

Informe del 1er Proyecto de Estadística - Equipo 6

Lázaro Raúl Iglesias Vera
Grupo C412

L.IGLESIAS@ESTUDIANTES.MATCOM.UH.CU

Miguel Tenorio Potrony
Grupo C412

M.TENORIO@ESTUDIANTES.MATCOM.UH.CU

Daniel Enrique Cordovés Borroto
Grupo C411

D.CORDOVESB@ESTUDIANTES.MATCOM.UH.CU

Resumen

Aplicar análisis estadístico a un set de datos referentes al rendimiento de estudiantes de una asignatura determinada y a su vez interpretar los comportamientos de muestras de diferentes tamaños, dada una población, usando los Estadísticos Descriptivos.

Abstract

Apply an statistical analysis to a set of data referent to the performance of students of a subject and the same time to interpret the behavior of some samples of different sizes given a population using Descriptive Statistics.

Palabras Clave: Estadísticos Descriptivos, Muestra, Variables.

Tema: Análisis Estadístico, Estadística Descriptiva, Estadística Inferencial.

1. Introducción

En este trabajo se planea realizar un análisis estadístico del desarrollo de un grupo de estudiantes en cierta asignatura, utilizando un subconjunto de variables que integra el set de datos donde son descritos los estudiantes de forma individual. También se estudia el comportamiento de los Estadísticos Descriptivos en varias muestras tomadas de una población normal y la vez se calculan intervalos de confianza para media y varianza.

2. Ejercicios

Se resolverán los ejercicios en el orden que están propuestos en la orientación [1].

Tanto en 2.1 como en 2.4 se utiliza y procesa el set de datos orientado para el equipo. Este contiene información para medir el rendimiento de estudiantes de educación secundaria en dos escuelas portuguesas. Contiene variables como los resultados en diferentes períodos, estado familiar y características económicas de cada estudiante. Todas estas variables están descritas en [2].

2.1 Ejercicio 1

2.1.1 INCISO A

Se consideraron las variables que miden los resultados de los tres períodos de cada estudiante como las más importantes, ya que interesaba mostrar la evolución del rendimiento en la asignatura analizada a través del

tiempo. En los datos se pueden apreciar otras variables, que podrían ser de interés, como el sexo de los estudiantes o el número de fracasos anteriores; pero dado el requisito de solo escoger tres de estas, se valoraron las notas.

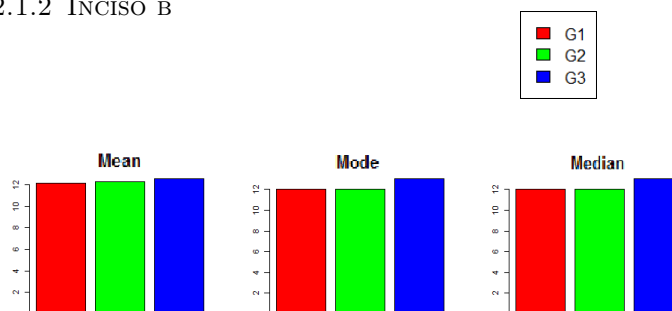
En la Figura 1, se muestran los Estadísticos Descriptivos de las variables.

Dato	Media	Moda	Mediana
G1	12.1125	12.0000	12.0000
G2	12.2382	12.0000	12.0000
G3	12.5157	13.0000	13.0000

Dato	Varianza	Desviación Estándar	CV
G1	6.5358	2.5565	0.2110
G2	6.0927	2.4683	0.2016
G3	8.6756	2.9454	0.2353

Figure 1: Estadísticos Descriptivos

2.1.2 INCISO B



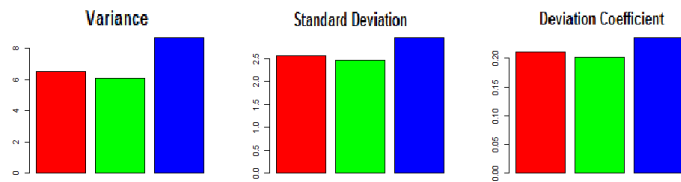


Figure 2: Estadísticos Descriptivos

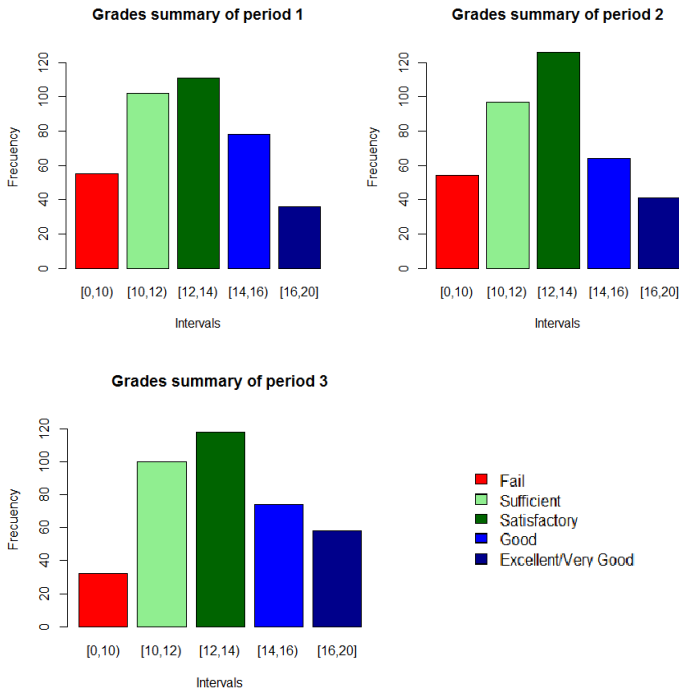


Figure 3: Notas por intervalo

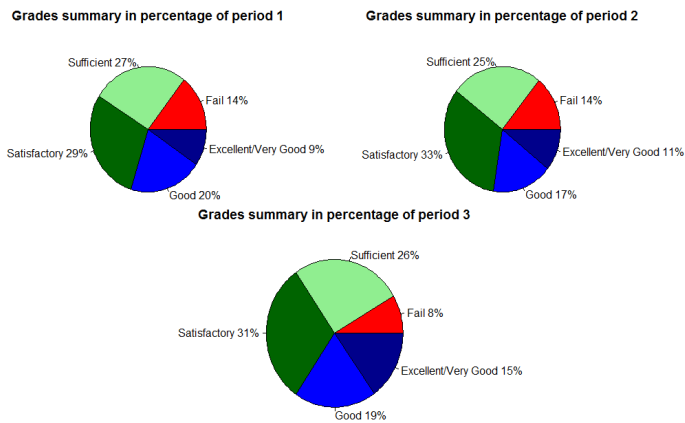


Figure 4: Porcentajes de los intervalos

2.1.3 INCISO C

La *media* (Figura 2 arriba a la izquierda) de las notas, aunque tiende a aumentar ligeramente a través de los periodos, está entre 12 y 13, lo cual quiere decir que el estudiante promedio en los tres periodos tiene una nota de suficiente.

La *moda* (Figura 2 arriba al centro) de las notas en los dos primeros periodos es 12, y en el tercer periodo 13, lo cual significa que la mayoría de los estudiantes

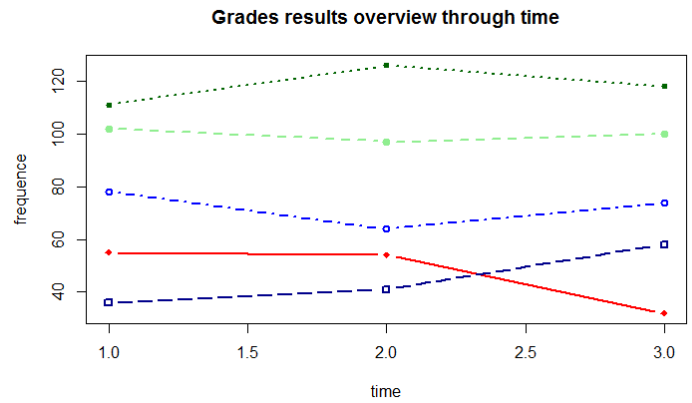


Figure 5: Evolución de los resultados

tienen nota de suficiente.

La *mediana* (Figura 2 arriba a la derecha) de las notas en los dos primeros periodos es 12, y en el tercer periodo 13, lo cual significa que el 50% de los estudiantes tienen nota de suficiente, resultado consistente con el análisis de la media y moda.

El *Coefficiente de Variación* (Figura 2 abajo a la derecha) está entre 16 y 26 en los tres periodos, lo cual quiere decir que los datos son heterogéneos.

Como es apreciable en las Figuras 3, 4 y 5, hay una tendencia de mejoría en las notas por cada periodo, siendo los indicadores más notables la disminución de los suspensos (en rojo) y el aumento de los excelentes (azul oscuro).

2.2 Ejercicio 2

2.2.1 INCISOS A, B, C

Se generó una población normal de tamaño 500 (con media 0 y desviación estándar 1). Acto seguido se generaron 8 muestras de distintos tamaños, 4 con reemplazo y 4 sin reemplazo, para analizar el comportamiento de los Estadísticos Descriptivos en las muestras y compararlos con los de la población. Se obtuvieron los resultados visibles en la Figura 6.

Cada gráfico muestra los valores de los Estadísticos Descriptivos correspondientes. En cada gráfico las barras indican el valor del estadístico en dicho gráfico para la muestra correspondiente. En total hay 8 muestras:

1. 20 elementos sin reemplazo (20)
2. 20 elementos con reemplazo (20R)
3. 20 elementos sin reemplazo (30)
4. 20 elementos con reemplazo (30R)
5. 20 elementos sin reemplazo (150)
6. 20 elementos con reemplazo (150R)

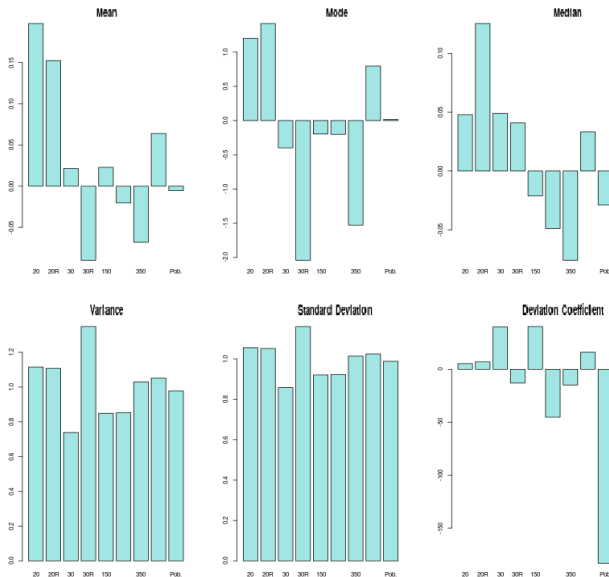


Figure 6: Análisis de Estadísticos Descriptivos en muestras

7. 20 elementos sin reemplazo (350)

8. 20 elementos con reemplazo (350R)

Pob. indica el valor del estadístico correspondiente en la población.

Como se esperaba, las muestras muestran un comportamiento similar a la población, su media es cercana a 0 y su desviación estándar a 1. También se muestran moda, mediana, varianza y coeficiente de desviación. Estos datos son de esas 8 muestras puntuales que se obtuvieron.

2.3 Inciso d

Se hallaron los intervalos de confianza para cada una de las muestras para la media y la varianza. Se obtuvieron los siguientes resultados.

Muestra	Inicio Media	Fin Media	Inicio Var	Fin Var
20	-1.004	-0.0670	0.5799	2.1391
20R	-0.7063	0.2161	0.5617	2.071
30	-0.6077	0.2026	0.7467	2.1276
30R	-0.5694	0.2165	0.7026	2.0020
150	-0.2118	0.1043	0.7875	1.2421
150R	-0.2644	0.0509	0.7833	1.2355
350	-0.1868	0.0267	0.8999	1.2113
350R	-0.2803	-0.0604	0.9542	1.2844

2.3.1 INCISO E

Se puede ver que entre las muestras de tamaño similar no existe una gran variación en los límites inferiores y superiores del intervalo de confianza para la media o varianza.

2.4 Ejercicio 3

Fueron analizadas las notas de los estudiantes de sexo masculino y femenino del tercer período para poder analizar las diferencias entre ambos grupos. Se asume que las observaciones provienen de una distribución normal.

Se puede apreciar en la Figura 7 que las medianas entre ambos grupos están cercanas (solapamiento entre las cuñas de las cajas), lo cual indica que las medias no deben ser muy diferentes; no obstante, se procede a realizar un test de hipótesis para su verificación.

En orden de realizar un test de hipótesis para las medias, se realiza una pruebas de hipótesis para la comparación de las varianzas, ya que estas son desconocidas.

H_1 : Las varianzas son diferentes

$$H_0 : \sigma_F^2 = \sigma_M^2$$

$$H_1 : \sigma_F^2 \neq \sigma_M^2$$

$$\alpha = 0.05$$

Se utilizó la función *var.test* del paquete de *R* *usefultools*. El resultado fue un p-value = 0.4158, lo cual significa que no se rechaza H_0 dado que el p-value > α . Por tanto, se asume que las varianzas son iguales (el test además ofrece como resultado la razón muestral entre las varianzas, la cual es 1.5139).

Para obtener los resultados de la prueba de hipótesis para la comparación de las medias de ambos grupos de estudiantes, se utilizó la función del lenguaje *R* *t.test*, asumiendo que las varianzas son desconocidas, pero iguales.

H_1 : Las medias son diferentes

$$H_0 : \mu_F = \mu_M$$

$$H_1 : \mu_F \neq \mu_M$$

$$\alpha = 0.05$$

El resultado fue un p-value = 0.8349, lo cual significa que no se rechaza H_0 dado que el p-value > α . Por consiguiente, se asume que las medias son iguales. La media muestral de los estudiantes es 11.6470 y 10.8235 para los grupos masculino y femenino respectivamente.

3. Conclusiones

Como resultado se puede decir que los datos procesados no difieren mucho entre sí, a pesar de ser heterogéneos, dado que las medidas de tendencia central así lo indican. Incluso, al estudiar las posibles diferencias de las notas de los estudiantes en cuanto al sexo, se demostró que no es posible afirmar que las medias son diferentes. Además se analizó el comportamiento de los Estadísticos Descriptivos en varias muestras de una población, se obtuvieron los resultados esperados para cada una de las muestras y para los intervalos de confianza de media y varianza.

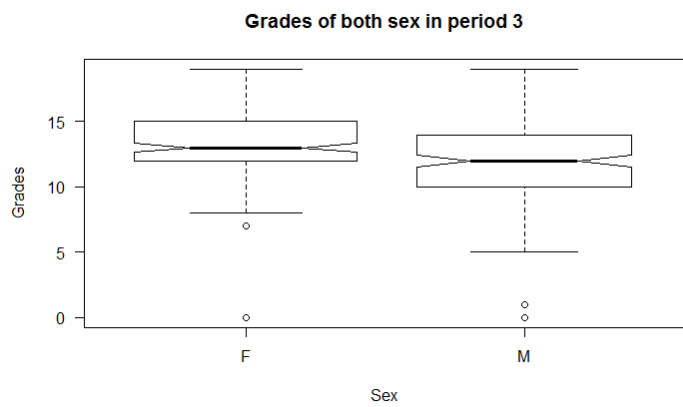


Figure 7: Gráfico de Caja y Bigote

References

- [1] *Proyecto Evaluativo Estadística Fase 1.* ([abrir](#))
- [2] [Artículo original](#)