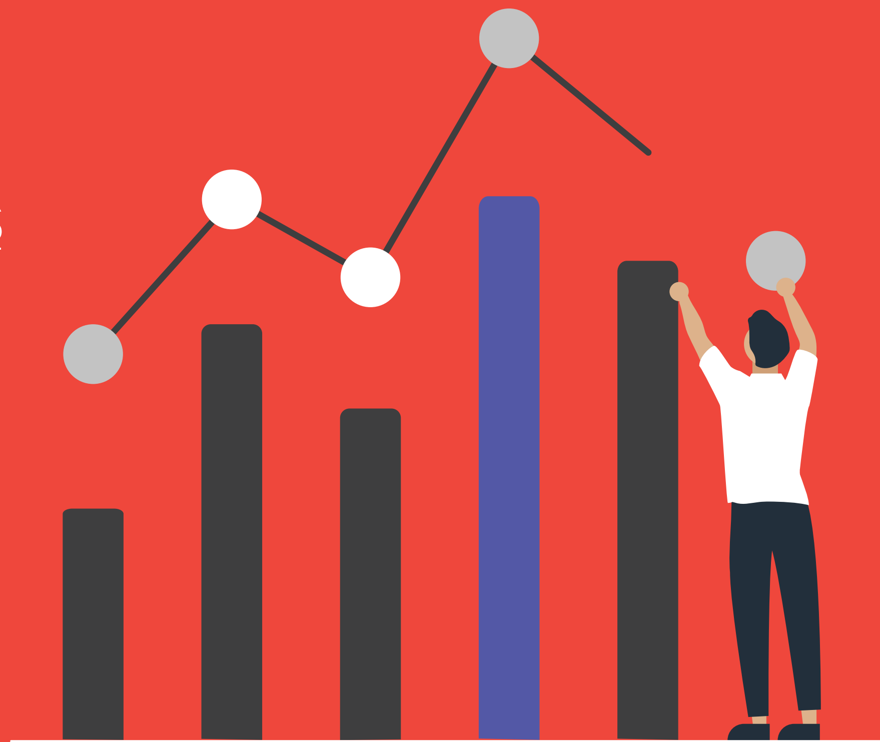
UNIVERSIDAD DE TALCA
CHILE



Estadística: medidas numéricas

MATEMÁTICAS

Vicerrectoría de Pregrado
CIVE Dirección de Ciclo de Vida Estudiantil
CIMA Unidad de Acompañamiento Estudiantil
PACE Programa de Acceso a la Educación Superior
Ministerio de Educación

RUTA DE APRENDIZAJE

El aprendizaje esperado en este documento es conocer formas númericas de describir datos cuantitativos, las medidas de tendencia central y de dispersión, para ubicar valores centrales de un grupo de datos y la variabilidad presente entre ellos.

Tipos de variable | Tablas de frecuencia | Gráficos | Medidas de tendencia central | Medidas de dispersión

ÍNDICE

Introducción

Una característica que se presenta en la recolección de datos es **la variabilidad**, por ejemplo, la estatura varía según la persona y el momento en el que se midió. Para conocer el comportamiento de la información obtenida, **se pueden establecer valores esperados** en distintos grupos y tener límites de variación habitual cuando se conoce la distribución de los datos estudiados.

Medidas de tendencia central

Tienen por objeto la obtención de un valor que resuma en sí todas las mediciones. La mayoría de ellos tratan de ubicar el centro de la distribución, de ahí su nombre (Taucher, 2014). En este documento revisarás el promedio (o media aritmética), la mediana y la moda.

Media aritmética

Se define como **la suma de los valores de todas las observaciones dividida por el número de observaciones.** Se representa por el símbolo $\bar{x}$ cuando se refiere a una muestra y por μ cuando se refiere a una población (Taucher, 2014). La fórmula general para la media puede escribirse como sigue.

- Media poblacional

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

El símbolo $\sum_{i=1}^{N} x_i$ indica que deben sumarse todos valores de la variable desde el primero hasta el último (Wayne, 1991) y N es la cantidad total de sujetos en la población.

- Media muestral

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

El símbolo $\sum_{i=1}^{n} x_i$ indica que deben sumarse todos valores de la variable de la muestra y n representa el tamaño de la muestra.
La media aritmética es una medida de ubicación muy utilizada, algunas de sus propiedades importantes son:

1. Todo conjunto de datos de intervalo, o de nivel de razón, posee una media.

2. Todos los valores se encuentran incluidos en el cálculo de la media.

3. La media es única, solo existe una media en un conjunto de datos.

4. La suma de las desviaciones de cada valor de la media es cero. Por ejemplo, la media de 3, 8 y 4 es 5, entonces:

$$\begin{aligned} \Sigma(x - \bar{x}) &= (3-5) + (8-5) + (4-5) \\ &= -2 + 3 - 1 \\ &= 0 \end{aligned}$$

De esta manera la media es un **punto de equilibrio** de un conjuto de datos (Lind et al, 2012).
A pesar de que la media es una medida muy utilizada, presenta una gran debilidad y es que **se ve afectada por datos atípicos**. Recuerda que el valor de cada elemento de una muestra, o población, se utiliza cuando se calcula la media, si uno o dos valores de estos son extremadamente grandes o pequeños comparados con la mayoría de los datos, la media podría no ser adecuada para representar los datos (Lind et al, 2012).

- Media de datos tabulados

Para aproximar la media aritmética de datos agrupados en una tabla de frecuencias es necesario el punto medio de la clase (marca de clase) y las frecuencias absolutas, se calcula de la siguiente manera:

$$\bar{x} = \frac{\Sigma f_i M}{n}$$

Donde:

$\bar{x}$ es la media muestral.
M es el punto medio de cada clase.
f_i es la frecuencia absoluta de cada clase.
n es el número de observaciones en la muestra.

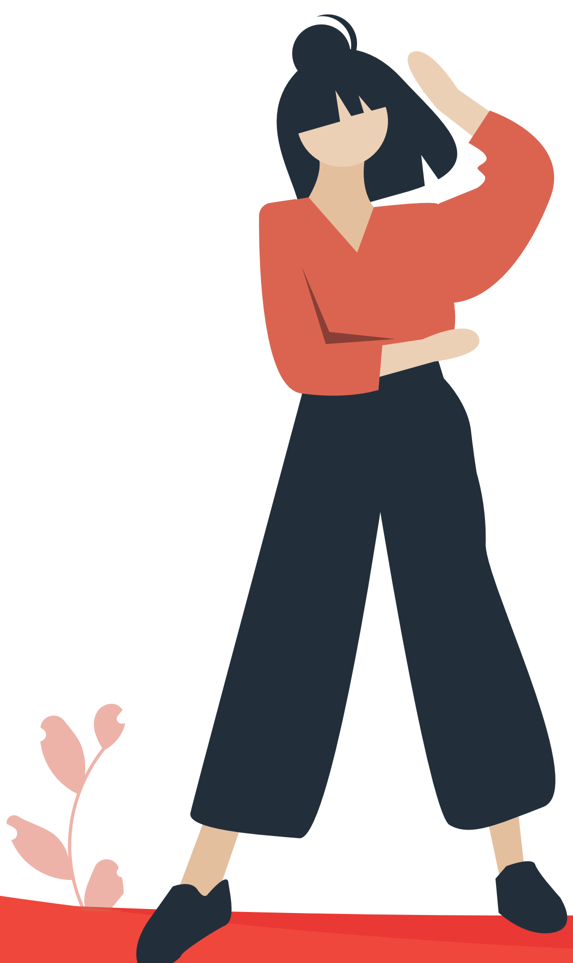
Mediana

Se define como el valor que deja igual número de observaciones (iguales o inferiores) por debajo de él, como valores iguales o superiores por encima de él, cuando los valores de la variable están ordenados según magnitud (Taucher, 2014).
Para su cálculo se debe considerar lo siguiente

- Si el número de valores es impar: la mediana será el valor que está en medio, cuando todos los valores se han arreglado en orden de magnitud.

- Si el número de valores es par: en este caso quedan dos observaciones en medio, por lo que la mediana será la media de ellas, cuando todas las observaciones se han dispuesto en el orden de su magnitud (Wayne, 1991).

La mediana tiene la ventaja de que no influyen en ella valores extremadamente grandes o pequeños.



Moda

La moda de un conjunto de valores es aquel valor que **ocurre con más frecuencia**. Si todos los valores son distintos no hay moda; por otra parte, un conjutno de valores puede tener más de una moda (Wayne, 1991). La moda es de especial utilidad para resumir datos de nivel nominal y ordinal.

La moda también tiene la ventaja de que no influyen en ella datos atípicos.

Recordando

Las medidas de tendencia central son útiles para posicionar los datos recolectados, pero al momento de interpretar los resultados, siempre es importante tener en consideración la existencia de datos extremos (atípicos) que puedan afectar la representatividad.

Medidas de dispersión

La dispersión de un conjuto de observaciones se refiere a la variedad que exhiben los valores de las observaciones. Si todos los vaores son iguales, no hay dispersión. La magnitud de la dispersión puede ser pequeña, cuando los valores, aunque distintos, están próximos entre sí. Si los valores están ampliamente desparramados, la dispersión es mayor (Wayne, 1991).

Rango

El rango o recorrido es la diferencia que existe entre el valor menor y el mayor de un conjunto de observaciones. Si se denota el recorrido por R, el valor mayor por x_L y el menor por x_S, el rango se calcula como:

$$R = x_L - x_S$$

Desviación media

La desviación media mide la cantidad media respecto de la cual los valores de una población o muestra varían. Se designa como DM y se calcula mediante la fórmula:

$$DM = \frac{\sum|x - \bar{x}|}{n}$$

Donde:

x es el valor de cada observación.
$\bar{x}$ es la media aritmética de los valores.
n es el número de observaciones en la muestra.
$||$ indica el valor absoluto.
(Lind et al, 2012)

Varianza y desviación estándar

La desviación estándar refleja la dispersión de los valores con respecto a la media, es grande cuando hay mucha dispersión y pequeña si hay poca (Taucher, 2014). Para poder calcularla, es necesario primero el calculo de la varianza, para esto, se resta la media a cada uno de los valores, se elevan al cuadrdado las diferencias y, a continuación se suman. Finalmente esta suma se divide entre el tamaño de la muestra menos 1 (Wayne, 1991). Luego, la fórmula de la variaza muestral está dada por:

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1}$$

Finalmente, para obtener la desviación estándar muestral, que representa la medida de dispersión en términos de las unidades originales, simplemente se calcula la raíz cuadrada de la varianza.

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

Recordando

Si se desea calcular la varianza y desviación estándar poblacional (denotada por σ), es necesaria la media poblacional (denotada por μ) y la división de la sumatoría debe ser por el tamaño de la población, así su fórmula está dada por:

$$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N} \quad \text{y} \quad \sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$$

- Desviación estándar de datos tabulados

Para calcular la desviación estándar de datos agrupados en una distribución de frecuencias, se utiliza la siguiente fórmula:

$$s = \sqrt{\frac{\Sigma f_i(M - \bar{x})^2}{n - 1}}$$

s es símbolo de la desviación estándar de la muestra.
$\bar{x}$ es la media muestral.
M es el punto medio de cada clase.
f_i es la frecuencia absoluta de cada clase.
n es el número de observaciones en la muestra.
(Lind et al, 2012)

PROBLEMAS RESUELTOS

A continuación, se presentan tres problemas resueltos con sus procedimientos, en estos problemas se sugiere hacer lo siguiente:

- Lee comprensivamente.
- Revisa el paso a paso.
- Destaca lo que te resulte importante.
- Destaca lo que te genere dudas y luego consulta al tutor.

1. Los siguientes valores son los niveles de glucosa en sangre extraída a 10 niños en ayunas, calcule su media, mediana, moda, rango y desviación estándar. (Ejercicio 1.5.1 página 40 de Bioestadística base para el análisis de las ciencias y la salud, Wayne, 1991)

Número	Valor
1	65
2	70
3	68
4	65
5	56

Número	Valor
6	72
7	65
8	63
9	62
10	65

a) Media
Paso 1: sumar todos los valores obtenidos

$$\bar{x} = \frac{65 + 70 + 68 + 65 + 56 + 72 + 65 + 63 + 62 + 65}{n}$$

$$\bar{x} = \frac{651}{n}$$

Paso 2: dividir por la cantidad de datos

$$\bar{x} = \frac{651}{10}$$

$$\bar{x} = 65{,}1$$

b) Mediana
Paso 1: se ordenan los datos de menor a mayor

56,62,63,65,65,65,65,68,70,72

Paso 2: ubicar el dato central

56,62,63,**65**,**65**,65,65,68,70,72

Como es una cantidad de datos par (10 datos) en el centro se ubican dos de ellos, así para obtener la mediana se calcula el promedio de los datos centrales:

$$M_e = \frac{65 + 65}{2}$$

$$M_e = 65$$

c) Moda
Se busca el dato que más se repita (puede ser más de uno)

El 65 es el número que más se repite (cuatro veces), por ende es la moda del conjunto de datos.

d) Rango

Para obtenerlo se resta el dato mayor con el menor.

R=72-56
R=16

d) Desviación estándar

Como es una muestra se utiliza la fórmula de la desviación estándar muestral:

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

Paso 1: a cada dato se le resta el promedio
Recordemos que el promedio calculado en el ítem a) es $\bar{x}$=65,1, luego la resta de cada dato con el promedio es:

Dato	$x_i - \bar{x}$
56	56-65,1= -9,1
62	62-65,1= -3,1
63	63-65,1= -2,1
65	65-65,1= -0,1
65	65-65,1= -0,1
65	65-65,1= -0,1
65	65-65,1= -0,1
68	68-65,1= 2,9
70	70-65,1= 4,9
72	72-65,1= 6,9

Paso 2: elevar al cuadrado cada resultado del paso anterior

Dato	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
56	-9,1	$(-9{,}1)^2 = 82{,}81$
62	-3,1	$(-3{,}1)^2 = 9{,}61$
63	-2,1	$(-2{,}1)^2 = 4{,}41$
65	-0,1	$(-0{,}1)^2 = 0{,}01$
65	-0,1	$(-0{,}1)^2 = 0{,}01$
65	-0,1	$(-0{,}1)^2 = 0{,}01$
65	-0,1	$(-0{,}1)^2 = 0{,}01$
68	2,9	$(2{,}9)^2 = 8{,}41$
70	4,9	$(4{,}9)^2 = 24{,}01$
72	6,9	$(6{,}9)^2 = 47{,}61$

Paso 3: sumar los resultados obtenidos en el paso anterior

Suma = 82,81+9,61+4,41+0,01+0,01+0,01+0,01+8,41+24,01+47,61
Suma = **176,9**

Paso 4: se divide por el tamaño de la muestra menos 1 (en este caso el tamaño de la muestra es 10).

$$s^2 = \frac{176{,}9}{10 - 1}$$

$$s^2 = \frac{176{,}9}{9}$$
$$s^2 = 19{,}6556$$

Recordando

En este paso obtuvimos la varianza muestral, para obtener la desviación estándar muestral necesitamos la raíz cuadrada de esta varianza.

Paso 5: se calcula la raíz cuadrada a la varianza.

$$s = \sqrt{19{,}6556}$$
$$s = 4{,}4334$$

La desviación estándar es 4,4334 este número indica la dispersión de los datos respecto a la media, mientras más pequeña sea la desviación más representativo de los datos es el promedio.

2. Los datos agrupados en una distribución de frecuencias que aparecen en seguida se basan en los datos de las ganancias de Applewood Auto Group. Determine la ganancia media por vehículo (Ejemplo 3.1 página 88 de Estadística aplicada a la economía y a los negocios; Lind, Wathen, Marchal, 2012)

Ganancia ($)	**Frecuencia**
[200-600[	8
[600-1000[	11
[1000-1400[	23
[1400- 1800[	38
[1800-2200[	45
[2200-2600[	32
[2600-3000[	19
[3000-3400]	4
Total	180

Recordemos que la fórmula de promedio para datos agrupados es:

$$\bar{x} = \frac{\Sigma f_i M}{n}$$

Paso 1: calcular la ganancia media (promedio) de cada intervalo (marca de clase de cada intervalo)

Se suman lo límites de cada intervalo y se dividen en dos, por ejemplo en el primer intervalo se tiene:

$$Ganancia\ media = \frac{200 + 600}{2}$$

$$Ganancia\ media = \frac{800}{2} = 400$$

En los otros intervalos tenemos:

Ganancia ($)	**Frecuencia**	**Ganancia Media (M)**
[200-600[	8	400
[600-1000[	11	800
[1000-1400[	23	1200
[1400- 1800[	38	1600
[1800-2200[	45	2000
[2200-2600[	32	2400
[2600-3000[	19	2800
[3000-3400]	4	3200
Total	180	

Paso 2: multiplicar cada ganancia media con su frecuencia absoluta.

Ganancia ($)	**Frecuencia**	**Ganancia Media (M)**	$\boldsymbol{M \cdot f}$
[200-600[	8	400	$400 \cdot 8 =$3200
[600-1000[	11	800	$800 \cdot 11 =$8800
[1000-1400[	23	1200	$1200 \cdot 23 =$27600
[1400- 1800[	38	1600	$1600 \cdot 38 =$60800
[1800-2200[	45	2000	$2000 \cdot 45 =$90000
[2200-2600[	32	2400	$2400 \cdot 32 =$76800
[2600-3000[	19	2800	$2800 \cdot 19 =$53200
[3000-3400]	4	3200	$3200 \cdot 4 = 12800$
Total	180		

Paso 3: sumar todos los resultados obtenidos al multiplicar la ganancia media con su frecuencia.

Ganancia ($)	Frecuencia	Ganancia Media (M)	$M \quad f$
[200-600[	8	400	3200
[600-1000[	11	800	8800
[1000-1400[	23	1200	27600
[1400- 1800[	38	1600	60800
[1800-2200[	45	2000	90000
[2200-2600[	32	2400	76800
[2600-3000[	19	2800	53200
[3000-3400]	4	3200	12800
Total	180		333200

Paso 4: se divide el total obtenido en el paso 3 por el total de frecuencias absolutas.

$$\bar{x} = \frac{333200}{180}$$

$$\bar{x} = 1851{,}11$$

Luego la media aritmética de la ganancia por vehículo es $1851.

3. Calcule la desviación estándar de la tabla de frecuencias de ganancias de vehículos del ejercicio anterior.

Recordemos que la fórmula de desviación estándar para datos agrupados es:

$$s = \sqrt{\frac{\Sigma f_i (M - \bar{x})^2}{n - 1}}$$

Paso 1: restar a la ganancia media de cada intervalo el promedio obtenido en el ejercicio 2 ($1851).

Ganancia ($)	Frecuencia	Ganancia Media (M)	$M - \bar{x}$
[200-600[	8	400	$400 - 1851 = -1451$
[600-1000[	11	800	$800 - 1851 = -1051$
[1000-1400[	23	1200	$1200 - 1851 = -651$
[1400- 1800[	38	1600	$1600 - 1851 = -251$
[1800-2200[	45	2000	$2000 - 1851 = 149$
[2200-2600[	32	2400	$2400 - 1851 = 549$
[2600-3000[	19	2800	$2800 - 1851 = 949$
[3000-3400]	4	3200	$3200 - 1851 = 1349$
Total	180		

Paso 2: elevar al cuadrado los resultados obtenidos en el paso 1.

Ganancia ($)	Frecuencia	Ganancia Media (M)	$M - \bar{x}$	$(M - \bar{x})^2$
[200-600[	8	400	-1451	$(-1451)^2 = 2105401$
[600-1000[	11	800	-1051	$(-1051)^2 = 1104601$
[1000-1400[	23	1200	-651	$(-651)^2 = 423801$
[1400- 1800[	38	1600	-251	$(-251)^2 = 63001$
[1800-2200[	45	2000	149	$(149)^2 = 22210$
[2200-2600[	32	2400	549	$(549)^2 = 301401$
[2600-3000[	19	2800	949	$(949)^2 = 900601$
[3000-3400]	4	3200	1349	$(1349)^2 = 1819801$
Total	180			

Paso 3: multiplicar los resultados obtenidos en el paso 2 con la frecuencia absoluta de cada intervalo.

Ganancia ($)	Frecuencia	Ganancia Media (M)	$M - \bar{x}$	$(M - \bar{x})^2$	$f \ (M \cdot \bar{x})^2$
[200-600[	8	400	-1451	2105401	$8 \cdot 2105401 =$16 843208
[600-1000[	11	800	-1051	1104601	$11 \cdot 1104601 =$12150611
[1000-1400[	23	1200	-651	423801	$23 \cdot 423801 =$9747423
[1400- 1800[	38	1600	-251	63001	$38 \cdot 63001 = 2394038$
[1800-2200[	45	2000	149	22210	$45 \cdot 22210 = 999045$
[2200-2600[	32	2400	549	301401	$32 \cdot 301401 = 9644832$
[2600-3000[	19	2800	949	900601	$19 \cdot 900601 = 17111419$
[3000-3400]	4	3200	1349	1819801	$4 \cdot 1819801 = 7279204$
Total	180				

Paso 4: sumar los resultados obtenidos en el paso 3.

Ganancia ($)	Frecuencia	Ganancia Media (M)	$M - \bar{x}$	$(M - \bar{x})^2$	$f \ (M \cdot \bar{x})^2$
[200-600[	8	400	-1451	2105401	$16\,843208$
[600-1000[	11	800	-1051	1104601	12150611
[1000-1400[	23	1200	-651	423801	9747423
[1400- 1800[	38	1600	-251	63001	2394038
[1800-2200[	45	2000	149	22210	999045
[2200-2600[	32	2400	549	301401	9644832
[2600-3000[	19	2800	949	900601	17111419
[3000-3400]	4	3200	1349	1819801	7279204
Total	180				76169780

Paso 5: dividir el resultado del paso anterior por el total de frecuencia menos 1.

$$s^2 = \frac{76169780}{180 - 1}$$
$$s^2 = \frac{76169780}{179}$$
$$s^2 = 425529{,}497$$

En este paso obtuvimos la varianza muestral.

Paso 6: calcular la raíz cuadrada del resultado obtenido en el paso anterior.

$$s = \sqrt{425529{,}497}$$
$$s = 652{,}33$$

Luego la desviación estándar es 652,33.

PROBLEMAS PROPUESTOS

A continuación, se presentan dos problemas propuestos para que puedas resolver y practicar, recuerda hacer lo siguiente:

- Resuélvelos siguiendo los pasos utilizados en los problemas resueltos.
- Si es necesario apóyate con los apuntes.
- Si surgen dudas, registrarlas para luego consultar con el tutor.
- ¡Buen trabajo!

1. Quince pacientes que realizaron visitas iniciales a un departamento sanitario municipal recorrieron las siguientes distancias:

Paciente	Distancia (millas)
1	5
2	9
3	11
4	3
5	12
6	13
7	12

Paciente	Distancia (millas)
8	6
9	13
10	7
11	3
12	15
13	12
14	15
15	5

(Wayne, 1991)

Calcula:

a) Media aritmética
b) Mediana
c) Moda
d) Rango
e) Desviación estándar

2. SCCoast, un proveedor de internet del sureste de Estados Unidos, elaboró una distribución de frecuencias sobre la edad de los usuarios de internet.

Determina
a) La media
b) La desviación estándar.

Edad (años)	Frecuencia
[10-20[	3
[20-30[	7
[30-40[	18
[40-50[	20
[50-60]	12

(Lind et al, 2012)

RESPUESTAS

1.
a) 9,4
b) 11
c) 12
d) 10
e) 4,22

2.
a) 40,167
b) 10,97

SÍNTESIS

Al revisar este documento es importante que recuerdes:

MEDIA ARITMÉTICA	Punto de equilibrio de un conjunto de datos.
MEDIANA	Punto medio de los valores una vez ordenados.
MODA	Valor de la observación que aparece con mayor frecuencia.
RANGO	Diferencia entre el valor mayor y el valor menor.
VARIANZA	Media aritmética de las desviaciones de la media elevadas al cuadrado.
DESVIACIÓN ESTÁNDAR	Medida de dispersión de los datos a la media.

BIBLIOGRAFÍA

Lind, D., Wathen, S., & Marchal , W. (2012). *Estadística aplicada a los negocios y la economía*. México: The McGraw-Hill Companies, Inc.

Taucher, E. (2014). *Bioestadística* . Ocho Libros Editores Ltda.

Wayne, D. (1991). *Bioestadística base para el análisis de las ciencias y la salud*. México : Limusa S.A. .

¿Quieres recibir orientación para optimizar tu estudio en la universidad?

CONTAMOS CON PROFESIONALES EXPERTOS EN
EL APRENDIZAJE QUE TE PUEDEN ORIENTAR

SOLICITA NUESTRO APOYO

Sitio Web de CIMA

Ver más fichas

Solicita más información

UNIVERSIDAD TALCA CHILE

CIMA UNIDAD DE ACOMPAÑAMIENTO ESTUDIANTIL