

# 1. Estadística Descriptiva

## 1.1. Muestreo

**Muestreo aleatorio simple** Contar con un listado de todas las unidades de la población (marco muestral), se numeran correctamente y se eligen mediante una tabla de números aleatorios la muestra. Es equiprobabilístico, la probabilidad es conocida y la misma para todos.

**Muestreo sistemático** Seleccionar individuos según una regla o proceso periódico. Se divide el total de la población por el tamaño de la muestra deseada y se obtiene la constante de muestreo. La primera unidad ( $r$ ) se extrae tomando al azar entre 1 y la constante de muestreo.  $S = \{r, r + k, r + 2k, \dots, r + nk\}$

**Muestreo estratificado** Se divide la población en estratos, niveles o grupos según criterios prefijados, los estratos son diferentes unos de otros. Se toma una submuestra a partir de cada estrato mediante un procedimiento aleatorio simple.

**Muestreo por conglomerados** Requiere elegir una muestra aleatoria simple de unidades heterogéneas entre sí de la población llamada conglomerados. Cada elemento de la población pertenece exactamente a un conglomerado. En este muestreo estos representan a toda la población. Los conglomerados deben formarse considerando una fuerte heterogeneidad entre sus elementos y homogeneidad entre conglomerados. Un conglomerado es una especie de miniaturas de la población.

**Muestreo por etapas** Se obtiene seleccionando primero una muestra de Unidades y luego una muestra aleatoria de elementos de Unidad seleccionada. La unidad de la primera etapa se llama UPM (unidad primaria de muestreo) la subsiguiente es la unidad secundaria de muestreo USM y así sucesivamente.

**Población** Es un conjunto cuyos elementos tienen una o más características comunes. Este conjunto puede estar formado por elementos animados o no, o por las mediciones que sobre esos elementos se realicen.

**Muestra** Es una parte o subconjunto de la población previamente especificada. Sus componentes deben tener la o las mismas características comunes que las de la población de donde ella proviene.

**Parámetro** caracteriza una población, por ejemplo el promedio de materias aprobadas de todos los alumnos de 2do año de las carreras de ingeniería es 3.

**Estimador** estadístico o estadístico caracteriza una muestra, por ejemplo, de un grupo de 25 alumnos de 2do año de la carrera de ingeniería es 3.

**Frecuencia acumulada** Es la frecuencia acumulada o sumada hasta el  $i$ -ésimo valor de la variable  $F_3 = \sum_{i=1}^3 f_i$

**Frecuencia relativa acumulada** es la frecuencia acumulada hasta cierto valor de la variable pero relativa al total de elementos observados  $H_i = F_i/n$

## 1.2. Medidas de posición y de dispersión

### Medidas de posición

- **MEDIA ARITMÉTICA** Es el valor que tomaría la variable si estuviese uniformemente repartida entre todos los individuos que forman la muestra.

$$M(x) = \frac{\sum_{i=1}^n X_i}{n} = \bar{X} \quad (1)$$

Si se tiene una muestra de tamaño  $n$  es estimador

$$\frac{\sum_{i=1}^N X_i}{N} = \mu \quad (2)$$

Si se tiene una población de tamaño  $N$  es parámetro.

Si se tienen los datos agrupados en intervalos o por valores distintos de la variable sea esta discreta o continua, se calcula la media aritmética ponderada.

$$\bar{X} = \frac{1}{\sum_{i=1}^k f_i} \sum_{i=1}^k x_i f_i \quad (3)$$

donde  $\sum_{i=1}^k f_i = n$  la media aritmética muestral

$$\mu = \frac{1}{\sum_{i=1}^k f_i} \sum_{i=1}^k x_i f_i \quad (4)$$

donde  $\sum_{i=1}^k f_i = N$  la media aritmética poblacional

- **MEDIANA** Es el valor central de la distribución, el valor del carácter a ambos lados del cual se reparten por mitades las observaciones
- **MODO** Es el valor del carácter que se representa con mayor frecuencia en la muestra o población, o sea, aquel al que le corresponde el mayor número de observaciones.
- **CUARTILES, DECILOS, CENTILES** Con el mismo sentido que la mediana, son los valores que dividen a las observaciones por cuartos, decimos y centesimos respectivamente, son 3, 9 y 99.
- **SEMI-RANGO** Es el valor central de los valores distintos de la variable, entre el valor mínimo y el valor máximo.

$$R = \frac{X_M + x_m}{2} \quad (5)$$

**Medidas de Dispersión** Medidas que informa sobre la variación o variabilidad de los datos.

- **RECORRIDO, RANGO O AMPLITUD** Es la diferencia entre la mayor y la menor de las observaciones.

$$R = X_{maximo} - X_{minimo} = X_M - x_m \quad (6)$$

Es la mas sencilla y directa medida de dispersión. No proporciona una medida de la variabilidad con respecto a alguna medida de posición, ni informa sobre la ubicación o dispersión de los datos ya que se calcula en base a los dos valores extremos.

- **VARIANCIA O VARIANZA** Se define como media de los cuadrados de los desvíos con respecto a la media aritmética de la variable

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^n (X_i - \mu) \quad (7)$$

Para un solo conjunto de datos no proporciona ayuda inmediata puesto que no puede interpretarse en terminos del problema (unidad elevada al cuadrado) Para un solo grupo, una viariabilidad pequeña, no representa nada ya que no hay punto de comparacion, no se puede decir si un valor pequeño es poco significativo o muy significativo.

- **DESVÍO ESTÁNDAR O DESVIACIÓN TÍPICA** Es la raíz cuadrada de la varianza, se toma en consideracion la raíz positiva. Para tener un significado cuantitativo de la magnitud de esta medida de dispersión será necesario referirse a un tipo de distribucion particular que se denomina poblacion o distribución normal. Cuando la muestra es grande, alrededor del 68 % de las observaciones están comprendidas entre  $x-s$  y  $x+s$ .
- **COEFICIENTE DE VARIACIÓN** Mide el porcentaje de la variabilidad relativa al promedio.

$$C.V = \frac{S_m}{\bar{X}} 100 \% \quad (8)$$

Resulta útil para comparar variabilidades de poblaciones o muestras diferentes porque tiene la ventaja de ser independiente de las unidades de medida

### 1.3. Propiedades de la media aritmética

**PROPIEDAD 1** La suma de los desvíos entre cualquier valor observado y el promedio es cero, cualquiera sea la distribución.

$$\sum_{i=1}^k (X_i - \bar{X})n_i = \sum_{i=1}^k x_i n_i - \bar{X} \sum_{i=1}^k n_i = \sum_{i=1}^k x_i n_i - N\bar{X} = \sum_{i=1}^k x_i n_i - \sum_{i=1}^k x_i n_i = 0 \quad (9)$$

**PROPIEDAD 2** La suma de los cuadrados de los desvios entre cada valor observado y el promedio es menor o igual que la suma de los cuadrados de los desvios con respecto a cualquier otro valor.

$$\sum_{i=1}^n (x_i - \bar{X})^2 \leq \sum_{i=1}^n (x_i - A)^2 \quad (10)$$

**PROPIEDAD 3** El promedio de un grupo de medias aritméticas, cada una de ellas ponderada por la cantidad de observaciones

que le dió origen, coincide en el promedio de las observaciones individuales.

**PROPIEDAD 4** La media aritmética de una constante es la misma constante

**PROPIEDAD 5** La media aritmética de una variable más (o menos) una constante es igual a la media aritmética de la variable más (o menos) la constante.

**PROPIEDAD 6** La media aritmética de una variable por una constante es igual a la media aritmética de la variable multiplicada por la constante.

**PROPIEDAD 7** La media aritmética de una variable que es suma (o resta) de otra variable original es igual a la suma (o resta) de las medias aritméticas de las variables originales.

## 1.4. Propiedades de la varianza

**PROPIEDAD 1** La varianza es una cantidad no negativa  $Var \geq 0$  puesto que se trata de una sumatoria de numeros positivos porque estan elevados al cuadrado.

**PROPIEDAD 2** La varianza de una constante es 0. Todos lo valores de la variable serán iguales a 0, los desvíos serán nulos y en consecuencia la varianza será 0.

**PROPIEDAD 3** Si a una variable se le suma una constante la varianza no cambia

$$Var(x + c) = Var(x) \quad (11)$$

Al sumar o restar una constante la distribución se desplaza pero no se vé afectada su dispersión.

$$Var(x + c) = Var(x) + Var(c) = Var(x) + 0 = Var(x) \quad (12)$$

**PROPIEDAD 4** Si los valores de la variable se multiplican (o dividen) por una constate cambia la dispersión puesto que al multiplicar el producto dependel del valor de la variable a considerar. Los desvíos se verán afectados y como estos se elevan al cuadrado, aparecerá la constante al cuadrado.

$$Var(xC) = C^2Var(x) \quad (13)$$

## 1.5. Diagrama de caja

El diagrama de BOXPLOT consiste en una caja cuyos bordes inferior y superior son los cuartiles 1° y 3° y la línea central representa la mediana. Las lineas que se extienden desde cada caja representan la distancia entre la mayor y la menor de las observaciones, los bigotes indican el rango de los datos. La longitud de la caja es la distancia entre el primer y el tercer cuartil, de forma que la caja contiene los datos centrales de la distribución. La línea dentro de la caja señala la posicion de la mediana. Si ésta cae cerca del final de la caja, indica la presencia de asimetría.

## 2. Probabilidad

**Experimento aleatorio** Denotado  $E_i$  y para cada uno de ellos el conjunto de resultados posibles se denominan  $S_i$

**Experimento** es una acción, proceso u operación en el que se obtienen resultados bien definidos y que conllevan a la observacion de estos resultados.

**Resultado** Es lo que se obtiene de un solo ensayo del experimento, es decir de una sola repetición del mismo.

**Ensayo** Es el acto que lleva a un resultado determinado, de entre los posibles resultados distintos del experimento

**Experimento o fenómeno aleatorio** Es el experimento en el cual el resultado se presenta al azar

**Espacio muestral** Es el conjunto formado por todos los resultados posibles de un experimento aleatorio. Lo denominamos  $S$ .

**Punto muestral** Es un punto del espacio muestral, uno de los resultados posibles del experimento aleatorio. Lo denominamos con  $a$  o  $s$  entonces  $a \in S$  ó  $s \in S$ .

**Principio de la multiplicación** En general, si los conjuntos  $A_1, A_2, \dots, A_k$  contienen  $n_1, n_2, \dots, n_k$  elementos existen  $n_1 \cdot n_2 \cdot n_3 \dots n_k$  maneras de combinar un elemento de  $A_1$ , uno de  $A_2$ , ... y uno de  $A_k$ .

**Principio de la adición** En general, si los conjuntos  $A_1, A_2, \dots, A_k$  contienen  $n_1, n_2, \dots, n_k$  elementos existen  $n_1 + n_2 + n_3 \dots + n_k$  maneras en que se puede dar el primero o el segundo o ... ó el k-ésima conjunto.

**Clase exhaustiva** Si se tienen los sucesos,  $A_1, A_2, \dots, A_k$  estos forman una clase exhaustiva de "S" si se cumplen con las siguientes condiciones.

- $A_j \cap A_i = \emptyset$  para cualquier  $j \neq i$ , esto es, si los k sucesos son mutuamente excluyentes.
- $\bigcup_{i=1}^{\infty} A_i$ , los k sucesos completan el espacio muestral "S"

**Suceso**  $A_i$  Un suceso A respecto a un espacio muestral S asociada al experimento E, es simplemente un conjunto de resultados posibles del mencionado experimento. Siendo  $A \in S$  puede estar formado por ningún elemento, uno, dos o coincidir con S. Entonces S es un suceso y  $\emptyset$  también lo es. Cualquier resultado individual también es un suceso.

$$A = \{a_1, a_2\} \quad (14)$$

$$A = \{\} \quad (15)$$

$$A = \{a_1, a_3\} \quad (16)$$

**Frecuencia relativa** Si ahora decimos que dado un posible resultado A, su frecuencia relativa es el número de veces que se da A en relación con el número de veces que se realiza o repite el experimento. Estamos definiendo  $\tilde{f}(a) = f(a)/n$  como frecuencia relativa del suceso A. Evidentemente  $\tilde{f}(a)$  no es una constante, a la larga existe una estabilidad de esa frecuencia relativa.

La frecuencia relativa  $\tilde{f}(A)$  tiene las siguientes propiedades

1.  $0 \leq \tilde{f}(A) \leq 1$
2.  $\tilde{f}(A) = 1$  si A ocurre cada vez (siempre) en las n repeticiones.
3.  $\tilde{f}(A) = 0$  si A nunca ocurre en las repeticiones
4. Si A y B son dos sucesos que se excluyen mutuamente entonces  $\tilde{f}(A \cup B) = \tilde{f}(A) + \tilde{f}(B)$
5.  $\tilde{f}(A)$  basado en n repeticiones converge en un valor fijo que podemos llamar  $P(A)$  cuando  $n \rightarrow \infty$

## 2.1. Teoría axiomática de la probabilidad

Sea E un experimento aleatorio y S el espacio muestral asociado con ese experimento aleatorio y sea P una función que asigna un número real a cada suceso  $A \in S$ . Llamamos probabilidad del suceso A y denotamos con  $P(A)$  al número real que satisface los siguientes axiomas:

1.  $0 \leq P(A) \leq 1$
2.  $P(S) = 1$
3. si  $A_1, A_2, A_3, \dots, A_k$  son sucesos que se excluyen mutuamente de par en par

$$\bigcup_{i=1}^{\infty} A_i = P(A_1) + P(A_2) + P(A_3) + \dots \quad (17)$$

Si A y B son mutuamente excluyentes

$$P(A \cup B) = P(A) + P(B) \quad (18)$$

Propiedades generales de  $P(A)$

1. Si  $\emptyset$  es el conjunto vacío entonces  $P(\emptyset) = 0$
2. Si A y A' son sucesos complementarios entonces  $P(A) = 1 - P(A')$
3. Sean A y B dos sucesos cualesquiera entonces  $P(A \cup B) = P(A) + P(B) - P(AB)$  en el caso de que sean mutuamente excluyentes  $P(A \cup B) = P(A) + P(B)$
4.  $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$

### Eventos Compuestos

1.  $P(A^c)$  Probabilidad de evento complementario
2.  $P(A \cup B)$  Probabilidad de que ocurran A o B
3.  $P(A \cap B)$  Probabilidad de que ocurran A y B
4.  $P(A/B)$  Probabilidad de que ocurra A si ya ha ocurrido B

**Evento complementario** Es el conjunto de todos los puntos muestrales que no pertenecen al conjunto. Dado A, es  $\bar{A}$  o  $A^c$  el complemento de A.

**Eventos mutuamente excluyentes** Son tales que la concurrencia de uno impide la ocurrencia del otro, en un mismo ensayo o experimento basico A y B son mutuamente excluyentes si  $AB \neq \emptyset$

**Probabilidad Condicional** Dados dos suceso A y B, la probabilidad de que ocurra A habiendo ocurrido B ( o de que A esté condicional a la aparicion previa de B) se escribe  $P(A/B)$

**Eventos probabilisticamente independientes** Son aquellos en que la ocurrencia de uno no modifica la probabilidad de ocurrencia del otro en más de un ensayo o experimento básico. A y B son independientes si  $P(A) = P(A/B)$  o  $P(B) = P(B/A)$

**Regla de la suma o de la adición** Sean A y B dos eventos definidos en S

$$P(A \cup B) = P(A) + P(B) - P(AB) \quad (19)$$

Si A y B son mutuamente excluyentes

$$P(A \cup B) = P(A) + P(B) - P(AB) \quad (20)$$

Para 3 eventos:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC) \quad (21)$$

**Regla del producto** Sean A y B dos sucesos definidos en S

$$P(A \cap B) = P(A)P(B/A) \quad (22)$$

o bien

$$P(A \cap B) = P(B)P(A/B) \quad (23)$$

Si A y B son independientes

$$P(A \cap B) = P(A)P(B) = P(B)P(A) \quad (24)$$

**Regla de Bayes o de la probabilidad de las causas**

$$P(B_1/A) = \frac{P(B_1 \cap A)}{P(A)} = \frac{P(B_1)P(A/B_1)}{\sum P(B_i)P(A/B_i)} \quad (25)$$

Supongamos que se tiene una clase exhaustiva de n sucesos  $B_i = \{B_1, B_2, \dots, B_n\}$  que por eso mismo constituye una partición de S. Los conjuntos  $B_i$  forman una particion si se da que:

1.  $B_i \cap B_j \neq \emptyset \forall i \neq j$
2.  $\bigcup B_i = 1$
3.  $P(B_i) > 0 \forall i$

La probabilidad del resultado final se obtiene como suma de las probabilidades de que el resultado de interes se de con cada una de las causas que lo original

$$P(A) = \sum P(B_i)P(A/B_i) \quad (26)$$

Si se quiere calcular la probabilidad de que habiendo obtenido A, este provenga de un  $B_i$  determinado, por ejemplo  $B_3$ , se quiere calcular la probabilidad de que el evento A, que ha ocurrido, sea el efecto de la causa  $B_3$ . Dicho de otra manera, bajo la hipótesis de que el efecto A ha sido observado se quiere calcular la probabilidad de que está actuando la causa  $B_3$  se tiene que:

$$P(B_3/A) = \frac{P(B_3 \cap A)}{P(A)} \quad (27)$$

$$P(B_3 \cap A) = P(B_3)P(A/B_3) \quad (28)$$

$$P(B_3/A) = \frac{P(B_3)P(A/B_3)}{\sum P(B_i)P(A/B_i)} \quad (29)$$

### 3. Distribuciones

**Distribución Probabilística** Es una distribución de probabilidades cada una de las cuales esta asociada con uno de los posibles valores diferentes de la variable aleatoria. Se cuenta con ella cuando se conocen todos los valores posibles de la variable aleatoria y las probabilidades asociadas con cada uno de estos valores posibles

**Función de probabilidad discreta** Verifica que

$$p(x_i) \geq 0 \forall i = 1, 2, \dots, n \quad (30)$$

$$\sum p(x_i) = 1 \quad (31)$$

Se define como

$$P_x(x) = \begin{cases} P(x_i) & \forall i = 1, 2, \dots, n \\ 0 & \text{otro caso} \end{cases} \quad (32)$$

**Función de probabilidad continua** Verifica que

$$f(x) \geq 0 \quad \forall x \in (-\infty, +\infty) \quad (33)$$

$$\int_{-\infty}^{\infty} f(x)dx = 1 \quad (34)$$

$$\text{Si } A : \left\{ \frac{x}{a} \leq x \leq b \right\} \quad P(A) \int_a^b f(x)dx \quad \forall a < b$$

Se define como

$$f_x(x) = \begin{cases} f(x_i) & \forall i = 1, 2, \dots, n \\ 0 & \text{otro caso} \end{cases} \quad (35)$$

Podemos definir la media aritmetica y la varianza de esa distribución en terminos de probabilidad que son:

$$\mu = \sum_{\forall x} [xp(x)] \quad (36)$$

$$\mu = \int_{-\infty}^{\infty} xf(x)dx \quad (37)$$

Esperanza de x o valor esperado de x para variable aleatoria discreta y continua respectivamente.

$$\sigma^2 = E(x - \mu^2)p(x) \quad (38)$$

$$\sigma = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx \quad (39)$$

Que denominamos Varianza de X para variable aleatoria discreta y continua respectivamente.

Dada X, variable aleatoria discreta, si  $y = H(x) = x^2$  será:

$$E(Y) = E(X^2) = \sum [x^2 p(x^2)] \quad (40)$$

y tenemos la Varianza expresada en terminos de esperanza:

$$Var(x) = \sigma^2 = E[(X - E(X))^2] = E(X^2) - [E(X)]^2 \quad (41)$$

**Distribución Hipergeométrica** Es el numero de formas distintas en que pueden seleccionarse las n personas de la muestra, de un total de N personas. N tomados de a n

$$\binom{N}{n} \quad (42)$$

La probabilidad de un grupo cualquiera o de una muestra cualquiera es:

$$\frac{1}{\binom{N}{n}} \quad (43)$$

Probabilidad de que salga uno cualquiera sea la muestra.

Ahora el numero de formas en que pueden seleccionarse  $Z_0$  entre los k que prefieren seleccionarse en el resto de miembros de la muestra entre los que no prefieren:

$$\binom{k}{Z_0} y \binom{N-k}{n-Z_0} \quad (44)$$

Entonces la distribucion hipergeométrica es el numero de casos favorables sobre el numero de casos posibles

$$\frac{\binom{k}{Z_0} y \binom{N-k}{n-Z_0}}{\binom{N}{n}} \quad (45)$$

**Distribución Poisson** Cada una de las variables aleatorias representa el numero total de ocurrencias de un fenomeno en un continuo. Expresa la probabilidad de un numero  $k$  de ocurrencias acaecidas en un continuo fijo, si estos eventos ocurren con una frecuencia media conocida y son independientes del continuo discurrido desde la ultima ocurrencia o suceso.

La distribucion de probabilidad de Poisson es una distribucion de probabilidades de una variable discreta que nos proporciona la probabilidad de que ocurra un determinado suceso un numero de veces  $k$  en un intervalo determinado de tiempo, espacio, volumen, etc.

El parametro  $\lambda$  es el numero medio de veces que ocurre el suceso en un intervalo continuo que puede ser: tiempo, espacio, volumen, etc. Su funcion de probabilidad es:

$$p(x = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad x = 0, 1, \dots \quad (46)$$

$p(xi) \geq 0$  se cumple ya que al ser  $\lambda$  el promedio de acierto que ocurren en un intervalo continuo es  $\lambda > 0$  entonces  $e^{-\lambda} > 0$ ,  $\lambda^x$  siempre es mayor que 0 y por propiedad de los numeros factoriales siempre  $x!$  es mayor que 0.

## 4. Regresion y Correlacion

Estudio de dos o mas variables con diferentes objetivos. Es un analisis que requiere la consideracion de 2 o mas variables cuantitativas en forma simultanea.

**Regresión simple** Interviene una sola variable independientes

**Regresión múltiple** Intervienen dos o mas variables independientes

**Regresión no lineal** La funcion que relaciona los parametros no es una combinacion lineal en los parametros

El objetivo de la regresión es hallar una funcion o modelo matematico para predecir y estimar el valor de una variable a partir de valores de otros punto

La variable  $Y$  es la dependiente(respuesta, predicha, endógena), es la variable que se desea predecir o estimar y.

La variable  $X$  es la independiente(predictor, explicativa, exogena), es la variable que provee las bases para estimar.

### 4.0.1. Regresión lineal simple

$$y = \alpha + \beta x + e \quad (47)$$

$$\mu_{y/x} = E(y/x) = \alpha + \beta x \quad (48)$$

La interpretacion de los coeficientes de regresion es:  $\alpha$ : es la ordenada al origen, indica el valor medio poblacional de la variable respuesta  $Y$  cuando  $x$  es cero. Si se tiene certeza de que la variable predictora  $x$  no puede asumir el valor 0, entonces la interpretacion no tiene sentido.

$\beta$  es la pendiente de la linea de regresion, indica el cambio o modificacion del valor medio poblacional de la variable  $y$  cuando  $x$  se incrementa en una unidad.

$e$  es un error aleatorio.

**Estimacion de la linea de regresion usando minimos cuadrados** Se debe minimizar el error cuadratico:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2 \quad (49)$$

entonces para minimizar se deriva respecto de  $\alpha$  y  $\beta$

$$\frac{\partial \sum e^2}{\partial \alpha} = 0 \quad (50)$$

$$\frac{\partial \sum e^2}{\partial \beta} = 0 \quad (51)$$

Se obtiene un par de ecuaciones normales para el modelo, cuya solución produce:

$$b = \frac{\sum_{i=1}^n X_i Y_i - \frac{\sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n}}{\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n}} \quad (52)$$

Y el parámetro  $a = \bar{Y} - b\bar{X}$ , entonces el modelo estimado es

$$a = \bar{Y} - b\bar{X} \quad (53)$$

La pendiente  $b$  indica el cambio promedio estimado en la variable respuesta cuando la variable predictora aumenta en una unidad adicional. La ordenada al origen  $a$  indica el valor promedio estimado de la variable respuesta cuando la variable predictora vale 0. Sin embargo carece de interpretación práctica si es irrazonable considerar que el rango de valores de  $x$  incluye a cero.

La variación de los  $Y_i$  se mide convencionalmente en términos de las desviaciones:

$$(Y_i - \bar{Y}_i) \quad (54)$$

La medida de la variación total  $SC_{tot}$  es la suma de las desviaciones al cuadrado  $\sum (Y_i - \bar{Y}_i)^2$

$$(Y_i - \bar{Y}_i) = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i) \quad (55)$$

Donde el miembro de la izquierda representa la desviación total, el primer término del otro lado de la igualdad es la desviación del valor ajustado por la regresión con respecto a la media general y el último es la desviación de la observación con respecto a la línea de regresión.

Si consideramos todas las observaciones y elevamos al cuadrado

$$\sum (Y_i - \bar{Y}_i)^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2 \quad (56)$$

Ahora el miembro de la izquierda es  $SC_{tot}$ , la suma de cuadrados total, el segundo  $SC_{reg}$ , la suma de cuadrados de la regresión y  $SC_e$  la suma de cuadrados del error.

Cada uno de estos cuadrados medios tiene una distribución Ji Cuadrado.

**Estimación de la variancia de los términos del error  $\sigma^2$**  Dado que los  $Y_i$  provienen de diferentes distribuciones de probabilidad con medias diferentes que dependen del nivel de  $x$ , la desviación de una observación  $Y_i$  debe ser calculada con respecto a su propia media estimada  $\hat{Y}_i$ . Por tanto, las desviaciones son las residuales:

$$Y_i - \hat{Y}_i = e_i \quad (57)$$

$$SC_e = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - a - bX_i)^2 = \sum_{i=1}^n e_i^2 \quad (58)$$

La suma de cuadrados del error tiene  $n - 2$  grados de libertad asociados con ella, ya que no se tuvieron que estimar dos parámetros.

$$CM_e = \frac{SC_e}{n - 2} = \frac{\sum_{i=1}^n e_i^2}{n - 2} \quad (59)$$

Donde  $CM$  es el cuadrado medio del error o cuadrado medio residual. Es un estimador insesgado de  $\sigma^2$

**Error estándar de la estimación  $S_e$  o  $S_{y/x}$**

Mide la dispersión o alejamiento promedio de los puntos con respecto a la recta estimada:

$$S^2_e = \frac{1}{n - 2} \sum (Y_i - \hat{Y}_i)^2 \quad (60)$$

$$S^2_e = \frac{1}{n - 2} \left( \sum (Y_i - \bar{Y})^2 - b \sum (X_i - \bar{X})(Y_i - \bar{Y}) \right) \quad (61)$$

**Prueba de hipótesis para el coeficiente de regresión  $\beta$**

$$H_0 : \beta = 0 \quad (62)$$

$$H_1 : \beta \neq 0 \quad (63)$$

y la variable pivotal es

$$y = \frac{b - \beta}{S_b} \approx t_{(n-2)} \quad (64)$$



## 5. ANOVA

El estadístico cuadrado medio dentro de grupos ( $CM_{dentro}$ ) o intravarianza mide la dispersion promedio de las observaciones individuales respecto a la media de su grupo.

Luego, si las unidades experimentales son asignadas aleatoriamente a los  $t$  grupos o si las muestras son muestras aleatorias de una poblacion homogénea, el cuadrado medio intragrupo sirve para estimar  $\sigma^2$ .

Los cuadrados medios intragrupos permanecen siempre como un residuo representando la variabilidad inherente a las unidades experimentales. Aunque el investigador controle todas las fuentes de variabilidad en su experimento, siempre quedará cierta variabilidad (residuo) en las respuestas de unidad en unidad.

El cuadrado medio de error proporciona una medida de la variación con la que hay que contar cuando se desea calcular diferencias significativas entre grupos. El cuadrado medio entre grupos ( $CM_{entre}$ ) o intervarianza mide las desviaciones entre las medias de cada grupo. Es decir, describe la dispersion entre las medias de cada grupo con respecto a la media total. Si no existe efecto de tratamiento, el cuadrado medio entre grupos es un estimador de  $\sigma^2$ .

Cuando el tratamiento modifica las medias de grupos el cuadrado medio entre grupos es un estimador de la varianza poblacional más la componente añadida por efecto del tratamiento.

Si en el conjunto de datos observados encontramos que  $CM_{entre}$  es muy superior al  $CM_{dentro}$ , existiran entonces sospechas fundadas como para creer que

$$\frac{(\sum(\mu_i - \mu)^2)}{(t - 1)} \neq 0 \quad (65)$$

Cuando ocurre esto, se debe admitir que no todas las medias poblacionales son iguales entre si.

Buscar un criterio estadístico para comparar  $CM_{entre}$  y  $CM_{dentro}$  de manera tal que si el valor calculado para el cociente  $\frac{CM_{entre}}{CM_{dentro}}$  excede un cierto limite, rechazamos la hipótesis de que son iguales y en caso contrario, la aceptamos.

Tal vez mas importante del analisis de la varianza es que mediante este procedimiento INFERIMOS la igualdad de medias a través de la igualdad de varianzas. El test F, de igualdad de varianzas, se transforma directamente en un test de igualdad de medias.

Un diseño estadístico de experimentos es un proceso de planificación del experimento mediante el cual se recolectan datos apropiados que al ser analizados por métodos estadísticos estos nos conducen a conclusiones metodológicas válidas.

### 5.1. Propósitos de ANOVA

‡ Proporcionar métodos que permitan obtener la mayor cantidad de información en forma objetiva, confiable y el mínimo costo.

‡ Controlar la variabilidad que está presente en todo experimento y que afecta a sus resultados.

### 5.2. Definiciones ANOVA

**Unidad experimental** Es la parte mas pequeña de material experimental a la que se le aplica el tratamiento y en donde se mide o registra la variable que se investiga.

**Tratamiento** Conjunto de condiciones experimentales o procedimentales que se van a aplicar a una unidad experimental en un diseño elegido y cuyos efectos van a ser registrados (respuesta).

**Observación** Valor que asume la variable respuesta en una determinada realización.

**Factor** variable controlada por el experimentador o variable independiente. Se estudia su efecto sobre la variable dependiente o respuesta.

**Nivel del factor** Es cada una de las categorías, valores o formas específicas del factor.

### 5.3. Principios básicos del diseño experimental

**Repetición** Es la reproducción o replica del experimento básico. Son observaciones de un mismo tratamiento en diferentes unidades experimentales. Las principales razones por las cuales es deseable la repetición:

1. Permite estimar el error experimental
2. Aumenta el alcance de la inferencia del experimento por selección y uso apropiado de las unidades experimentales

**Aleatorizacion** Consiste en la asignacion al azar de los tratamientos a las unidades experimentales con el propósito de asegurar que un determinado tratamiento no presente sesgo.

La aleatorizacion hace validos los procesos de inferencia estadística y los supuestos asegurando conclusiones confiables. Los errores están distribuidos independientemente.

**Control Local** Consiste en tomar medidas (acciones del experimentador) para hacer el diseño experimental mas eficiente de tal manera que pueda permitir la reduccion del error experimental y así hacerla mas sensible a cualquier prueba de significacion. el error experimental puede reflejar variacion del material experimental, errores de medicion u observaciones, errores de experimentacion. Para reducir el error experimental:

- El uso de un diseño experimental apropiado
- La seleccion minuciosa del material a usar (lo mas homogeneo posible)
- El incremento del numero de repeticiones en el experimento
- El perfeccionamiento de la tecnica experimental y el mayor cuidado al dirigir el experimento
- La utilizacion de la informacion proporcionada por variables relacionadas a la variable en Estudio

## 5.4. Diseño Completamente Aleatorizado

En este experimento los tratamientos en estudio se distribuyen al azar en forma irrestricta sobre todas las unidades experimentales siendo el numero de repeticiones por tratamiento igual o diferente. Este diseño se emplea cuando la variabilidad en todo el material experimental es relativamente pequeño y uniformemente distribuido.

**Ventajas** Facil de planear y analizar. Flexible en el empleo del número de tratamientos y repeticiones. Permite tener dentro del analisis de varianza el maximo numero de grados de libertad para las sumas de cuadrados del error.

**Desventajas** La necesidad de homogeneidad del material experimental es difícil de encontrar en experimentos de campo, por lo que su uso se restringe a experimentos de laboratorio.

## 5.5. Modelo I

$$Y_{ij} = \mu_i + \epsilon_{ij} \quad i = 1, \dots, t \quad j = 1, \dots, r \quad (66)$$

Donde:

$Y_{ij}$  Es el valor observado de la j-ésima observacion del i-ésimo tratamiento.

$\mu_i$  Es el valor medio poblacional del i-ésimo tratamiento

$\epsilon_{ij}$  Es el error aleatorio de la j-ésima observacion correspondiente al i-ésimo tratamiento

O si en el modelo anterior denotamos a  $\mu_i = \mu + \tau_i$  se obtiene la siguiente forma alternativa del modelo:

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \quad i = 1, \dots, t \quad j = 1, \dots, r \quad (67)$$

Donde:

$Y_{ij}$  Es el valor observado de la j-ésima observacion del i-ésimo tratamiento.

$\mu$  Es el valor medio poblacional general

$\tau$  Es el efecto medio poblacional del i-ésimo tratamiento

$\epsilon_{ij}$  Es el error aleatorio de la j-ésima observacion correspondiente al i-ésimo tratamiento

## 5.6. Supuestos para el modelo

1. Para cada tratamiento la variable aleatoria respuesta  $Y_{ij}$  se distribuye normalmente. Para cada  $T_i$   $Y_{ij} : N(\mu_i, \sigma_i^2)$
2. Las  $Y_{ij}$  son variables aleatorias independientes entre si.  $Cov(Y_{ij}, Y_{kl}) = 0 \quad \forall i \neq k; \forall j \neq l$
3. Las varianzas de los t tratamientos son iguales (homocedasticidad)  $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_t^2 = \sigma^2$

## 6. Preguntas

† Diga cual de las siguientes expresiones es falsa cuando nos referimos al ANOVA:

- a) Los tratamientos son una fuente de variación.
- b) Los grados de libertad de la variación del error es el numero de unidades experimentales menos uno.

- c) Las poblaciones tienen iguales varianzas.
- d) Se calcula el cociente de dos varianzas muestrales y se lo compara con un valor de F.
- e) Un cuadrado medio es una varianza estimada.
- f) La región crítica para la regla de decisión es unilateral derecha.

† Si el suceso B está incluido en A o dicho de otra manera, está dentro de A, señale la respuesta correcta:

- a)  $P(A) < P(B)$
- b)  $P(A \cap B) = P(B)$
- c)  $P(A \cap B) = 0$
- d) Ninguna de las respuestas es correcta

† ¿Qué medidas descriptivas (tendencia central, dispersión, posición) se pueden leer

- a) en el gráfico de caja y bigote (box-plot)?
- b) en el histograma?
- c) en el polígono de frecuencias absolutas?
- d) en la ojiva (de frecuencias acumuladas)?

† Indique la o las expresiones correctas. En un histograma de frecuencias absolutas, lo que representa a la frecuencia correspondiente a un intervalo es:

- a) la altura del rectángulo.
- b) la altura correspondiente a la marca de clase.
- c) el área del rectángulo.
- d) la altura correspondiente a la marca de clase.
- e) ninguna de las anteriores.

† ¿Cuándo dos sucesos son mutuamente excluyentes?

† Defina o explique claramente que es:

- 1) Cuartilo
- 2) Modo
- 3) Función de cuantía

† Indique si cada afirmación es correcta o incorrecta. En los casos en que la afirmación sea incorrecta justifique su respuesta.

- 1)  $R^2$  representa la fracción de la variación total en Y que no está explicada.
- 2) Un gran fabricante de automóviles ha tenido que retirar varios modelos de su línea 1993 debido a problemas de control de calidad que no fueron descubiertos con los procedimientos finales de inspección aleatoria. Este es un ejemplo de Error de tipo II.

† Una cierta población distribuida normalmente tiene una desviación estándar conocida de 1.0 ¿Cual es el ancho total de un intervalo de confianza de 95 % para la media de la población?

- a) 1.96
- b) 0.98
- c) 3.92
- d) No se puede determinar de la información dada

† Si  $P(A \cap B) = P(A)P(B/A)$  es porque:

- a) A y B son mutuamente excluyentes? Si/No
- b) A y B son sucesos dependientes? Si/No

- c) A y B son sucesos independientes? Si/No

† Suponga que en un corral hay doce machos y ocho hembras y se seleccionan al azar y se separan para ser tratados 5 de los animales. La siguiente expresión, da la probabilidad de obtener:

$$\frac{\binom{12}{2}\binom{8}{3}}{\binom{20}{5}} \quad (68)$$

- a) Al menos dos machos en los cinco ensayos
- b) Exactamente dos machos en cinco ensayos
- c) Exactamente dos machos en seis ensayos
- d) Exactamente dos machos en cinco ensayos
- e) Exactamente tres hembras en cuatro ensayos
- f) Ninguna de las anteriores. ¿Cual?

† Indique si cada una de las expresiones es correcta para expresarse como conclusión o interpretación, después de la construcción de un intervalo de confianza para un promedio población. (Al construir el intervalo empleando un coeficiente de confianza de 0.95, se encontró el límite inferior 10 y el límite superior 25)

- a)  $P(10 < \mu < 25) = 0,95$  SI / NO, porque...
- b)  $P(\mu = 17,5) = 0,95$  SI / NO, porque...
- c)  $0 < \mu < 25$  con un coeficiente de confianza 0.95 SI / NO, porque

† Explique que significa estadísticamente " $\beta$ ":

- a) en un problema de regresión lineal simple
- b) en un problema de Dócima o prueba de hipótesis

† Señale con una cruz cuales son las características de una distribución Hipergeométrica:

- a) Variable Continua
- b) Variable discreta
- c) Tiene dos resultados posibles
- d) Tiene dos resultados posibles (mutuamente excluyentes)
- e) Tiene mas de dos resultados posibles
- f) El experimento básico se repite un numero finito de veces
- g) El experimento básico se repite hasta el éxito
- h) Las probabilidades se mantienen constantes
- i) Las probabilidades no se mantienen constantes

† Una variable de Poisson se caracteriza porque:

- a) Es muy pequeña la probabilidad de un suceso elemental (o simple)
- b) Los sucesos elementales son independientes entre si en cada experimento
- c) Cuenta el que se repita cierto numero de veces un suceso elemental a lo largo de un continuo
- d) Se debe verificar todo lo anterior simultáneamente
- e) Cuenta en que momento ocurre por primera vez un suceso elemental

† Dados dos sucesos A y B el teorema del producto se expresa:

$$P(A \cap B) = P(A \cup B) = P(A) \times P(B) \quad (69)$$

- a) cuando A y B son independientes
- b) siempre

- c) nunca
- d) cuando  $P(A)$  y  $P(B)$  son independientes
- e) cuando A y B son mutuamente excluyentes

† ¿Cual de los siguientes es el primer paso para calcular la mediana de un conjunto pequeño de datos?

- a) Calcular la mediana de orden
- b) Ordenar los datos
- c) Determinar las frecuencias relativas de los valores de los datos
- d) Sumar los valores de la variable y dividir por la mitad
- e) Encontrar el valor que deja la mitad de los valores a cada lado de el
- f) Ninguno de los anteriores

† La diferencia entre una variable aleatoria con distribución binomial y una distribución hipergeométrica:

- a) a la cantidad de observaciones
- b) a que en las dos hay dos resultados posibles
- c) a la aleatoriedad
- d) a ninguno de los 3.

† Si un estimador es insesgado la esperanza del estimador coincide con el valor del parámetro.

- a) Siempre
- b) A veces
- c) Nunca
- d) Ninguna de las anteriores

† La probabilidad de no rechazar una hipótesis nula cuando es cierta es:

- a) el nivel de confianza
- b) el tamaño de la región critica
- c) el nivel de sesgo
- d) ninguna de las tres

† Razonar para cuales de los siguientes problemas la distribución binomial es un modelo adecuado:

- a) Determinación de la probabilidad de que un agente de ventas lleve a cabo 2 ventas en 5 entrevistas independientes si la probabilidad es 0.25 de que el agente lleve a cabo una venta en una entrevista determinada
- b) Determinación de la probabilidad de que no mas de 1 de 10 artículos producidos por una maquina sea defectuoso cuando los artículos se seleccionan a través del tiempo y se sabe que la proporción de defectuosos aumenta con el desgaste de la maquina con el tiempo.

† ¿Cuando dos sucesos son mutuamente excluyentes?

† ¿Cuando dos sucesos son probabilísticamente independientes?

† En cada uno de los casos siguientes se pide:

- a) Reconocer de que aplicación de usos de  $\chi^2$  se trata.
- b) escribir en palabras la hipótesis nula y alternativa.

Un científico social selecciono una muestra de 140 personas y las clasifico de acuerdo a su nivel de ingreso y si jugaron o no al Bingo durante el ultimo mes. La tabla de contingencia obtenida fue:

	Ingreso			
jugó?	bajo	medio	alto	TOTAL
jugó	46	28	21	95
no jugó	14	12	19	45
TOTAL	60	40	40	140

Un científico social selecciono muestras de personas de diferente nivel de ingreso y las clasifico de acuerdo a si jugaron o no al bingo durante el ultimo mes. La tabla de contingencia obtenida fue:

	Ingreso			
jugó?	bajo	medio	alto	TOTAL
jugó	46	28	21	95
no jugó	14	12	19	45
TOTAL	60	40	40	140

† Si la variable aleatoria discreta toma solo valores 1, 2, 5, 8, 9 la función de distribución acumulada,  $F(x)$ :

- a) Sólo toma valores mayores o iguales que cero y menores o iguales que uno
- b) Puede calcularse a partir de la función de masa de probabilidad
- c) Es posible calcularla para el valor de la variable 5,39
- d) Todas las anteriores

† Definir error de tipo I y error de tipo II, ¿cuando se los utiliza?, ¿como se los designa?

† Diferencias entre homogeneidad e independencia en cuanto a su valor esperado, la región crítica y muestreo.

† ¿Qué condiciones se deben cumplir para que haya una función de cuantía?

† Nombrar los supuestos de regresión simple

† Nombrar las condiciones para que se cumpla ANOVA

† A partir de una tabla de ANOVA, ¿Qué hipótesis nula y alternativa puede probar?

- a)  $H_0 : \beta = 0,833$  con  $H_1 : \beta \neq 0,833$ ?
- b)  $H_0 : \beta = 0$  con  $H_1 : \beta > 0$ ?
- c)  $H_0 : \beta = 0$  con  $H_1 : \beta \neq 0$ ?
- d) Otra ...

† Responda VERDADERO si la oración que se le presenta es SIEMPRE verdadera. Si esa oración no es siempre verdadera reemplace la palabra en negrita por palabra/s que hagan siempre verdadera la afirmación.

- a) La **media** de los datos siempre divide a estos en mitades, la mitad mas chicos y la mitad mas grandes que ella
- b) La suma de los cuadrados de las desviaciones de la variable respecto de la media  $\sum(x_i - \bar{x})^2$ , **a veces es negativa**

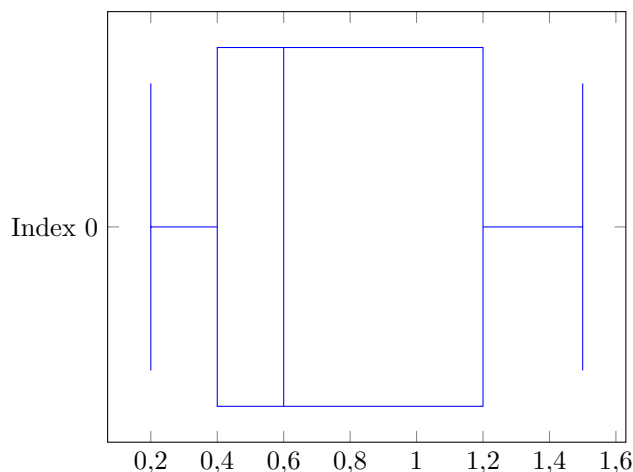
† Coloque verdadero (V) o falso (F). Justifique cuando sea falso.

- a) Al calcular la mediana se toman en consideración todos los valores de la variable
- b) En un box plot, la linea central de la caja es la media aritmética

† Se tienen los cinco datos necesarios para construir un diagrama de caja y bigote.

29    8    2    6    15

- a) La linea central de la caja es: La mediana, la media aritmética, el modo, la varianza?
- b) El borde derecho de la caja es el: primer cuartilo (Q1), segundo cuartilo (Q2), tercer cuartilo (Q3)?
- c) Ubique los cinco datos en el diagrama.

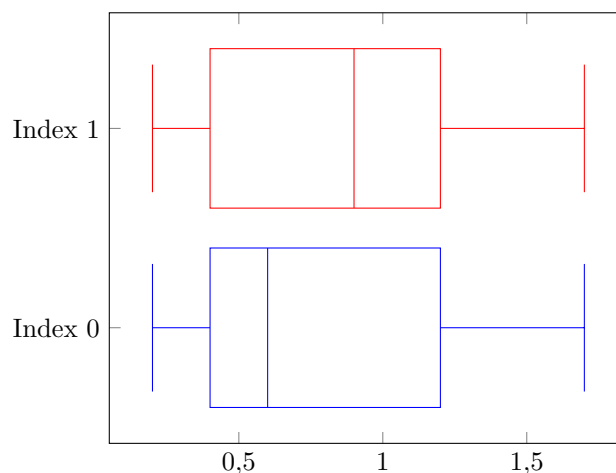


† Se ha medido la longitud de los 20 ejemplares capturados en ultimo torneo de pesca, obteniéndose:

11	14.6	8.4	7.4	2.8	15.5	9.9	12.3	14.2	2.3
10.7	4.6	9.0	5.6	7.7	10.1	11.8	12.7	10.4	14.8

- la mediana de los datos es  $(2.3-10.7)/2 = 6.5$ 
  - a) SI, porque ...
  - b) NO, porque ...
  - c) NO PUEDO CALCULARLA, porque ...
- la media aritmética de los datos podría ser 15.6
  - a) SI, porque ...
  - b) NO, porque ...
  - c) NO PUEDO CALCULARLA, porque ...

† Analice las cajas:



Es cierto que ahora SI, en la caja del segundo diagrama, hay un 25% a cada lado de la mediana y en el primero NO? Diga SI o NO y justifique.

† Seleccione LA respuesta correcta. Señale con una cruz X.

- ☐ En una distribución Hipergeométrica, los dos resultados posibles no son mutuamente excluyentes porque no son independientes.
- ☐ La diferencia fundamental a tener en cuenta para saber si se emplea un modelo Binomial o uno Hipergeométrico es analizar si las repeticiones son o no son independientes.
- ☐ La característica común de las tres distribuciones discretas (binomial, hipergeométrica, poisson) es que, además de tratarse de variables discretas, el numero de repeticiones del experimento básico es conocido de antemano.

† Marque las/las respuestas correctas:

- ☐ En una función de distribución para variable discreta la altura de la gráfica en cada punto de la variable representa la probabilidad en ese punto.
- ☐ En una función de distribución para variable continua la altura de la gráfica en cada punto de la variable representa la probabilidad acumulada hasta ese valor de la variable.
- ☐ En una función de densidad la altura de la gráfica en cada punto de la variable representa la probabilidad acumulada hasta ese valor de la variable.

† ¿Cual es la variable en un esquema o modelo binomial?

- ☐ El numero de repeticiones del experimento básico
- ☐ El porcentaje de éxitos
- ☐ El numero de éxitos en las n repeticiones del experimento básico.

☐ El numero de las repeticiones hasta el primer éxito

☐ Ninguna de las anteriores

† Sea la siguiente tabla:

variable	0	10	100	500
probabilidad	0.45	...	0.4	0.05

Complete la tabla para que sea función de probabilidad.

† Diga cual de las siguientes afirmaciones es falsa:

☐ Si de una población en la cual cada repetición del experimento básico tiene dos resultados mutuamente excluyentes se extrae una muestra de tamaño 10, con reposición, la variable numero de éxitos se distribuye segun una distribución hipergeométrica.

☐ Si de una población se extrae una muestra de tamaño 10, sin reposición, la variable numero de éxitos se distribuye segun una distribución hipergeométrica, si en cada repetición del experimento básico hay dos resultados mutuamente excluyentes.

☐ Si en un experimento es correcto emplear un modelo binomial cuando las extracciones se hacen sin reposición, es porque se conoce el porcentaje de éxitos en la población o porque la población es lo suficientemente grande.

† ¿Cual es la diferencia entre una distribución binomial y una hipergeométrica?

☐ La variable en la binomial es discreta y en la hipergeométrica es continua.

☐ En la distribución binomial la probabilidad de éxito es constante en las diferentes repeticiones y en la hipergeométrica cambia en cada repetición

☐ En la distribución hipergeométrica la probabilidad de éxito es constante en las diferentes repeticiones y en la binomial cambia en cada repetición

☐ No hay diferencias entre las dos distribuciones

† De acuerdo con el Teorema Central del Limite: ¿Podemos esperar que la media de la distribución muestral de medias se aproxime a la media de la población?

† Señale la o las expresión/es correcta/s

☐ El estimador puntual del parámetro esta encerrado en el intervalo de confianza

☐ El parametro que se estima esta encerrado en el intervalo de confianza

☐ Ninguna de las anteriores.

† Al construir el intervalo para la media población empleando un coeficiente de confianza de 0.95, se entro que el limite inferior es 10 y el limite superior es 25. ¿Es correcto decir que: Se tiene una probabilidad de 0.95 de que esos números contendrían al verdadero valor del parámetro?

† En el siguiente conjunto de datos vericar:

4	9	3	6	2	7
---	---	---	---	---	---

▪  $E(X + c) = E(x) + c$

▪  $E(cX) = cE(x)$

▪  $\sigma^2(X + c) = \sigma^2(X)$

▪  $\sigma^2(cX) = c^2E(X)$

† Si la variable aleatoria discreta toma sólo valores 1, 2, 5, 8, 9, la función de distribución acumulada, F(x)

▪ a) Solo toma valores mayores o iguales que cero, y menores o iguales que uno.

▪ b) Puede calcularse a partir de la función de masa de probabilidad f(x)

▪ c) Es posible calcularla para el valor de la variable 5,39.

▪ d) Todas las anteriores

† Señale la o las expresión/es correcta/s



- ☐ El estimador puntual del parámetro está encerrado en el intervalo de confianza
- ☐ El parámetro que se estima está encerrado en el intervalo de confianza
- ☐ Ninguna de las anteriores

† El gerente de una empresa de transportes recibe un lote de paquetes a transportar y desea comparar la variabilidad relativa entre el peso de los paquetes y el volumen que ocupan ¿que debe de utilizar? Señale la/s respuesta correcta.

- Las desviaciones típicas.
- Los rangos.
- Los coeficientes de variación.
- La diferencia de las medias.
- La diferencia de las varianzas.

† Indique si las siguientes expresiones son verdaderas o falsas.

- ☐ En una tabla de distribución de frecuencias se utilizan intervalos mutuamente excluyentes
- ☐ El valor de la mediana se encuentra afectado por la presencia de valores extremos
- ☐ En un conjunto de datos el valor calculado de la mediana siempre se encuentra entre los valores de la media aritmética y la moda.
- ☐ La sumatoria de las diferencias entre cada uno de los valores con respecto a la media aritmética siempre debe ser igual a cero.

† Señale cual de las siguientes afirmaciones es falsa:

- La media aritmética es siempre el centro de gravedad de la distribución.
- En una distribución continua simétrica, media y mediana coinciden.
- La media aritmética cambia cuando cambia algún dato.
- La mediana no siempre cambia cuando lo hace algún dato.
- En las distribuciones continuas simétricas todas las medidas de centralización coinciden.

† Seleccione la respuesta correcta. Cuando a todos los datos de una muestra se les suma una constante.

- a) La media no varía
- b) La media queda incrementada en esa constante
- c) El desvío estándar no varía
- d) El desvío estándar queda aumentado en esa constante

† Entre el primero y el tercer cuartilo se encuentra:

- ☐ El 100 %
- ☐ La mitad de la distribución
- ☐ El 25 % de la distribución
- ☐ El 10 % de la distribución
- ☐ Ninguna de las anteriores

† Indique si las siguientes afirmaciones son verdaderas o falsas

- a) La función de probabilidad  $f(x)$  de una variable aleatoria continua  $X$ , siempre y sin restricciones, asume valores iguales o mayores a cero.
- b) Si de una población dicotómica se extrae una muestra con reemplazamiento, la distribución de probabilidad que resulta es la hipergeométrica
- c) Cuando se utiliza la distribución binomial en un estudio de opinión se está suponiendo que la población es grande.
- d) La distribución hipergeométrica se puede aproximar por una distribución binomial cuando la muestra se extrae de una población grande.

† Marque la respuesta correcta. Comparar la media con la mediana de un conjunto de datos te da una idea de lo esparcidos que se encuentran los valores del conjunto de datos.

- La media y la mediana tienen que coincidir para saber esto
- Si la media es mayor que la mediana los datos están mal
- Si la media es menor que la mediana los datos están mal
- Cuando la media y la mediana distan mucho los datos están muy desperdigados

† Sea  $B$  un suceso contenido en el suceso  $A$ ,  $B \in A$ , ¿Que opción es correcta?

- $P(B) < P(A)$
- $P(B \cup A) = P(B) + P(A)$
- $P(A \cap B) = 0$
- Ninguna de las anteriores

† Si  $F$  es una función de distribución de una variable aleatoria continua, entonces cumple:

- $F(x)$  está acotado entre 0 y 1.
- $F$  es estrictamente creciente
- $\int_0^{+\infty} = 1$
- ninguna de las anteriores

† Mencione las condiciones de la distribución Binomial e indique ¿Bajo que condiciones la distribución Binomial puede ser aproximada por la distribución de Poisson?

† Indique si las siguientes afirmaciones son Verdaderas o Falsas. De acuerdo al teorema central del límite:

- a) Cuando se disminuye el tamaño de la muestra, se disminuye el error estándar de la media (la desviación estándar de la media muestral)
- b) Podemos esperar que la media de la distribución muestral de medias se aproxime a la media de la población
- c) Habrá más dispersión en la distribución de las medias que en la población

† Explique los posibles errores a cometer en la decisión de una prueba de hipótesis. Su relación con el nivel de significación.

† Enuncie los supuestos teóricos para el modelo de regresión lineal simple

† Defina coeficiente de correlación y coeficiente de determinación. Explique en que caso los utiliza y como interpreta a cada uno de ellos

† En un estudio de análisis de la Varianza el investigador forma cinco grupos y toma cuatro observaciones en cada uno de ellos. ¿Cuántas unidades experimentales hay? ¿Cuántos tratamientos? Escriba las hipótesis en términos de los efectos y en términos de los promedios

† Señale la respuesta correcta. Si dos sucesos son independientes:

- a) No pueden ocurrir a la vez.
- b) Siempre ocurre uno u otro, pero no ambos a la vez.
- c) Siempre ocurre al menos uno de los dos.
- d) Si pasa uno, el otro no puede ocurrir.
- e) Todo lo anterior es falso.

† Mencione qué es un intervalo de confianza y relacione amplitud con nivel de confianza.

† Enumere cuales son las semejanzas y las diferencias entre una prueba de independencia y una prueba de homogeneidad.

† Escriba los modelos y supuestos en regresión lineal simple.

† Indique la diferencia entre  $r$  y  $R^2$ , ¿Cuándo se los utiliza y como se interpretan?

† Marque la respuesta correcta. Comparar la media con la mediana de un conjunto de datos te da una idea de lo esparcidos que se encuentran los valores del conjunto de datos.

- a) La media y la mediana tienen que coincidir para saber esto.
- b) Si la media es mayor que la mediana los datos están mal.
- c) Si la media es menor que la mediana los datos están mal-
- d) Cuando la media y la mediana distan mucho los datos están muy desperdigados.

† Si  $F$  es una función de distribución de una variable aleatoria continua, entonces cumple:

- a)  $F(x)$  está acotada entre 0 y 1.
- b)  $F$  es estrictamente creciente.
- c)  $\int_0^\infty F(x).dx = 1$
- d) Ninguna de las anteriores.

† Mencione las condiciones de la distribución Binomial e indique.¿Bajo qué condiciones la distribución Binomial puede ser aproximada por la distribución de Poisson?

† Indique si las siguientes afirmaciones son Verdaderas o Falsas De acuerdo al teorema central del límite.

- a) Cuando se disminuye el tamaño de la muestra, se disminuye el error estandar de la media (la desviación estándar de la media muestral)
- b) Podemos esperar que la media de la distribución muestral de medias se aproxime a la media de la población.
- c) Habrá más dispersión en la distribución de las medias que en la población

† Explique los posibles errores a cometer en la decisión de una prueba de hipótesis. Su relación con el nivel de significación.

† En un estudio de análisis de la Varianza el investigador forma cinco grupos y toma cuatro observaciones en cada uno de ellos. ¿Cuántas unidades experimentales hay?.¿Cuántos tratamientos? Escriba las hipótesis en términos de los efectos y en términos de los promedios. † Seleccione la/s respuesta/s correcta/s: Para comparar correctamente, desde un punto de vista descriptivo, dos o más variables debe usarse:

- a) Desvío estándar.
- b) Amplitud.
- c) El coeficiente de Variación.
- d) La covarianza.

† Cuando todos los datos de una muestra se les suma una constante:

- a) La media no varía.
- b) La media queda incrementada en esa constante.
- c) La varianza no varía.
- d) La varianza queda aumentada en esa constante.
- e) Ninguna de las anteriores.

† Se desea investigar el PH en el crecimiento de cierto microorganismo en un medio específico. Para ellos se prueban en orden aleatorio 24 microorganismos, 6 para cada uno de los 4 niveles de PH elegidos por el investigador y se mide el crecimiento en micrones.

- a) Escriba el modelo general e indique los elementos que lo componen.
- b) Defina las unidades experimentales y las repeticiones de cada una.
- c) ¿Cuales y cuantos son los tratamientos?

† Explique la diferencia entre intervalo de confianza y prueba de hipótesis. ¿Qué es el nivel de confianza? ¿y el nivel de significación?

† ¿Qué mide y como se interpreta un coeficiente de correlación negativo? ¿Cómo deben ser las variables para poder interpretar el coeficiente?

† ¿Cómo se calcula e interpreta el coeficiente de determinación? ¿Cómo deben ser las variables para poder interpretar este coeficiente?

† Diga cual de las siguientes expresiones es un supuesto teórico cuando nos referimos a ANOVA

- a) los tratamientos son una fuente de variación.
- b) Las poblaciones tienen iguales varianzas.
- c) Se calcula el cociente de dos varianzas muestrales y se los compara con un valor de F.
- d) Un cuadrado medio es una varianza estimada.
- e) La región crítica para la regla de decisión es uniteral derecha.