	0,030 316	-0,343	0,732

Selección de Modelo

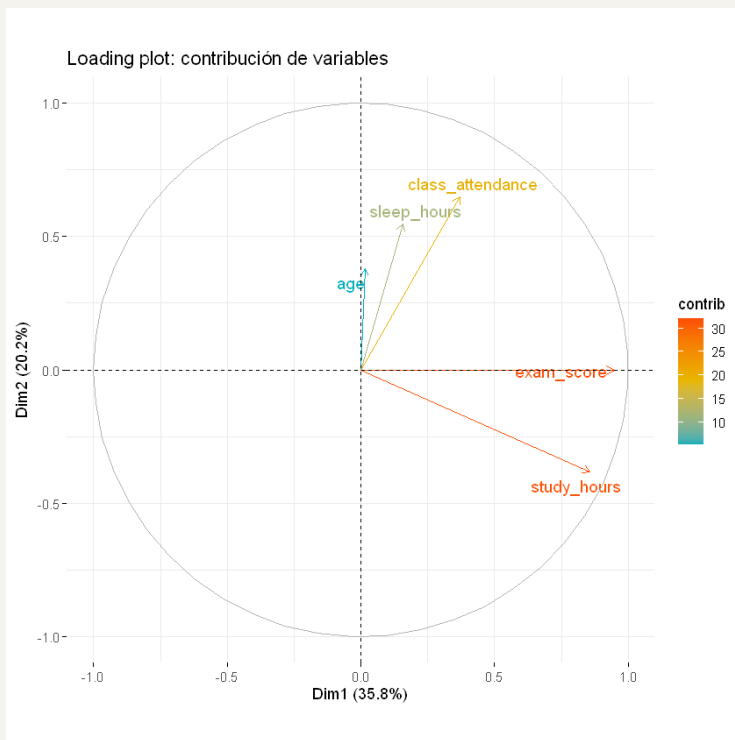
Variable	Coefficiente	Error Est.	<i>t</i>	<i>p</i>
Intercepto	15,112 242	0,461 829	32,720	<.001***
Horas de sueño	1,446 783	0,039 966	36,200	<.001***
Método de estudio (ref: coaching)				
Estudio en grupo	-7,696 005	0,219 482	-35,060	<.001***
Método mixto	-4,849 712	0,219 881	-22,060	<.001***
Videos online	-8,825 371	0,217 450	-40,590	<.001***
Auto-estudio	-9,473 642	0,217 319	-43,590	<.001***
Horas de estudio	5,896 161	0,029 992	196,590	<.001***
Asistencia a clases	0,339 524	0,004 005	84,770	<.001***
Calidad del sueño (ref: regular)				
Buena	4,586 318	0,169 671	27,030	<.001***
Mala	-4,815 546	0,169 255	-28,450	<.001***
Calificación instalaciones (ref: alta)				
Baja	-7,819 068	0,170 139	-45,960	<.001***
Media	-3,764 709	0,169 383	-22,230	<.001***
Métricas del modelo				
Error estándar residual	9.788 (19988 gl)			
R^2 múltiple	0.7322			
R^2 ajustado	0.7321			
Estadístico <i>F</i>	4968 (11, 19988 gl), $p < 0,001$			
Observaciones	20000			

PCA

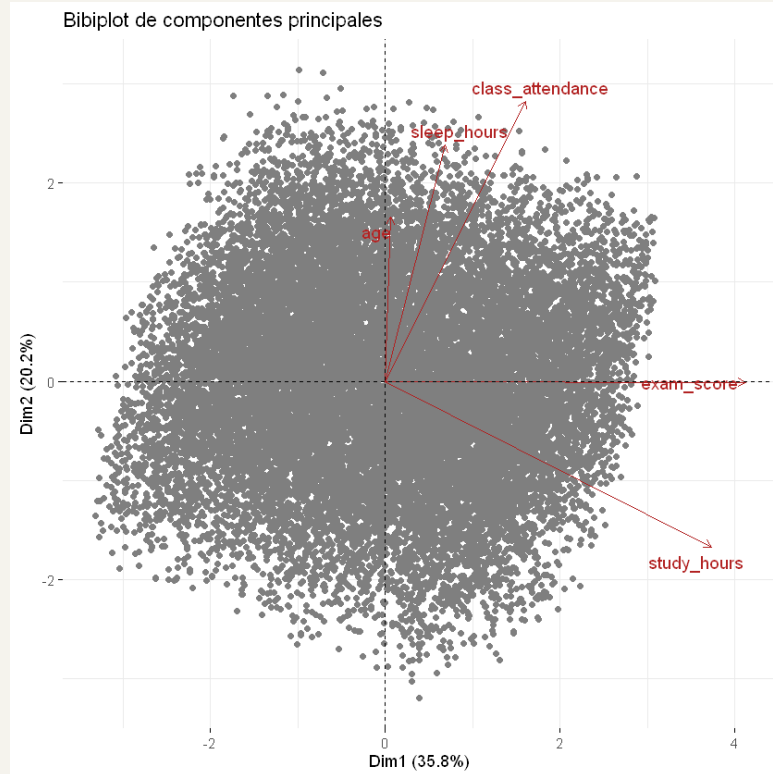
	Componentes Principales				
	PC1	PC2	PC3	PC4	PC5
Estadísticos de los componentes					
Eigenvalor	1,791 988	1,010 838	1,001 437	0,989 227	0,206 510
% Varianza	35,839 77	20,216 75	20,028 74	19,784 54	4,130 20
% Varianza acumulada	35,839 77	56,056 52	76,085 26	95,869 80	100,000 00
Eigenvectores					
Edad	0,011 097	0,377 637	0,788 564	0,485 216	-0,000 550
Horas estudio	0,639 938	-0,381 818	0,067 079	0,172 786	-0,640 578
Asistencia clases	0,275 730	0,644 591	0,088 314	-0,651 822	-0,275 326
Horas sueño	0,117 831	0,544 153	-0,604 833	0,556 626	-0,119 825
Puntaje examen	0,707 421	-0,002 407	-0,006 729	-0,002 569	0,706 752

PCA

pca_obj

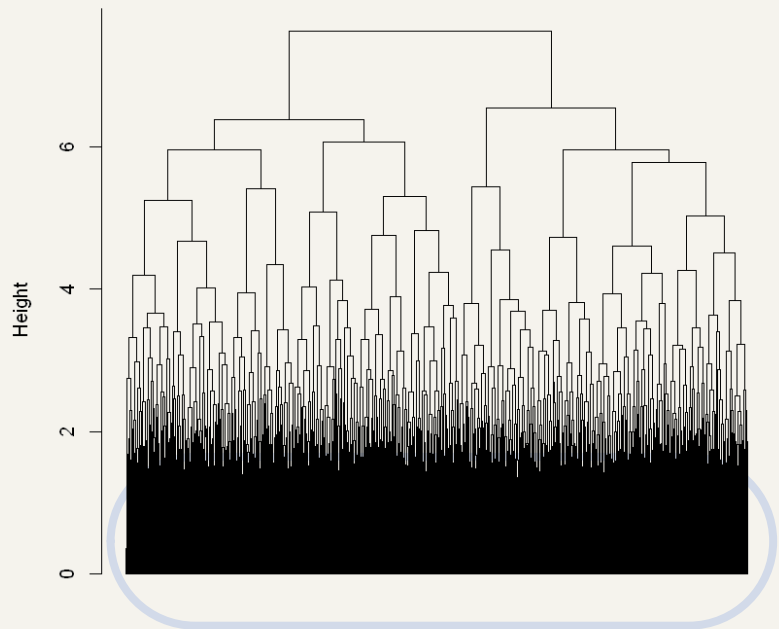


PCA

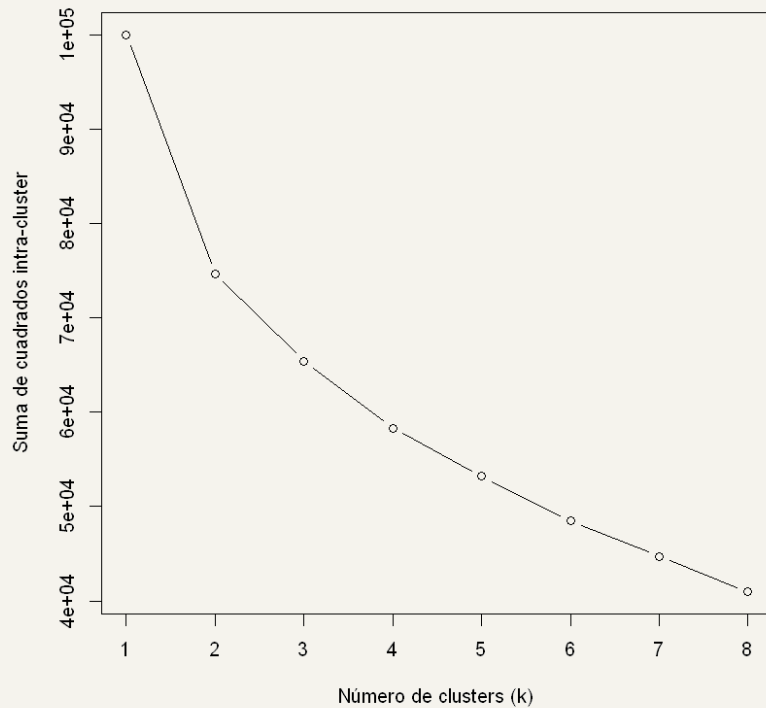


Clustering

Dendrograma de clustering jerárquico (completo)

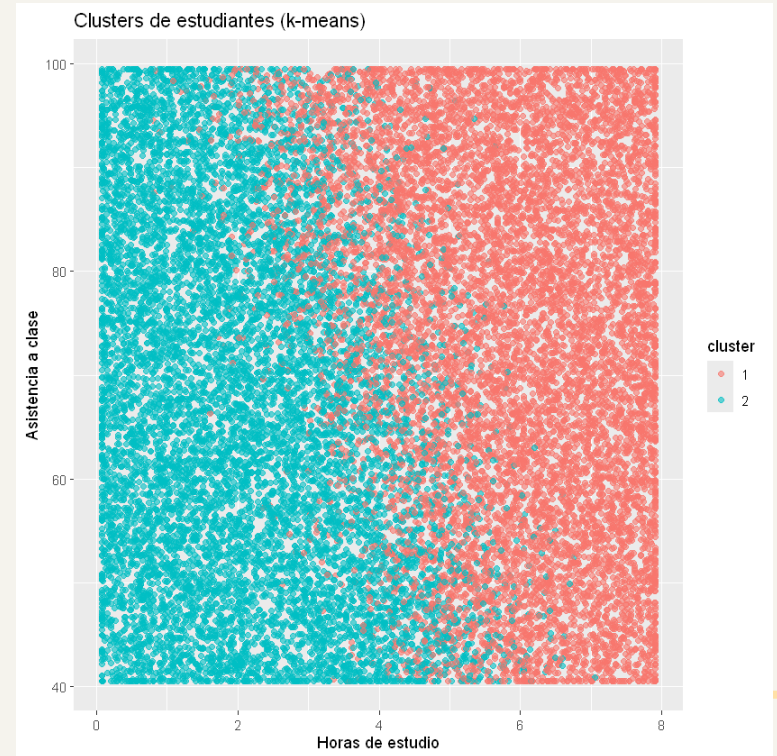


Método del codo



K-Means

Variable	Cluster 1	Cluster 2
Edad	20.50	20.45
Horas de estudio	5.81	2.15
Asistencia a clases (%)	73.16	66.78
Horas de sueño	7.13	6.89
Puntaje examen	76.92	47.68



Plots Multivariados

