



**FACULTAD DE BIOLOGÍA**

**UNIVERSIDAD MICHOACANA  
DE SAN NICOLÁS DE  
HIDALGO**

**FUNCIONES Y ESTADÍSTICA**

**DESCRIPTIVA**

## Actividades prácticas para reforzar el aprendizaje del temario de asignaturas teóricas

### I. DERIVACIÓN DE FUNCIONES

#### Introducción

Un modelo matemático es una descripción matemática de un fenómeno del mundo real. Por ejemplo, el crecimiento de las poblaciones de animales, la concentración de un producto en una reacción química, el funcionamiento de las neuronas y la dinámica intracelular. Los modelos matemáticos representan un problema médico o biológico de una manera objetiva. Definen una serie de relaciones matemáticas entre las mediciones cuantitativas (del problema) y sus propiedades. El modelado matemático permite a los biólogos comprender la complejidad de los sistemas biológicos, hacer predicciones y realizar experimentos virtuales que serían difíciles o imposibles de realizar en la vida real.

#### Objetivo de la actividad

Calcular e interpretar la relación y el grado de asociación entre variables, para que el alumno pueda aplicarlo a problemas biológicos.

#### Material

- Notas de clase
- Base de datos nicotina.csv
- Base de datos grillos.csv
- Software para el análisis de datos (Excel, Google Sheets, Python, R, etc)

#### Actividades

Con base a las notas vistas en clase resolver las siguientes actividades. Recuerda que las bases de datos que se deberán utilizar, así como las notas de clase las podrás encontrar en la carpeta [PRÁCTICA 2](#).

1. La cantidad de aguacates (en toneladas) producidas en una huerta es proporcional al número

de árboles en el huerto. Si el huerto consta de 255 árboles y produjo 145 toneladas.

- Calcula la constante de proporcionalidad. ¿Qué unidades tiene la constante de proporcionalidad?
- Expresa la producción  $P$  como función del número de árboles  $x$ .
- Elabora la tabla correspondiente para  $x = 0, 50, 100, 150, 200$ .
- Esboza la gráfica de  $P(x)$ .

2. La función de costo,  $C(x)$ , describe el costo de producción en (miles de pesos) en una granja de truchas para producir  $x$  toneladas de trucha al año. Se determinó en un estudio que dicha función tenía la forma

$$C(x) = -0.005x^2 + 12.5x + 75$$

- Evalúa el costo cada 250 toneladas, es decir en  $x = 0, 250, \dots, 1500$
- Esboza la gráfica de la función costo  $C(x)$  en el intervalo  $[0, 1500]$ .
- ¿Aproximadamente cuántas toneladas tiene el costo máximo? ¿Aproximadamente cuál es el costo máximo?

3. Durante el siglo XVII las ballenas del Atlántico Norte casi se extinguieron por causa del exterminio de la industria pesquera Europea. En un periodo de 50 años, la población descendió de 250,000 ejemplares a sólo 2000. Suponiendo una función lineal de decremento de la población.

- Determina la función que describe la población en relación del tiempo transcurrido en años.
- ¿Aproximadamente cuántas ballenas desaparecían cada año durante ese periodo?
- ¿Después de cuánto tiempo la cantidad de ballenas se había reducido a la mitad?

Aunque aún existen en nuestros días las ballenas en el Atlántico Norte, éstas están condenadas a la extinción.

4. En cierto modelo, la tasa de crecimiento de la biomasa es  $f(x)$  y es función de la biomasa  $x$  de un recurso natural  $f(x)$  está dada por

$$f(x) = rx \left(1 - \frac{x}{K}\right)$$

donde  $r$  es la tasa intrínseca de crecimiento;  $K$  es el nivel de saturación. Para una población de truchas  $K = 5000$ ,  $r = 0.05$ .

- (a) Esboza la gráfica de  $f(x)$  utilizando los valores  $x = 1000, 1500, \dots, 3500, 4000$ .
- (b) ¿En qué intervalo de  $1000 < x < 4000$  es creciente la función  $f(x)$ ? ¿dónde es decreciente?
- (c) Para qué valor de  $x$ ,  $f(x)$  alcanza un valor máximo? ¿Cuál es el valor máximo de  $f(x)$ ?
- (d) Para qué valores de  $x$ , la tasa de crecimiento  $f'(x)$  es 0.

5. Una población de golondrinas que arriba a una comunidad se incrementó de 5600 ejemplares a 5630 al cabo de un año. Suponiendo un periodo de reproducción anual a tasa porcentual anual constante. Determina dicha tasa  $r$ .

6. Durante la revolución mexicana la población en México disminuyó de 15 millones en 1910, a 13.9 millones de habitantes en 1920. Suponiendo un modelo exponencial anual  $P(t) = P_0(1 - r)^t$ , donde  $t$  son los años transcurridos desde 1910 y  $P_0$  es la población de 1910 y  $r$  es la tasa porcentual anual de decrecimiento.

a) Determina la tasa porcentual anual de decrecimiento,  $r$ .

b) Esboza la gráfica de la función usando  $t = 0, 1, 2, \dots, 10$ .

8. Si la población de un país crece de 16 millones a 18.5 millones en 10 años. Calcula la tasa de crecimiento en el modelo exponencial  $P = P_0e^{rt}$ .

9. Calcula la porción de material radioactivo presente en una muestra de 10 gr de  $Rn_{86}^{222}$  después de 240 horas.

10. Un fósil tiene 30 % del carbono 14 que contenía originalmente. Calcula la edad del fósil.

10. La cantidad de bacterias en un cultivo crece a razón media constante de 1000 bacterias por día. Si después de 4 días el cultivo tiene 20,000 bacterias ¿Cuántas bacterias había al inicio del experimento?

11. Define los siguientes conceptos

a) Correlación

b) Diagrama de dispersión

c) Coeficiente de correlación lineal

d) Coeficiente de correlación producto momento de Pearson

12. Se realizó un estudio clínico para investigar la eficacia del fármaco Dozenol para tratar el insomnio. Se descubrió que existe una correlación entre la cantidad ingerida de Dozenol y la duración del sueño. Con base en este análisis estadístico, ¿podemos concluir que el Dozenol es la causa del sueño? ¿Por qué?

13. Mientras diez osos se encontraban anestesiados, unos investigadores midieron las distancias (en pulgadas) alrededor del pecho de los osos y los pesaron (en libras). Se utilizó Minitab para calcular el valor del coeficiente de correlación lineal, que resultó ser  $r = 0.993$ .

a) ¿Existe una correlación lineal significativa entre tamaño del pecho y el peso? Explique.

b) ¿Qué proporción de la variación del peso puede explicarse por la relación lineal entre el peso y el tamaño del pecho?

14. Cuando la nicotina es absorbida por el cuerpo, se produce cotinina. Por consiguiente, la medición de cotinina es un buen indicador de cuánto fuma una persona. A continuación, se incluye el reporte del número de cigarrillos fumados por día y las cantidades medidas de nicotina (en ng/mL). (Los valores provienen de sujetos seleccionados al azar de la National Health Examination Survey). Utilizando la base de datos nicotina.csv de la carpeta Funciones\_Estadística:

a) Crear una gráfica de dispersión

b) Usar un ajuste lineal para los datos. ¿Cuál es la ecuación de ajuste?

c) Realizar una gráfica con el ajuste lineal correspondiente

d) ¿Existe una correlación lineal? Explique el resultado.

15. Una aplicación clásica de la correlación es la asociación entre la temperatura y el número de veces que un grillo chirría en un minuto. A continuación se indican los números de chirridos en un minuto y las temperaturas correspondientes en grados Fahrenheit (según datos de The Song of Insects, de George W. Pierce, Harvard University Press). Utilizando la base de datos grillos.csv de la carpeta Funciones\_Estadística:

a) Crear una gráfica de dispersión

- b) Usar un ajuste lineal para los datos. ¿Cuál es la ecuación de ajuste?
- c) Realizar una gráfica con el ajuste lineal correspondiente
- d) ¿Existe evidencia suficiente para concluir que existe una relación entre el número de chirridos en un minuto y la temperatura?

## **BIBLIOGRAFÍA**

- Stewart, J., REDLIN, L., & WATSON, S. (2010). Precálculo. Matemáticas para el cálculo. Cengage Learning Editores, SA.
- Stewart, J., & Romo, J. H. (2017). cálculo. Cengage Learning.

## **RÚBRICA DE EVALUACIÓN**

<b>Criterios</b>	<b>Puntaje</b>
Redacción y ortografía	10
Metodología en el análisis de datos	60
Interpretación y presentación de resultados	30

## **FORMA DE ENTREGA**

La presente práctica deberá entregarse mediante Classroom.

## II. DERIVADAS E INTEGRALES

### Introducción

Las derivadas e integrales son conceptos fundamentales en la biología y la medicina. Las derivadas se utilizan para estudiar fenómenos biológicos como el crecimiento celular, la propagación de enfermedades y la dinámica de poblaciones. También son útiles para analizar funciones fisiológicas, como el ritmo. Las integrales se utilizan para obtener la distancia recorrida en función de la velocidad y aceleración. En biología, las integrales definidas se aplican para calcular la probabilidad de extinción de una especie animal. Las derivadas son la pendiente de una función, mientras que las integrales son el área bajo la curva. Las derivadas pueden proporcionar un valor instantáneo preciso de la tasa de cambio.

### Objetivo de la actividad

Plantear hipótesis biológicas en términos de modelos matemáticos que involucran funciones y que el alumno pueda aplicarlo a la resolución de problemas biológicos.

### Material

- Notas de clase
- Tabla de derivación
- Tabla de integración
- Wolfram Alpha

### Actividades

Con base a las notas vistas en clase resolver las siguientes actividades. Recuerda que puedes comprobar tus resultados usando <https://www.wolframalpha.com/>. Recuerda que las tablas de derivación e integración, así como las notas que se deberán utilizar las podrás encontrar en la carpeta [PRÁCTICA 3](#).

1. Una población de bacterias crece de acuerdo a la ley de evolución  $P(t) = 3000 - 600t + 30t^2$ . Donde  $t$  es el número de horas transcurridas desde iniciado el experimento. Mediante

aproximaciones sucesivas estima la tasa instantánea de variación para  $t_0 = 5, 10, 15$ .

2. La concentración  $c(t)$  (en mg por cc) de cierta droga en el flujo sanguíneo a tiempo  $t$  (en min), está determinada por  $c(t) = 0.8 + 0.72t - 0.9t^2$ .

a) Calcula la derivada  $c'(t)$ . Indica cuáles son sus unidades.

b) Calcula la rapidez con la que cambia la concentración de droga medio minuto después de administrada.

c) Calcula la rapidez con la que cambia la concentración de droga un minuto después de administrada.

d) Evalúa  $c(t)$  y  $c'(t)$  en  $t = 0, 0.1, 0.2, \dots, 0.9, 1.0$ . Esboza la gráfica de  $c(t)$  y de  $c'(t)$ .

e) Calcula el tiempo  $t_0$  para el cual  $c'(t_0) = 0$ .

f) Indica en qué intervalo dentro de  $0 \leq t \leq 1$ , la concentración es creciente. En cuál es decreciente.

3. La cantidad de contaminantes  $Q$  (en  $\text{km}^3$ ) presente en un lago  $t$  años después de que se vierten desechos en él, decae de manera exponencial, de acuerdo a la fórmula

$$Q(t) = Q_0 e^{-rt}$$

donde  $Q_0$  es la cantidad original de desechos vertidos y  $r$  es una constante de degradación.

a) Si 5 años después de vertidos  $2 \text{ km}^3$  de desechos en un lago, estos se degradan a  $1.5 \text{ km}^3$ . Calcula la constante  $r$ .

b) Calcula la razón instantánea de decrecimiento 10 años después de vertidos  $2 \text{ km}^3$  de desechos en un lago.

c) Calcula cuánto tiempo tomará remover 90% de la contaminación del lago. Es decir, después de cuánto tiempo permanecerán 10% de los desechos aún sin degradar.

d) Calcula después de cuánto tiempo se habrá removido el 99% de la contaminación.

e) Determina que pasa con la contaminación para  $t$  muy grande.

4. En un lago la población de peces es  $P$ ,  $H$  es un parámetro conocido como la razón de pesca, es decir la cantidad de peces que se extraen por unidad de tiempo (peces/año). La tasa de crecimiento de la población de peces como función de la población está dada por

$$R(P) = 2P - 0.02P^2 - H$$

Para cada uno de los siguientes valores del parámetro  $H = 75, 100, 200$  esboza la gráfica de  $R(P)$  contra  $P$ . Encuentra el valor de  $P$  para el cuál la tasa de crecimiento es máxima.

5. La tasa de crecimiento de la biomasa en un lago es  $f(x) = 2.5x + 2$  toneladas por día, donde  $x$  son los días transcurridos. Calcula la biomasa acumulada en el lago desde inicio del día 3 hasta el final del día 10.

6. Un río lleva un caudal constante de 20 litros por segundo. Una presa capta todo ese caudal. Calcula el agua acumulada en la presa durante 10 días. Suponiendo que el río experimenta un crecimiento lineal del caudal en época de lluvias dado por  $v(x) = 20 + .0005x$ , donde  $x$  son los días transcurridos desde el inicio de la época de lluvias. Calcula la cantidad de agua acumulada en la presa durante los primeros 10 días de la época de lluvias.

7. Una colonia de bacterias originalmente es de 20000 bacterias y crece con una rapidez de  $50x$  bacterias cada hora, donde  $x$  son las horas transcurridas desde el inicio de las observaciones. Calcula la cantidad acumulada de células durante una hora.

8. Una población crece a una tasa constante de 1300 individuos por año. Si originalmente había 50000 individuos. ¿Cuántos individuos habrá dentro de 15 años?

9. Una población de bacterias se reproduce a una tasa de  $50e^{.05t}$  bacterias/hora, donde  $t$  son las horas transcurridas desde el inicio del experimento. Calcula el incremento de la población durante 24 horas.

## BIBLIOGRAFÍA

- Stewart, J., REDLIN, L., & WATSON, S. (2010). Precálculo. Matemáticas para el cálculo. Cengage Learning Editores, SA.
- Stewart, J., & Romo, J. H. (2017). cálculo. Cengage Learning.

## RÚBRICA DE EVALUACIÓN

<b>Criterios</b>	<b>Puntaje</b>
Redacción y ortografía	10
Metodología en el análisis de datos	60
Interpretación y presentación de resultados	30

### **FORMA DE ENTREGA**

La presente práctica deberá entregarse mediante Classroom.

### III.INTEGRACIÓN

- a. La tasa de crecimiento de la biomasa en un lago es  $f(x) = 2.5x + 2$  toneladas por día, donde  $x$  son los días transcurridos. Calcula la biomasa acumulada en el lago desde inicio del día 3 hasta el final del día 10.
- b. Un río lleva un caudal constante de 20 litros por segundo. Una presa capta todo ese caudal. Calcula el agua acumulada en la presa durante 10 días. Suponiendo que el río experimenta un crecimiento lineal del caudal en época de lluvias dado por  $v(x) = 20 + .0005x$ , donde  $x$  son los días transcurridos desde el inicio de la época de lluvias. Calcula la cantidad de agua acumulada en la presa durante los primeros 10 días de la época de lluvias.
- c. Una colonia de bacterias originalmente es de 20000 bacterias y crece con una rapidez de  $50x$  bacterias cada hora, donde  $x$  son las horas transcurridas desde el inicio de las observaciones.
- d. Calcula la cantidad acumulada de células durante una hora.
- e. Una población crece a una tasa constante de 1300 individuos por año. Si originalmente había 50000 individuos. ¿Cuántos individuos habrá dentro de 15 años?
- f. Una población de bacterias se reproduce a una tasa de  $50e^{.05t}$  bacterias/hora, donde  $t$  son las horas transcurridas desde el inicio del experimento. Calcula el incremento de la población durante 24 horas.

## IV. CONCEPTOS Y TERMINOLOGÍA BÁSICA

### TEMA: CONCEPTOS Y TERMINOLOGÍA BÁSICA

#### Introducción

La bioestadística es una disciplina científica que emplea los diferentes métodos de análisis de la estadística para abordar los objetos de estudio o los problemas de la biología y de la salud para así obtener datos importantes y poder representarlos e interpretarlos. La bioestadística se aplica en todas las áreas de las ciencias que se dedican al estudio de la vida y la salud, como la biología, la medicina, la farmacia o las ciencias agropecuarias. La bioestadística se divide en dos grandes ramas: la bioestadística descriptiva y la bioestadística analítica o inferencial.

#### Objetivo de la actividad

Adquirir el lenguaje formal básico de la estadística, que permita comprender conceptos para abordar la resolución de problemas biológicos.

#### Material

- Notas de clase

#### Actividades

Con base a las notas vistas en clase resolver las siguientes actividades. Recuerda que las notas de clase las podrás encontrar en la carpeta [PRÁCTICA 1](#).

1. Define los siguientes conceptos
- |                        |                                    |
|------------------------|------------------------------------|
| a) Datos               | h) Datos discretos                 |
| b) Estadística         | i) Datos continuos                 |
| c) Población           | j) Nivel de medición nominal       |
| d) Censo               | k) Nivel de medición ordinal       |
| e) Parámetro           | l) Nivel de medición de intervalo  |
| f) Estadístico         | m) Nivel de medición de razón      |
| g) Datos cuantitativos | n) Muestra de respuesta voluntaria |

2. Describe cada uno de los siguientes tipos de muestreo y da un ejemplo.

- a) Aleatorio simple
- b) Estratificado
- c) Polietápico
- d) Sistemático

3. ¿Qué es una muestra de respuestas voluntarias, y por qué generalmente no es adecuada para los métodos estadísticos?

4. El New York Times publicó un artículo que incluía la siguiente afirmación: “Por fin, el chocolate ocupa el lugar que merece en la pirámide alimenticia, junto a sus vecinos de clase alta: el vino tinto, las frutas, los vegetales y el té verde. Varios estudios, reportados en el Journal of Nutrition revelaron que, después de comer chocolates, los sujetos a prueba incrementaron los niveles de antioxidantes en su sangre. El chocolate contiene flavonoides, antioxidantes asociados con la disminución del riesgo de enfermedades cardíacas y embolias. Mars, Inc., la empresa de dulces, y la Chocolate Manufacturers Association financiaron gran parte de la investigación”. ¿Qué es incorrecto en este estudio?

5. En un estudio del fármaco Lipitor contra el colesterol, a 270 pacientes se les dio un placebo, y 19 de esos 270 pacientes reportaron dolor de cabeza. ¿Qué porcentaje de este grupo placebo reportó dolor de cabeza? De los 270 pacientes del grupo placebo, el 3.0% reportó dolor de espalda. ¿Cuál es el número real de pacientes que reportó dolor de espalda?

6. Determine si la descripción dada corresponde a un estudio observacional o a un experimento.

a) Emily Rosa, de 9 años de edad, se convirtió en la autora de un artículo en el Journal of the American Medical Association, después de poner a prueba a terapeutas de contacto profesionales. Usando una mampara de cartón, ella colocaba la mano encima de la mano del terapeuta, quien debía de identificar la mano que Emily había elegido.

b) Tratamiento contra la sífilis. Ha surgido una gran controversia en torno del estudio de pacientes con sífilis que no recibieron un tratamiento que los habría curado. Su salud fue vigilada

por años después de que se descubrió que padecían esa enfermedad.

c) Control de calidad. La Food and Drug Administration de Estados Unidos elige al azar una muestra de grageas de aspirina Bayer, y mide la exactitud de la cantidad de aspirina en cada gragea.

d) Brazaletes magnéticos. A los pasajeros de un barco de crucero se les dan brazaletes magnéticos, que aceptan usar en un intento por disminuir o eliminar los efectos del mareo.

7. En el artículo “Cardiovascular Effects of Intravenous Triiodothyronine in Patients Undergoing Coronary Artery Bypass Graft Surgery” (Journal of the American Medical Association, vol. 275, núm. 9), los autores explican que los pacientes fueron asignados a uno de tres grupos: 1. un grupo tratado con triyodotironina, 2. un grupo tratado con una píldora de sal normal y dopamina, y 3. un grupo placebo al que se le dio una píldora de sal normal. Los autores resumen el diseño muestral como un “experimento prospectivo, aleatorio, a ciegas doble, placebo y controlado”. Describa el significado de cada uno de estos términos en el contexto de este estudio.

8. SEMARNAT le ha comisionado a usted para muestrear las plantas que crecen alrededor de todo el río Chiquito de Morelia. Describa procedimientos para obtener una muestra de cada tipo: aleatoria, sistemática, de conveniencia, estratificada, por conglomerados.

9. Usted planea realizar un experimento para probar la eficacia del Sleepeze, un nuevo fármaco que se supone que reducirá el efecto del insomnio. Usted usará una muestra de sujetos que han sido tratados con el fármaco y otra muestra de sujetos a los que se les dio un placebo.

a) ¿Qué es el “estudio a ciegas” y cómo se utilizaría en este experimento?

b) ¿Por qué es importante el uso del estudio a ciegas en este experimento?

c) ¿Qué es un diseño completamente aleatorio?

d) ¿Qué es un diseño rigurosamente controlado?

e) ¿Qué es la réplica y por qué es importante?

10. Si un experimento produce datos que son de naturaleza continua, ¿los datos también deben ser cuantitativos o pueden ser cualitativos?

## **BIBLIOGRAFÍA**

- Triola, M. F. (2004). Probabilidad y estadística. Pearson educación.
- Wayne W. Daniel. "Bioestadística base para el análisis de las ciencias de la salud". Limusa. México. 2005.
- Sokal R. y Rohlf F. J. "Biometría Principios y Métodos Estadísticos en la investigación biológica". H Blume ediciones. España. 1979. 832 págs.
- Spiegel Murray R. "Estadística Teoría y Problemas Resueltos". McGraw-Hill. México. 1979.

## **RÚBRICA DE EVALUACIÓN**

<b>Criterios</b>	<b>Puntaje</b>
Redacción y ortografía	10
Metodología en el análisis de datos	60
Interpretación y presentación de resultados	30

## **FORMA DE ENTREGA**

La presente práctica deberá entregarse mediante Classroom.

## V. ORGANIZACIÓN Y PRESENTACIÓN DE LOS DATOS

### Introducción

El análisis estadístico de datos biológicos es una necesidad para que las conclusiones sean aceptadas por la comunidad científica. Los gráficos estadísticos son potentes herramientas para la visualización de datos que permiten representar de manera accesible información compleja. Consiguen presentar la información al usuario o lector de manera clara y precisa, facilitando la comparación y la comprensión de la evolución de distintas variables.

### Objetivo de la actividad

Aprender a elaborar, analizar e interpretar tablas y gráficos, para presentar resultados de problemas biológicos.

### Material

- Notas de clase
- Base de datos basura.csv
- Base de datos geiser.csv
- Software para el análisis de datos (Excel, Google Sheets, Python, R, etc)

### Actividades

Con base a las notas vistas en clase resolver las siguientes actividades. Recuerda que las bases de datos que se deberán utilizar, así como las notas de clase las podrás encontrar en la carpeta [PRÁCTICA 4](#).

1. Define los siguientes concepto

- |   |  |
|---|--|
| a) Metadatos  | f) Marcas de clase                     |
| b) Distribución de frecuencias (o tabla de frecuencias) | g) Anchura de clase                    |
| c) Límites de clase inferiores                          | h) Histograma                          |
| d) Límites de clase superiores                          | i) Gráfica de ojiva                    |
| e) Fronteras de clase                                   | j) Histograma de frecuencias relativas |

2. ¿Qué utilidad tiene para los biólogos y biólogas utilizar datos agrupados? ¿Para qué es útil para los biólogos y biólogas usar un histograma para la visualización de datos?
3. Describe el proceso para crear una tabla de frecuencias.
4. Describe el proceso para crear un histograma.
5. Utilizando el archivo basura.csv que se encuentra en la carpeta Funciones\_Estadistica, construya una tabla de datos agrupados para la columna de Plástico (debe incluir las frecuencias, frecuencias relativas, las frecuencias porcentuales y las frecuencias acumuladas), un histograma, un polígono de frecuencias, una gráfica de ojiva y una gráfica de tallo y hojas además especifique lo siguiente:
  - a) Indique el número de clases
  - b) Indique los límites superiores e inferiores
  - c) Indique las marcas de clase
  - d) Interprete la tabla de datos agrupados
  - e) Interprete el histograma, el polígono de frecuencias, la gráfica de ojiva y una gráfica de tallo y hojas. Los pesos indicados en la base se encuentran en libras.
6. Mediante el uso de un histograma, ¿qué criterio se puede utilizar para determinar si los datos tienen una distribución aproximadamente normal? ¿Este criterio es totalmente objetivo o implica un juicio subjetivo.
7. ¿Cuál sería una explicación razonable para que exista un hueco dentro de un histograma?
8. Con base al archivo .csv que se encuentra disponible en la carpeta, crear una gráfica de barras y una de pastel. ¿Cuál gráfica es más apropiada para representar esta información? ¿Por qué?
9. Con base a las alturas de las erupciones del géiser Old Faithful recopiladas en el archivo geiser.csv en la carpeta Funciones\_Estadistica, crear:

- a) Utilice las alturas para construir una gráfica de tallo y hojas. ¿Qué sugiere esta gráfica acerca de la distribución de las alturas?
- b) Construya una gráfica de puntos con las alturas. ¿Qué sugiere esta gráfica acerca de la distribución de las alturas?
- c) Utilice las alturas para construir una ojiva. ¿Cuántas erupciones tuvieron una altura menor de 120 pies?

## **BIBLIOGRAFÍA**

- Reyes Pedro. "Bioestadística". Limusa. México. 2005.
- Triola, M. F. (2004). Probabilidad y estadística. Pearson educación.
- Wayne W. Daniel. "Bioestadística base para el análisis de las ciencias de la salud". Limusa. México. 2005.

## **RÚBRICA DE EVALUACIÓN**

<b>Criterios</b>	<b>Puntaje</b>
Redacción y ortografía	10
Metodología en el análisis de datos	60
Interpretación y presentación de resultados	30

## **FORMA DE ENTREGA**

La presente práctica deberá entregarse mediante Classroom.

## VI. PROCESO DE REDUCCIÓN DE LA INFORMACIÓN

### Introducción

La reducción de datos se puede llevar a cabo de múltiples formas: selección y extracción de características, discretización, selección y generación de instancias, entre otras. La estadística descriptiva sustituye o reduce el conjunto de datos obtenidos por un pequeño número de valores descriptivos, como el promedio, la mediana, la media geométrica, la varianza y la desviación típica. Actualmente, existen diferentes lenguajes de programación y softwares científicos que permiten analizar grandes cantidades de datos.

### Objetivo de la actividad

Dominar los conceptos esenciales de la Estadística descriptiva, así como aplicar los métodos y procedimientos descriptivos estadísticos en el procesamiento de la información científica, con el apoyo del software para facilitar el análisis de datos.

### Material

- Notas de clase
- Base de datos osos.csv
- Software para el análisis de datos (Excel, Google Sheets, Python, R, etc.)

### Actividades

Con base a las notas vistas en clase resolver las siguientes actividades. Recuerda que las bases de datos que se deberán utilizar, así como las notas de clase las podrás encontrar en la carpeta [PRÁCTICA 5](#).

1. Define los siguientes conceptos

- |                                |                             |
|--------------------------------|-----------------------------|
| a) Medida de tendencia central | f) Rango                    |
| b) Moda                        | g) Desviación estándar      |
| c) Media                       | h) Varianza                 |
| d) Mediana                     | i) Coeficiente de variación |
| e) Mitad de rango              | j) Percentil                |

k)

### Sesgo

2. Se realizó un experimento para determinar si una deficiencia de dióxido de carbono en la tierra afecta los fenotipos de los guisantes (chícharos). A continuación se indican los códigos de los fenotipos: 1 amarillo claro, 2 verde claro, 3 = amarillo rugoso y 4 verde rugoso. ¿Se pueden obtener medidas de tendencia central para estos valores? ¿Los resultados tienen algún sentido?

2, 1, 1, 1, 1, 1, 1, 4, 1, 2, 2, 1, 2, 3, 3, 2, 3, 1, 3, 1, 3, 1, 3, 2, 2

3. Moscas de la fruta. Más adelante se indican las longitudes (en milímetros) del tórax de una muestra de moscas de la fruta machos. Las moscas de la fruta (*Drosophila*) son algunos de los sujetos favoritos de experimentación de los investigadores porque poseen una composición cromosómica sencilla, se reproducen con rapidez, tienen una gran cantidad de descendientes y son fáciles de cuidar. (Éstas son características de las moscas de la fruta y no necesariamente de los investigadores). Si nos enteramos de que las medidas se obtuvieron de moscas que volaban sobre una manzana que se encontraba en una mesa de cocina en Pocatello, ¿la media sirve como un estimado razonable de todas las moscas de la fruta que residen en Estados Unidos?

0.72, 0.90, .84, 0.68, 0.84, 0.90, 0.92, 0.84, 0.64, 0.84, 0.76

4. Se realizó un experimento para probar la longevidad de árboles tratados con fertilizantes. El experimento se lleva a cabo durante un periodo fijo de cinco años. (Se dice que la prueba se censura a los cinco años). Los resultados muestrales (en años) son 2.5, 3.4, 1.2, 5, 5 (donde 5 indica que el árbol aún estaba vivo al final del experimento). ¿Qué puede concluir acerca de la vida media de los árboles?

5. IMC y género. Es bien sabido que los hombres tienden a pesar más y a ser más altos que las mujeres. El índice de masa corporal (IMC) es una medida que se basa en el peso y en la estatura. A continuación se muestran los valores de IMC de hombres y mujeres elegidos de manera aleatoria. ¿Parece existir una diferencia en la variación entre los dos conjuntos de datos?

Hombres: 23.8, 23.2, 4.6, 6.2, 3.5, 24.5, 1.5, 31.4, 26.4, 22.7, 27.8, 8.1

Mujeres: 19.6, 23.8, 9.6, 29.1, 25.2, 21.4, 22.0, 27.5, 33.5, 0.6, 29.9, 17.7

6. Si un conjunto de datos indica la longevidad (en días) de moscas de la fruta, ¿qué unidades se utilizan para la desviación estándar? ¿Qué unidades se utilizan para la varianza?

7. Investigadores de la Universidad de Pennsylvania realizaron experimentos con álamos. A continuación se indican los pesos (en kg) de los álamos que no recibieron tratamiento y de los álamos tratados con fertilizantes y riego. ¿Parece que existe una diferencia entre las dos desviaciones estándar?

Sin tratamiento: 0.15, 0.02, 0.16, 0.37, 0.22

Con fertilizantes y riego: 2.03, 0.27, 0.92, 1.07, 2.38

8. Un conjunto de datos consta de 20 valores bastante cercanos. Se incluye otro valor, pero este nuevo dato es un valor extremo (muy alejado de los otros valores). ¿Cómo afecta el valor extremo a la desviación estándar? ¿No hay ningún efecto? ¿Hay un pequeño efecto? ¿Hay un gran efecto?

9. Con base al conjunto de datos del archivo osos.csv que se encuentra en la carpeta Funciones\_Estadística, para cada una de las columnas con calcular e interpretar las siguientes medidas de tendencia central y de dispersión:

a) Moda

b) Media

c) Mediana

d) Mitad de rango

e) Rango

f) Desviación estándar

g) Varianza

h) Coeficiente de variación

j) ¿Para todas las columnas es posible obtener estas medidas? ¿Por qué?

La EDAD está en meses, MES es el mes de la medición (1 5 enero), SEXO está codificado como 1 = macho y 2 = hembra, cabezaL es la longitud de la cabeza (pulgadas), cabeza es la anchura de la cabeza (pulgadas), CUELLO es la circunferencia del cuello (pulgadas), ESTAT es la estatura del cuerpo (pulgadas), PECHO es la circunferencia torácica (pulgadas) y el PESO está medido en libras

10. Con base al conjunto de datos del archivo basura.csv que se encuentra en la carpeta Funciones\_Estadística, utilizando la columna Papel crear una tabla de datos agrupados y un histograma, además calcular e interpretar las siguientes medidas de tendencia central y de dispersión:<a) Moda

b) Media

c) Mediana

d) Rango

e) Desviación estándar

f) Varianza

## BIBLIOGRAFÍA

- Reyes Pedro. "Bioestadística". Limusa. México. 2005.
- Triola, M. F. (2004). Probabilidad y estadística. Pearson educación.
- Wayne W. Daniel. "Bioestadística base para el análisis de las ciencias de la salud". Limusa. México. 2005.

## RÚBRICA DE EVALUACIÓN

Criterios	Puntaje
Redacción y ortografía	10
Metodología en el análisis de datos	60
Interpretación y presentación de resultados	30

## FORMA DE ENTREGA

La presente práctica deberá entregarse mediante Classroom.

## **ANEXO**

### **MATERIAL DIGITAL Y DESCARGA**

Para obtener el material digital (bases de datos, tutoriales y notas) correspondiente a la materia de Funciones y Estadística Descriptiva, consulta el siguiente enlace o escaneando el código QR:

[https://drive.google.com/drive/folders/1YagEcxCBp7\\_pwmJ24UQgIkcZ-c5\\_KDwA?usp=sharing](https://drive.google.com/drive/folders/1YagEcxCBp7_pwmJ24UQgIkcZ-c5_KDwA?usp=sharing)

