# IDENTIFICACIÓN DE LA VARIABILIDAD DE UN PROCESO Y CONTROL DEL ERROR EXPERIMENTAL

**Estadística Comparativa**

- La estadística comparativa es aquella parte de la estadística que se propone comparar dos o más poblaciones. Existen algunas herramientas para hacer comparaciones.
- Las más conocidas son las pruebas de hipótesis y el análisis de varianza, pero existen muchas más.

**Prueba de hipótesis**

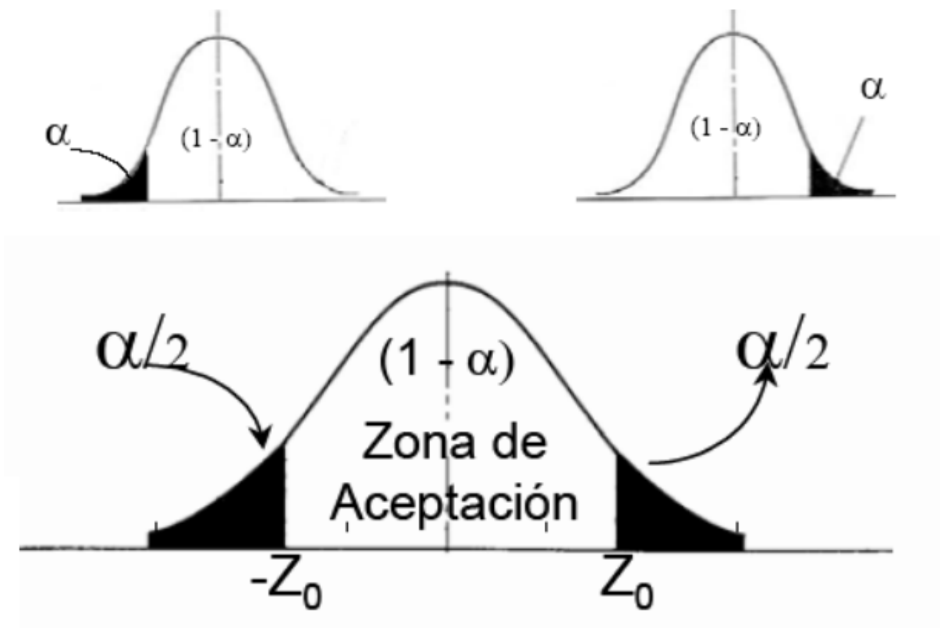
- Hipótesis estadística a una asunción sobre una población que está siendo muestreada.
- Test hipótesis simplemente una regla mediante la cual esta hipótesis se acepta o se rechaza.
- Regla basada generalmente en un estadístico muestral llamado estadístico de prueba, ya que se lo usa para probar la hipótesis.
- La región crítica de un estadístico de prueba consiste en todos los valores del estadístico donde se hace la decisión de rechazar $H_0$
- 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 nula verdadera, estamos cometiendo un error de tipo I. La probabilidad de cometer un error del tipo I se llama $\alpha$.
- Si aceptamos una hipótesis nula falsa, estaremos cometiendo un error del tipo II, y la probabilidad de cometerlo se la denomina $\beta$.

| | Investigador | |
|---|---|---|
| Hipotesis nula | Se acepta Ho | Se rechaza Ho |
| Ho es verdadrea | Decision correcta | Error tipo I |
| Ho es falsa | Error tipo II | Decision orrecta |

- Uno de los objetivos de las pruebas de hipótesis es diseñar test en donde $\alpha$ y $\beta$ sean pequeños, pero la mayor ventaja que nos dan las pruebas de hipótesis es precisamente esto: Poder medir $\alpha$ y $\beta$ , de tal forma que nosotros podamos medir la incertidumbre, remplazando palabras vagas como "pudo" "tal vez" posiblemente" que nosotros ponemos al "ojímetro" por un número que denota cuanto es la posibilidad de equivocarnos.
- Para probar una hipótesis la expresamos en su forma nula ($H_0$), y formulamos una hipótesis alterna ($H_1$) que aceptaremos al rechazar $H_0$.
    - Poner de tal forma que no haya diferencias.
- Debe de ser una hipótesis simple, por ejemplo:
    - "No hay diferencia entre las medias"
    - "La media de la población es igual a 10"
    - "Los caracteres son independientes"
    - "Las varianzas son homogéneas(iguales)“
- Ambas hipótesis deben ser distintas y mutuamente excluyentes.
- Queremos saber si dos poblaciones, una con peso promedio de muestreo 14.0 g y otra con 15.0 g tienen igual media de peso.

- "14.0 no es lo mismo que 15.0".
- ¿Cuándo son iguales entonces?
- Cuando 1ª piscina pesa ¿14.2? ¿14.5g? ¿14.8? ¿14.9? ¿14.99? porque no ¿14.85? o ¿14.849?

- Esto no lo podemos saber al ojo o por "feeling"
- Método estadístico dice si son o no diferente con un (1-) x 100% de confianza.
- % de confianza (1- $\alpha$) x 100 %:
  - % confianza de no estar cometiendo un error tipo I
  - Si repetimos infinitamente el experimento, al menos este % de veces nos daría el mismo resultado.
- Y el área crítica(W) es el área de la curva en donde $H_0$ va a ser rechazada en $\alpha$x100% de las veces.



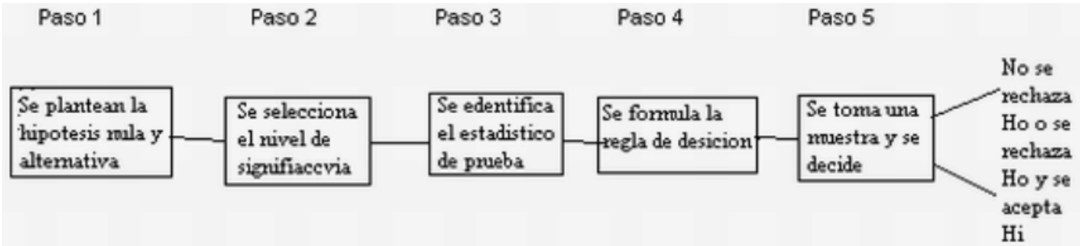

- El valor de $\alpha$ al cual decidimos nosotros aceptar o rechazar $H_0$ va a depender únicamente de nosotros, que tanto deseamos estar seguros de no equivocarnos. Si el resultado de equivocarme va a resultar en que Yo me muera me inclino por valores de bajos (0.00000000001), si el resultado de equivocarme va a resultar en que a mi perro le salgan canas me iría por un valor más alto (x ej.$\alpha$ =0.1). En general se usan valores entre 0.1 y 0.01, siendo el más común el de 0.05.

Pasos

1. Expresar claramente la hipótesis nula ($H_0$) y su alterna ($H_1$) en los libros para pruebas más comunes.
2. Especificar el nivel de significancia y el tamaño de la muestra (n), $\alpha$ generalmente 0.05 o 0.01.
3. Escoger estadístico para probar $H_0$ (libros), ojo con asunciones y restricciones que involucra
4. Determinar distribución muestral estadístico cuando $H_0$ es verdadera (libros)
5. Designar región crítica donde $H_0$ va a ser rechazada en 100 x $\alpha$ % de muestras cuando es verdadera (formula en libros)
6. Escoger muestra(s) aleatoria(s) de tamaño n, (proceso mecánico). Medir variable de interés.

7. Calcular estadístico de prueba: remplazar datos obtenidos en la fórmula del libro (calculadora o PC).
8. Comparar el estadístico calculado con el teórico y decidir con base en resultado y guiándonos por la zona crítica o de rechazo si:

a) Aceptamos $H_0$.
b) Rechazamos $H_0$ (y aceptamos $H_1$).
c) No tomamos ninguna decisión (si pensamos que los datos no son concluyentes)




Pruebas de una población

- Empezaremos con las pruebas unimuestrales en las que tratamos de probar si un parámetro calculado a partir de un estadístico es igual o no a un valor predeterminado o a un parámetro poblacional conocido.

- Estudiaremos cuatro pruebas en este capítulo:
  - Una media con varianza conocida (Excel)
  - Una media con varianza desconocida
  - Una varianza
  - Una proporción

Una media con varianza conocida (Excel)

- $\alpha = 1$- ABS (PRUEBA.Z(Rango, $\mu$))
  - Devuelve el valor de $\alpha$ para una cola

$$H_1: \mu < \mu; \mu < \mu$$

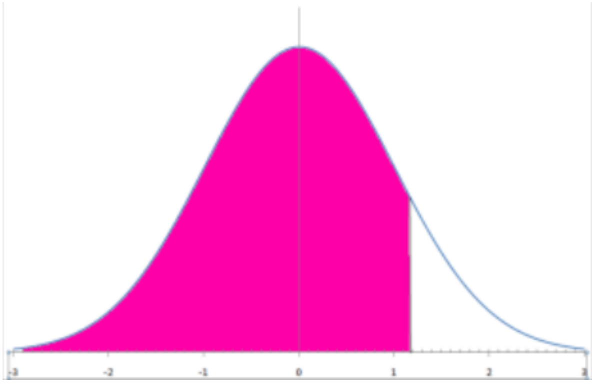
- Si es menor que $\alpha$ esperado rechazamos $H_0$
- Rango: la muestra ($n < 30$)
- $\mu$: La medida con la que se quiere comparar
- Equivale a:

$\alpha$= DISTR.NORM. ESTAND(-ABS($\bar{x} - \mu/(\sigma/\sqrt{n})$))

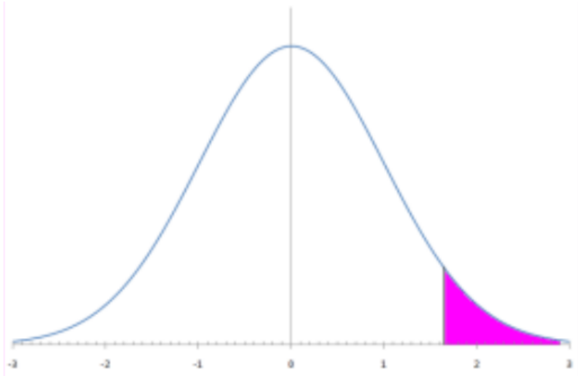
- Para dos colas
  - $H_1: \mu < \mu; \mu < \mu$
  - $\alpha = (1$- ABS (PRUEBA.Z(Rango, $\mu$)))*2

=ABS(PRUEBA.Z(Rango $\mu$)

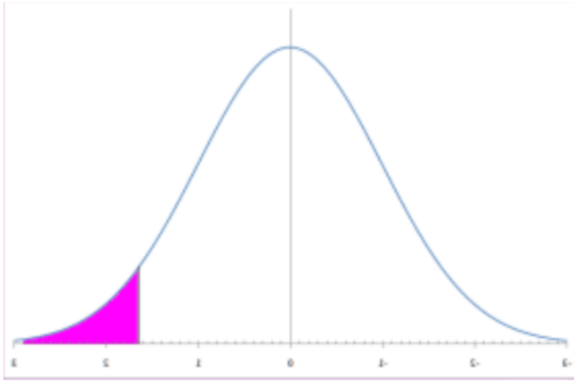
  - 1- $\alpha$ (rechazo cuando es mayor)



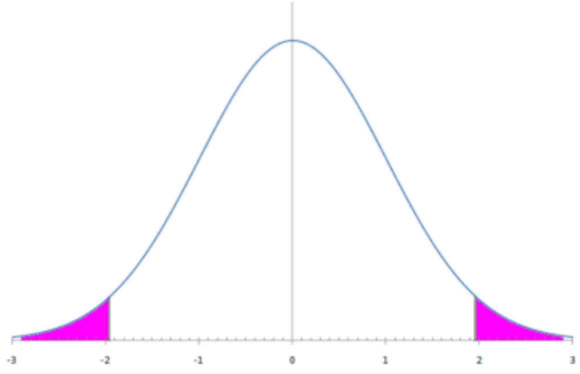
- =1-ABS(PRUEBA.Z(Rango, $\mu$))
  - $\alpha$ (rechazo cuando es menor)



- =DISTR.NORM. ESTAND(-ABS($\bar{x} - \mu/(\sigma/\sqrt{n})$))
  - $\alpha$ (rechazo cuando es menor)



- =(1-ABS(PRUEBA.Z (Rango, $\mu$)))*2)
    - $\alpha$ (rechazo cuando es menor)




Caso General Z

- Cálculo Z muestra a mano
- Probabilidad $\alpha$
    - Rechazo cuando es menor
    - $\alpha$=DISTR.NORM.ESTAND(-ABS(Z))
        - 1 cola
    - $\alpha$=(DISTR.NORM.ESTAND(-ABS(Z)))*2
        - 2 colas
- Región de Rechazo: $\mathrm{Z}_{(\mathrm{p})}; \mathrm{p} = \alpha; \mathrm{p} = \alpha/2$
    - Rechazo cuando entra en región. (Depende $H_1$)
    - =DISTR.NORM.ESTAND.INV(p)

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 cinco casos:
  - Dos varianzas independientes,
  - Dos medias independientes con varianzas conocidas
  - Dos medias independientes con varianzas desconocidas e iguales
  - Dos medias independientes con varianzas desconocidas y desiguales
  - Dos proporciones.

Prueba F para 2 varianzas

- =PRUEBA.F(Rango 1, Rango 2)
  - $1 - \alpha/2$ (rechazo cuando es mayor)

- =PRUEBA.F(Rango 1, Rango 2)
  - $\alpha/2$ (rechazo cuando es menor)

- =DISTR.F$(\mathrm{F}, v_1, v_2)$*2
  - $\alpha/2$ (rechazo cuando es menor)
  - $v_1 > v_2$

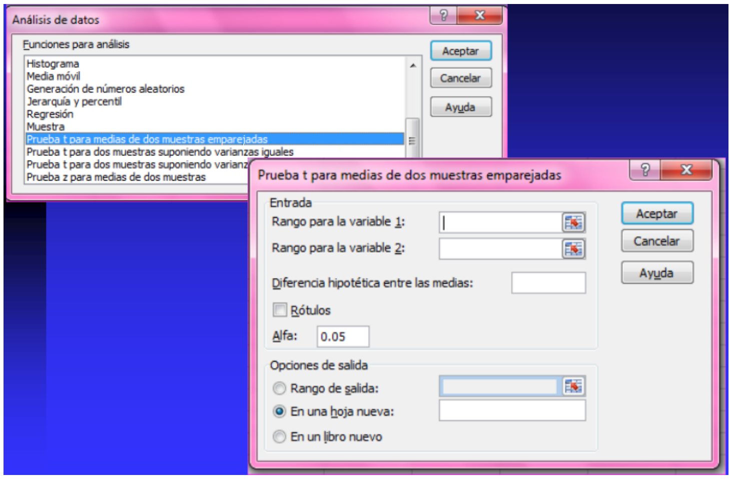
Caso General F

- Cálculo F a mano
- Probabilidad de a
  - a =DISTR.F$(\mathrm{F}, v_1, v_2)$*
  - 1 cola
- Región de Rechazo: $\mathrm{F}_{(\mathrm{p})}: \mathrm{p} = \alpha; \mathrm{p} = \alpha/2$
  - Rechazo cuando entra en región. (depende $H_1$)
  - =DISTR.F.INV$(\mathrm{p}, v_1, v_2)$

Prueba t para 2 Medias

- =PRUEBA.T(matriz 1; matriz 2; colas; tipo)
    - Devuelve a (rechazo cuando es mayor)
    - En 1 cola va de 0 a 0.5. Solo evalúa mayor.
    - En 2 colas va de 0 a 1.
    - Matriz 1 y Matriz 2 son datos de mi muestra
    - # Colas depende de $H_1; \neq: 2 < \text{o} >: 1$
    - Tipos:
        - 1: Muestras pareadas
        - 2: Varianzas iguales
        - 3: Varianzas desiguales

Pruebas de Muestras Dependientes




*Fuente:* Marcillo, F.(2010).

**Métodos de Muestreo**

Una muestra es un conjunto representativo de una población. Es decir, conjunto de unidades, una porción del total, que nos representa la conducta del universo en su conjunto.

Una muestra, en un sentido amplio, no es más que eso, una parte del todo que llamamos universo y que sirve para representarlo.

**Leyes del método del muestreo**

El método de muestreo se basa en ciertas leyes que le otorgan su fundamento científico, las cuales son:

- Ley de los grandes números: si en una prueba, la probabilidad de un acontecimiento o suceso es P, y si éste se repite una gran cantidad de veces, la relación entre las veces que se produce el suceso y la cantidad total de pruebas (es decir, la frecuencia F del suceso) tiende a acercarse cada vez más a la probabilidad P.
- Ley de la regularidad estadística: un conjunto de n unidades tomadas al azar de un conjunto N, es casi seguro que tenga las características del grupo más grande.
- Ley de la inercia de los grandes números: esta ley es contraria a la anterior. Se refiere al hecho de que, en la mayoría de los fenómenos, cuando una parte varía en una dirección, es probable que una parte igual del mismo grupo varíe en dirección opuesta.
- Ley de la permanencia de los números pequeños: si una muestra suficientemente grande es representativa de la población, una segunda muestra de igual magnitud deberá ser semejante a la primera; y, si en la primera muestra se encuentran pocos individuos con características raras, es de esperar encontrar igual proporción en la segunda muestra.

**Tipo de Muestras**

- **<u>Muestreo aleatorio simple</u>:** la forma más común de obtener una muestra es la selección al azar. es decir, cada uno de los individuos de una población tiene la misma posibilidad de ser elegido. Si no se cumple este requisito, se dice que la muestra es viciada. Para tener la seguridad de que la muestra aleatoria no es viciada, debe emplearse para su constitución una tabla de números aleatorios.

- **Muestreo estratificado:** una muestra es estratificada cuando los elementos de la muestra son proporcionales a su presencia en la población. La presencia de un elemento en un estrato excluye su presencia en otro. Para este tipo de muestreo, se divide a la población en varios grupos o estratos con el fin de dar representatividad a los distintos factores que integran el universo de estudio. Para la selección de los elementos o unidades representantes, se utiliza el método de muestreo aleatorio.

- **Muestreo por cuotas:** se divide a la población en estratos o categorías, y se asigna una cuota para las diferentes categorías y, a juicio del investigador, se selecciona las unidades de muestreo. La muestra debe ser proporcional a la población, y en ella deberán tenerse en cuenta las diferentes categorías. El muestreo por cuotas se presta a distorsiones, al quedar a criterio del investigador la selección de las categorías.

- **Muestreo intencionado:** también recibe el nombre de sesgado. El investigador selecciona los elementos que a su juicio son representativos, lo que exige un conocimiento previo de la población que se investiga.

- **Muestreo mixto:** se combinan diversos tipos de muestreo. Por ejemplo: se puede seleccionar las unidades de la muestra en forma aleatoria y después aplicar el muestreo por cuotas.

- **Muestreo tipo:** la muestra tipo (master simple) es una aplicación combinada y especial de los tipos de muestra existentes. Consiste en seleccionar una muestra "para ser usada" al disponer de tiempo, la muestra se establece empleando procedimientos sofisticados; y una vez establecida, constituirá el módulo general del cual se extraerá la muestra definitiva conforme a la necesidad específica de cada investigación.

**Clasificación de las muestras**

- **Muestra accidental.** Es aquella que se obtiene sin ningún plan preconcebido; las unidades elegidas resultan producto de circunstancias fortuitas. Si entrevistamos a los primeros 50 transeúntes que pasan por cierta calle o medimos la profundidad del mar a lo largo de un trayecto entre dos puntos cualesquiera, estaremos en presencia de una muestra accidental; los datos obtenidos podrán o no representar al universo en estudio. El investigador no puede saber hasta qué punto sus resultados podrán proyectarse, con confiabilidad, hacia el conjunto más amplio que desea conocer.

- **Muestra por cuotas.** Consiste en predeterminar la cantidad de elementos de cada categoría que habrán de integrar la muestra. Así podemos asignar una cuota de 50 hombres y 50 mujeres a una muestra de 100 individuos, asumiendo que ésa es la distribución de la población total. Por más que esa presunción llegue a ser válida, no deja de existir cierta arbitrariedad en este modo de proceder, por lo que la rigurosidad estadística de las muestras por cuotas se reduce considerablemente.

- **Muestra intencional.** Las unidades se eligen en forma arbitraria, designando a cada unidad según características que para el investigador resulten de relevancia. Se emplea, por lo tanto, el conocimiento y la opinión personal para identificar aquellos elementos que deben ser incluidos en la muestra. Se basa, primordialmente, en la experiencia de alguien con la población. Estas muestras son muy útiles y se emplean frecuentemente en los estudios de caso, por más que la posibilidad de generalizar conclusiones a partir de ellas, sea en rigor nula. En algunas oportunidades se usan como guía o muestra tentativa para decidir cómo tomar una muestra aleatoria más adelante.

**Muestras Aleatorias**

Los procedimientos más usuales para la obtención de muestras aleatorias son:

- **Azar simple.** Este procedimiento se inicia confeccionando una lista de todas las unidades que configuran el universo, numerando correlativamente cada una de ellas. Luego, mediante cualquier sistema (tabla de números al azar, programas de computación), se van sorteando al azar estos números hasta completar el total de unidades que deseamos que entren en la muestra. De este modo, la probabilidad que cada elemento tienen de aparecer en la muestra es exactamente la misma. Si cada uno de los elementos que integran la población no tiene la misma posibilidad de ser elegido, se habla entonces de una muestra viciada. Este método nos garantiza una selección completamente aleatoria, pero resulta muy lento y costoso, pues nos obliga a elaborar listas completas de todas las unidades de interés, lo que a veces es sencillamente imposible. Por este motivo, sólo se emplea cuando los universos son relativamente pequeños. Este método no será adecuado si, por ejemplo, queremos sacar una muestra de todas las personas analfabetas que existen en un país. En cambio, si nuestra intención es extraer una muestra del universo de todos los alumnos que ingresan a una universidad en un determinado año, resultará muy adecuado.

- **Azar sistemático.** También se requiere de un listado completo de las unidades que integran el universo en estudio. Luego se efectúan las siguientes operaciones:
    - Se calcula la constante K, que resulta de dividir el número total de unidades que componen el universo por el número de unidades que habrán de integrar la muestra:

$K = N/n$

Donde:

$N$ = número total de unidades que componen el universo.

n = número total de unidades que integrarán la muestra.

- **Muestras por conglomerados.** Esta técnica tiene utilidad cuando el universo que se requiere estudiar admite ser subdividido en universos menores de características similares a las del universo total. Se procede a subdividir el universo en un número finito de conglomerados y, entre ellos, se pasa a elegir algunos que serán los únicos que se investigarán; esta elección puede realizarse por el método del azar simple o por el del azar sistemático. Una vez cumplida esta etapa, puede efectuarse una segunda selección, dentro de cada uno de los conglomerados elegidos, para llegar a un número aún más reducido de unidades muestrales.

La ventaja de esta técnica es que obvia la tarea de confeccionar el listado de todas las unidades del universo. Su desventaja mayor radica en que, al efectuarse el muestreo en dos etapas, los errores muestrales de cada una se van acumulando, lo que da un error mayor que para los métodos anteriores. La técnica de conglomerados suele utilizarse cuando queremos extraer muestras de los habitantes de un conjunto geográfico amplio, por ejemplo, una gran ciudad o un conjunto de pueblos, por lo que se procede a tomar cada pueblo o grupo de manzanas como un conglomerado independiente; del mismo modo, se la utiliza para conocer las reservas forestales y marinas, para estudiar las estrellas y otros casos semejantes.

- **Tamaño de la muestra y error muestral.** Recordemos que la muestra descansa en el principio de que las partes representan al todo y, por tal, refleja las características que definen a la población de la cual fue extraída, lo cual nos indica que es representativa. Es decir, que, para hacer una generalización exacta de una población, es necesario tomar una muestra representativa. Por lo tanto, la validez de la generalización depende de la validez y tamaño de la muestra.

Cuando trabajamos con muestras, generalmente se presentan dos tipos de errores:

- **Error de muestreo:** Es el error natural de estimación provocado por trabajar con muestras y no con la población y puede ser controlado por el investigador.
- **Error de no muestreo**
  - No respuesta (si no responden, puede haber sesgo en la información)
  - Respuesta inexacta o falsa
  - Sesgo de selección (sustituciones fortuitas)

**Referencia:**

Marcillo, F.(2010). Curso Práctico de Bioestadística Con Herramientas De Excel. [Diapositivas de Power Point]. Escuela Superior Politécnica del Litoral. https://www.researchgate.net/publication/41151207_Curso_Practico_de_Bioestadistica_Con_Herramientas_De_Excel_-_Clase_1

Mena, Oscar. (2020). Métodos de Muestreo. [Diapositivas de Power Point]. Centro de Investigación y Estudios Avanzados del IPN. Recuperado de: https://www.academia.edu/36925547/METODOS_DE_MUESTREO