



## **GUIA DE ESTUDIO PARA EL EXAMEN DE ADMISIÓN A LA MAESTRIA EN BIOTECNOLOGÍA**

### **ESTADÍSTICA**

1. Revisar conceptos estadísticos tales como: escalas de medición, estadística descriptiva, distribuciones e hipótesis.
  - 1.1. Escalas de medición: Identificar las diferentes escalas en las que se miden las variables en investigación.
  - 1.2. Estadística descriptiva: Desarrollar un análisis descriptivo a partir de datos estadísticos. Revisar el concepto proporción.
  - 1.3. Distribuciones: Revisar el concepto de distribución de densidades poblacionales y probabilidad.
  - 1.4. Hipótesis y estadísticos descriptivos: Revisar el planteamiento de hipótesis y su inferencia con base en estadísticos descriptivos.

### **BIBLIOGRAFÍA**

- Castillo, M. A. (2013) Estadística aplicada, México: Trillas
- Clifford, B. R. y Taylor, R. A. (2008) Bioestadística, México: Pearson.
- Daniel, W. W. (2006) Bioestadística. Base para el análisis de las ciencias de la salud, México: Limusa-Wiley.
- Márquez-De Cantú, M. J. (1991) Probabilidad y estadística para ciencias químico-biológicas, México: Mc Graw Hill.
- Quevedo-Urías, H. y Pérez-Salvador, B. R. (2008) Estadística para ingeniería y ciencias, México: Grupo Editorial Patria.



## Guía de estudio AMPLIADA para el examen de admisión en Estadística

**Revisar conceptos estadísticos tales como: escalas de medición, estadística descriptiva, distribuciones e hipótesis**

**1.1. Escalas de medición:** Identificar las diferentes escalas en las que se miden las variables en investigación.

**Ejemplos:** ¿En qué escala se miden las variables en los siguientes contextos de investigación? Después de cada variable que se encuentra subrayada, escribe, sobre la línea entre paréntesis el nombre de la escala de medición en la que está evaluada: nominal, ordinal, de intervalo o de razón.

**Ejemplo 1.1.1.** Se estudiaron tres cepas de *Aspergillus niger*, con el fin de investigar con cuál tipo de cepa (nominal) se obtiene un mejor crecimiento (razón).

**Ejemplo 1.1.2.** Se evalúa el pH (intervalo) y la viscosidad (intervalo) obtenida en los vasos de yogurt a través del tiempo (intervalo), realizándose evaluaciones cada 2 horas durante 40 horas en total.

**Ejemplo 1.1.3.** Se evalúa la intensidad de la coloración en: muy baja, baja, intermedia, alta y muy alta (ordinal) de la carne de truchas a las que se les ha venido alimentando de manera especial adicionando una cantidad determinada de cempasúchil para obtener trucha salmonada.

**1.2. Estadística descriptiva:** Desarrollar un análisis descriptivo a partir de datos estadísticos.  
Revisar el concepto proporción

**Ejemplo 1.2.1.** Se realizó un estudio para explorar el tipo de yogurt de mayor preferencia en 4 regiones del país. Los cuatro tipos seleccionados fueron:

- a) Yogurt cremoso (adicionado de crema)
- b) Yogurt aflanado (semisólido)
- c) Yogurt para beber
- d) Yogurt bajo en calorías

Se encuestaron 925 consumidores adultos, a los que se les preguntó cuál de estos tipos de yogurt prefieren consumir.

Los datos de las frecuencias obtenidas para la preferencia de cada tipo de yogurt por región fueron las siguientes:

| Región    | Tipo de yogurt |          |            |                  | Total |
|-----------|----------------|----------|------------|------------------|-------|
|           | Cremoso        | Aflanado | Para beber | Bajo en calorías |       |
| Norte     | 59             | 47       | 46         | 71               | 223   |
| Occidente | 69             | 78       | 46         | 34               | 227   |
| Centro    | 65             | 63       | 66         | 61               | 255   |
| Sureste   | 58             | 42       | 68         | 52               | 220   |

a) Si se considera que la preferencia depende de la región del país, construye la tabla de frecuencias de cada tipo de yogurt (en %) por región, considera el total por región como el 100%.

### Respuesta

| REGIÓN    | TIPO DE YOGURT |          |            |                  | Total |
|-----------|----------------|----------|------------|------------------|-------|
|           | Cremoso        | Aflanado | Para beber | Bajo en calorías |       |
| Norte     | 26.46 %        | 21.08 %  | 20.63 %    | 31.84 %          | 100 % |
| Occidente | 30.40 %        | 34.36 %  | 20.26 %    | 14.98 %          | 100 % |
| Centro    | 25.49 %        | 24.71 %  | 25.88 %    | 23.92 %          | 100 % |
| Sureste   | 26.36 %        | 19.09 %  | 30.91 %    | 23.64 %          | 100 % |

b) ¿En qué regiones se presenta una proporción menor a 15 % en la preferencia de algún tipo de yogurt? ¿De qué tipo de Yogurt se trata?

**Respuesta:** Solo se presenta en la región Occidente, donde el yogurt bajo en calorías tuvo una proporción de 14.98 %

c) ¿En qué región tiene una mayor aceptación el yogurt para beber?

**Respuesta:** En la región sureste

d) Si tú fueras productor de un yogurt bajo en calorías ¿dónde esperarías mayor aceptación?

**Respuesta:** En la región norte

e) ¿Y en donde menos?

**Respuesta:** En la región occidente

f) ¿En qué región se presentan una distribución más homogénea en las proporciones de preferencia en los yogurts?

**Respuesta:** En la región centro.

g) ¿Qué probabilidad tendrías de vender yogurt cremoso en el sureste del país? ¿En qué región tendrías más probabilidad de vender este tipo de yogurt?

**Respuesta:** La probabilidad es de 26.36 %, la región occidente es la de mayor probabilidad de aceptación del yogurt cremoso con p= 30.4 %

h) Si tu fueras un productor de solamente elabora yogurt para beber y bajo en calorías, ¿En qué región tendrías más aceptación global de ambos productos y en cual menos?

**Respuesta:** Cómo interesa la aceptación global de yogurt para beber y el bajo en calorías, se suman sus porcentajes para cada región. La mayor aceptación la tendrían en la región sureste con 54.55% y la menor en la región occidente con 35.24%.

**Ejemplo 1.2.2.** Se evaluó la producción enzimática a 16 lotes en los que se realizó, durante 41 horas, un proceso de fermentación sólida a partir de la cáscara de naranja que se desecha cuando se extrae el jugo inoculada con *Aspergillus niger*. El proceso de fermentación se llevó a cabo aplicando 4 temperaturas diferentes. Los datos recabados se presentan a continuación:

| Temperatura (°C) | Producción Enzimática  |
|------------------|------------------------|
| 33               | 63.8, 51.4, 57.1, 60   |
| 34               | 58.7, 60.5, 62.4, 59   |
| 35               | 76.2, 76, 76.3, 68.3   |
| 36               | 80.1, 81.2, 79.8, 82.1 |

- a) Calcula el promedio general y la varianza general de la producción enzimática

**Respuesta:** Media= 68.31 y Varianza = 105.67

- b) Calcula las producciones enzimáticas promedio, las desviaciones típicas y los valores extremos, utilizando los datos de cada nivel de temperatura. Llena con los estadísticos obtenidos en una tabla

**Respuesta:**

| Temperatura (°C) | Media | Desviación típica | Mínimo valor | Máximo valor |
|------------------|-------|-------------------|--------------|--------------|
| 33               | 58.1  | 5.19              | 51.4         | 63.8         |
| 34               | 60.15 | 1.69              | 58.7         | 62.4         |
| 35               | 74.2  | 3.94              | 68.3         | 76.3         |
| 36               | 80.8  | 1.06              | 79.8         | 82.1         |

- c) ¿En qué niveles de temperatura las medias quedaron por debajo de la media general?

**Respuesta:** A 33 y 34 °C de temperatura

- d) ¿En qué nivel de temperatura las mediciones de producción enzimática son más homogéneas?

**Respuesta:** En 36°C, ya que la varianza es menor

### 1.3. Distribuciones: Revisar el concepto de distribución de densidades poblacionales y probabilidad.

**Ejemplo 1.3.1.** El 70% de los alimentos congelados de una marca determinada se venden antes de su fecha de caducidad en los supermercados del D.F. Hoy, 1° de mayo del 2010, se dejaron 20 alimentos con fecha de caducidad: 30-julio-10, contesta los siguientes incisos:

- a) Plantea la función de densidad de la variable X: número de alimentos que se venden antes del 30-jun-2010, de los 20 que se dejaron hoy.

Función binomial:  $f(X) = P[X \leq x] = C_{20}^x (.70)^x (.30)^{20-x}$ , donde  $C_{20}^x$  son las combinaciones de 20 en x.

- b) ¿Con qué probabilidad menos de 15 se venden antes del 30-jun-2010?

$P[X < 15] = P[X \leq 14] = \underline{0.5836}$

c) ¿Con qué probabilidad se venden a lo más 16 antes del 30-jun-2010?

$$P[X \leq 16] = \underline{0.8929}$$

d) ¿Con qué probabilidad se venden más de 13, antes del 30-jun-2010?

$$P[X > 13] = 1 - P[X \leq 13] = \underline{0.6080}$$

e) ¿Con qué probabilidad se venden al menos 15, antes del 30-jun-2010?

$$P[X \geq 15] = 1 - P[X < 15] = 1 - P[X \leq 14] = \underline{0.4164}$$

**Ejemplo 1.3.2.** Se toma una muestra aleatoria de 15 tortillas de harina de una empresa que las produce la cual garantiza que únicamente el 10% tienen algún defecto (están incompletas, dobladas, etc.). Contesta los siguientes incisos utilizando la distribución binomial:

a) ¿Cuál es la probabilidad de que exactamente 2 sean defectuosas?

**Respuesta:** Sea X el número de tortillas defectuosas de las 15 elegidas al azar, entonces la

$$P[X=2] = \underline{0.2669}$$

b) ¿Cuál es la probabilidad de que a lo más 3 sean defectuosas?

**Respuesta:**  $P[X \leq 3] = \underline{0.9444}$

**Ejemplo 1.3.3.**  $X_1, X_2, X_3, \dots, X_{36}$  es una muestra aleatoria de la variable X: Producción de una proteína, que tiene distribución normal con media 5 y varianza 4. Encuentra: la desviación típica y el error típico (desviación típica de la media)

**Respuesta:**

La desviación típica de los datos es = 2

El error típico (desviación típica de la media muestral) es  $(4/36)^{0.5} = 2/6 = 0.333$

a) Calcula  $P(\bar{X} \leq 5.5)$

**Respuesta:**  $P(\bar{X} \leq 5.5) = \underline{p(Z \leq 5.55 - 5/0.333) = p(Z \leq 1.5) = 0.9332}$

b) Calcula  $P(3.8 \leq X \leq 4.8)$

**Respuesta:**  $P(3.8 \leq X \leq 4.8) = \underline{0.1859}$

Encuentra las constantes g y h tales que

c)  $P(X \leq g) = .95$   $P(\bar{X} \geq h) = .025$

**Respuesta:**  $g = \underline{8.29}$  y  $h = \underline{5.6533}$

- d) Si esta muestra proviene de una población con 500 unidades, y solo nos interesan aquellas cuyo contenido de proteína entre 3.8 y 4.8. ¿Qué porcentaje de unidades serían las que tomarías? ¿Qué número de unidades se esperaría que fueran las de interés?

**Respuesta:** Por el inciso b) la probabilidad de tener unidades entre 3.8 y 4.8 es 0.1859, por lo que el porcentaje de muestras que se tomarían es 18.59%. Esto representa  $0.1859 \times 500 = 92.95$ , así que se esperaría tener 93 unidades en el rango de proteína de 3.8 y 4.8.

#### 1.4. Hipótesis y estadísticos descriptivos: Revisar el planteamiento de hipótesis y su inferencia con base en estadísticos descriptivos.

**Ejemplo 1.4.1.** Un investigador mide el rendimiento de nitrógeno molecular en dos reactores que tienen diferentes condiciones de medio de cultivo, condiciones A y B. ¿Dan los datos suficiente evidencia para probar que son mejores las condiciones B?

- a) Plantea las hipótesis

**Respuesta:**  $H_0: \mu_A \geq \mu_B$  VS.  $H_a: \mu_A < \mu_B$

- b) Utilizando los estadísticos descriptivos que se muestran en la tabla contesta los incisos que se plantean al final de ésta.

**Estadísticos descriptivos**

|             | Número de casos | Media     | Desviación típica |
|-------------|-----------------|-----------|-------------------|
| Condición A | 15              | 0.2833333 | 0.139983          |
| Condición B | 15              | 0.7166666 | 0.1737678         |

- a) ¿Qué prueba vas a utilizar? **Respuesta:** T-Student para comparar las medias de dos poblaciones independientes.
- b) ¿Qué valor toma el estadístico de prueba? **Respuesta:**  $T = -7.5213$
- c) ¿Qué decisión toma respecto a  $H_0$ , con un nivel de significancia de 0.01?
- ( x )Rechazar ( )No rechazar
- d) Da la conclusion.

**Respuesta:** Los datos dieron evidencia, con una significancia del 0.01, que el rendimiento promedio de nitrógeno molecular en las condiciones B es mayor que en las de A.

**Ejemplo 1.4.2.** Un fabricante de pan compra costales de harina a un nuevo proveedor que establece que el peso promedio del costal es de 50.1 Kg; sin embargo, se ha detectado que posiblemente el peso sea menor a lo indicado. Para verificar esta sospecha, se seleccionan aleatoriamente una muestra de 15 costales de los lotes surtidos recientemente, obteniendo los estimadores de la media de 49.4 y de la varianza de 1.2. ¿se puede confirmar la sospecha?

- a) Plantea las Hipótesis. **Respuesta:**  $H_0: \mu \geq 50.1$      $H_a: \mu < 50.1$
- b) ¿Qué prueba estadística vas a usar? **Respuesta:** prueba T-Student para una sola población
- c) ¿Qué valor toma el estadístico de prueba? **Respuesta:**  $t = -2.47$ , con 14 grados de libertad
- d) ¿Qué decisión tomas respecto a  $H_0$ , con un nivel de significancia de 0.05? **Respuesta:** El valor de tablas a una significancia de 0.05 y 14 g.l es -1.76
- ( X )Rechazar    (   )No rechazar
- e) ¿Cuál es la conclusión? **Respuesta:** Se confirma la sospecha, el peso promedio es menor a 50.1