

Tarea 1

Ainé Fernández 188680

Ejercicio 1: Normal Bivariada

Gráfica de superficie

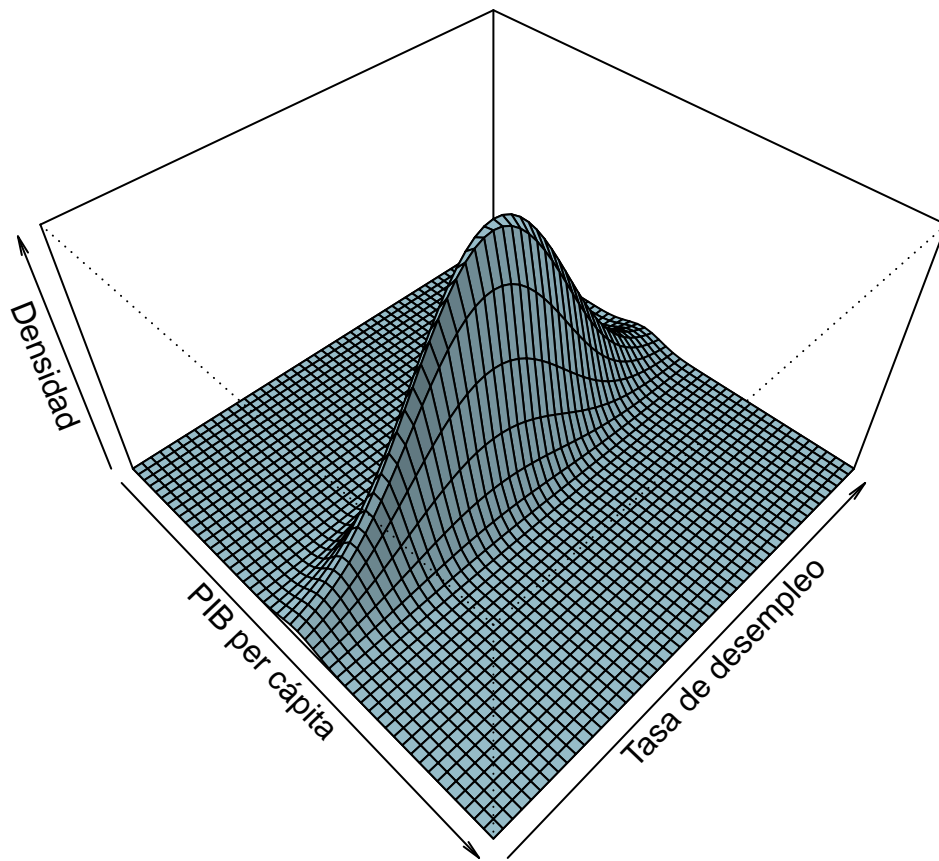


Figure 1: Gráfica de superficie

Estadísticos Descriptivos

Table 1: Estadísticos descriptivos

	Estadístico	PIB Per Cápita	Tasa Desempleo
1	Media	40,081.860	5.169
2	Varianza	13,694,989.000	3.913
3	Desviación estándar	3,700.674	1.978
4	Mínimo	29,994.340	-0.224
5	Máximo	47,188.230	10.536
6	Covarianza	-1,201.140	-1,201.140
7	Correlación	-0.164	-0.164

¿Qué indican las medias de las variables simuladas? Compara con los valores teóricos de las medias establecidas en el ejercicio

Las medias de las variables simuladas son muy similares a los valores teóricos de la distribución normal bivariada.

Table 2: Comparación de medias valores teóricos vs medias de variables simuladas

Variable	Teórica	Simulada
PIB per cápita	40000	40081.860
Tasa de desempleo	5	5.169

Esto muestra que se cumple la Ley de los Grandes Números, la cual establece que, a medida que aumenta el tamaño de la muestra, la media muestral se aproxima cada vez más a la media teórica o poblacional.

¿Cómo interpretas la covarianza negativa entre el PIB per cápita y la tasa de desempleo? ¿Es coherente con la teoría económica?

Existe una relación negativa entre estas dos variables, lo que significa que, si el PIB per cápita aumenta, la tasa de desempleo tiende a disminuir y viceversa. Esto es coherente con la Ley de Okun, que establece que un aumento del 1% en la tasa de desempleo está asociado con una disminución del 2% en el PIB. Aunque la Ley de Okun no se deriva de una predicción teórica específica, los datos empíricos suelen confirmar su validez.

¿Qué puede inferirse sobre la correlación entre el PIB per cápita y la tasa de desempleo?
¿Considera que es una relación fuerte o débil?

Sugiere que hay una relación negativa entre las dos variables, pero el valor de la correlación de -0.164 indica que esta relación es débil, ya que está muy cerca de cero, lo que señala una baja asociación entre ambas variables.

Gráficas

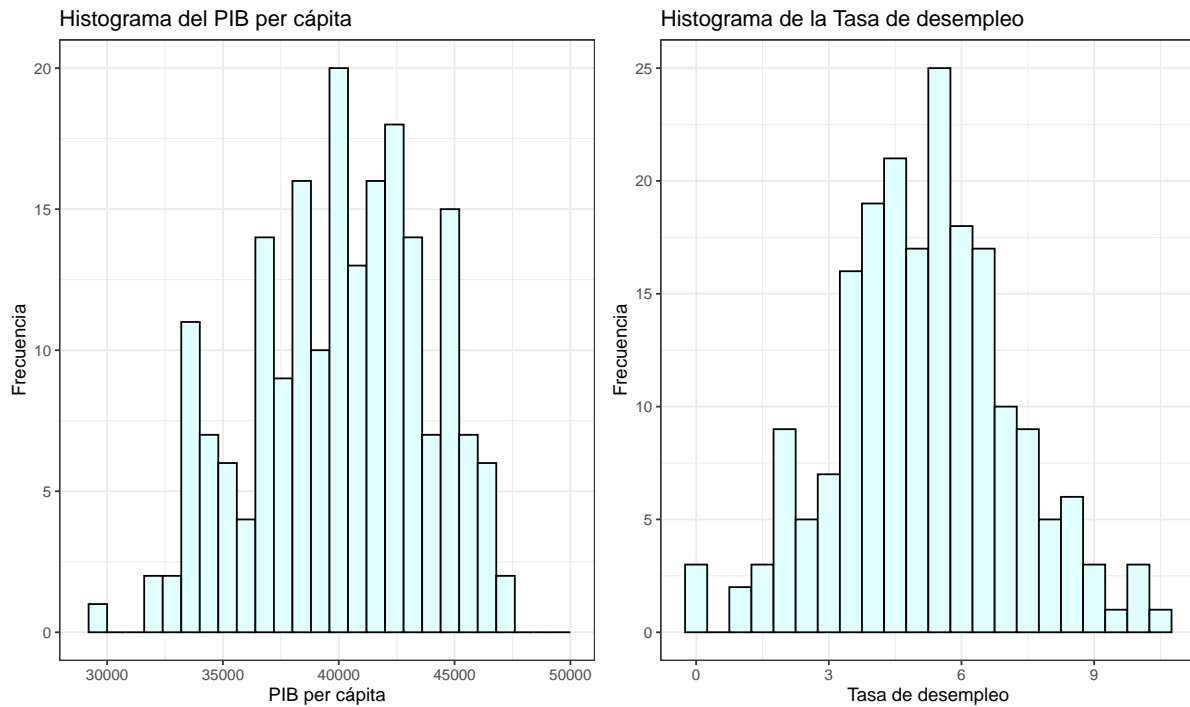


Figure 2: Histogramas de las variables simuladas

Scatter plot bivariado

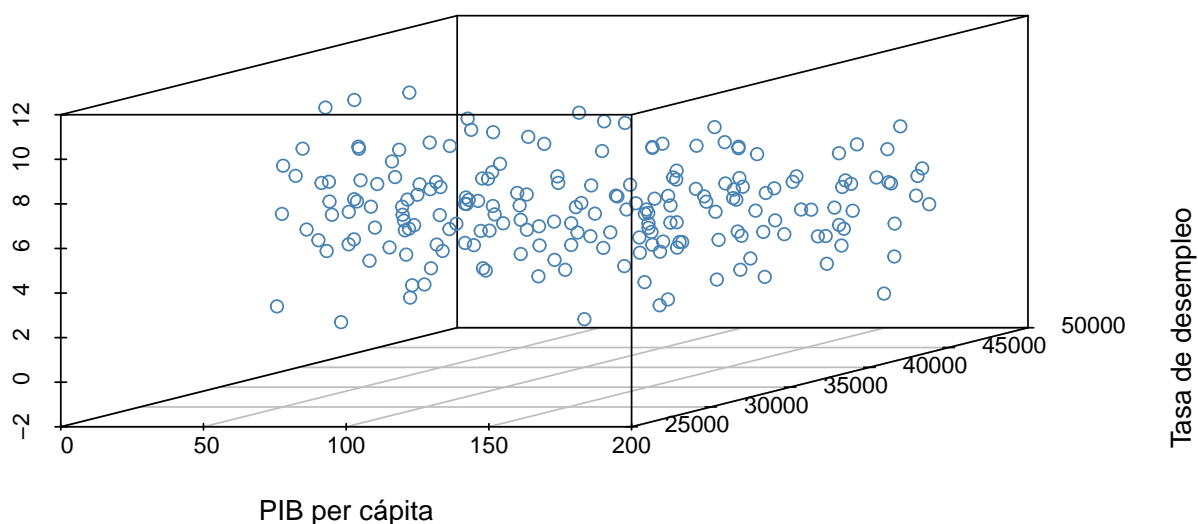


Figure 3: Scatterplot bivariado

¿Qué sugiere el histograma bivariado sobre la relación entre el PIB per cápita y la tasa de desempleo? Comente sobre la densidad de puntos y cómo se distribuyen en la gráfica.

Los puntos parecen estar distribuidos de manera uniforme en la parte superior del gráfico, lo que podría indicar que no existe una relación clara entre las dos variables. Esto se respalda con la baja correlación calculada previamente

Cálculo de probabilidades

Table 3: Cálculo de Probabilidad

PIBpercapita	Tasadedesempleo	Probabilidad
Entre 15,000 y 18,000	Menor a 3%	3.16e-11

Ejercicio 2: Statistics for Starbucks

EDA

Table 4: Data summary

Name	starbucks
Number of rows	1147
Number of columns	15
Column type frequency:	
character	4
numeric	11
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
product_name	0	1	8	47	0	93	0
size	0	1	4	7	0	11	0
trans_fat_g	0	1	1	3	0	7	0
fiber_g	0	1	1	2	0	12	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
milk	0	1	2.51	1.68	0	1.0	2.0	4	5	
whip	0	1	0.25	0.43	0	0.0	0.0	0	1	
serv_size_m_l	0	1	461.34	172.18	0	354.0	473.0	591	887	
calories	0	1	228.39	137.67	0	130.0	220.0	320	640	
total_fat_g	0	1	6.19	5.97	0	1.0	4.5	10	28	
saturated_fat_g	0	1	3.88	4.01	0	0.2	2.5	7	20	
cholesterol_mg	0	1	15.24	17.97	0	0.0	5.0	30	75	
sodium_mg	0	1	139.65	93.09	0	70.0	135.0	200	370	
total_carbs_g	0	1	37.72	23.26	0	20.0	37.0	53	96	
sugar_g	0	1	34.99	22.46	0	18.0	34.0	49	89	
caffeine_mg	0	1	91.86	78.11	0	30.0	75.0	150	475	

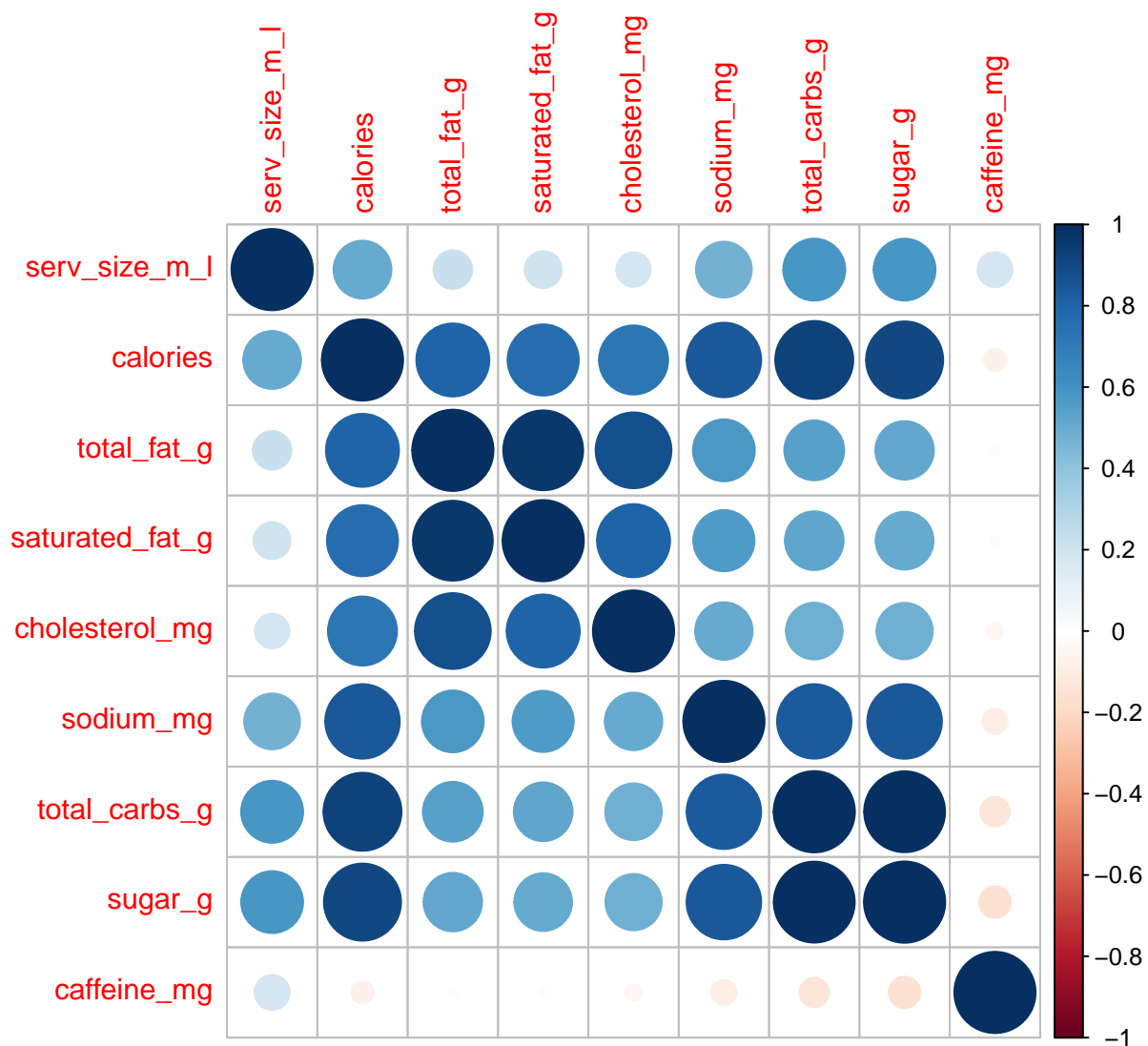


Figure 4: Correlaciones

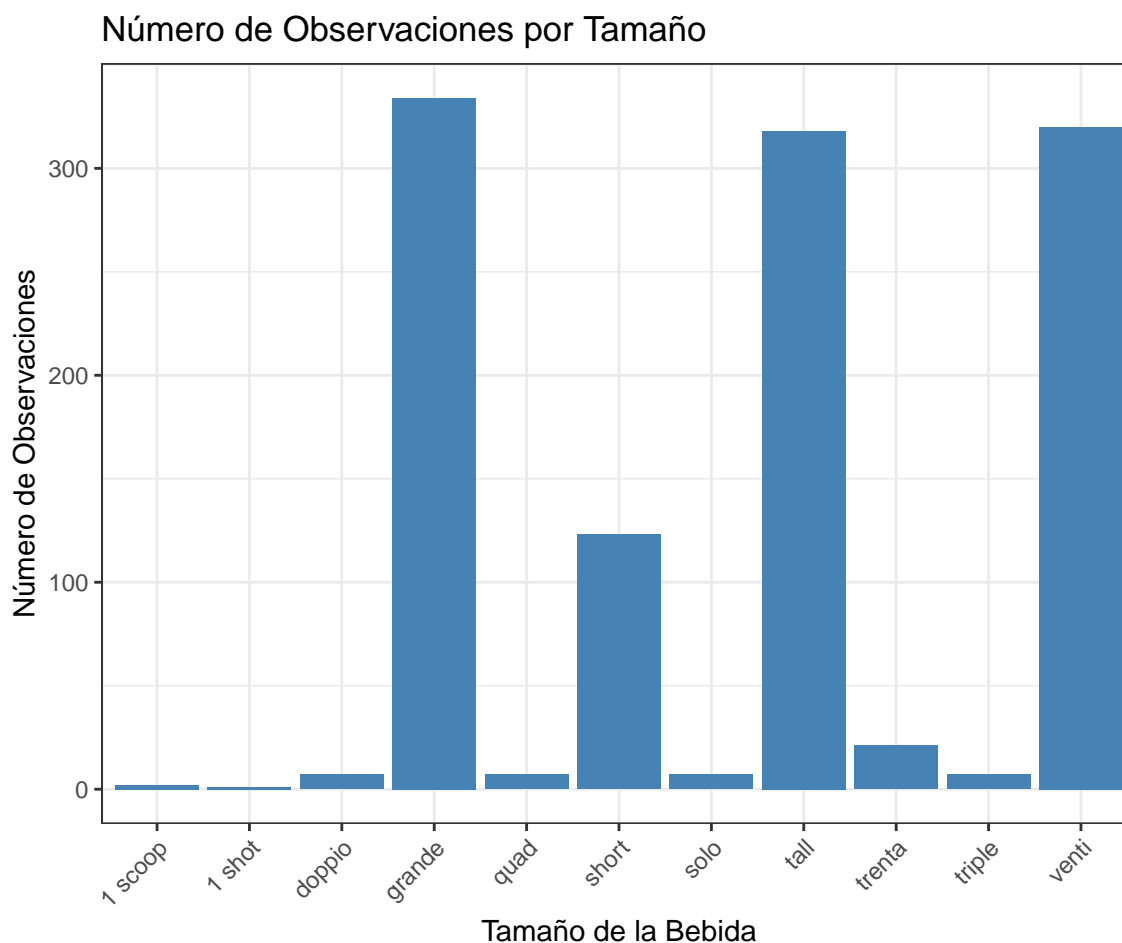


Figure 5: Número de observaciones por bebida

Lo primero que debemos notar es que la base de datos no tiene observaciones faltantes, lo cual es bastante bueno para el análisis. Luego, encontramos que hay 11 variables numéricas, pero Milk y Whip son categóricas, ya que representan el tipo de leche y si se añadió crema batida o no.

Otra observación interesante es que las calorías tienen una alta correlación con todas las otras variables, excepto con la cafeína, que parece tener una correlación negativa pero débil. De hecho, la cafeína no parece tener una relación fuerte con ninguna de las otras variables, y su relación débil con ellas parece ser negativa. Mientras tanto, las demás variables parecen estar fuertemente correlacionadas de manera positiva entre sí.

Por otro lado, hay muy pocas observaciones del tamaño Trenta, mientras que hay más de los tamaños Tall, Grande y Venti, los tamaños comunes en Starbucks

Estadísticos descriptivos

Table 7: Estadísticos descriptivos

Statistic	N	Mean	Pctl(25)	Median	Pctl(75)	St. Dev.	Min	Max
calories	1,147	228.4	130	220	320	137.7	0	640
sugar_g	1,147	35.0	18	34	49	22.5	0	89
caffeine_mg	1,147	91.9	30	75	150	78.1	0	475
total_fat_g	1,147	6.2	1.0	4.5	10.0	6.0	0.0	28.0

Gráficos

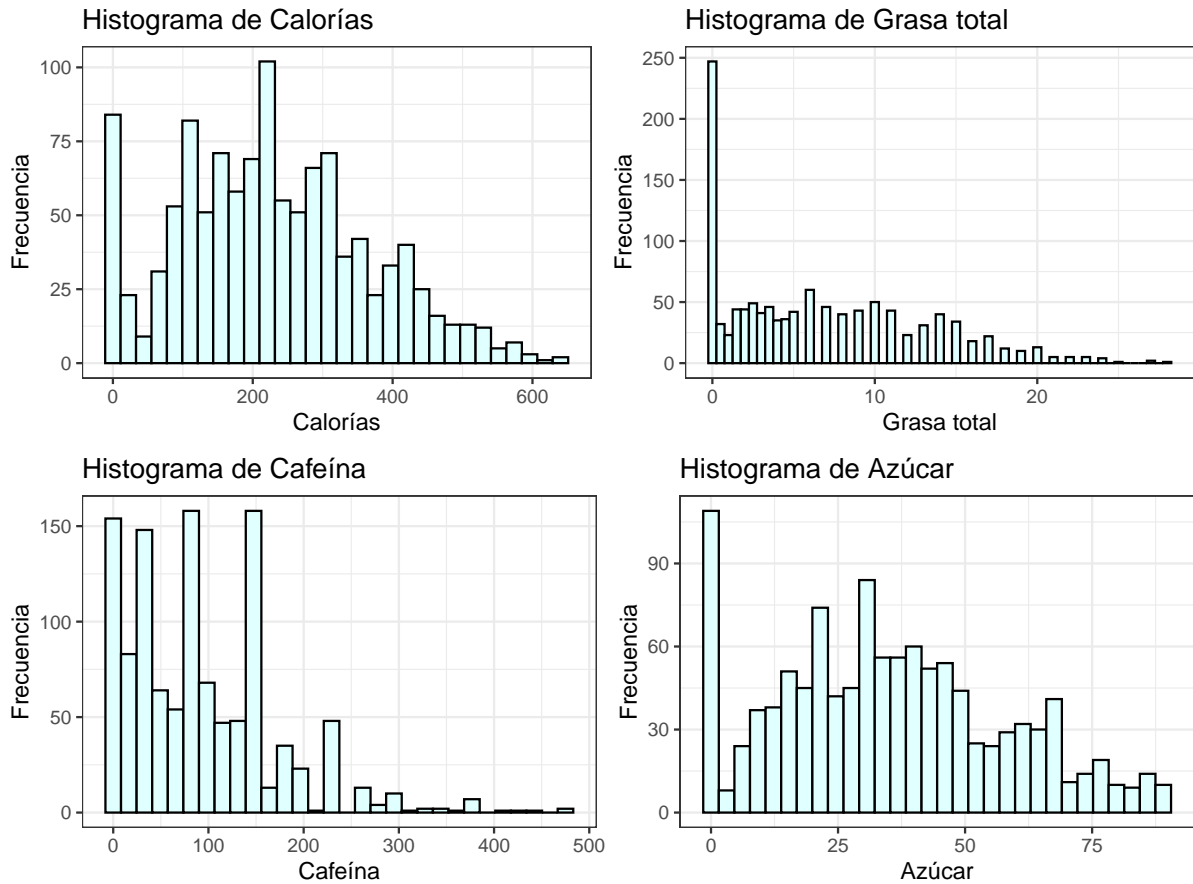


Figure 6: Histogramas de las variables de interés

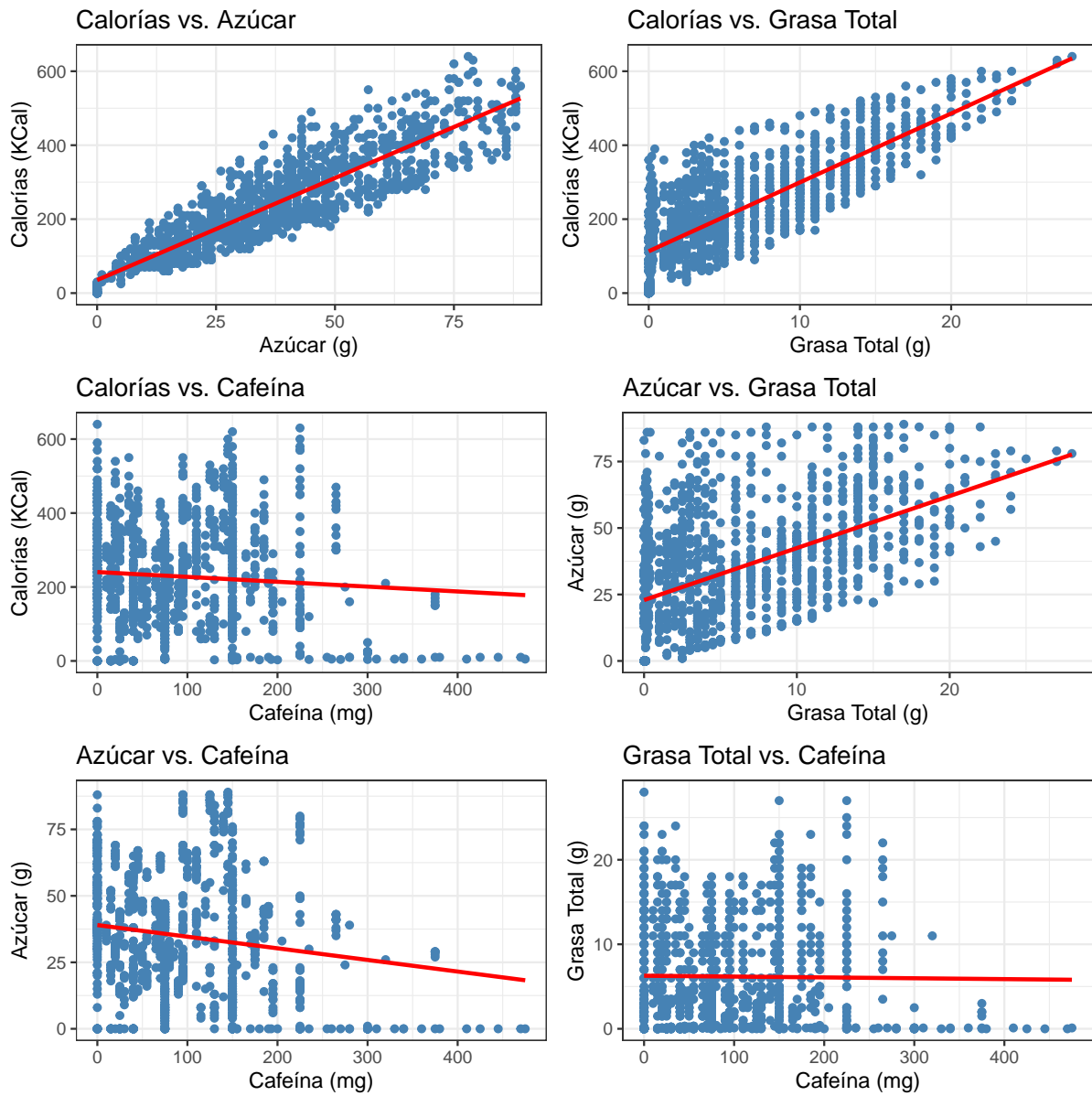


Figure 7: Diagramas de dispersión de variables de interés

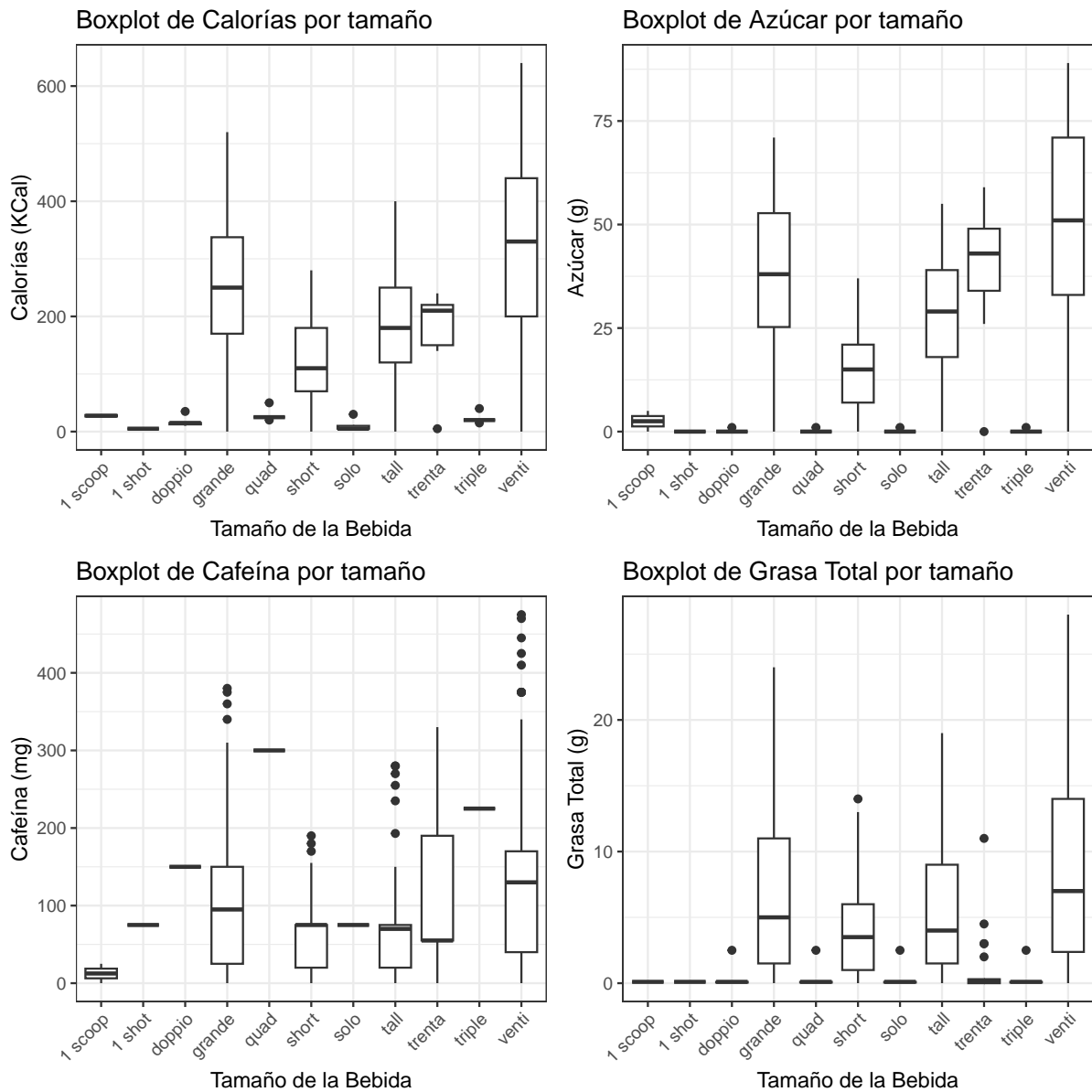


Figure 8: Diagramas de caja y brazo de variables de interés

Intervalos de confianza

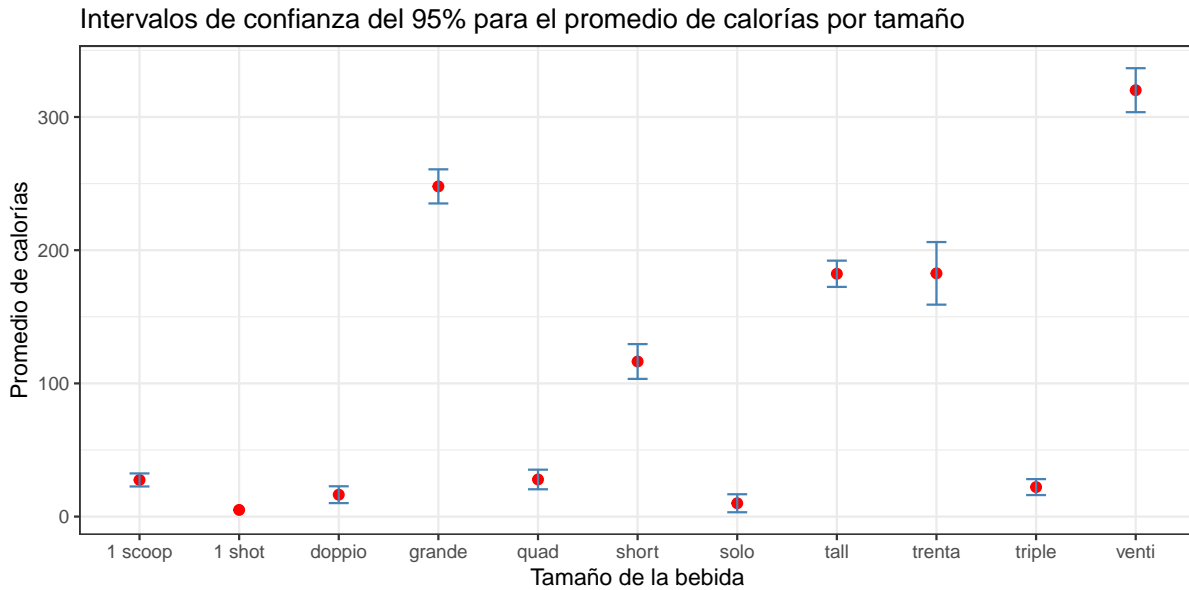


Figure 9: Intervalos de confianza del 95% para el promedio de calorías por tamaño

Los puntos rojos en el gráfico representan la estimación puntual del promedio de las calorías. Podemos observar que en los tamaños pequeños de bebida hay menos variabilidad en el promedio de las calorías, y generalmente son medias más bajas. El tamaño Trenta, que es el de mayor capacidad, muestra mucha variabilidad, pero su media no es tan alta como la del tamaño Venti. También es cierto que no se tienen tantas observaciones del tamaño Trenta. En general, podemos concluir que las calorías aumentan conforme al tamaño: los tamaños con mayores capacidades tienden a tener más calorías.

Prueba de hipótesis

Hipótesis nula (H0): El contenido promedio de grasa en la leche sin grasa es igual o mayor que en la leche entera.

Hipótesis alternativa (H1): El contenido promedio de grasa en la leche sin grasa es menor que en la leche entera.

Table 8: Prueba de hipótesis Media de Grasa Total

	MediaNonFat	MediaWhole	EstadisticoT	Valor_p	Conclusión
t	3.518468	10.13105	-12.04193	1.680664e-28	Rechazar H0

Podemos concluir que, basándonos en esta prueba, hay suficiente evidencia para decir que el contenido de grasa total en los cafés con leche descremada **NO** es igual o mayor que en los cafés con leche entera.

Bootstrap y Jackknife

Table 9: Correlación entre Cafeína y Calorías

Variables	Correlación
Cafeína y Calorías	-0.0746

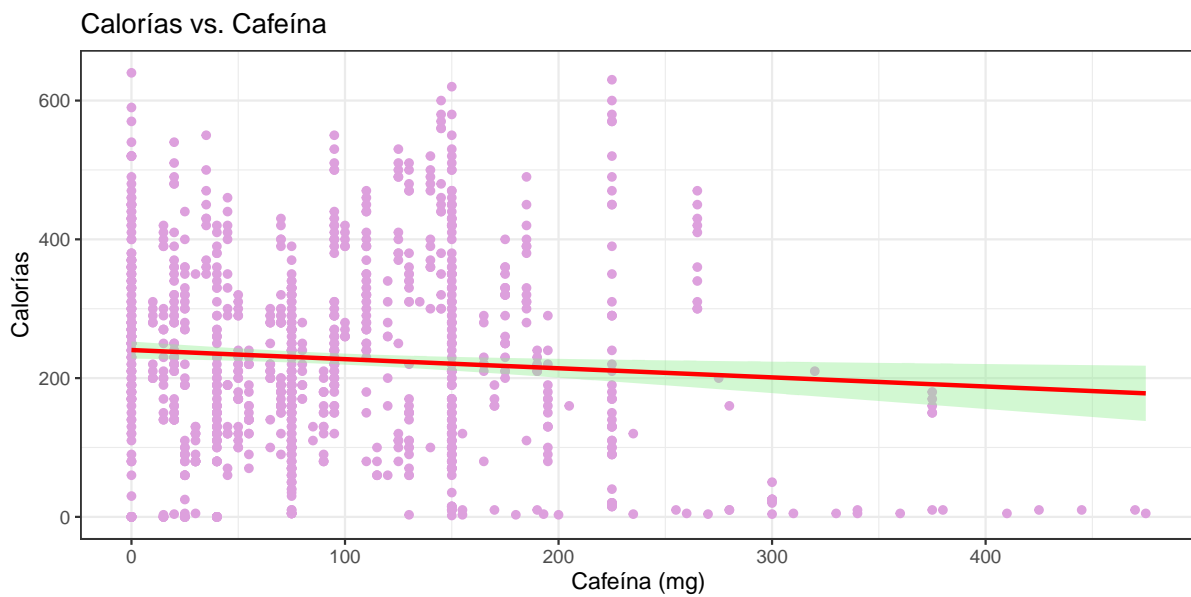


Figure 10: Relación entre calorías y cafeína

Table 10: Resultados Jackknife vs Bootstrap

Método	DesviaciónEstándar
Bootstrap (50% datos)	0.0481
Bootstrap (100% datos)	0.0333
Jackknife (50% datos)	0.0010
Jackknife (100% datos)	0.0010

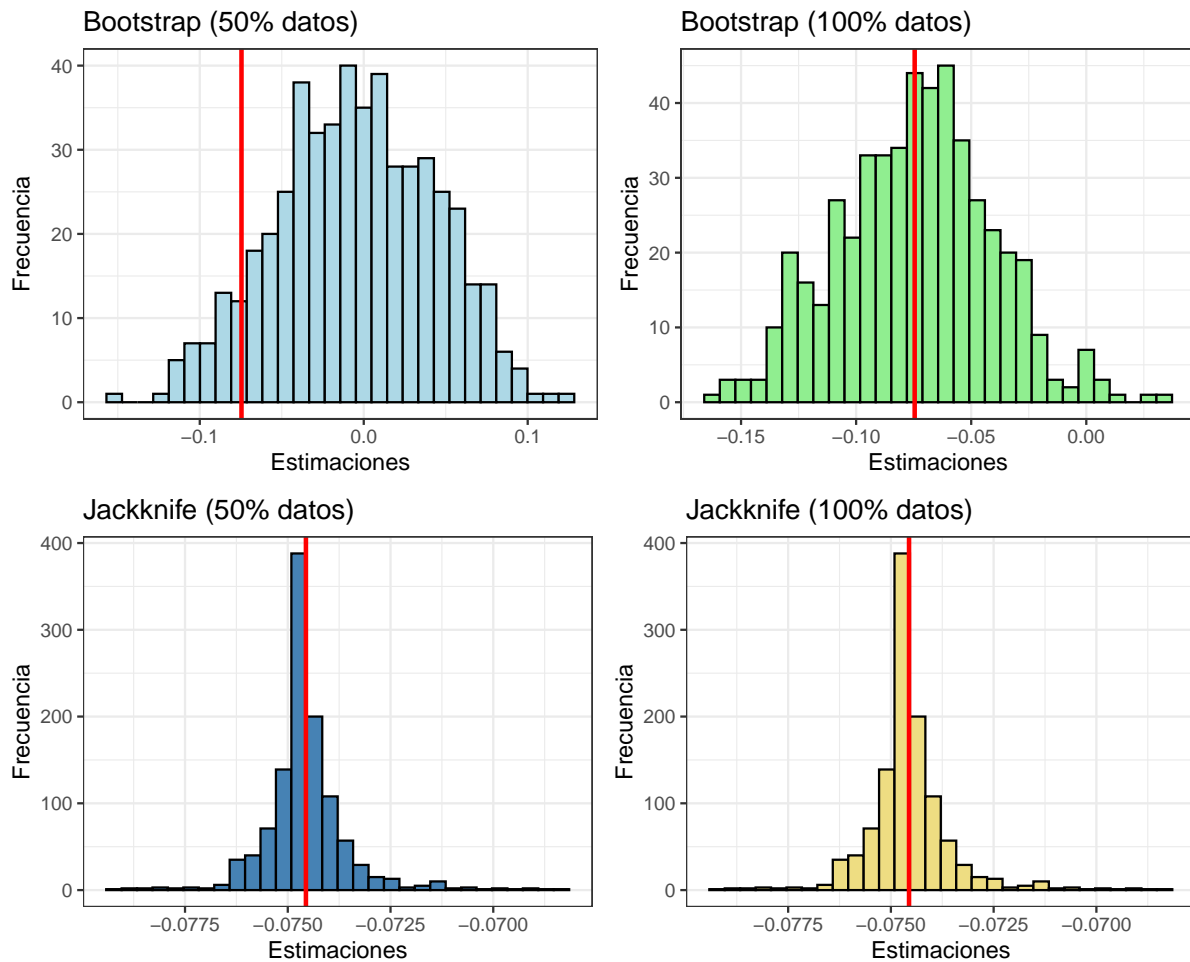


Figure 11: Bootstrap vs Jackknife

Podemos observar que los estimadores de Jackknife tienen menos varianza que los de Bootstrap al tener las estimaciones más concentradas hacia el valor real de la correlación, esto concuerda con las desviaciones estándar calculadas.

Referencias

Investopedia. (s.f.). Ley de Okun. Investopedia. Recuperado el 10 de septiembre de 2024, de <https://www.investopedia.com/terms/o/okunslaw.asp>