**ESCUELA SUPERIOR POLITECNICA DEL LITORAL**

**CENTRO DE EDUCACION CONTINUA**

**MANUAL PRACTI­CO DE**

**ES­TA­DIS­TICA BASICA Y DISE­ÑO EXPE­RI­MENTAL APLICA­DOS A LA**

**ACUI­CULTURA**

1

**Fabrizio Marcillo M.**

**1992**

**INDICE**

 Pág.

- Introducción v

 CAPITULO I

1. Manejo de datos.-

 1.1.- Datos.- 1

 1.2.- Variables.- 1

 1.3.- Poblaciones y muestras.- 3

 1.4.- Distribución de frecuencias.- 4

 1.5.- Gráficos.- 6

 1.6.- Estadísticos.- 8

 1.6.1.- Estadísticos de cen­tralización.- 9

 1.6.2.- Estadísticos de dispersión.- 10

 CAPITULO II

2.- Teoría de probabilidades.- 12

 2.1.- Espacio Muestral 13

 2.2.- Ley de la multiplicación 13

 2.3.- Permutaciones y combinaciones 13

 2.4.- Definición de probabilidad 14

 2.5.- Teoremas básicos 14

 CAPITULO III

3.- Distribución Normal- 16

 3.1.- Características.- 16

 3.2.- Distribución normal tipificada 18

 3.3.- Aplicaciones. 18

 CAPITULO IV

4.- Otras Distribuciones.- 19

 4.1.- Distribución "t" de Student.- 19

 4.2.- Ji-cuadrado.- 21

 4.3.- "F"de Fisher - Schnedecor.- 22

 4.3.- Distribución binomial.- 22

 CAPITULO V

5.- Bondad de ajuste.- 24

 5.1.- Generalidades 24

 5.2.- Prueba Ji cuadrado 25

 CAPITULO VI

6.- Estimación.- 26 6.1.- Puntual.- 26

 6.2.- Estimación por intervalos.- 28

 6.2.1.- Estimas de medias por

 intervalos. 28

 6.2.2.- De varianzas.- 29

 CAPITULO VII

7.- Pruebas de Hipótesis 30

 7.1.- Generalidades.- 30

 7.2.- Una Población.- 32

 7.2.1.- Una Media con varianza

 conocida.- 32

 7.2.2.- Una media con varianza desconocida.- 33

 7.2.3.- Una varianza.- 34

 7.3.- Dos poblaciones independientes.- 35

 7.3.1.- Dos varianzas.- 35

 7.3.2.- Medias, varianzas conocidas.- 37

 7.3.3.- Medias, varianzas desconocidas e iguales.- 38

 7.3.4.- Medias, varianzas desconocidas y desigua­les.- 39

 7.4.- Pruebas en muestras dependientes.- 40

 7.5.- Pruebas de proporciones.- 41

 7.6.- Pruebas de independencia.- 41

 CAPITULO VIII

8.- Diseño experimental 46

 8.1.- Generalidades.- 46

 8.2.- Tipos de diseño.- 48

 8.2.1.- Un factor completamente aleatorio.- 48

 8.2.2.- Un factor, bloques aleatorios.- 49

 8.2.3.- Cuadrado Latino.- 50

 8.2.4.- Experimentos Factoriales.- 50

 8.3.- Determinación del tamaño muestral.- 51

 8.4.- Datos atípicos.- 52

 CAPITULO IX

9.- Análisis de varianza 54

 9.1.- Generalidades.- 54

 9.2.- ANOVA de una vía.- 55

 9.3.- ANOVA de dos vías.- 58

 9.4.- Restricciones.- 60

 9.5.- Comparaciones múltiples.- 61

 9.5.1.- Prueba de rango múltiple de Duncan.- 61

 CAPITULO X

10.- Ajustes de curvas 63

 10.1.- Diagrama de dispersión.- 64

 10.2.- Método de los mínimos cuadrados.- 64

 10.3.- Coeficiente de correlación.- 66

 10.4.- Uso de la regresión lineal para comparaciones

 múltiples de dos variables cuantitati­vas.- 66

 10.5.- Regresiones no lineales.- 68

 INTRODUC­CION

Este manual es una reco­pilación de varios procesos y métodos esta­dísticos ex­presados de una manera práctica e inteligi­ble para personas con pocos conocimientos de matemáti­ca teórica.

No es intención de esta obra dar una explicación deta­llada y profunda de los procesos matemáticos que invo­lucran las suposicio­nes y cálculos estadísticos, sino más bien dar una guía práctica sobre los cálculos a rea­lizar en determinada situación, dando una idea gene­ral de lo que es diseño experimental.

Los métodos descritos en este libro son de estadística general, y por lo tanto pueden ser usados en distintas situaciones; sin embargo, como el curso está programado especialmente para ser usado en acuicultura, los ejer­cicios acompañados en el anexo corresponden a esta área.

Los tópicos a tratar en este texto entrarán dentro de 4 grupos principales :

- Conceptos y definiciones básicas de estadística, así como ciertas suposiciones que vamos a hacer a lo largo del mísmo.

- Teoría de probabilidades y distribuciones teóricas de las mísmas.

- Diseño experimental, basado en gran parte en la lógica y en el sentido común, así como en los conceptos básicos y la teoría que vayamos revisando.

- Cálculos aritméticos y tablas numéricas, que realizadas de forma más bien rutinaria y mecánica después de haber definido el problema que tenemos, nos darán los materiales sobre los cuales se van a basar las inferencias y medir la incertidumbre.

Los objetivos básicos del libro son entonces: Dar bases para reconocer un problema identificando el tipo de cálculo a realizar. Y, fomentar una forma de pensar clara y disciplinada, especialmente cuando se trata de recolectar e interpretar información numérica.

Definiremos **Estadística** como: " La ciencia pura y aplicada (no exacta) que crea, desarrolla y aplica técnicas de modo que pueda evaluarse la incertidumbre.

La misma no es nueva , ya que desde edades antiguas se la usaba, principalmente en los conocidos censos.

Posteriormente (siglo XVII) se desarrolló la teoría de probabilidades, basada en los juegos de azar.

Muchas teorías, principalmente de carácter biológico como las de Mendel o Darwin tuvieron bases estadísticas.

Sin embargo la mayoría de los métodos de estadística moderna que se utilizan actualmente, no se desarrollaron hasta mediados del siglo XIX y principios del XX, principalmente para uso en biología, agricultura y genética.

Es interesante este punto, ya que al ser desarrollados estos métodos precisamente para ser usados en ciencias biológicas, ellos toman en cuenta la variabilidad propia de poblaciones naturales en sus cálculos y tablas.

**CAPITULO I**

**MANEJO DE DATOS**

**1.1.- Datos.-**

Llamamos **dato** a cualquier **observación** (valor numérico o cualitati­vo) que mida una característica o atributo (**variable**) de individuos. En otras palabras, a los valo­res experimentales que va a tomar una variable determinada.

Los datos estadísticos, obtenidos de **muestras**, experi­mentos o cualquier colección de mediciones, a menudo son tan numerosos que carecen de utilidad a menos que sean condensados o reducidos a una forma más adecuada. Por ello, en esta sección nos ocuparemos del agrupa­mien­to de datos, así como de ciertos estadísticos o medidas que representarán el significado general de nues­tros datos.

**1.2.- Variables.-**

Llamamos **variable** a una propiedad con respecto a la cual los individuos de una muestra se diferencian en algo verificable.

Variables **mensurables** son aquellas cuyos diferentes valores pueden ser expresados de manera numérica; por ejemplo, el peso y la longitud.

Variables **ordinales** son aquellas que pueden ser expre­sadas en orden de magnitud; por ejemplo, los grados de madurez o de llenura.

**Atributos** son aquellas variables que no pueden ser medidas, pero pueden ser expresadas cualitativamente. Ellos representan propieda­des; por ejemplo, el color.

Variables **discretas** son aquellas cuyo conjunto de posi­bles valores son fijos, y no pueden tomar valores in­termedios.

Definimos como variables **continuas** a aquellas cuyo conjunto de posibles valores puede alcanzar un número infinito entre dos valores cuales quiera.

Llamamos variables **independientes** a aquellas cuyo valor no depende de otra variable; matemáticamente, K varia­bles aleatorias son independientes, si y solo si se cumple lo siguiente:

 F(x1,x2,..xn)=F1(x1).F2(x2)..Fn(xn)

para todos los valores de estas variables en los cuales las funciones están definidas.

Llamamos variables **dependientes** a aquellas cuyo valor va a depender de otra función o, matemáticamente, si no cumple la condición anterior.

**1.3.- Población y muestras.-**

Llamamos **población** (estadística) al grupo de individuos bajo estudio, osea al conjunto de objetos, mediciones u observa­ciones del cual tomamos nuestra **muestra**.

Una población puede ser finita o infinita, dependiendo de su tamaño.

Conociendo la **distribución de frecuencias** de algúna característica (variable) de la población, es posible describirla por medio de una **función de densidad**, la cual a su vez vendrá caracterizada por ciertos **parámetros**.

Al ser la población completa generalmente muy grande para ser estudiada en su totalidad, resulta más conveniente estudiar solo un subconjunto de dicha población.

Sea una variable aleatoria dada (X); los valores de esta variable aleatoria (X1, X2,...Xn) forman una mues­tra de la variable X, si ellas son independientes y **siguen la misma distribución de X**, en otras palabras, si representa fielmente a X.

El objetivo de los muestreos es obtener información sobre las **distribuciones de frecuencia** de la población o más preciso de los **parámetros poblacionales**.

**1.4.- Distribución de frecuencias.-**

La distribución de frecuencias es una operación median­te la cual dividimos un conjunto de datos en un número de clases apropiadas, mostrando también el número de elementos en cada clase.

La primera etapa en la construcción de una distribución de frecuencias consiste en decidir cuántas clases uti­lizar, y elegir los límites de cada clase. En general, el número de clases depen­derá del número y rango de los datos. Matemáticamente, el número de intervalos (k) viene dado por la siguiente fórmula:

 

1

aunque siempre hay que ver qué tan bien representa ésto la veracidad de los datos. Empíricamente, se recomienda usar un número de intervalos no menor que 5 o mayor que 15.

Llamamos intervalo de representación al intervalo donde se representan los datos.

Intervalo real son los verdaderos límites del intervalo de representación, y viene dado por el punto medio entre los límites de dos inter­valos de representación consecutivos.

Definimos la marca de clase como el punto medio entre el límite superior y el inferior de un intervalo de representación.

La frecuencia es la cantidad de ocurrencias de datos dentro de un intervalo de representación.

La frecuencia relativa es la relación entre la frecuen­cia de un intervalo y la frecuencia total expresada en porcentaje.

La frecuencia acumulada y acumulada relativa es la sumatoria del número de ocurrencias o porcentajes de todos los intervalos menores o iguales al presente.

Por ejemplo, con los siguientes datos de longitud cefá­lica en:

*Tilapia nilótica*:

 29.0 29.0 27.0 28.2 28.9 26.1 28.4

 29.4 28.2 26.0 29.5 27.4 25.3 26.0

 27.0 27.6 26.0 29.7 29.8 26.3 28.5

 30.2 27.0 28.0 31.0 25.6 33.4 28.0

 29.0 28.0 29.5 26.4 27.3 29.3 26.0

 28.0 26.0 29.5 29.5 29.4 26.6 26.4

 28.0 27.7 28.1 27.6 26.8 27.0 29.3

 28.0 27.0 31.0 27.0 27.0 28.9 29.3

Construímos una tabla parecida a la si­guiente:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  IntervaloRepresentación |  Interval o Real |  Frecuenc ia |  Frecuenc ia Relativa |  Frecuenci a Acumulada |  F. Acumulada Relativa |  Marca de Clase |
|  24 - 25 |  23.5 - 25.5 |  2 |  3.57 % |  2 |  3.57 % |  24.5 |
|  26 - 27 |  25.5 - 27.5 |  19 |  33.93 % |  21 |  37.50 % |  26.5 |
|  28 - 29 |  27.5 - 29.5 |  29 |  51.79 % |  50 |  89.29 % |  28.5 |
|  30 - 31 |  29.5 - 31.5 |  5 |  8.93 % |  55 |  28.22 % |  30.5 |
|  32 - 33 |  31.5 - 33.5 |  1 |  1.78 % |  56 |  100.00 % |  32.5 |

**1.5.- Gráficos.-**

1

**FIGURA # 1.-** Histograma de fre­cuencias absolutas.

Las propiedades de las distribu­ciones de frecuencias relaciona­das con su forma se hacen más evidentes por medio de gráficos.

La forma más común de re­presen­tar una distribución de frecuen­cia es el histo­grama, en el cual el área de los rectángulos re­pre­sen­tan las frecuencias de clase, y sus bases se ex­tienden en las fronteras de los interva­los reales. En este tipo de gráfico, las marcas de clase están situadas en la mitad del rango del rec­tángulo. Me­diante este gráfico po­de­mos represen­tar la fre­cuencia o la frecuencia relativa (Figu­ras # 1 y # 2), pero no la frecuencia acumulada o acumu­lada re­lati­va.

Otros gráficos similares a los histogramas son los dia­gramas de barras (figu­ra # 3); aquí, las alturas y no las áreas representan las frecuencias de clase, y no se pretende fijar ningu­na escala horizontal continua; en otras pala­bras, el ancho de las ba­rras no interesa. Por esto se pue­den graficar tanto las fre­cuencias ab­solutas o rela­tivas, así como las acumula­das.

2**FIGURA # 2.-** Histograma de fre­cuencias relativas.

Otra forma de presentar las distribuciones de fre­cuen­cia en forma gráfica es el polí­gono de frecuen­cias. En él, las frecuen­cias de cla­ses son grafi­cadas so­bre las marcas de clase y uni­das mediante lí­neas rec­tas. Ade­más, agregamos va­lores co­rres­pondien­tes a cero en los puntos límites de la dis­tribución.

Con estos gráficos podemos re­presentar indistintamen­te las frecuencias netas o acumuladas; sin embargo, cuando graficamos estas últimas, en vez de utili­zar las marcas de clase como abscisas utiliza­mos el límite supe­rior del in­ter­valo real de frecuencia (Fi­gu­ras # 4 y 5).

3**FIGURA # 3.-** Gráfico de ba­rras de frecuencia absoluta

Para presentar datos de frecuen­cias relatívas, el grá­fico de sectores o "pie" se usa con mucha frecuencia (Figura # 6). Este corres­ponde a un círculo dividi­do en varios secto­res, correspondiendo cada uno a un intervalo, y en donde el área de cada sector es proporcional a la fre­cuencia relativa.

**1.6.- Estadísticos.-**

La mayoría de las investi­ga­cio­nes estadísticas se pro­ponen generalizar a partir de la in­formación conteni­da en **muestras aleatorias** acerca de la **pobla­ción** de donde fueron obte­nidas.

4**FIGURA # 4.-** Polígono de fre­cuen­cias absolutas.

En general, trataremos de hacer infe­rencias sobre los **parámetros de las poblaciones** (por ejemplo la media  o la varianza 2 ).

Para efectuar tales infe­rencias utili­zaremos **esta­dís­ti­cos mues­treales** como y s2; es decir cantidades calcu­la­das con base en **datos** u observaciones de la **muestra**.

5**FIGURA # 5.-** Polígono de frecuencias acumuladas

**1.6.1.- Estadísticos de cen­tra­liza­ción.-**

Podemos definir al **paráme­tro** **media poblacional**  como la media aritmética de todos los individuos de nuestra población, y re­presenta la esperanza ma­temática de nuestra varia­ble aleatoria:

2

6**FIGURA # 6.-** Gráfico de sectores

Este **parámetro** no lo cono­cemos, y no lo conoceremos nunca a no ser que mues­treá­ramos la pobla­ción completa. Es por esto que para estimar este paráme­tro utilizamos el **estadís­tico** **prome­dio** o media mues­treal.

Nuestro promedio va a estar dado por la media aritmé­tica de los datos de nuestra muestra. Su fórmula es:

 

3

Otros estadísticos de centrali­zación son la moda y la mediana.

La **moda** corresponde a la marca de clase del intervalo de clases con mayor frecuencia. En términos aproxima­dos, es el valor que mas encontramos en nuestro mues­treo.

La **mediana** correspondería al valor del dato que se encuentra más cercano a la mitad si ordenáramos nues­tros datos, o al valor del dato que tiene igual número de datos mayores de el que menores de él.

La mediana viene dada por el valor del dato número (n+1)/2 cuando n es impar y por la media del dato# (n/2) y el dato# (n/2 +1) cuando n es par.

**1.6.2.- Estadísticos de dispersión.-**

Las medidas de centralización nos dan una idea de hacia dónde están distribuidos nuestros datos, pero no de cómo están distribuídos. Para ésto tenemos las medidas de dispersión.

El **parámetro** **varianza poblacional** 2 mide el promedio de los cua­drados de las desviaciones de todos los valo­res de una variable de una población con respecto a la media po­blacional. Este parámetro equivale a la si­guiente fórmula:

 

4

La **varianza empírica** s2 es el **estadístico** mediante el cual hacemos estimaciones de nuestro **parámetro** varianza poblacional. Debido a que la varianza empírica es un estimador sesgado de la varianza poblacional, el cálcu­lo de la misma va a estar dada por:

 

5

Nótese que a medida que el tamaño de la muestra (n) aumenta, la diferencia entre 2 y s2 disminuye.

La **desviación típica** o desviación estándar ( o s), no es otra cosa que la raíz cuadrada positiva de la va­rianza.

El **rango** es la diferencia entre el valor de nuestro mayor dato y el valor de nuestro menor dato.

**CAPITULO II**

**TEORÍA DE PROBABILIDADES**

Eventos que son "comunes" o "improbables" son aquellos cuya probabi­lidad de ocurrencia son grandes o pequeñas, respectivamente.

Sin darnos cuenta, nosotros calculamos empíricamente la probabilidad de todas los sucesos que nos rodean; así, determinamos que tan "común" o "raros" son ciertos acontecimientos.

Los estadísticos remplazan las palabras informativas pero impre­cisas como "con dificultad", "pudo" o "casi con seguridad" por un número que va de 0 a 1, lo cual indica de forma precisa que tan probable o improbable es el evento.

Logicamente, haciendo inferencias a partir de muestras sobre una población, es decir de una parte sobre el todo, no podemos esperar llegar siempre a resultados correctos, pero la estadísti­ca nos ofrece procedimientos que nos permiten saber cuantas veces acertamos "en promedio". Tales enunciados se conocen como **enunciados probabilísticos**.

**2.1.- Espacio Muestral.-**

Llamamos **espacio muestreal** al conjunto universal de una población o a todos los valores probables que nuestra variable aleatoria puede tomar.

Enumerar todos los posibles sucesos del espacio mues­treal de una población dada puede ser tedioso y, en algunos casos, hasta imposible. En la práctica, la mayo­ría de los problemas no requieren de una enumera­ción completa, y basta con saber el número total de sucesos posibles del espacio muestreal.

Para contar los números de sucesos posibles sin tener que enumerar­los uno por uno, existen como reglas gene­rales la ley de la multiplicación y las reglas de las permutaciones y combinaciones.

**2.2.- Ley de la multiplicación**

Si A1, A2,... An son *n* acciones distintas que se pueden realizar de K1, K2,... Kn maneras, el total de maneras que se puede efectuar A1, A2,... An viene dado por:

 

6

**2.3.- Permutaciones y combinaciones**

Cualquier disposición **ordenada** de *r* elementos de un conjunto con *n* elementos (donde *r*  *n*) se llama permu­tación.

El número total de permutaciones de *r* objetos escogidos entre un conjunto de *n* objetos viene dado por:

 

7

Lo mismo que no es más que un caso especial de la ley de la multiplicación.

Una elección de *r* elementos de un conjunto de *n* elemen­tos distin­tos, **sin atender a un orden** se llama combina­ción.

El número total de combinaciones de *r* objetos escogidos entre un conjunto de *n* objetos viene dado por:

 

8

**2.4.- Definición de probabilidad**

Si un evento puede ocurrir de ***N*** maneras mutuamente exclusivas e igualmente posibles, y si ***n*** de ellas tie­nen una característica ***E***, entonces, la posibilidad de ocurrencia de ***E*** es la fracción ***n/N*** y se indica por:

 

9

En otras palabras, la probabilidad de que ocurra un evento determinado (exito) es la razón de el número de exitos posibles por el número de total de eventos distintos posibles (exitos mas fracsos).

Cabe anotar que la definición de exito o fracaso no tiene ninguna relación con la bondad del suceso, y es asignada de acuerdo a nuestras necesidades.

En general, para sucesos en los cuales el tamaño del espacio muestral nos sea desconocido o infinito, cuando no podamos enumerar la cantidad total de exitos, o en el caso de que las ***N*** maneras en que puede ocurrir el evento no sean igualmente posibles, se define la proba­bilidad de un evento como la proporción de las veces que eventos de la misma clase ocurren al repetir muchas veces el experimento en condiciones parecidas, y esta es la definición que más usaremos en este texto.

**2.5.- Teoremas básicos**

- La probabilidad de un evento cualquiera va a estar en el rango de cero a uno.

 0  P(E)  1

- La suma de la probabilidad de ocurrencia de un evento mas la probabilidad de no ocurrencia del mismo es igual a uno.

 P(E) + P(E) = 1

- Para dos eventos cualesquiera A y B, la probabilidad de que ocurra A o B viene dado, por la probabilidad de que ocurra A, mas la probabilidad de que ocurra B, menos la probabilidad de que ocurran ambos.

 P(A o B) = P(A) + P(B) - P(AB)

**CAPITULO III**

 **DISTRIBUCIÓN NORMAL**

Una distribución o **densidad de probabilidad** de una variable aleatoria **x** es la función de distribución de la probabilidad de dicha variable o, en otras palabras, la probabilidad de que dicha variable tome ciertos valores.

Estas distribuciones de probabilidad pueden ser discre­tas o continuas, de acuerdo con el tipo de variable al cual representen.

Hay infinidad de distribuciones de densidad, una para cada población, pero se han definido ciertas distribuciones "modelo" más comunes como la **Normal**, **binomial**, **Ji-cuadrado**, **"t" de Student**, **F de Fisher**; a las cuales podemos aproximar estas distribuciones particulares.

**3.1.- Características.-**

En el siglo XVIII, los científicos observaron con sor­presa que los grados de distribución muestreal de va­rios estadísticos, tales como la media, seguían una distribu­ción continua que denominaron "**curva normal de erro­res**", y le atribuyeron reglas de la probabilidad.

La ecuación de densidad de probabilidad normal viene dada por:

 

10

Para - < x < , y su representación gráfica (figura # 7 ) corresponde a la campana de Gauss.

Entre sus características tenemos:

- El área debajo de la curva entre 2 puntos dados re­presenta la probabilidad de que ocurra un hecho entre esos dos puntos;

- Su dominio va de menos infinito a más infinito;

- Es simétrica con respecto a su media;

- Tiene dos colas y es asintótica al eje x por ambos lados;

- El valor del área debajo de toda la curva es igual a 1;

7**FIGURA # 7.-** Curva normal (=0, =1)

- El centro de la curva está representado por la media poblacional ().

- Para cualquier curva normal, el área de - a + es igual a 0.6827; de -2 a +2 de 0,9545 y de -3 a +3 de 0,9973;

- La distribución muestreal de varios esta­dísticos, tales como la media, tie­nen una distri­bu­ción aproximadamente normal e indepen­diente de la configura­ción de la po­blación.

 Muchos aspectos naturales, como el peso de una espe­cie bioacuá­tica o la altura de personas de una pobla­ción, son característi­cas que presentan una dis­tribu­ción normal.

**3.2.- Distribución normal tipificada**

Llamamos distribución normal tipificada a la distribu­ción especial que representa a todas las variables aleatorias normales y que es la distribución de otra variable normal llamada ***Z***.

En donde ***Z*** va a ser igual a:

 

11

Y se la conoce como variable aleatoria estandarizada.

Esta función se caracteriza por tener media igual a cero (0) y desviación tipificada igual a uno (1).

La distribución tipificada se acostumbra a representar en forma acumulada. La tabla de la probabilidad acumu­lada para varios valores de ***Z*** permite encontrar la probabilidad de un intervalo dado a y b, tomando la diferencia entre probabilidades acumuladas de los valo­res de ***Z*** correspondientes sin tener que construir una distribu­ción de probabilidad.

Para buscar la probabilidad de que ***Z*** sea mayor a x, buscamos en la tabla # 1, localizando primero la fila que corresponda al primer entero y su primer decimal, y luego buscamos la columna que corresponda al segundo decimal.

**3.3.- Aplicaciones.**

Parte de la importancia práctica de esta distribución teórica de probabilidad estriba en que muchos fenómenos biológicos presentan datos distribuidos de manera tan suficientemente Normal que su distribución es la base de gran parte de la teoria estadística uada por los biólogos.

Ya que muchos fenómenos naturales, así como ciertos estadísticos como la media, siguen una distribución Normal o aproximadamente normal, podemos hacer algunas suposiciones a partir de ella. Entre ellas tenemos:

- Estimaciones de intervalos de confianza para la media.

- Pruebas de hipótesis con respecto a medias.

- Aproximaciones a otras distribuciones de probabili­dad.

**CAPITULO IV**

**OTRAS DISTRIBUCIONES**

Además de la distribución Normal estudiaremos otras distribuciones de probabilidad que nos serán de utili­dad a lo largo del curso.

Estudiaremos La distribución Ji-cuadrado, "t" de Stu­dent y la "F" de Fisher - Schnedecor; además definiremos la distribución binomial.

Existen muchas otras distribuciones de probabilidad, tanto continuas como discretas, pero su estudio no entrará en este manual.

**4.1.- Distribución "t" de Student.-**

Desarrollada con base en distribuciones de frecuencia empíricas en 1908 por William Gosset, conocido por el sobrenombre de "Student". Un cervecero-estadístico que encontraba ciertas dificultadaes al usar la distribución Normal en muestras pequeñas. Esta mísma tabla ya había sido calculada matemáticamente en 1875 por un astrónomo alemán, el cual sin embargo no le había encontrado utilidad práctica.

El problema recide en que la distribución muestreal de la media se ajusta muy bien a la distribución Normal cuando se conoce . Si **n** es grande, esto no presenta ningún problema, aun cuando  sea desconocida, por lo que en este caso es razonable sustituirla por ***s***. Sin embargo, en el caso de usar valo­res de **n** < 30, o sea en el caso de pequeñas mues­tras, esto no funciona tan bien.

Definiendo el estadístico ***t***:

 

12

Se puede probar que siendo el promedio de una muestra tomada de una población normal con media  y varianza 2, el estadístico ***t*** es el valor de una variable alea­toria con distribución "t" de student y parámetro  (grados de libertad) = n-1.

8**Figura # 8.-** Curva "t" de stu­dent.

Como puede apreciarse en la figura # 8 , la distri­bu­ción "t" es muy similar a la distribución normal, y entre sus característi­cas tenemos:

- Tiene media igual 0, es asintó­tica al eje x y su domi­nio va de - a +;

- El área bajo la curva desde - a + es igual a 1;

- Al igual que la distribu­ción Normal estándar, esta distribu­ción tiene media 0, pero su varianza depen­de del parámetro , deno­minado grados de liber­tad;

- La varianza de la distri­bución "t" excede a uno, pero se apro­xima a ese número cuando ; y,

- Al aumentar el valor de , la distribución "t" de stu­dent se aproxima a la distribución Normal, es más, para tamaños muestrea­les de 30 ó más, la dis­tribución Normal ofrece una excelente aproxima­ción a la distribución "t".

Entre las aplicaciones de esta distribución tenemos la estimación de intervalos de confianza para medias a partir de muestras pequeñas y las pruebas de hipótesis basadas en muestras < 30.

En la tabla # 2 se encuentran los valores de t, a la dere­cha de los cuales se encuentra un (100 x )% del área de la curva.

Para buscar el valor de t para  = n-1 en la tabla, primero localizamos la columna del co­rrespon­diente valor de  y la fila correspondiente al valor de . La intersección de la fila y la columna nos dará el valor de t.

**4.2.- Ji-cuadrado.-**

La distribución Ji-cuadrado es una función de densidad de probabi­lidad que sigue aproximadamente una distribu­ción gamma con  =  /2 y  = 2, y que representa la distribución muestreal de la varianza.

Definimos el estadístico Ji-cuadrado (2) como:

 

13

9**Figura # 9.-** Curva ji-cuadrado.

Entre sus características tene­mos:

- Es asimétrica y asintó­tica al eje x por la dere­cha;

- Su dominio va de 0 a + (Figu­ra # 8);

- El área bajo la curva desde 0 a + es igual a 1;

- Tiene parámetro  = n-1;

- Al incrementarse  se aproxima a la distribución normal; y,

- Representa la distribución muestreal de la varianza.

Entre las aplicaciones de esta distribución tenemos la determina­ción de intervalos de confianza para varian­zas, las pruebas de hipótesis para una varianza, el ajuste de datos a una distribución dada conoci­da y las pruebas de independencia.

En la tabla # 3 se dan algunos valores de 2 para va­rios valores de , donde 2 es tal que el área bajo la curva, a su derecha, es igual a . En esta tabla, la columna del lado izquierdo contiene los valores de , el primer renglón consta de áreas  en la cola del lado derecho de la distribución Ji-cuadrada y la tabla pro­piamente dicha la constituyen los valores de 2.

**4.3.- "F"de Fisher - Schnedecor.-**

Otra distribución de probabilidad muy usada es la "F" de Fisher - Schnedecor, la cual está relacionada con la distribución beta, y representa la distribución mues­treal de la razón de dos varianzas. Es decir que se obtiene de la razón de dos distribuciones Ji-cuadrado.

Definimos el estadístico F como:

 

14

El cual es el valor de una variable aleatoria que tiene distribu­ción F con parámetros 1=n1-1 y 2=n2-1.

10**Figura # 10.-** Curva F.

Su distribución gráfica la pode­mos apreciar en la figu­ra # 10.

Entre sus propiedades te­nemos:

- Es asimétrica, y asintó­tica al eje x por el lado derecho;

- Su dominio va de 0 a +;

- El área bajo la curva desde 0 a + es igual a 1; y,

- Tiene parámetros 1=n1-1 y 2=n2-1.

La tabla # 4 contiene los valores de F de  = .01 y .05 para varias combina­ciones de 1 y 2.

Entre las aplicaciones de esta distribución están las pruebas de hipótesis entre 2 varianzas y los análisis de varianza y covarianza.

**4.4.- Distribución binomial.-**

Muchos ensayos poseen solo dos resultados posibles, por ejemplo un animal sobrevive o no a cierto tratamiento, posee o no cierta característica. Estos fenómenos presentan generalmente una distribución de densidad asociada con la distribución **binomial**.

Entre las aplicaciones de la distribución binomial tenemos los intervalos de confianza para poporciones y las pruebas de hipótesis para proporciones.

**CAPITULO V**

 **BONDAD DE AJUSTE**

**5.1.- Generalidades**

Para usar las diferentes distribuciones estudiadas, en nuestros problemas prácticos, debemos primero de asegu­rarnos que nuestros datos se aproximen a una de estas distribuciones dadas.

Hablamos de bondad de ajuste cuando tratamos de compa­rar una distribución de frecuencia observada con los correspondientes valores de una distribución esperada o teórica.

Estudiaremos la prueba ji-cuadrado para bondad de ajus­te, la cual sirve tanto para distribuciones discre­tas como para distribuciones continuas.

También existe una prueba no paramétrica denomi­nada de Kolmogo­roff-Smirnov, la cual sólo sirve para distribu­ciones continuas, y no será estudiada por nosotros.

**5.2.- Prueba Ji cuadrado**

Para realizar esta prueba debemos de seguir los siguientes pasos:

- Ordenamos nuestros datos separándolos en k rangos o clases, cuidando de que en cada rango (i) la frecuencia observada (oi) sea > 4;

- Contamos las frecuencias observadas en cada clase (oi);

- Decidimos la distribución a la cual queremos ajustar nuestros datos, expresando la hipótesis nula y su alterna:

 **H0 =** Existe un buen ajuste.

 **H1 =** No existe un buen ajuste.

- Calculamos la frecuencia teórica esperada ei para cada in­tervalo i, siendo ésta igual al producto del tamaño muestreal N por la probabilidad de dicho rango obte­nida de la tabla correspondiente;

- Calculamos el estadístico 2:

 

15

 El cual sigue una distribución ji-cuadrada con  = k - m - 1 grados de libertad, en donde k es el número de intervalos y m es el número de parámetros estimados. Por ejemplo, para la distribución normal utilizaremos m=2 (media y varian­za); y,

- Si el valor de 2 calculado es menor que el corres­pondiente valor de 2() de la tabla # 3, concluimos que existe un buen ajuste; de lo contrario, no.

 W={2 > 2}

**CAPITULO VI**

 **ESTIMACIÓN**

**6.1.- Puntual.-**

En esencia, la estimación puntual se refiere a la elec­ción de un estadístico calculado a partir de datos muestreales, respecto al cual tenemos alguna esperanza o seguridad de que esté "razonable­mente cerca" del pará­metro que ha de estimar.

Llamamos estimador insesgado a un estadístico cuyos va­lores promedios son iguales a los del parámetro que trata de estimar.

De dos estadísticos dados, **1 y **2, podemos decir que el más eficiente estimador de ** es aquel cuya varianza de distribución muestreal es menor.

Para poblaciones normales, el estimador más eficiente de  es el promedio ().

Sin embargo, debemos recordar que cuando empleamos una media muestreal para estimar la media de una población, la probabilidad de que la esti­mación sea en realidad igual a  es prácticamente nula. Por esta razón es conveniente acompañar la estimación puntual de  con una afirmación de cuán cerca podemos razonablemente esperar que se encuentre la estimación. Esto viene dado por:

 

16

para un porcentaje de confianza de 100 x 1- y n30.

Para el caso de tener una muestra de tamaño menor que 30, pero pudiendo suponer razonablemente que estamos mues­treando una población normal, E viene dado por la si­guiente fórmula:

 

17

para un porcentaje de confianza de 100  (1-) y para  = n-1 grados de libertad.

En lo que respecta a la varianza poblacional, el esti­mador insesgado más eficiente es la varianza muestreal:

 

18

Y, para proporciones, el estimador insesgado mas eficiente del **parámetro** proporción poblacional (**p**) es el estadístico proporción muestral (**x/n**).

 

19

En donde **x** es el número de observaciones con un caracter determinado y **n** es el número total de observaciones (x + x).

**6.2.- Estimación por intervalos.-**

Dado que no puede esperarse que las estimaciones pun­tuales realmente coincidan con las cantidades que in­tentan estimar, a veces es preferible reemplazarlas con estimaciones por intervalos, esto es, con intervalos en los cuales podemos esperar con un grado razonable de certeza que contengan al parámetro en cuestión.

**6.2.1.- Estimas de medias por intervalos.-**

Para poblaciones con distribución normal, conociendo la varianza poblacional , o en su defecto teniendo una muestra grande (n  30),el intervalo de confianza para  viene dado por:

 

20

con un nivel de confianza de 100  (1-)%.

Para muestras aleatorias pequeñas (n<30), tomadas de una población presumiblemente normal y de varianza desconocida, el intervalo de confianza para las medias vendrá dado por:

 

21

con un nivel de confianza de 100  (1-)% y n-1 grados de libertad.

**6.2.2.- De varianzas.-**

El intervalo de confianza para la varianza en poblacio­nes normales viene dado por:

 

22

con un nivel de confianza de 100  (1-)% y n-1 grados de libertad.

**6.2.3.- De proporciones.-**

Para proporciones, el intervalo de confiansa de la proporción poblacional **p** calculada a partir de la proporción muestral **x/n** vendrá dado por:

 

23

con un nivel de confianza de 100  (1-)% y n-1 grados de libertad.

**6.3.- Gráfico de intervalos.-**

11**FIGURA # 11.-** Gráfico de inter-valos confianza.

Los intervalos de confianza pueden representarse gráfi­camente (Figura # 11), elaborando un gráfico de lineas para las medias, y repre­sentando en el mismo me­diante barras los límites superior e inferior del intervalo de confianza.

Este tipo de gráfico tiene la ventaja de permitir ob­servar la precisión de nuestra estimación.

**CAPITULO VII**

 **PRUEBAS DE HIPÓTE­SIS**

**7.1.- Generalidades.-**

Llamamos hipótesis estadística a una asumción sobre una pobla­ción que está siendo muestreada.

Un test de hipótesis es simple­mente una regla mediante la cual esta hipótesis se acepta o se rechaza.

Esta regla está basada generalmente en un estadístico muestreal llamado estadístico de prueba, ya que se lo usa para probar la hipótesis.

La región crítica de un estadístico de prueba consiste en todos los valores del estadístico donde se hace la decisión de rechazar H0.

Debido a que las pruebas de hipótesis están basadas en estadísticos calculados a partir de *n* observaciones, la decisión tomada está sujeta a posibles errores.

Si rechazamos una hipótesis verdadera, estamos come­tiendo un error del tipo I. La probabilidad de cometer un error del tipo I se llama .

Si aceptamos una hipótesis falsa, estaremos cometiendo un error del tipo II, y la probabilidad de cometerlo se la denomina .

Uno de los objetivos de las pruebas de hipótesis es diseñar tests en donde  y  sean pequeños.

Para probar una hipótesis, generalmente la expresamos en su forma nula (**H0**), y formulamos una hipótesis al­terna (**H1**) que aceptaremos al rechazar la nula.

Ambas hipótesis deben ser distintas y mutuamente exclu­yentes.

Los pasos básicos para efectuar la mayoría de las prue­bas de hipótesis son los siguientes:

1.- Expresar claramente la hipótesis nula (H0) y su alterna (H1);

2.- Especificar el nivel de significancia  y el tamaño de la muestra (n);

3.- Escoger un estadístico para probar H0, tomando en cuenta las asumciones y restricciones que involucran usar este estadístico;

4.- Determinar la distribución muestreal de este esta­dístico cuando H0 es verdadera;

5.- Designar la región crítica de la prueba, en la cual H0 va a ser rechazada en 100% de las muestras cuando H0 es verdadera;

6.- Escoger una (dos) muestra(s) aleatoria(s)de tama­ño n;

7.- Calcular el estadístico de prueba; y,

8.- Comparar el estadístico calculado con el teórico y decidir:

 a) aceptar H0.

 b) rechazar H0 (y aceptar H1).

 c) no tomar ninguna decisión.

**7.2.- Una población.-**

Llamadas también pruebas unimuestrales. Aquí tratamos de probar la igualdad de un parámetro poblacional cono­cido ** con un parámetro poblacional calculado a partir de un estadístico **o.

Estudiaremos tres pruebas en este capítulo: una media con va­rianza conocida, una media con varianza descono­cida y una varianza.

**7.2.1.- Una media con varianza conocida.-**

Utilizamos esta prueba para comparar una media pobla­cio­nal conocida () con la calculada a partir de el prome­dio de una muestra cuya varianza conocemos(0); o, que en su defecto, su tamaño sea  30, por lo que supondremos que la varianza poblacional sea igual a la muestreal.

Las hipótesis a probar son:

**H0 =** 0 = 

**H1 =** 0  

El estadístico de prueba usado es:

24

donde Z sigue una ley normal tipificada (0,1).

La región crítica o de rechazo (W) viene dada por:

 \_Z\_  Z(/2)

Se pueden probar también estas hipótesis alternas con sus respecti­vas regiones de rechazo (W):

**H0 =** 0 = 

**H1 =** 0 <  W = {Z  -Z()}

ó,

**H0 =** 0 = 

**H1 =** 0 >  W = {Z  Z()}

**7.2.2.- Una media con varianza desconocida.-**

Utilizamos esta prueba para comparar una media pobla­cio­nal conocida () con la calculada a partir de el prome­dio de una muestra cuya varianza no conocemos (0), y cuyo tamaño sea < 30. En este caso trabajaremos con la va­rianza muestreal en vez de la poblacional, pero usare­mos el estadístico de prueba "t".

Las hipótesis a probar son:

**H0 =** 0 = 

**H1 =** 0  

El estadístico de prueba usado es:

 

25

donde t sigue una ley "t" de student con n-1 grados de libertad.

La región crítica o de rechazo (W) viene dada por:

 \_t\_  t(/2)

Se puede probar también estas hipótesis alternas con sus respecti­vas regiones de rechazo (W):

**H0 =** 0 = 

**H1 =** 0 <  W = {t < -t()}

ó,

**Ho =** 0 = 

**H1 =** 0 >  W = {t  t}

**7.2.3.- Una varianza.-**

Utilizamos esta prueba para comparar una varianza po­bla­cio­nal conocida (2) con una calculada a partir de la varianza muestral de una población muestreada (20), supo­niendo con cierto grado de confianza que esté nor­mal­mente distribuida. En este caso trabajamos con el esta­dístico de prueba 2.

Las hipótesis a probar son:

**Ho =** 20 = 2

**H1 =** 20  2

El estadístico de prueba usado es:

 

26

donde 2 sigue una ley Ji - cuadrado con  = n-1 grados de liber­tad.

La región crítica o de rechazo (W) viene dada por:

 2(/2)  2 2(1-/2)

Se puede probar también estas hipótesis alternas con sus respecti­vas regiones de rechazo (W):

**Ho =** 20 = 2

**H1 =** 20 < 2 W = {2  2(1-)}

ó,

**Ho =** 20 = 2

**H1 =** 20 > 2 W = {2  2}

**7.3.- Dos poblaciones independientes.-**

Llamadas también pruebas bimuestreales, son usadas cuan­do queremos comparar dos estadísticos poblacionales calculados a partir de muestras de esas poblaciones.

En este capítulo estudiaremos cuatro casos: dos varian­zas indepen­dientes, dos medias independientes con va­rianzas conocidas, dos medias independientes con va­rianzas desconocidas e iguales y dos medias indepen­dientes con varianzas des­co­nocidas y desi­guales.

**7.3.1.- Dos Varianzas.-**

Utilizamos esta prueba para comparar dos varianzas po­bla­cio­nales (21 y 22), calculadas a partir de las varianzas muestrales (s21 y s22) de poblaciónes muestreadas, supo­niendo con cierto grado de confianza que estén normal­mente distribuidas. Consideramos s21 a la mayor de las dos.

En este caso trabajamos con el esta­dístico de prueba *F*.

Las hipótesis a probar son:

**Ho =** 21 = 22

**H1 =** 21  22

El estadístico de prueba usado es:

 

27

donde *F* sigue una ley F de Fisher - Schnedecor con n1-1 y n2-1 grados de libertad.

La región crítica o de rechazo (W) viene dada por:

 *F* > F(/2)

Se puede probar también esta hipótesis alterna con su respectiva región de rechazo (W):

**Ho =** 21 = 22

**H1 =** 21 > 22 W = {*F*  F()}

**7.3.2.- Medias, varianzas conocidas.-**

Utilizamos esta prueba para comparar dos medias pobla­cio­nales calculadas a partir del prome­dio de dos mues­tras cuyas varianzas conocemos, o en su de­fec­to,cuyo ta­maño individual sea  30, en donde supondre­mos que la varianza poblacional sea igual a la mues­tral.

Las hipótesis a probar son:

**Ho =** 1 = 2

**H1 =** 1  2

El estadístico de prueba usado es:

 

28

donde *Z* sigue una ley normal tipificada (0,1).

La región crítica o de rechazo (W) viene dada por:

 \_*Z*\_  Z(/2)

Se puede probar también esta hipótesis alterna con su respectiva región de rechazo (W):

**Ho =** 1 = 2

**H1 =** 1 > 2 W = {*Z*  Z()}

**7.3.3.- Medias, Varianzas desconocidas e iguales.-**

Utilizamos esta prueba para comparar dos medias pobla­cio­nales calculadas a partir del prome­dio de dos mues­tras cuyas varianzas no conocemos, y cuyos tama­ños sean < 30, siempre y cuando hayamos demostrado con anterio­ridad, mediante una prueba F, que las varianzas po­bla­cionales de ambos son iguales.

En este caso trabajaremos con la va­rianza mues­tral en vez de la poblacional, pero usaremos el esta­dístico de prueba "t".

Las hipótesis a probar son:

**Ho =** 1 = 2

**H1 =** 1  2

El estadístico de prueba usado es:

 

29

donde *t* sigue una ley "t" de student con n-1 grados de libertad.

La región crítica o de rechazo (W) viene dada por:

 \_t\_  t(/2)

Se puede probar también esta hipótesis alterna con su respectiva región de rechazo (W):

**Ho =** 1 = 2

**H1 =** 1 > 2 W = {t  t()}

**7.3.4.- Medias, varianzas desconocidas y desiguales.-**

En el caso de que no se haya demostrado la igualdad de varianzas entre las poblaciones mediante una prueba F, no sería necesario realizar un test de medias, ya que las poblaciones son tan heterogéneas respecto a sus varianzas.

Si, a pesar de esto, se desea realizar una prueba de medias, (problema de Behrens - Fisher), se puede reali­zar de varias maneras, siendo una de ellas la prueba de Smith - Satterthwaite, usando el estadístico:

 

30

el cual sigue una ley "t" de student con grados de libertad igual a:

 

31

Las hipótesis y áreas críticas son las mismas usadas en la prueba "t" bimuestreal.

**7.4.- Pruebas en muestras dependientes.-**

Esta prueba se aplica a muestras de poblaciones depen­dientes la una de la otra, como en el caso de pruebas en una misma población antes y después de un tratamien­to.

La lógica de la prueba se reduce a determinar las dife­rencias en los individuos de la muestra, y a probar si la media de estas diferencias son iguales a cero.

Las hipótesis a probar son:

Ho:  = 0

H1:   0

El estadístico de prueba a usar es:

 

32

donde *t* sigue una ley "t" de student con  = n-1 grados de libertad; es el promedio de las diferencias entre ambas muestras; sd es la desviación estandar muestral de las diferencias y n es el número de diferencias.

La región de rechazo (W) viene dada por:

 \_*t*\_  t(/2)

Otras formas de tratar muestras dependientes es mediante un diseño de bloques aleatorios o mediante un análisis de covarianza.

**7.5.- Pruebas de proporciones.-**

Se usan para comparar proporciones o porcentajes.

Exis­ten dos casos. El primero es comparar una proporción poblacional conocida **p** con una calculada a partir de un estadístico **p0**, donde el producto np debe ser mayor o igual que cuatro.

Aquí las hipótesis a probar son:

Ho: p0 = p

H1: p0  p

Y el estadístico de prueba usado es:

 

33

el cual sigue una ley normal tipificada (0,1), p es la proporción conocida; q = 1- p y po es la pro­por­ción calculada a partir de la muestra.

El área crítica o de rechazo (W) viene dada por:

 *\_Z\_* > Z(/2)

Se pueden probar también estas hipótesis alternas con sus respecti­vas regiones de rechazo (W):

**Ho =** p0 = p

**H1 =** p0 < p W = {*Z* < -Z()}

ó,

**Ho =** p0 = p

**H1 =** p0 > p0 W = {*Z* > Z()}

para el otro caso se tienen dos proporciones calcula­das a partir de muestras. En este caso las hipótesis a probar son:

**Ho =** p1 = p2

**H1 =** p1  p2

El estadístico de prueba usado es:

 

34

donde *Z* sigue una ley normal tipificada (0,1), y p es la proporción de ambas muestras juntas:

 

35

La región crítica o de rechazo (W) viene dada por:

 \_*Z*\_  Z(/2)

Se puede probar también esta hipótesis alterna con su respectiva región de rechazo (W):

**Ho =** p1 = p2

**H1 =** p1 > p2 W = {*Z*  Z()}

**7.6.- Pruebas de independencia.-**

En el caso de que queramos determinar la dependencia o independen­cia de dos variables **x** y **y** entre si utiliza­mos la prueba ji-cuadrado para independencia.

Para realizar la misma seguimos los siguientes pasos:

- Enunciamos la hipótesis nula y su alterna:

 **H0** : Los caracteres son independientes.

 **H1** : Los caracteres son dependientes.

- Construimos una tabla de doble clasifica­ción, divi­diéndola en **s** filas y **r** columnas, en donde **r** es el número de modalidades del carácter **x** y **s** es el número de modalidades del carácter **y**, poniendo en cada casi­lla fila por columna (ij), el número de observaciones que poseen el carácter **x** con la moda­lidad **i** y el carácter **y** con la modalidad **j** (nij):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  y x | 1 .. |  .. j ..  |  .. r |  (ni.) total |
|  1 . |  n11 |  n1j |  n1r |  (n1.)  n1j |
|  i . |  ni1 |  nij |  nir |  (ni.)  nij |
|  s  |  ns1 |  nsj |  nsr |  (ns.)  nsj |
|  (n.j) total |  (n.1)  ni1 |  (n.j)  nij |  (n.r)  nir |  N  nij  |

- Calculamos los valores esperados (npij) para cada casillero fila por columna (ij), mediante la fórmula:

 

36

- Calculamos el valor del estadístico 2:

 

37

 el cual sigue una ley ji-cuadrado con =(r-1)(s-1) gra­dos de libertad.

- La región de rechazo viene dada por:

 W = {2 > 2()}

**CAPITULO VIII**

 **DISEÑO EXPERIMENTAL**

**8.1.- Generalidades.-**

Llamamos investigación a la "búsqueda sistemática de la verdad no descubierta" (Leedy, 1974), poniendo especial atención a la palabra "sistemática".

Un experimento puede ser definido como "un estudio donde cierta(s) variable(s) independiente(s) es(son) manipulada(s), y su(s) efecto(s) en una o más variables independiente(s) determinado(s), siendo los niveles de esta(s) variable(s) independientes asignados al azar" (Hicks, 1982).

El experimento debe incluir una enunciación espe­cifi­ca del problema que se quiere solucionar, o sea que se debe describir exactamente lo que se quiere averiguar. No es bueno enunciar el problema en forma general, sino más bien en una forma en que todos los puntos queden claros y que nos muestren la forma en que la investiga­ción debe ser llevada.

El enunciado del problema debe incluir la referencia a uno o más criterios (variables dependientes) usados para determinar el problema. Se debe determinar si estos criterios pueden ser medidos, con cuánta preci­sión y con qué medios.

El enunciado también debe definir los factores (varia­bles independientes) que afectarán a estos criterios y si éstos van a ser mantenidos fijos o variados a cier­tos niveles o al azar, el número de factores , el tipo y su disposición, y también si estos factores van a ser combinados y en qué forma, todo lo cual determinará el tipo de cálculo estadístico a realizar.

Los factores (variables independientes) básicamente pue­den ser manejados así:

- Controlados rígidamente y mantenidos fijos a lo largo del experimento, con lo cual los resultados obtenidos son válidos solamente para estas condiciones fijas.

- Controlados a niveles fijos de interés.

- Aleatorios para promediar el efecto de variables que no pueden ser controladas.

Un principio básico es maximizar el efecto de la(s) variable(s) de interés, minimizar la varianza del error y controlar ciertas variables a niveles específicos.

El diseño del experimento involucra la cantidad de observaciones que se van a realizar, el orden en el cual se va a efectuar (que debería ser aleatorio para promediar las diferencias de ciertas variables que no podemos controlar, así como para poder asumir que los errores en medición son independientes), y tambien el orden de aleatoriedad a usar.

Al haber decidido esto se debe de mantener lo más ape­gado posible al plan trazado.

Debemos además estar al tanto de las restricciones del modelo que estamos usando y expresar el problema en forma de una hipótesis que podamos probar, y su hipóte­sis alterna.

Debe de tomarse en cuenta en el diseño de forma muy especial las restricciones de tipo logístico y finan­cieras, ya que habrá que hacer un compromiso entre estas y lo óptimo en terminos matemáticos.

El análisis de datos es el último paso e incluye la realización mecánica de pasos ya decididos como son:

- Recolección y procesamiento de datos.

- Cálculo de ciertos estadísticos de prueba usados para hacer una decisión.

- La toma de la decisión en sí.

- La exposición de los resultados.

Los resultados deben ser expuestos de forma clara, precisa e intelegible, y los cálculos estadísticos pue­den ser incluidos en los apéndices, sólo de ser necesa­rio.

**8.2.- Modelos de diseño.-**

Existen innumerables modelos de diseño de experimentos, no siendo el propósito de este manual el abarcarlos todos; sin embargo se explicarán algunos modelos de diseños bási­cos, lo que facilitará la comprensión de otros diseños que se encuentren en otros libros más avanzados.

**8.2.1.- Diseño de un factor, completamente aleatorio.-**

Siempre que sólo se varíe un factor, ya sea cualitativo o cuantitativo, fijo o al azar, el experimento se con­sidera **de un factor**.

Dentro de este factor existirán varios niveles o trata­mientos, cuyo número lo designaremos como **k**. El objeti­vo del experimento es determinar si existen diferencias entre los efectos de los tratamientos.

Si el orden de experimentación aplicado a los distintos tratamientos es completamente aleatorio, de forma que se consideren aproximadamente homogéneas las condicio­nes en que se está trabajando, llamaremos a este **diseño** **completamente aleatorio**.

Cada tratamiento tendrá ni observaciones, sin importar que el tamaño de las muestras no sea igual de un trata­miento a otro.

El modelo matemático vendrá dado por:

 

1

Lo que significa que cada iésima observación del jésimo tra­tamiento va a ser igual a la media poblacional () más un efecto del jésimo tratamiento (j) mas un error alea­torio para el mismo ( ij).

Este diseño se lo resuelve generalmente mediante un ANOVA de una vía con la hipótesis nula:

H0 : j = 0 ; para todos los j.

**8.2.2.- Diseño de un factor, bloques aleatorios.-**

En el caso de que no todas las observaciones que poda­mos realizar para los **n** tratamientos puedan considerar­se homogéneas, y es más, podamos considerar que hay fuen­tes conocidas de variabilidad, nos podemos librar de esta variabilidad dividiendo las observaciones de cada clasificación en bloques.

En el caso de que en cada bloque haya una observación de cada tratamiento, y siempre y cuando los tratamien­tos sean asignados al azar dentro de cada bloque, deno­minamos a este **diseño en bloques aleatorios**.

El modelo matemático usado vendrá dado por:

 

2

en donde j es el efecto del jésimo bloque y i es el efecto del iésimo tratamiento.

Este diseño se lo resuelve mediante un ANOVA de dos vías, en donde la hipótesis fundamental a probar es:

H0 : i = 0 , para todos los tratamientos,

pudiéndose también probar:

H0 : j = 0 , para todos los bloques,

la cual, al rechazarse, nos indicará que el criterio de clasificación en bloques ha sido correcto.

**8.2.3.- Cuadrado Latino.-**

En el caso de que queramos eliminar dos fuentes de variación conocidas, en un experimento de un solo fac­tor, utilizaremos el **diseño de cuadrado latino**, el cual es un diseño en el cual cada tratamiento aparece una y sólo una vez en cada fila (1a fuente de variación), y una y sólo una vez en cada columna (2a fuente de varia­ción).

Debemos recordar que aunque trabajamos con dos restric­ciones a la aleatoriedad, seguimos trabajando con solo un factor.

Este diseño es posible únicamente cuando el tamaño de cada una de las restricciones es igual al número de tratamientos.

El tipo de ANOVA a usarse es el de tres vías.

En caso de tener tres restricciones a la aleatoriedad, estaremos frente a un diseño de **cuadrado greco-latino**.

**8.2.4.- Experimentos Factoriales.-**

Llamamos **experimento factorial** a aquel en el cual todos los niveles de un factor (variable independiente 1) son comparados con todos los niveles de otro(s) factor(es) (varia­ble independiente 2(n) ).

El modelo matemático usado vendrá dado por:

 

3

en donde ij es el efecto del iésimo tratamiento del 1er factor con el jesimo tratamiento del 2do factor y k el efecto de la kesima repetición.

La diferencia de un experimento factorial con otro dise­ñado en bloques, es que en el diseño en bloques nos inte­resa solamente los efectos de **un factor** aunque lo sepa­remos en bloques para **eliminar las interferencias de otro factor**. En un experimento factorial, en cambio, estamos interesados en **evaluar el efecto de estos dos o más factores**, así como sus posibles **interacciones**.

Una interacción ocurre cuando un cambio en un factor produ­ce un distinto cambio en nuestra variable a un nivel de otro factor que a otro nivel de este otro factor.

Algunas de las ventajas de los diseños factoriales son:

- Es posible mayor eficiencia que con el diseño de experimentos de un factor.

- Todos los datos son usados para calcular todos los efectos.

- Se adquiere información sobre posibles interacciones entre los dos tratamientos.

**8.3.- Determinación del tamaño muestreal.-**

Uno de los principales factores que hay que considerar en nuestro diseño experimental es el del tamaño de la muestra.

En general, es recomendable tomar la muestra tan grande como sea posible; sin embargo, podemos calcular matemá­ticamente el número mínimo necesario con base en lo siguiente:

- La mínima diferencia que deseamos detectar ().

- La variación existente en la población ().

- La máxima probabilidad de error que deseamos tomar ( y ).

Para estimaciones puntuales, nuestro tamaño muestral vendrá dado por:

 

4

Para pruebas de hipótesis tenemos la fórmula:

 

5

para ensayos de una cola, y la misma usando Z(/2) y Z(/2) para ensayos de dos colas.

A pesar de esto, los tamaños muestreales son general­men­te escogidos de forma arbitraria, debido a limita­ciones de orden económico y práctico.

**8.4.- Datos atípicos.-**

Llamamos datos atípicos a aquellos datos que debido a su lejanía de los otros datos de nuestra muestra, pode­mos considerarlos como no pertenecientes a la pobla­ción, o como datos que desvían nuestra muestra del verdadero valor de la población.

Es importante tener un criterio objetivo matemático ya que considerar subjetivamente los mismos puede tener serias consecuencias.

Existen varios criterios para determinar que números son atípicos, generalmente en base del tamaño de la muestra y de su varianza. Nosotros estudiaremos el criterio de Chauvenet.

Este criterio considera atípicos a los datos que se alejan de la media más de 1/2n de lo que lo haría una pobla­ción normal.

Calculamos para cada dato el estadístico:

 

6

y, si el valor del mismo es mayor que el valor corres­pondiente en la tabla # 5, consideramos al dato atípico y lo eliminamos.

Hay que tener cuidado de no a realizar este procedi­mien­to más de una vez en una muestra, porque corremos el riesgo de que al disminuir nuestra desviación estan­dar eliminemos datos que no sean atípicos.

Otros criterios para determinar datos atípicos son el test de Dixon y el de Grubb.

**CAPITULO IX**

 **ANÁLISIS DE VARIANZA**

**9.1.- Generalidades.-**

Al estudiar las pruebas de hipótesis, veíamos que éstas se usaban para comparar una o dos poblaciones.

En el caso de querer comparar más de dos poblaciones usando pruebas bimuestreales, tendríamos que determinar las diferencias entre cada par posible. Esto, además de ser tedioso, aumentaría la posibilidad total de un error hasta niveles prohibitivos.

Sin embargo, tenemos el aná­lisis de varianza (ANOVA), el cual no es otra cosa que una prueba de hipótesis para más de dos muestras, en donde tratamos de averi­guar si las poblaciones que muestreamos son todas igua­les en lo que respecta a sus medias, o si en su defec­to, hay por lo menos una que sea distinta.

Así, las hipótesis básicas que aplicaremos en un ANOVA serán:

Ho : 1 = 2 = ... = n (todas las medias son iguales)

H1 :  i tal que   i (al menos hay una media desi­gual)

Sin embargo, existen ANOVAs que no sólo comparan trata­mientos, sino bloques; e incluso hay los que comparan varios factores y su efecto conjunto, por lo que estas hipótesis básicas pueden ser complementadas con otras, dependiendo del caso.

**9.2.- ANOVA de una vía.-**

Utilizamos un ANOVA de una vía cuando queremos comparar las medias en un experimento diseñado por azar simple.

Las hipótesis a probar son:

Ho : 1 = 2 = ... = n (todas las medias son iguales)

H1 :  i tal que   i (al menos hay una media desi­gual)

Para facilidad en la realización de los cálculos, uti­li­zaremos las sgtes tablas:

 T R A T A M I E N T O S

 1 2 . . . . . . j . . . . . k

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  Y11 |  Y12 |  . |  . |  . |  Y1j |  . |  . |  . |  Y1k |  |
|  Y21 |  Y22 |  . |  . |  . |  Y2j |  . |  . |  . |  Y2k |  |
|  . |  . |  |  |  |  . |  |  |  |  . |  |
|  . |  . |  |  |  |  . |  |  |  |  . |  |
|  Yi1 |  Yi2 |  . |  . |  . |  Yij |  . |  . |  . |  Yik |  |
|  . |  . |  |  |  |  . |  |  |  |  . |  |
|  . |  Yn22 |  |  |  |  |  |  |  |  . |  |
|  . |  |  |  |  |  Ynjj |  |  |  |  . |  |
|  . |  |  |  |  |  |  |  |  |  . |  |
|  Yn1l |  |  |  |  |  |  |  |  |  Ynkk |  |
|  |  |  |  |  |  |  |  |  |  |  TOTALES |
|   Yil |   Yi2 |  . |  . |  . |   Yij |  . |  . |  . |   Yik |   Yij |
|  n1 |  n2 |  . |  . |  . |  nj |  . |  . |  . |  nk |  N |
|   Y2i1 |   Y2i2  |  . |  . |  . |   Y2ij |  . |  . |  . |   Y2ik |   Y2ij |

En donde **k** es el número de tratamientos; n1, n2, ... nk, los tamaños de las muestras de cada uno de los k tra­ta­mientos;  Yij, la suma de todas las muestras; N es el número total de muestras; y,  Y2ij la suma de los cua­drados de todas las muestras.

La siguiente es la tabla del ANOVA:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  FUENTE |  Gdos. de lbtad. |  Suma de Cuadrados (SC) |  Suma media de cuadrados |  *F* calcul. |
| TRATAMIENTO |  K - 1 |  ( Yij)2/nj - ( Yij)2/N |  SCT/(K - 1) |  SMCT/SMCE |
| ERROR |  N - K |  SC TOTAL - SCT |  SCE / (N - K) |  |
| TOTAL |  N - 1 |   Y2ij - ( Yij)2/N |  |  |

En donde la relación **SMCT / SMCE** es nuestro estadístico de prueba llamado *F*, y que sigue una distribución "F" de Fis­her - Schnedecor con 1 = K-1 y 2 = N-K grados de liber­tad.

La zona crítica o de rechazo (W) viene dada por:

 *F*  F()

**9.3.- ANOVA de dos vías.-**

Utilizamos un ANOVA de dos vías cuando queremos compa­rar las medias de un experimento con dos criterios de clasificación como, por ejemplo, con un diseño por blo­ques aleatorios; es decir, cuando tene­mos otra fuente de variación conocida, además de nuestro tratamiento.

De esta forma tomamos en cuenta el error que podría haber al considerar iguales a dos poblaciones con el mismo tratamiento, pero realizado en distintas condi­ciones.

Se lo puede utilizar también en bloques incompletos, pero este caso no será estudiado en este manual.

También lo podemos usar para probar muestras dependien­tes como, por ejemplo, las pruebas de antes y después.

Las hipótesis básicas a probar son:

**Ho** : 1 = 2 = ... = n (   de los trat. son igua­les)

**H1** :  i tal que   i (  de los trat. desi­gual)

Es decir, probar los efectos de los tratamientos.

Pudiéndose también probar:

**Ho'** : a = b = ... = k (   de los bloques son igua­les)

**H1'** :  j tal que   j (  de los bloques desi­gual)

O sea, probar los efectos de los bloques.

Las tablas usadas en este caso son las siguientes:

 T R A T A M I E N T O

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  BLOQUES  |  A |  B |  . |  . |  . |  i |  . |  . |  . |  n |  Ti |
|  1 |  YA1 |  YB1 |  . |  . |  . |  Yi1 |  . |  . |  . |  Yn1 | nYi1 |
|  2 |  YA2 |  YB1 |  . |  . |  . |  Yi2 |  . |  . |  . |  Yn2 | nYi2 |
|  . |  . |  . |  . |  . |  . |  . |  . |  . |  . |  . |  . |
|  . |  . |  . |  . |  . |  . |  . |  . |  . |  . |  . |  . |
|  j |  YAj |  YBj |  . |  . |  . |  Yij |  . |  . |  . |  Ynj | nYij |
|  . |  . |  . |  . |  . |  . |  . |  . |  . |  . |  . |  . |
|  . |  . |  . |  . |  . |  . |  . |  . |  . |  . |  . |  . |
|  k |  YAk |  YBk |  . |  . |  . |  Yik |  . |  . |  . |  Ynk |  n Yik |
|  T.j |  kYAj |  kYBj |  . |  . |  . | kYij |  . |  . |  . | kYnj |  Yij |
|  T2.j |  kY2A j |  k Y2Bj |  . |  . |  . |  kY2 ij |  . |  . |  . |  kY2 nj |  Y2ij |

En donde **n** es el número de tratamientos; **k** el número de bloques; y, **N** (n x k), el número total de observacio­nes.

 Yij es la suma de todas las N observaciones; y, Y2ij, la suma de todas estas N observaciones al cua­dra­do. nYij es la suma de los n tratamientos del blo­que número **j** y kYij es la suma de los k bloques del tra­ta­miento número **i**.

La tabla de ANOVA usada para los cálculos es la siguiente:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  FUENTE |  Gdos. de lbtad. |  Suma de Cuadrados  (SC) |  Suma media de cuadrados (SMC) |  *F* calcul. |
| TRATAMIENTO |  n - 1 | n (k Yij)2/k - ( Yij)2/N |  SCT/(K - 1) |  SMCT/SMCE |
| BLOQUE |  k - 1 | k (n Yij)2/n - ( Yij)2/N |  SCB / (k-1) |  SMCT/SMCE |
| ERROR |  (n-1) (k-1) |  SC TOTAL-SCT-SCB |  SCE / (n-1)(k-1) |  |
| TOTAL |  (nk) - 1 |   Y2ij - ( Yij)2/N |  |  |

En donde la relación **SMCT / SMCE** es nuestro estadístico de prueba *F* para diferencias entre tratamientos, y sigue una distribución "F" de Fisher - Schnedecor con 1 = (n-1) y 2 = (n-1)(k-1) grados de libertad.

La relación **SMCB / SMCE** es nuestro estadístico de prue­ba *F'*para diferencias entre bloques, el cual sigue una distribución "F" de Fisher - Schnedecor con 1 = (k-1) y 2= (n-1)(k-1) grados de libertad.

La región de rechazo para cualquiera de los dos esta­dísticos viene dada por:

 *F*  F ()

para los grados de libertad correspondientes.

**9,4.- ANOVA multifactorial.-**

Utilizamos un ANOVA multifactorial cuando queremos comparar los efectos de mas de 1 tratamiento, así como los efectos de las interacciones de esos tratamientos.

Es decir, cuando tenemos un diseño factorial.

En este capítulo veremos como se realiza un ANOVA de 2 factores.

Definiremos  como el efecto del factor # 1,  como el efecto del factor # 2,  como el efecto de las repeticiones y () como el efecto de las interacciones de ambos factores.

Las hipótesis a probar son:

**H0 :** 1=2=...=a=0 (todos los efectos del 1er factor son 0)

**H1 :**  i tal que i0 (al menos 1 efecto del 1er factor no es 0)

Pudiendose probar también:

**H0' :** 1=2=...=b=0 (todos los efectos del 2do factor son 0)

**H1' :**  j tal que j0 (al menos 1 efecto del 2do factor no es 0)

y:

**H0'' :** 1=2=...=r=0 (todos los efectos de las réplicas son 0)

**H1'' :**  k tal que k0 (al menos 1 efecto de las réplicas no es 0)

asi como:

**H0''' :** Todos los ()ij son iguales.

**H1''' :** Al menos un ()ij no es igual.

La tabla de datos sigue el siguiente modelo:

 ╔═════════════╦════════════╦═══════════════╤═════════════════╤══════════════╦════════════╗

 ║ FACTOR 1 ║ FACTOR 2 ║ REPLICA 1 .. │ .. REPLICA k .. │ .. REPLICA r ║ TOTAL ║

 ╠═════════════╬════════════╬═══════════════╪═════════════════╪══════════════╬════════════╣

 ║ NIVEL 1 ║ NIVEL 1 ║ Y111 │ Y112 │ Y113 ║ rY11k ║

 ║ . ║ . ║ │ │ ║ ║

 ║ . ║ . ║ │ │ ║ ║

 ║ NIVEL 1 ║ NIVEL j ║ Y1j1 │ Y1jk │ Y1jr ║ rY1jk ║

 ║ . ║ . ║ │ │ ║ ║

 ║ . ║ . ║ │ │ ║ ║

 ║ NIVEL 1 ║ NIVEL b ║ Y1b1 │ Y1bk │ Y1br ║ rY1bk ║

 ║ . ║ . ║ │ │ ║ ║

 ║ . ║ . ║ │ │ ║ ║

 ║ NIVEL i ║ NIVEL j ║ Yij1 │ Yijk │ Yijr ║ rYijk ║

 ║ . ║ . ║ │ │ ║ ║

 ║ . ║ . ║ │ │ ║ ║

 ║ NIVEL a ║ NIVEL b ║ Yab1 │ Yabk │ Yabr ║ rYabk ║

 ║ ║ ║ │ │ ║ ║

 ╚═════════════╩════════════╬═══════════════╪═════════════════╪══════════════╬════════════╣

 ║ │ │ ║ ║

 ║ abYij1 │ abYijk │ abYijr ║ abrYijk ║

 ║ │ │ ║ ║

 ╚═══════════════╧═════════════════╧══════════════╩════════════╝

En donde **a** es el número de tratamientos o niveles del factor 1, **b** el número de niveles del factor 2, **n** es el número de tratamientos conjuntos a x b y **r** el número de réplicas o repeticiones.

El orden de un experimento factorial se denota como **a** x **b**.

Primero realizamos un anova de 2 vias, siguiendo la tabla de ANOVA:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  FUENTE |  Gdos. de lbtad. |  Suma de Cuadrados  (SC) |  Suma media de cuadrados (SMC) |  *F* calcul. |
| REPETICION |  r - 1 | r (n Yijk)2/n - ( Yijk)2/N |  SCR/(r - 1) |  SMCR/SMCE |
| TRATAMIENTOS |  ab - 1 | n (r Yijk)2/r - ( Yijk)2/N |  SCT / (ab-1) |  SMCT/SMCE |
| ERROR |  (ab-1) (r-1) |  SC TOTAL-SCT-SCR |  SCE / (ab-1)(r-1) |  |
| TOTAL |  (abr) - 1 |   Y2ijk - ( Yijk)2/N |  |  |

Luego de lo cual realizamos la subdivisión de la suma de cuadrados de los tratamientos en sus componentes **A**, **B** y **AxB** :

 FACTOR 2

╔══════════╦════════════╦══════════════╦════════════╦══════════╗

║ FACTOR 1 ║ NIVEL 1.. ║ .. NIVEL j.. ║ ..NIVEL b ║ TOTAL 2 ║

╠══════════╬════════════╬══════════════╬════════════╬══════════╣

║ ║ ║ ║ ║ ║

║ NIVEL 1 ║ r Y11k ║ r Y1jk ║ r Y1bk ║ brY1jk ║

║ . ║ ║ ║ ║ ║

║ . ║ ║ ║ ║ ║

║ . ║ ║ ║ ║ ║

║ ║ ║ ║ ║ ║

║ NIVEL i ║ r Yi1k ║ r Yijk ║ r Yibk ║ brYijk ║

║ . ║ ║ ║ ║ ║

║ . ║ ║ ║ ║ ║

║ . ║ ║ ║ ║ ║

║ ║ ║ ║ ║ ║

║ NIVEL a ║ r Ya1k ║ r Yajk ║ r Yabk ║ brYajk ║

║ ║ ║ ║ ║ ║

╠══════════╬════════════╬══════════════╬════════════╬══════════╣

║ TOTAL 1 ║ arYi1k ║ arYijk ║ arYibk ║abrYijk║

║ ║ ║ ║ ║ ║

╚══════════╩════════════╩══════════════╩════════════╩══════════╝

Y aplicamos las formulas:

 

7

 

8

y

 

9

Luego de lo cual construimos la tabla final de ANOVA:

╔════════════╦══════════╦══════════╦═══════════╦═════════════════╗

║ ║ Gdos. de║ ║ ║ ║

║ FUENTE ║ libertad║ SC ║ SMC ║ *F* calculado ║

╠════════════╬══════════╬══════════╬═══════════╬═════════════════╣

║ ║ ║ ║ ║ ║

║ Repetición ║ r - 1 ║ SCR ║ SCR/(r-1) ║ SMCR / SMCE ║

╠════════════╬══════════╬══════════╬═══════════╬═════════════════╣

║ Tratamient.║ ║ ║ ║ ║

║ ║ ║ ║ ║ ║

║ A ║ a - 1 ║ SCA ║ SCA/(a-1) ║ SMCA / SMCE ║

║ ║ ║ ║ ║ ║

║ B ║ b - 1 ║ SCB ║ SCB/(b-1) ║ SMCB / SMCE ║

╠════════════╬══════════╬══════════╬═══════════╬═════════════════╣

║ Interacción║ ║ ║ SC(AxB)/ ║ ║

║ AxB ║(a-1)(b-1)║ SC(AxB) ║ (a-1)(b-1)║ SMC(AxB)/SMCE ║

╠════════════╬══════════╬══════════╬═══════════╬═════════════════╝

║ ║ ║ ║ SCE/ (ab- ║

║ Error ║(ab-1)(r-1) ║ SCE ║ 1)(r-1) ║

╠════════════╬══════════╬══════════╬═══════════╝

║ ║ ║ ║

║ Total ║ abr - 1 ║ SC Total ║

╚════════════╩══════════╩══════════╝

En donde la relación **SMCR/SMCE** es nuestro estadístico de prueba *F* para diferencias entre réplicas, y sigue una distribución "F" de Fisher - Schnedecor con 1 = r-1 y 2 = (ab-1)(r-1) grados de libertad.

La relación **SMCA/SMCE** es nuestro estadístico de prueba *F'* para diferencias debidas al tratamiento 1, y sigue una distribución "F" de Fisher - Schnedecor con 1 = a-1 y 2 = (ab-1)(r-1) grados de libertad.

La relación **SMCB/SMCE** es nuestro estadístico de prueba *F''* para diferencias debidas al tratamiento 2, y sigue una distribución "F" de Fisher - Schnedecor con 1 = b-1 y 2 = (ab-1)(r-1) grados de libertad.

Y, la relación **SMC(AxB)/SMCE** es nuestro estadístico de prueba *F'''* para diferencias debidas a la interacción de ambos tratamientos, y sigue una distribución "F" de Fisher - Schnedecor con 1 = (a-1)(b-1) y 2 = (ab-1)(r-1) grados de libertad.

La región de rechazo para cualquiera de los cuatro estadísticos viene dada por:

 W = { *F*  F() }

Para los grados de libertad correspondientes.

**9.5.- Restricciones**

 Los análisis de varianza se basan en ciertas suposicio­nes que debemos de tomar en cuenta al realizarlos. Estas son:

- El proceso está bajo control y puede ser repetido.

- Todas las muestras deben ser aleatorias e independien­tes.

- Las poblaciones de las cuales muestreamos deben de seguir una ley Normal de distribución.

- Todas las poblaciones que estamos probando deben de tener varianzas iguales.

Algunos libros discuten estas suposiciones y qué se puede hacer si éstas no se cumplen.

Algunos estudios han demostrado que la falta de norma­lidad no afecta seriamente el análisis cuando el tamaño de la muestra es igual para todos los tratamientos.

En el caso de que la varianza y la media sean depen­dientes entre sí, existen tablas que indican cómo rom­per esta dependencia; sin embargo este caso no será estudiado por nosotros.

Para probar la igualdad de varianzas podemos usar va­rias pruebas. Una forma fácil de hacerlo es observar el rango de cada tratamien­to. Se saca un rango promedio , el cual se lo multiplica por un factor *D*4 extraído de la tabla # 6 para distintos tamaños de muestra. Si el producto *D*4 es mayor que todos los rangos, se puede asumir con alguna seguridad que las varianzas son homo­géneas.

Otras pruebas existentes son la prueba *C* de Cochran y la prueba *F* de Bartley - Box.

**9.6.- Comparaciones múltiples.-**

Después de realizar nuestro ANOVA, sabremos si todos nuestros tratamientos son iguales o si existe al menos uno que sea distinto; pero, usualmente queremos saber más: Cuál tratamiento es mejor y cuál es peor? Cuál tratamiento difiere de los otros y cuáles son iguales entre sí?

Existen muchas formas de responder estas preguntas, entre ellas tenemos las pruebas de los con­trastes orto­gonales, la prueba de Scheffé la prueba de Tukey, la prueba de rango múltiple de Duncan y la prueba de Student Newman Keuls o S-N-K. Aquí sólo estudiaremos estas dos últimas.

**9.6.1.- Prueba Student- Newman-Keuls (SNK).-**

Es recomendada para comparar factores cua­litativos y para ser usada cuando la decisión de cuál comparación debe hacerse se la realiza **después** de que los datos han sido examinados.

Las comparaciones se hacen mediante amplitudes múltiples basandose en los resultados.

Los pasos a seguir son los siguientes:

1.- Ordene los k medias de menor a mayor.

2.- Determine el **MSCE** y sus grados de libertad de la tabla de ANOVA.

3.- Obtenga el error estándar de la media para cada tratamiento mediante la fórmula:

 

10

4.- Busque en la tabla # 7 de rangos studentizada los valo­res de los rangos significativos (rp) a un nivel , con 2 grados de libertad de la tabla de ANOVA y

p = 2,3,..k, y apunte estos k-1 rangos.

5.- Multiplique estos rangos por Sxj para lograr un grupo de k-1 rangos menos significativos.

6.- Pruebe todas las k(k-1)/2 posibles diferencias entre pares de medias con sus respectivos valores de rangos menos significativos.

Los valores de diferencias  que el rango de mínima significancia indican diferencias significativas.

**9.5.2.- Prueba de rango múltiple de Duncan.-**

Es una prueba menos potente que la SNK y menos prudente.

Se diferencia de la SNK en que usa niveles de confianza variables en vez de fijos, dependiendo del número de medias.

El procedimiento general a seguir es el mismo, sin embargo usamos la tabla # 7.

**CAPITULO X**

 **AJUSTE DE CURVAS**

Llamamos regresión a un problema en el cual fijamos valo­res dados de una variable independiente (x), y realiza­mos observaciones en una variable dependiente (y) de ésta.

El propósito de este estudio es lograr una ecuación para predecir **y** a partir de **x**, dentro de un rango espe­cífico.

En un análisis de correlación se mide, para cada mues­tra los valores de **x** y **y**; éstos son graficados para en­contrar relaciones entre ellos. Además se calculan algunos estadísticos para determi­nar la fuerza de la relación, aunque un alto valor de correlación no indica necesariamente que x es causado por y, o viceversa.

1**Figura # 12.-** Diagrama de dispersión.

Por lo tanto, la regresión puede ser usada para experi­mentos reales, mientras que la correla­ción se usa para estudios *ex post fac­to*, es decir, para ana­li­zar datos obtenidos con ante­rioridad.

**10.1.- Diagrama de disper­sión.-**

Llamamos diagrama de dis­persión a un gráfico (fi­gura # 12) en el cual van a estar represen­tados, mediante pun­tos, los valores de nuestros pares de va­riables (x,y).

Este diagrama sirve para darnos una idea visual del tipo de relación que exis­te entre ambas variables, y debe de ser hecho antes de iniciar cualquier cálculo para evitar trabajos inne­cesarios.

**10.2.- Método de los míni­mos cuadrados.-**

Llamamos regresión lineal a un experimento donde trata­mos de relacionar dos variables **x** y **y**, mediante una ecuación de la recta, esto es:

 

11

2**Figura 13.-** Recta de regresión por el método de los mínimos cuadrados.

en donde **a** es la intersec­ción de la recta con el eje Y, y **b** es la pendiente de la recta.

Para encontrar los valores de **a** y **b** de la recta que más se acer­ca a nuestros datos experimenta­les, uti­lizamos el método de los mínimos cuadrados; es de­cir, vamos a tomar la rec­ta para la cual los cua­drados de las dife­ren­cias entre los puntos experi­men­tales (x,y) y los pun­tos calculados (x',y') sea mínima.

Las fórmulas para calcular los coeficientes **a** y **b** son:

 

12

 

13

**10.3.- Coeficientes de correlación.-**

Llamamos coeficiente de determinación (r2) a la propor­ción de la variación en la variable **y** que puede ser atribuida a una **regresión lineal** con respecto a la variable x. Se lo calcula mediante la fórmu­la:

 

14

Su raíz cuadrada positiva (r) se la conoce como coefi­ciente de correlación de Pearson y es un estimador del parámetro **coeficiente de correlación poblacional** .

Llamamos **eta cuadrado (2)** a la relación entre la suma de cuadrados entre tratamientos (SCT) y la suma de cuadrados total (SC Total) del ANOVA, y representa a la máxima variación total que puede ser atribuida a **cual­quier** regresión de **y** con respecto de **x**; o sea, la máxi­ma correlación que podemos esperar en una curva o mode­lo matemático que pase por todas las **y** para cada valor de **x**.

**10.4.- Uso de la regresión lineal para comparaciones de dos variables cuantitativas.-**

Cuando estamos comparando datos de dos variables cuan­titativas y nuestro ANOVA nos da diferencias muy sig­nificativas, podemos recurrir a la regresión lineal para determinar cómo responde nuestra variable depen­diente **y** a las variaciones dentro de un rango de **x**.

El primer paso a seguir es calcular la ecuación lineal con sus constantes **a** y **b**.

Con base en esto elaboramos una tabla con los valores de **x**, , **y'**(calculada a partir de la ecuación para cada valor de **x**),**y** ( **- y'**) y **(y)2**.

Con base en los datos de nuestra tabla de **x** y **y**, elabo­ra­mos nuestra nueva tabla de ANOVA:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  FUENTE |  Gdos. de lbtad. |  Suma de Cuadrados  (SC) |  Suma media decuadrados (SMC) |  *F* calcul. |
| TRATAMIENTO |  k - 1 |  15 | SCT/(K-1)  | SMCT/SMCE  |
| LINEAL |  1 |  SCT - SCDL |  SCL / 1 |  SMCL/ SMCE |
| DESV. DE LIN. | K-2 |   16 |  SCDL / (k - 2) |  SMCDL /SMCE |
| ERROR |  N - k |  SC Total - SCT |  SCE / N - k |  |
| TOTAL |  (nk) - 1 |  17  |  |  |

En donde las hipótesis a probar son:

**Ho** : No hay diferencias atribuibles a regresión lineal.

**H1** : Sí hay diferencias atribuibles a regresión lineal.

Para el estadístico *F* = SMCL / SMCE con 1 = 1 y 2 = N-k grados de libertad.

y:

Ho : No hay desviaciones de la regresión lineal calcu­la­da.

H1 : Si hay desviaciones de la regresión lineal calculada.

Para el estadístico *F* = SMCD / SMCE con 1 = k-2 y 2 = N-k grados de libertad.

**10.5.- Regresiones no lineales .-**

Además de la regresión lineal existen otros tipos de relaciones posibles entre las variables **x** y **y**. En los fenómenos naturales de crecimiento poblacional es muy común la regresión exponencial de la forma:

 

18

en donde **a** es conocido como "índice de Falton", y **b** es el índice de crecimiento relativo.

Es característico de esta relación que, al graficarse sus pares de datos (x,y) en un papel semilogarítmico, el resultado sea una línea recta.

Podemos, de la misma forma, enderezar los datos numéricamen­te reemplazando esta ecuación por:

 

19

Osea, linealizando la curva, después de lo cual ten­dre­mos un caso de regresión lineal.

Otro caso de regresión no lineal o curvilínea, ocurre cuan­do la relación entre ambas variables sigue una curva de grado 2 ó mayor, o sea que sigue una ecuación polino­mial de la forma :

 

20

El ajuste de curvas polinomiales también se utiliza para obtener aproximaciones mediante regresión, cuando nuestro modelo lineal no tiene suficiente fuerza. Pero, no estudiaremos este caso en nuestro manual.

**10.6.- Pruebas de hipótesis respecto a no correlación.-**

El objeto de esta prueba es determinar si existe o no correlación entre dos variables, esto es, si el coeficiente  es o no igual a 0, siempre y cuando sigan una distribución aproximadamente normal.

Las hipótesis a probar son:

**H0 :**  = 0 (No existe correlación entre las dos variables)

**H1 :**   0 (Si existe correlación entre las dos variables)

El estadístico de prueba **t** se lo calcula en base a r:

 

21

y sigue una distribución "t" de student con n-2 grados de libertad.

La región de rechazo viene dada por :

 W = { t > t(/2)

Para los grados de libertad correspondientes.

**10.7.- Análisis de covarianza.-**

Al estudiar el diseño de bloques aletorios, decíamos que podíamos librarnos del error experimental debido a variables identificables y controlables. Cuando estas variables concomitantes son identificables y mesurables; pero no controlables, estamos frente a un

análisis de covarianza.

El método mediante el cual analizamos los datos es una combinación del método de regresión lineal y del ANOVA, cuyo modelo lineal va a estar dado por:

 Yij =  + i + xij + ij

En donde i es el efecto del iesimo tratamiento, ij el valor del error aleatorio, y xij es la pendiente de la ecuación de regresión lineal.

La hipótesis a probar será:

**H0 :** 1 = 2 = .. k = 0 (no hay efecto del tratamiento)

**H1 :**  i tal que i  0 (si hay efecto del tratamiento)

Los cálculos se realizan de la siguiente manera:

- Se calcula la suma de cuadrados del tratamiento, error y total del ANOVA de una via para las x.

- Se calcula la suma de cuadrados del tratamiento, error y total delANOVA de una via para las y.

- Se calcula la suma de productos del tratamiento, error y total mediante las siguientes fórmulas:

 

22

 

23

 

24

 

25

- Se calcula la suma de cuadrados del tratamiento, error y total para las y ajustadas mediante las fórmulas:

 

26

 

27

 

28

Construimos seguidamente nuestra tabla de análisis de covarianza:

╔══════════════╦══════════╦══════════╦═════════╦═══════════╦═════════════╦═══════════════════╗

║ ║ ║ ║ ║ ║ Grados de ║ ║

║ FUENTE ║ SCx ║ SCy ║ SP ║ SCy' ║ libertad ║ SMC ║

╠══════════════╬══════════╬══════════╬═════════╬═══════════╬═════════════╬═══════════════════╣

║ ║ ║ ║ ║ ║ ║ ║

║ TRATAMIENTO ║ SCTx ║ SCTy ║ SPT ║ SCTy' ║ k - 1 ║ SCTy'/(k-1) ║

╠══════════════╬══════════╬══════════╬═════════╬═══════════╬═════════════╬═══════════════════╣

║ ║ ║ ║ ║ ║ ║ ║

║ ERROR ║ SCEx ║ SCEy ║ SPE ║ SCEy' ║ nk - k - 1 ║ SCEy'/(nk-k-1) ║

╠══════════════╬══════════╬══════════╬═════════╬═══════════╬═════════════╬═══════════════════╝

║ ║ ║ ║ ║ ║ ║

║ TOTAL ║ SCTotx ║ SCToty ║ SPTot ║ SCToty' ║ nk - 2 ║

╚══════════════╩══════════╩══════════╩═════════╩═══════════╩═════════════╝

Nuestro estadístico de prueba *F* vendrá dado por la relación **SMCTy'/SMCEy'**, y sigue una distribución "F" de Fisher - Schnedecor con 1=k-1 y 2=nk-k-1 grados de libertad.

**- Tablas estadísticas**

1.- Función de distribución Normal ***Z***()

2.- Valores de *t*

3.- Valores de2

4.- Valores de *F*

**TABLA # 5**

Valores de Chauvenet

╔═══════╤════════╦═══════╤════════╗

║ n │ xm -/s║ n │ xm -/s║

╠═══════╪════════╬═══════╪════════╣

║ 2 │ 1.15 ║ 14 │ 2.10 ║

╟───────┼────────╫───────┼────────╢

║ 3 │ 1.38 ║ 16 │ 2.15 ║

╟───────┼────────╫───────┼────────╢

║ 4 │ 1.53 ║ 18 │ 2.20 ║

╟───────┼────────╫───────┼────────╢

║ 5 │ 1.64 ║ 20 │ 2.24 ║

╟───────┼────────╫───────┼────────╢

║ 6 │ 1.73 ║ 25 │ 2.33 ║

╟───────┼────────╫───────┼────────╢

║ 7 │ 1.80 ║ 30 │ 2.39 ║

╟───────┼────────╫───────┼────────╢

║ 8 │ 1.86 ║ 40 │ 2.50 ║

╟───────┼────────╫───────┼────────╢

║ 9 │ 1.91 ║ 50 │ 2.58 ║

╟───────┼────────╫───────┼────────╢

║ 10 │ 1.96 ║ 60 │ 2.64 ║

╟───────┼────────╫───────┼────────╢

║ 11 │ 2.00 ║ 80 │ 2.74 ║

╟───────┼────────╫───────┼────────╢

║ 12 │ 2.04 ║ 100 │ 2.81 ║

╟───────┼────────╫───────┼────────╢

║ 13 │ 2.07 ║ 1000 │ 3.48 ║

╚═══════╧════════╩═══════╧════════╝

**TABLA # 6**

Rango de diferencias *D4*

╔═════════╤═══════╤═══════╤═══════╤═══════╤═══════╤═══════╤═══════╤═══════╤════════╗

║ │ │ │ │ │ │ │ │ │ ║

║ **n** │ 2 │ 3 │ 4 │ 5 │ 6 │ 7 │ 8 │ 9 │ 10 ║

╟─────────┼───────┼───────┼───────┼───────┼───────┼───────┼───────┼───────┼────────╢

║ │ │ │ │ │ │ │ │ │ ║

║ ***D4*** │ 3.267 │ 2.575 │ 2.282 │ 2.155 │ 2.004 │ 1.924 │ 1.864 │ 1.816 │ 1.770 ║

╚═════════╧═══════╧═══════╧═══════╧═══════╧═══════╧═══════╧═══════╧═══════╧════════╝

7.- Percentiles de los rangos Studentizados.

8.- Valores críticos de *D* de Duncan

**- Ejercicios varios aplicados a la Acuicultura.-**

**1.-** En un muestreo de peso en una camaronera se obtu­vieron los siguientes resultados (en gramos):

16 15 16 15 15 15 15 15 17 17 15 16 16 17 16 16 16 16 15 16 16 16 16 16 17 17 16 16 16 16 16 16 18 18 18 19 18 19 18 18 18 18 19 18 19 18 18 18 18 19 19 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 19 18 19 19 19 19 18 18 18 19 19 18 18 19 19 19 19 18 20 21 20 20 20 21 21 21 22 22 21 21 22 21 20 22 22 21 21 21 21 21 21 21 20 22 20 21 20 20 20 21 21 21 21 22 22 21 21 21 22 22 21 20 21 21 22 22 22 22 22 22 21 21 26 24 24 24 25 25 25 26 25 26 24 24 23 25 26 24 24 24 25 25 25 26 24 25 25 25 24 25 25 25 24 26 29 28 28 29 27 28 29 28 28 28 27 28

28 28 27 28 27 28 29 30 28 29 29 29 28 29 28 28 28 28 29 29

Con base en estos datos:

**a.-** Construya una tabla de distribución de frecuencias.

**b.-** Grafique mediante un histograma la frecuenci rela­tiva.

**c.-** Grafique mediante un gráfico de barras la frecuen­cia relativa.

**d.-** Grafique mediante un polígono de frecuencias la frecuencia relativa y la frecuencia acumulada relativa.

**e.-** Grafique mediante un gráfico de sectores la fre­cuencia relativa.

**f.-** Calcule el promedio.

**g.-** Calcule la mediana.

**h.-** Calcule la desviación estándar empírica.

**i.-** Calcule la varianza empírica.

**j.-** Calcule el rango.

**k.-** Construya una distribución de frecuencias usando los rangos comerciales (colas por libra).

**l.-** Realize los mismos gráficos que se hicieron para la otra distribución de frecuencias.

**m.-** Compare y comente la utilización del gráfico de barras y el histograma en ambos casos.

**2.-** Antes de cosechar una piscina de cultivo de cama­rón, tomamos una muestra de 230 animales la misma que pesa 3,956 gramos y cuya varianza es 3.78; al cosechar la misma obtenemos de la empacadora una liquidación de la cual calculamos un peso promedio de 16.83 y una varianza de 3.96. En estos datos, señale:

**a.-** El promedio.

**b.-** La media.

**c.-** La varianza empírica.

**d.-** La varianza poblacional.

**3.-** Calcule de cuantas maneras pueden suceder los si­guientes eventos:

**a.-** Elegir dos peces de un tanque que contiene 25 pe­ces.

**b.-** Elegir un pez y luego otro de un tanque conte­niendo 25 peces.

**c.-** Contestar un examen de verdadero o falso que tenga 12 pregun­tas.

**d.-** Ordenar una cena de un menú que tenga 10 entradas, 6 sopas, 7 platos fuertes y 9 postres.

**e.-** Repartir una mano de cuarenta a una persona.

**f.-** Obtener "ronda" en una primera mano de cuarenta.

**4.-** Calcule la probabilidad de:

**a.-** Sacar una ronda en una primera mano de cuarenta

**b.-** Sacar dos caras en el lanzamiento de dos monedas.

**c.-** Sacar cara y sello en el lanzamiento de dos mone­das.

**d.-** Sacar cara en la primera moneda y sello en la se­gunda moneda.

**5.-** Explique que entendería Ud. si un in forme metereo­lógico dice "hay 50% de probabilidad de que llueva".

**6.-** En los dos ejercicios anteriores, calcule la proba­bilidad de no ocurrencia de los eventos descritos.

**7.-** Obtenga de la tabla de distribución Normal los valores de ***Z*** para los siguientes valores de :

**a.-** 1.28

**b.-** 1.64

**c.-** 1.96

**8.-** Obtenga los valores de ***t*** para  = 1, 3, 9 y 15 grados de libertad para los siguientes valores de :

**a.-** .1

**b.-** .5

**c.-** .25

**d.-** .01

**e.-** .005

**9.-** Obtenga los valores de ***2*** para  = 1, 3, 9 y 15 gra­dos de libertad para los siguientes valores de :

**a.-** .1

**b.-** .5

**c.-** .25

**d.-** .01

**e.-** .005

**10.-** Obtenga los valores de ***F*** con  = .5 y  = .1 para los siguientes grados de libertad:

**a.-** 1 = 3, 2 = 5

**b.-** 1 = 12, 2 = 2

**c.-** 1 = 5, 2 = 3

**d.-** 1 = 15, 2 = 15

**11.-** Determine si los valores del ejercicio # 1 siguen una ley Normal.

**12.-** Calcule el error (E) que podemos esperar con res­pecto a la media poblacional en el promedio del ejerci­cio # 1 con un 90% y un 95% de confianza.

**13.-** Calcule el intervalo de confianza para la media en el ejercicio # 1 con 90% y 95% de confianza.

**14.-** Con base en los siguientes datos de longitud total de postlarvas de *P. vannamei* (en mm.), calcule los in­tervalos de confianza para la media y para la varianza con un 95% de confian­za:

7.5 7.2 7.1 6.9 7.2 8.0 7.5 7.4 6.8 7.8 7.7 7.2 7.2 7.4 7.6 6.9 7.9 7.8 7.6 7.5 7.4 6.9 7.0

**15.-** Con base en los siguientes datos de longitud total diarias de larvas de *P. vannammei*, construya un gráfico de las medias con sus respectivos intervalos de con­fianza:

 DIA L O N G I T U D E S T O T A L E S ( e n m i c r a s )

 1 618 614 612 607 617 618 612 607 615 613 615

 2 1,355 1,349 1,307 1,380 1,303 1,323 1,295 1,352 1,398 1,368 1,353

 3 1,355 1,349 1,307 1,380 1,303 1,323 1,576 1,590 1,620 1,708 1,649

 4 2,271 2,252 2,307 2,385 2,223 2,239 2,334 2,305 2,079 2,190 2,267

 5 2,972 2,983 3,108 2,629 2,478 2,553 2,669 3,184 2,638 2,501 2,485

 6 3,257 3,155 3,323 3,113 3,118 3,123 2,987 2,956 2,856 3,097 3,140

 7 3,755 3,576 3,649 3,748 3,350 3,597 3,135 3,515 3,506 3,598 3,354

 8 4,194 4,079 3,922 4,139 3,657 4,111 3,609 3,731 3,706 3,755 3,790

 9 4,679 4,591 4,348 4,204 4,257 4,511 4,194 4,157 3,926 4,063 4,192

 10 4,548 4,596 4,398 4,724 4,619 4,848 4,422 4,545 4,371 4,721 4,808

 11 4,884 4,932 5,047 5,281 5,098 5,215 5,007 4,834 4,261 4,490 4,637

 12 5,037 5,141 5,133 5,378 5,460 5,477 5,228 5,186 4,952 4,661 4,882

 13 5,344 5,588 5,445 5,674 5,637 5,915 5,433 5,579 5,349 5,424 5,307

 14 5,839 5,624 5,755 5,840 5,186 5,790 5,574 5,678 5,405 5,445 5,620

 15 5,768 5,875 5,974 6,163 5,680 5,803 5,592 5,785 5,696 5,652 5,638

 16 6,291 6,065 6,222 6,168 6,079 6,209 5,864 5,832 6,065 5,947 5,989

 17 6,604 6,433 6,539 6,339 6,353 6,234 6,054 6,189 6,173 6,111 6,031

 18 6,972 6,935 6,837 6,378 6,513 6,481 6,689 6,662 6,578 6,360 6,144

 19 7,386 7,283 7,436 6,800 6,866 7,080 6,781 6,841 6,857 6,573 6,375

 20 7,700 7,688 7,674 7,234 7,322 7,265 7,081 7,007 6,870 6,696 6,684

**16.-** En una muestra aleatoria de 100 camarones tomados de una piscina de cultivo se obtiene un peso promedio de 17.3 gramos con varianza = 3,7. Decida si existen diferencias significativas (=0.5) entre esta piscina y el peso programado de cosecha = 18 gramos.

**17.-** Determine si la media de un embarque de alevines del cual tomamos una muestra de 15 alevines, y cuyas longitudes (en mm.) son:

10.3 11.3 11.4 11.3 10.6 10.8 10.2 10.6 10.8 11.2 10.2 10.4 10.6 10.5

10.9

es de menos de 11 mm. con una significancia del 95%.

**18.-** Determine en el ejemplo anterior si la varianza poblacional es de 0.19,  = 0.5.

**19.-** Determine si el embarque de alevines del ejercicio # 17 tiene la misma varianza ( = 0.5) que otro del cual tomamos la siguiente muestra:

10.6 10.8 10.2 11.4 11.5 11.3 10.5

10.7 10.7 10.2 11.4 10.2 10.9

**20.-** Se desea saber si el crecimiento semanal promedio de una camaronera en dos años consecutivos fué el mismo o si hubo diferencias significativas ( = 0.05) entre ellos. En el año 1, en 51 cultivos se obtuvo un creci­miento promedio de 0.98 con varianza 0.135; en el si­guiente año el crecimiento en 47 cultivos fué de 0.92 con varianza 0.191. Realize los cálculos.

**21.-** Determine si hay diferencias significativas ( = 0.5) entre los embarques de los ejercicios #17 y 19.

**22.-** Se desea probar el efecto de polvillo de arroz en el crecimiento de *Chaetoceros gracilis*, para lo cual se realizan tres cultivos con este tratamiento y tres sin el.

Los siguientes son los datos de concentración final en los seis cultivos:

Con polvillo : 1'238,000 1'354,000 1'452,500

Sin polvillo : 1'154,000 1'430,500 1'338,500

Con una significancia del 95% demuestre si hay diferen­cias entre ambos tratamientos.

**23.-** Se desea determinar la eficiencia de cierto quími­co como bactericida en tanques de larvas, para lo cual se realizan los siguientes contajes de bacterias en el agua de 12 tanques antes y después de su aplicación:

|  |  |
| --- | --- |
|  ANTES |  DESPUES |
|  15.200  16.300  10.100  8.300  6.400  28.100 4.300 6.800 9.300 11.400 12.800 34.000  |  8.400  9.300  6.500  4.200  3.100  12.300  2.400  4.500  9.400  4.200  6.300  11.100   |

Determine si hay diferencias significativas ( = 0.01) debidas al tratamiento.

**24.-** Cierta camaronera compra solamente semilla que tenga al menos un 60% de *P. vannamei*. Si un provedor trae un lote, el cual da como resultados del análisis 50%, 65%, 62% y 55% en 4 muestras; indique si se debe­ría comprar esta larva.

**25.-** Se prueban dos tipos de dietas para larvas de camarón en 10 tanques, dando como resultado en la dieta A: 72.35%, 82.25%, 76.44%, 74,09%, 74.72% y 76.70% de sobrevi­vencia y en la B: 74.96%, 67.95%, 87.29% y 62.92%. Demuestre si existen diferencias signifi­cativas ( = 0.5) con respecto a sobrevivencia entre ambos alimentos.

**26.-** Determine si existen datos atípicos en los ejerci­cios # 17 y 19.

**27.-** Se obtienen los siguientes resultados de creci­miento semanal (en gramos) usando 4 distintos alimentos (A, B, C, D) en 19 piscinas de camarón:

|  |  |  |  |
| --- | --- | --- | --- |
|  A |  B |  C |  D |
|  0.74 0.76 0.75 0.74 0.75 0.76 |  0.68 0.71 0.71 0.72 |  0.75 0.77 0.77 O.76 0.75 |  0.72 0.74 0.73 0.73  |

Demuestre si existen diferencias significativas ( = 0.5) en crecimiento entre alimentaciones.

**28.-** Se prueban tres alimentos distintos (A, B y C) en cuatro camaroneras situados en diferentes zonas, dando como resultado los siguientes crecimientos:

Camaronera # 1: A= 0.86, B= 0.94, C= 1.12

Camaronera # 2: A= 0.54, B= 0.62, C= 0.71

Camaronera # 3: A= 0.71, C= 0.61, C= 0.78

Camaronera # 4: A= 0.94, C= 0.98, C= 0.99

Pruebe si existen diferencias significativas ( = 0.5) en creci­mientos entre alimentos.

Use un ANOVA de dos vias y uno de una via y compare los resultados. Discuta las diferencias.

Pruebe la convenienciade usar el diseño en bloques demostrando las diferencias existentes en crecimiento entre camaroneras.

**29.-** Realize pruebas de rango múltiple de Duncan y SNK a los ejercicios anteriores.

**30.-** Los siguientes datos son relativos a los residuos de cloro en un tanque después de haber sido tratado con productos químicos:

 Horas ppm de Cl-

 2 1.8

 4 1.5

 6 1.4

 8 1.1

 10 1.1

 12 0.9

Ajuste una linea de mínimos cuadrados con la que poda­mos predecir los residuos de cloro en función de las horas después de haberse realizado el tratamiento.

BIBLIOGRAFIA

- Capa H. (1991). *Estadística Básica*. Cimacyt.

- Hicks C. (1982). *Fundamental concepts in the design of expe­riments*. CBS College Publishing.

- Leedy P. (1974). *Practical Research: Planning and design*. Macmillan.

- Miller I., Freund J. (1986). *Probabilidad y estadís­tica para ingenieros*. Prentice - Hall.

- Möller F. (1979). *Manual of methods in aquatic enviroment research. Part 5.- Statistical tests.* FAO Fisheries technical paper No. 182.

- Steel R., Torrie J. (1985). *Bioestadistica: Principios yprocedimientos.* McGraw - Hill.