



UNIVERSIDAD AUTÓNOMA DEL PERÚ
DEPARTAMENTO DE FORMACIÓN GENERAL

TEXTO

ESTADÍSTICA

LIMA-PERÚ

TÍTULO DE LA OBRA
ESTADÍSTICA

AUTORES

Mg. Edward F. Huamaní Alhua
Lic. Gilberth Pesantes Calderón
Mg. Juan C. Oruna Lara

Universidad Autónoma del Perú
Panamericana Sur km 16.3 Villa El Salvador

PRIMERA EDICIÓN MARZO 2016

Tiraje: 1800 Ejemplares

EDITOR GENERAL

Universidad Autónoma del Perú
Panamericana Sur km 16.3 Villa El Salvador

DISEÑO Y DIAGRAMACIÓN

Universidad Autónoma del Perú

DISEÑO CARATULA

Universidad Autónoma del Perú

Ley 26905 Biblioteca Nacional Perú

Hecho el depósito Legal en la Biblioteca Nacional del Perú N° 2016-03688

ISBN 978-612-4286-04-9

IMPRESO EN

IMPRESIÓN ARTE PERÚ S.A.C
Jr. Recuay 375-A, Breña
Telf: 3323401 RPC: 986601361

Derechos reservados conforme a Ley N° 822.

Queda terminantemente prohibida la reproducción total o parcial de esta obra por cualquier medio, ya sea electrónico, mecánico, químico, óptico, incluyendo el sistema de fotocopiado, sin autorización escrita de los autores, quedando protegidos los derechos de propiedad intelectual y auditoría por la legislación peruana.

ÍNDICE

PRESENTACIÓN	7
INTRODUCCIÓN A LA ASIGNATURA	8
UNIDAD 1: METODO CIENTÍFICO Y ORGANIZACIÓN DE DATOS	9
1.1. FUNDAMENTOS DE INVESTIGACIÓN	11
1.1.1. LA CIENCIA.....	11
1.1.2. CONOCIMIENTO CIENTÍFICO	11
1.1.3. LA ESTADÍSTICA.....	12
1.1.4. CLASIFICACIÓN DE LA ESTADÍSTICA.....	12
1.1.5. POBLACIÓN.....	13
1.1.6. MUESTRA	13
1.1.7. TIPOS DE MUESTREO.....	14
1.1.8. PARÁMETRO Y ESTADÍGRAFO	17
1.1.9. VARIABLE	18
1.1.10. EL MÉTODO CIENTÍFICO.....	19
1.1.11. LA INVESTIGACIÓN CIENTÍFICA	20
1.1.12. ETAPAS DEL MÉTODO ESTADÍSTICO	20
GUÍA DE PRÁCTICA N°1.....	25
1.2. MARCO METODOLÓGICO DE INVESTIGACIÓN.....	32
1.2.1. MARCO METODOLÓGICO	32
1.2.2. VARIABLE ESTADÍSTICA	32
1.2.3. OPERACIONALIZACIÓN DE VARIABLES	36
1.2.4. TIPO DE ESTUDIO	37
1.2.5. DISEÑO DE ESTUDIO	37
1.2.6. TÉCNICAS DE RECOLECCIÓN DE DATOS.....	37
1.2.7. ESCALAS DE MEDICIÓN	38
1.2.8. SUMATORIAS.....	41
GUÍA DE PRÁCTICA N°2.....	43
1.3. TABLAS DE DISTRIBUCIÓN DE FRECUENCIAS.....	48
1.3.1. TABLA DE DISTRIBUCIÓN DE FRECUENCIAS.....	48
1.3.2. TABLA DE DISTRIBUCIÓN DE FRECUENCIAS DE UNA VARIABLE CUALITATIVA.....	49
1.3.3. TABLA DE FRECUENCIAS DE VARIABLES CUANTITATIVAS DISCRETAS	50
1.3.4. TABLA DE FRECUENCIAS DE VARIABLES CUANTITATIVAS CONTINUAS	51
GUÍA PRÁCTICA N° 3	53
1.4. REPRESENTACIONES GRÁFICAS.....	59
1.4.1. GRÁFICOS ESTADÍSTICOS.....	59
1.4.2. PARTES DE UN GRÁFICO	59
1.4.3. CLASES DE GRÁFICOS.....	59

GUÍA DE PRÁCTICA N°4	66
UNIDAD 2: MEDIDAS ESTADÍSTICAS DESCRIPTIVAS Y ANÁLISIS DE REGRESIÓN - CORRELACIÓN	73
2.1. MEDIDAS DE TENDENCIA CENTRAL	74
2.1.1. DEFINICIÓN	74
2.1.2. MEDIA ARITMÉTICA	74
2.1.3. MODA	75
2.1.4. MEDIANA	77
2.1.5. COMPARACIÓN ENTRE LA MEDIA, MODA Y MEDIANA	79
2.1.6. MEDIA PONDERADA	80
2.1.7. CUARTILES	81
2.1.8. DECILES	82
2.1.9. PERCENTILES	82
GUÍA PRÁCTICA N° 5	84
2.2. MEDIDAS DE DISPERSIÓN	90
2.2.1. DEFINICIÓN	90
2.2.2. RANGO O RECORRIDO	90
2.2.3. VARIANZA Y DESVIACIÓN ESTÁNDAR	91
2.2.4. COEFICIENTE DE VARIACIÓN	93
GUÍA PRÁCTICA N°6	94
2.3. ANÁLISIS DE CORRELACIÓN Y REGRESIÓN LINEAL SIMPLE	99
2.3.1. INTRODUCCIÓN	99
2.3.2. ANÁLISIS DE CORRELACIÓN	99
2.3.3. TIPOS DE CORRELACIÓN	99
2.3.4. DIAGRAMA DE DISPERSIÓN	100
2.3.5. COEFICIENTE DE CORRELACIÓN	101
2.3.6. ANÁLISIS DE REGRESIÓN	102
2.3.7. REGRESIÓN LINEAL SIMPLE	103
2.3.8. MÉTODO DE MÍNIMOS CUADRADOS PARA ESTIMAR LOS COEFICIENTES DE REGRESIÓN	103
2.3.9. RECTA DE MÍNIMOS CUADRADOS	104
2.3.10. COEFICIENTE DE DETERMINACIÓN (R^2)	106
GUÍA DE PRÁCTICA N°7	107
UNIDAD 3: PROBABILIDADES Y DISTRIBUCIONES MUESTRALES	115
3.1. PROBABILIDAD BÁSICA	117
3.1.1. PROBABILIDAD Y EVENTOS	117
3.1.2. EXPERIMENTO	117
3.1.3. EXPERIMENTO ALEATORIO	118
3.1.4. ESPACIO MUESTRAL Y EVENTOS	118

3.1.5. TÉCNICAS DE CONTEO	119
3.1.6. CONCEPTO CLÁSICO DE PROBABILIDAD	121
GUÍA DE PRÁCTICA N°8.....	123
3.2. PROBABILIDAD CONDICIONAL.....	127
3.2.1. CONCEPTO	127
3.2.2. REGLA DE MULTIPLICACIÓN	128
3.2.3. TEOREMA DE LA PROBABILIDAD TOTAL	128
3.2.4. TEOREMA DE BAYES	129
GUÍA DE PRÁCTICA N°9.....	130
3.3. DISTRIBUCIÓN DE PROBABILIDAD CONTINUA.....	135
3.3.1. VARIABLE ALEATORIA	135
3.3.2. TIPOS DE VARIABLES ALEATORIAS	136
3.3.3. VARIABLE ALEATORIA DISCRETA.....	136
3.3.4. VARIABLE ALEATORIA CONTINUA	139
3.3.5. DISTRIBUCIÓN NORMAL	141
3.3.6. DISTRIBUCIONES MUESTRALES	143
3.3.7. DISTRIBUCIÓN MUESTRAL DE LA MEDIA	144
3.3.8. DISTRIBUCIÓN MUESTRAL DE UNA PROPORCIÓN	146
GUÍA DE PRÁCTICA N°10	148
3.3. ESTIMACIÓN DE PARÁMETROS Y TAMAÑO DE MUESTRA.....	153
3.4.1. ESTIMACIÓN DE PARÁMETROS.....	153
3.4.2. INTERVALO DE CONFIANZA.....	153
3.4.3. INTERVALO DE CONFIANZA PARA LA MEDIA POBLACIONAL μ DE UNA POBLACIÓN NORMAL SI LA VARIANZA POBLACIONAL σ^2 ES CONOCIDA	154
3.4.4. INTERVALO DE CONFIANZA PARA UNA PROPORCIÓN POBLACIONAL	155
3.4.5. TAMAÑO DE MUESTRA.....	155
GUÍA DE PRÁCTICA N°11	158
UNIDAD 4: PRUEBAS DE HIPÓTESIS	165
4.1. PRUEBA DE HIPÓTESIS EN UNA MUESTRA PARA MEDIA Y PROPORCIÓN POBLACIONAL	166
4.1.1. INTRODUCCIÓN	166
4.1.2. HIPÓTESIS ESTADÍSTICA.....	166
4.1.3. PRUEBA DE HIPÓTESIS.....	167
4.1.4. ERRORES TIPO I Y TIPO II	167
4.1.5. PASOS DE UNA PRUEBA DE HIPÓTESIS.....	168
4.1.6. PRUEBA DE HIPÓTESIS PARA LA MEDIA CON VARIANZA POBLACIONAL σ^2 CONOCIDA	170
4.1.7. PRUEBA DE HIPÓTESIS PARA UNA PROPORCIÓN	171
GUÍA DE PRÁCTICA N°12	172

4.2. PRUEBA DE HIPÓTESIS EN DOS MUESTRAS INDEPENDIENTES PARA MEDIA Y PROPORCIÓN POBLACIONAL	178
4.2.1. PRUEBA DE HIPÓTESIS PARA LA DIFERENCIA DE MEDIAS.....	178
4.2.2. PRUEBA DE HIPÓTESIS PARA DIFERENCIA DE PROPORCIONES	179
GUÍA DE PRÁCTICA N°13	181
4.3. PRUEBA DE INDEPENDENCIA CHI-CUADRADO DE DOS VARIABLES CUALITATIVAS .	187
4.3.1. PRUEBAS NO PARAMÉTRICAS	187
4.3.2. PRUEBA CHI-CUADRADO PARA LA INDEPENDENCIA DE DOS VARIABLES	187
GUÍA DE PRÁCTICA N°14	193
ANEXOS	198
REFERENCIAS BIBLIOGRÁFICAS	203

PRESENTACIÓN

Con mucho beneplácito se presenta a la comunidad universitaria el presente Texto: Estadística, este ha sido elaborado por docentes del curso de Estadística de la Universidad Autónoma del Perú, con la finalidad de orientar y facilitar el aprendizaje significativo de nuestros estudiantes.

Para la elaboración de este Texto, se han consultado diversas fuentes, las cuales han sido contextualizadas a las necesidades del curso y a las características metodológicas que la universidad aplica como estrategia del desarrollo cognoscitivo de nuestros estudiantes.

La Estadística, nace de las necesidades reales del hombre. La variada y cuantiosa información relacionada con éste y que es necesaria para la toma de decisiones, hace que la estadística sea hoy, una importante herramienta de trabajo.

Este Texto pretende, en nuestros estudiantes, sentar las bases teórico prácticas sobre Cultura Estadística tan fundamental en los tiempos actuales de abundante información y con urgencia de una adecuada toma de decisiones en todo plano de la vida.

Para un mejor aprovechamiento se ha dividido el presente texto en 4 unidades: Unidad I: Método Científico y Organización de Datos, Unidad II: Medidas Estadísticas Descriptivas y Análisis de Regresión y Correlación, Unidad III: Introducción a Probabilidades y Distribuciones Muestrales y en Unidad IV: Pruebas de Hipótesis.

El Texto de Estadística, se ha estructurado con aspectos teóricos y prácticos de estadística con el fin de que nuestros estudiantes desarrollen sus capacidades cognitivas y las habilidades numéricas que le permitan tomar decisiones en el actual contexto de real incertidumbre.

INTRODUCCIÓN A LA ASIGNATURA

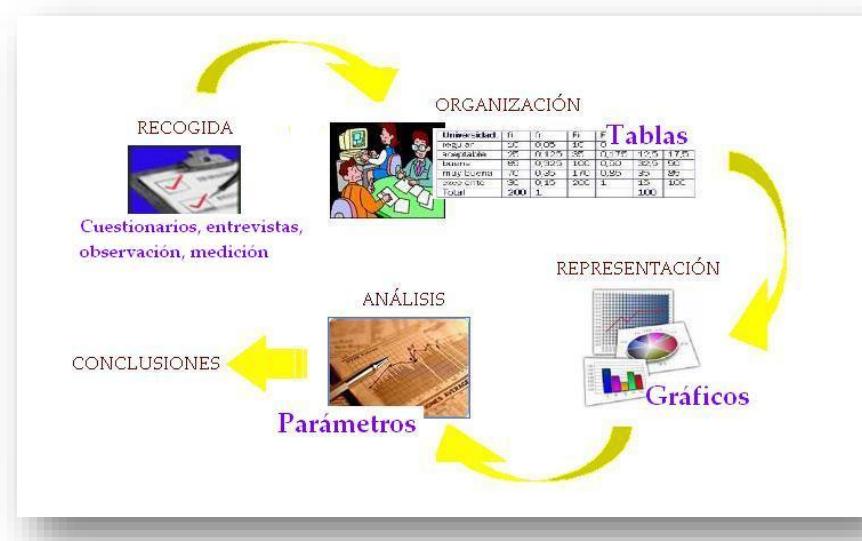
En nuestros días, la estadística se ha convertido en método efectivo para describir con exactitud los valores de datos económicos, políticos, sociales, psicológicos, biológicos y físicos, y sirve como herramienta para relacionar y analizar dichos datos. El objetivo que se persigue es ya no sólo reunir y tabular los datos, sino sobre todo conocer el proceso de interpretación de esa información.

La estadística es una palabra de uso común; se emplea en periódicos, noticieros de radio y televisión, y por personas de diversas ocupaciones. Los comentaristas deportivos hablan de las estadísticas del juego de fútbol. Los noticiarios hablan de las estadísticas de criminalidad, de producción, o de educación. La palabra se encuentra arraigada en la cultura popular por lo cual toda persona independientemente de su profesión debería tener conocimientos de estadística.

La creciente complejidad de las actividades de los negocios en años recientes ha incrementado definitivamente el uso de la estadística para tomar decisiones en cualquier nivel de la administración. Los hombres de negocios utilizando la estadística pueden producir los volúmenes de ventas, medir las reacciones de los consumidores ante nuevos productos, tomar decisiones de cómo invertir el presupuesto para publicidad, determinar el mejor método para utilizar las habilidades y aptitudes de sus empleados. El hombre de negocios utiliza encuestas estadísticas para determinar la reacción del público ante sus nuevos productos.

Teniendo en cuenta lo manifestado anteriormente es necesaria la estadística por su base científica al tomar decisiones. La Estadística puede dar respuesta a muchas de las necesidades que la sociedad actual nos plantea. Su tarea fundamental es la reducción de datos, con el objetivo de representar la realidad y transformarla, predecir su futuro o simplemente conocerlo.

UNIDAD 1: METODO CIENTÍFICO Y ORGANIZACIÓN DE DATOS



<http://bit.ly/1SIBma1>

Contenido Temático

- ❑ Fundamentos de Investigación
- ❑ Marco Metodológico de investigación
- ❑ Tablas de distribución de frecuencias
- ❑ Representaciones Gráficas

1.1. FUNDAMENTOS DE INVESTIGACIÓN

1.1.1. LA CIENCIA

Es el conjunto de conocimientos razonablemente relacionados con el fin de alcanzar la verdad.

El ideal de la ciencia es llegar a conocer y explicar todo.

La ciencia es un sistema de conocimientos ordenados de la que se deducen principios y leyes generales, cuya veracidad se comprueba y se puntualiza constantemente, por consiguiente es falible.

Para cumplir con sus propósitos la ciencia emplea mediciones, especifica condiciones de observación, efectúa experimentos y persigue la generalización. Es un estudio sistemático que se caracteriza por que se corrige a sí misma

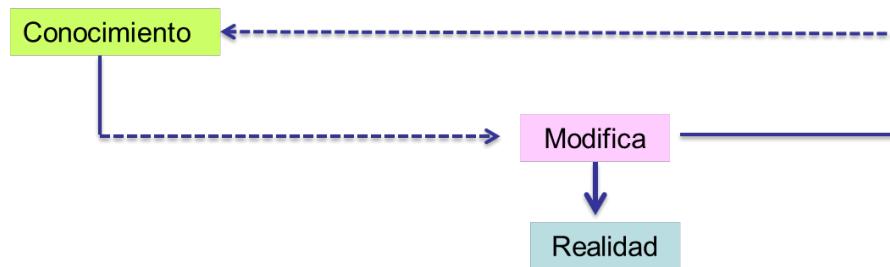


Los tres elementos anteriores permiten toda relación científica, hasta el punto que no puede suprimirse uno de ellos, pues no podríamos concebir la ciencia sin base en la realidad, y esta se torna en ciencia por la investigación.

1.1.2. CONOCIMIENTO CIENTÍFICO

Conocer es una actividad por medio de la cual el hombre adquiere certeza de la realidad, y que se manifiesta como un conjunto de representaciones sobre las cuales tenemos certeza de que son verdaderas.

El conocimiento científico es una de las formas que tiene el hombre para otorgarle un significado con sentido a la realidad.



<http://es.slideshare.net/PEDROHUERGO/>

El conocimiento científico resulta de observar, descubrir, explicar o predecir la realidad, convirtiéndose en un conocimiento sistemático.

1.1.3. LA ESTADÍSTICA

La estadística es una ciencia que nos proporciona métodos y procedimientos de recolección, organización, representación, análisis e interpretación de datos para la toma de decisiones y predicción de fenómenos.



<http://cibertareas.info/graficas-y-estadisticas.html>

1.1.4. CLASIFICACIÓN DE LA ESTADÍSTICA

La Estadística se clasifica en:

ESTADÍSTICA DESCRIPTIVA. Trata de la recolección, clasificación, presentación y descripción de los datos, sin sacar conclusiones sobre un grupo mayor. El campo de validez de las conclusiones obtenidas se extiende únicamente al conjunto de unidades observadas.

ESTADÍSTICA INFERENCIAL. Proceso a través del cual se obtienen conclusiones sobre una población, a través de la información que proporciona una muestra. La confianza de tal extrapolación dependerá de la **representatividad de la muestra**. Generalmente el análisis inferencial se lleva a cabo para realizar predicciones, mostrar relaciones de causa y efecto, así como para probar hipótesis y teorías científicas.



<https://bibliotecadeinvestigaciones.files.wordpress.com/>

1.1.5. POBLACIÓN

Es la totalidad de individuos, elementos o medidas que poseen alguna característica común susceptible de ser estudiada. Tiene las siguientes características:

- Homogeneidad:** que todos los miembros de la población tengan las mismas características.
- Tiempo:** se refiere al período de tiempo donde se ubicaría la población de interés.
- Espacio:** se refiere al lugar donde se ubica la población de interés. Un estudio no puede ser muy abarcador, hay que limitarlo a un área o comunidad en específico.
- Cantidad:** se refiere al tamaño de la población. La falta de recursos y tiempo también nos limita la extensión de la población.

Ejemplos

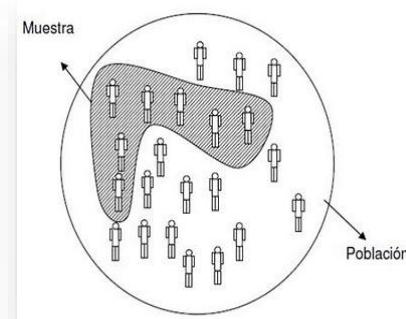
- Todos los clientes de la empresa de cable Telecom en la provincia de Lima.
- El total de alumnos del colegio “Pedro Ruiz Gallo” en el distrito de Chorrillos
- La totalidad de Empresas del sector Minero del Perú en el año 2012

1.1.6. MUESTRA

Es un subconjunto de elementos seleccionados de una población, lo ideal es que sea un subconjunto representativo de toda la población, es decir que refleje las características esenciales de la misma y se pueda realizar generalizaciones.

Las razones para trabajar con muestras son:

Ahorro de tiempo, ahorro de dinero, facilidades operativas.



<https://estadistikids.wordpress.com/2012/06/22/>

Ejemplos:

- 50 clientes de la empresa de cable Telecom en la provincia de Lima
- 250 alumnos del colegio “Pedro Ruiz Gallo” en el distrito de Chorrillos
- 11 empresas del sector minero de Perú escogidas al azar.

DATO. Valor de la variable asociado con un elemento de la población o muestra. Puede ser un número, una palabra o un símbolo.

UNIDAD ESTADÍSTICA. Los individuos u objetos de una población que tienen una característica medible.

1.1.7. TIPOS DE MUESTREO

Los tipos de muestreo de manera general pueden ser Probabilísticas y No Probabilísticas

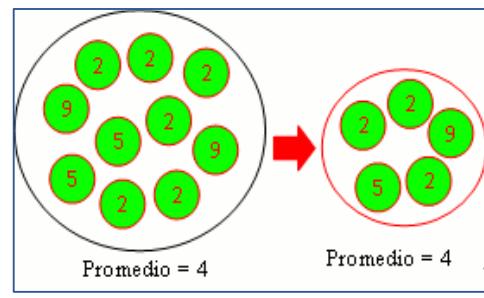
A. MUESTREO PROBABILÍSTICO. Cuando cada unidad o elemento de la población tienen una determinada probabilidad de ser incluida en la muestra. Los principales muestreos de este tipo son:

A1. MUESTREO ALEATORIO SIMPLE (M.A.S.). Cuando todos y cada uno de los elementos de la población tienen igual probabilidad debe ser incluidos en la muestra. Se caracteriza por:

- Sencillo y de fácil comprensión
- Se requiere de antemano un listado completo de toda la población
- Cuando se trabaja con muestras pequeñas es posible que no represente a la población adecuadamente

Para seleccionar una muestra aleatoria simple se debe tener en cuenta las siguientes recomendaciones:

- a) Enumerar los elementos de la población del 1 hasta N (N es el tamaño de la población)
- b) Utilizar algún procedimiento para seleccionar los n elementos de la población que conformaran la muestra. Puede ser un sorteo, Tabla de números aleatorios o algún programa computacional que genere números aleatorios.



<http://bit.ly/1Qo1YY3>

Ejemplo1: en un colegio se desea obtener una muestra de 20 alumnos del quinto año de secundaria. En un hospital se desea obtener una muestra de 40 pacientes con tuberculosis.

Ejemplo2: una empresa tiene 120 trabajadores y se quiere extraer una muestra aleatoria simple de 30 trabajadores. Para ello se numeran los trabajadores del 1 al 120, se sortean 30 números de entre los 120. Entonces, la muestra estará formada por los 30 trabajadores a los que les correspondan los números obtenidos

A2. MUESTREO SISTEMÁTICO: ($K=N/n$). Es un procedimiento de selección por el cual el primer elemento de la muestra es elegida al azar entre las K primeras unidades poblacionales y luego el resto de las unidades se seleccionan cada K-ésima unidad o elemento de la población. Se caracteriza por:

- Rapidez y facilidad de selección de la muestra
- No siempre es necesario tener un listado de toda la población

- Cuando la población está ordenada siguiendo una tendencia conocida, asegura una cobertura de unidades de toda la población

Con el muestreo sistemático se logra mayor eficiencia si las unidades que se hallan próximas tienen mayor uniformidad que las unidades que se encuentran alejadas entre sí. El muestreo Sistemático es especialmente útil en auditorías, cuando la información relevante se registra en forma ordenada, como en la memoria de una computadora o en un archivo de tarjetas.

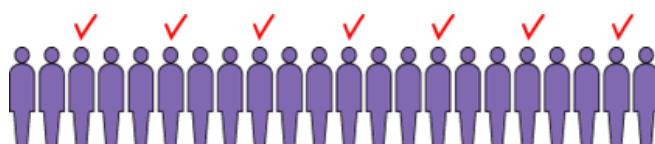


<http://bit.ly/1QpsKZJ>

Ejemplo: una empresa tiene 120 trabajadores y se quiere extraer una muestra aleatoria sistemática de 30 trabajadores. Para ello se numeran los trabajadores del 1 al 120, Se calcula el intervalo constante entre cada individuo mediante:

$$\frac{N(\text{Población})}{n(\text{muestra})} = \frac{120}{30} = 4$$

Se sortea un número del 1 al 4. Supongamos que sale el número 2; entonces el primer trabajador seleccionado para la muestra será el número 2, los siguientes trabajadores se obtendrían sumando 4, hasta llegar a tener 30 trabajadores. Los trabajadores seleccionados para la muestra serían los que se correspondan a los números: 2, 6, 10, 14, 18,

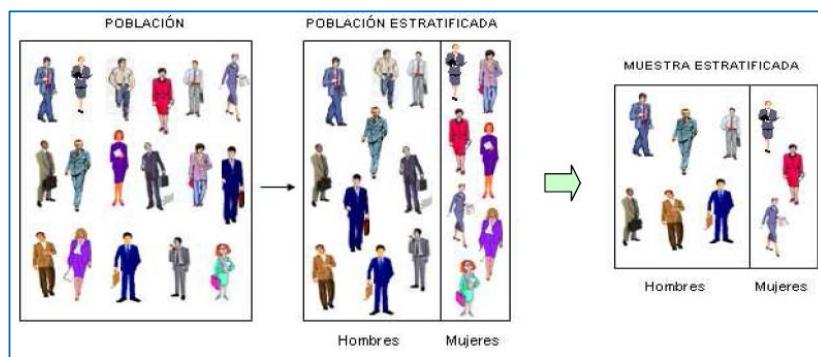


<http://bit.ly/21cwful>

A3. MUESTREO ESTRATIFICADO. El procedimiento consiste en dividir a la población en grupos llamados **estratos**. Dentro de cada estrato los elementos deber ser lo más homogéneo posibles con respecto a las características de la(s) variable(s) en estudio. Los estratos deben ser homogéneos dentro de sí y heterogéneos entre ellos. Se caracteriza por:

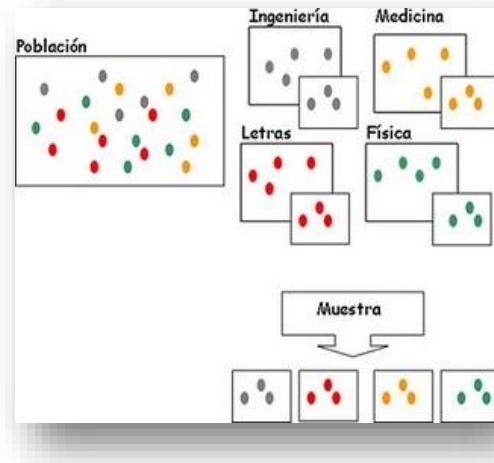
- Tiende a asegurar que la muestra represente adecuadamente a la población en función de las variables seleccionadas
- Se obtiene estimaciones más precisas.

Ejemplo1: los conos de Lima Metropolitana, los niveles de secundaria, la estratificación según el sexo.



<http://bit.ly/1KYHdzE>

Ejemplo 2, si se quisiera estudiar las actitudes políticas de los estudiantes de una universidad, se podría subdividir en estratos de acuerdo con el tipo de estudios que cursen, suponiendo que estas actitudes van a ser diferentes entre quienes siguen ingeniería, letras, medicina u otras carreras. Luego, se efectuaría un muestreo dentro de cada sub universo así definido para, finalmente, realizar un análisis integrando los resultados de todas las sub muestras.



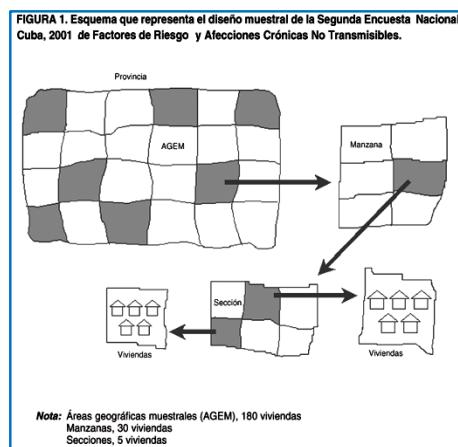
<http://bit.ly/1KYH0fN>

A4. MUESTREO POR CONGLOMERADOS. Es un sistema de muestreo en el que las unidades de análisis de la población se consideran conglomerados o unidades primarias. Se considera como unidad de muestreo al conglomerado y extraemos una muestra de conglomerados a partir del cual se estimará los parámetros poblacionales. El número de unidades elementales se denomina tamaño del conglomerado. Los conglomerados deben ser heterogéneos dentro de sí y homogéneos entre ellos. Se caracteriza por:

- Es muy eficiente cuando la población es muy grande y dispersa
- Reduce costos.
- No es necesario tener un listado de toda la población, solo de las unidades primarias.
- Se puede utilizar como marco como áreas geográficas cuyas características ya están ya muy delimitadas.



Ejemplos: Los distritos del cono sur, los colegios estatales del distrito de San Juan de Lurigancho, las Viviendas de una manzana, fábricas de producción de harina de Chimbote.



<http://bit.ly/1oOjLXO>

B. MUESTREO NO PROBABILÍSTICO. Se basa en opinión (criterio o juicio) personal del investigador. Donde el investigador con su experiencia designa cuales elementos forman parte de la muestra, sin embargo, debe evitarse, ya que no puede hacerse ninguna inferencia válida si la muestra se eligió usando este tipo de muestreo.

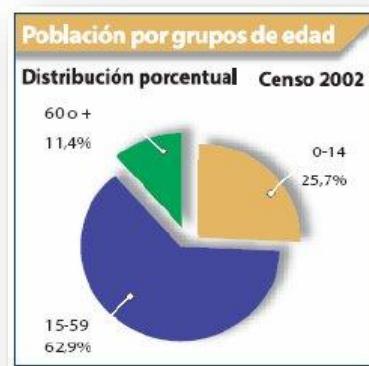
1.1.8. PARÁMETRO Y ESTADÍGRAFO

PARÁMETRO. Es una medida de resumen que nos describe alguna característica de la población, para calcular dicho valor es necesario utilizar todos los valores de la población. Entre los parámetros más conocidos tenemos:

- Media poblacional simbolizado por μ
- Varianza poblacional simbolizado por σ^2
- Proporción poblacional simbolizado por P

Ejemplos:

- El costo promedio de una casa en el distrito de Villa el Salvador
- El porcentaje de empleados que tienen automóvil dentro de una empresa.



<http://bit.ly/1orTGS8>

ESTADÍSTICO O ESTADÍGRAFO. Es una medida que nos describe alguna característica de interés y cuyo valor es calculado utilizando sólo los valores de los elementos o unidades de una muestra. Entre los estadísticos más conocidos tenemos:

- Media muestral simbolizado por \bar{x}
- Varianza muestral simbolizado por s^2
- Proporción muestral simbolizado por p

Ejemplo

- La venta promedio mensual de 10 empresas elegidas al azar del ramo textil.
- El salario promedio de una muestra de los Gerentes de una empresa.
- El porcentaje de Clientes que prefieren Pepsi en un grupo elegido al azar.



<http://bit.ly/1PJPG8N>

1.1.9. VARIABLE

Una variable es cualquier característica de los elementos de una población susceptible de tomar diferentes valores. Todo aquello que puede ser medido, observado o manipulado durante un estudio.

Es una propiedad o característica que puede ser percibida (o medida) y que cambia de un sujeto u objeto a otro o en el mismo sujeto u objeto a lo largo del tiempo.



Ejemplos:

- Estado Civil de una persona: {Casado, Soltero, Viudo}
- El número de hijos de una familia: {0, 1, 2, 3, ...}
- La altura de los alumnos: {1,62 ; 1,74; ...}
- Marca de TV que prefiere un cliente: {LG, Samsung, Sony, Panasonic}
- Raza de perros: Bulldog, Chow, Terrier

<http://bit.ly/20ZU6kk>

TIPOS DE VARIABLES

a. **Variable Cualitativa:** Son variables cuyos valores consisten en categorías de clasificación y responde a una cualidad o atributo. la característica o variable que se estudia no es numérica.

Ejemplos: Procedencia (Costa, Sierra, Selva), sexo (Masculino, Femenino), condición social, causas de accidentes laborales, ciudad donde vive, estado civil, etc.

b. **Variable Cuantitativa:** Cuando La variable se registra en forma numérica. Es aquella que se obtiene de medir y por lo tanto se expresa mediante un número acompañado del nombre de la unidad de medida.

Ejemplos: Número de pisos por edificio, Estatura, Peso, los gastos de un municipio, los sueldos de los gerentes, etc.



<http://www.anuncios.com.pe/polos-publicitarios/>

1.1.10. EL MÉTODO CIENTÍFICO

La palabra método proviene del griego «Méthodos» que significa hacer algo siguiendo un camino para alcanzar un fin determinado o una meta. Es un conjunto finito y ordenado de normas regulativas que adecuadamente observadas, conducen al logro de un fin o meta o al menos la facilitan.

El método científico está constituido por un conjunto de reglas metódicas que regulan el proceso de cualquier investigación que merezca ser calificada de científica. Es decir es una sucesión de pasos que debemos dar para resolver un problema y descubrir así nuevos conocimientos.



<http://1bachcarla.blogspot.pe/>



<http://es.slideshare.net/edisoncoimbra/>

EJEMPLO DE APLICACIÓN DEL MÉTODO CIENTÍFICO

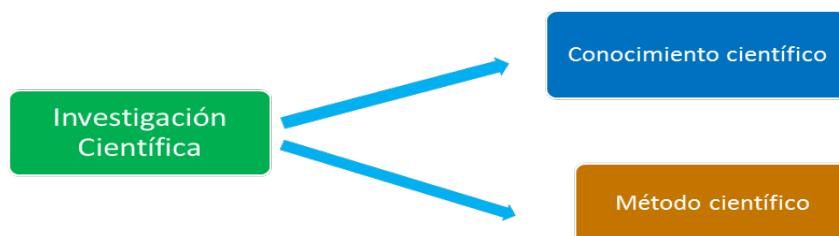
- **Problema:** mi computadora no funciona correctamente.
- **Observación:** mi computadora se apaga sola y abre páginas web que yo no identifico.
- **Hipótesis:** la computadora ha sido infectada con un virus.
- **Experimentación:** compro un antivirus original, después lo instalo en mi computadora, luego paso el antivirus pero antes de terminar el escaneo la computadora se apaga sola.
- **Conclusión:** la computadora tenía un virus tan nuevo que el antivirus no logró eliminarlo.
- **Resultado:** la computadora ahora está bien, le coloque un antivirus nuevo.



<https://computerdatasystem.wordpress.com/>

1.1.11. LA INVESTIGACIÓN CIENTÍFICA

Es un proceso que consiste en la búsqueda de nuevos conocimientos aplicando el método científico con el propósito de encontrar la verdad o falsedad de conjeturas o hipótesis. La investigación cumple rigurosamente diversos pasos o etapas en la búsqueda de esa verdad.



<http://slideplayer.es/slide/123281/>

La investigación científica es la materialización del método científico. En este proceso de investigación científica, se tiene en cuenta la siguiente secuencia:



1.1.12. ETAPAS DEL MÉTODO ESTADÍSTICO

Las Etapas del método estadístico al realizar una investigación es similar al método científico, consta de los siguientes pasos:

A. PLANIFICACIÓN DEL ESTUDIO.

Esta etapa implica:

Aquí se determina lo que se va a investigar, abarca:

- Planteamiento del problema en estudio**, consiste en definir la naturaleza e importancia del problema que se estudia ya sea en una empresa, colegio, universidad, etc. indicando la variable que se va a medir.
- Determinación de los objetivos** es decir que es lo que desea saber u obtener con la investigación. También se plantean en algunos casos hipótesis o conjeturas.
- Definir la población y muestra con la cual se va a trabajar**, y con qué recursos se cuentan para la ejecución del estudio.



<http://jehovaestaccontigo.blogspot.pe/>

- d) **Revisión de antecedentes**, es decir se debe explorar toda publicación relacionada con nuestro estudio. El investigador debe tener en cuenta los antecedentes porque el estudio que hará será la continuación de otras investigaciones realizadas anteriormente.

Ejemplo de Planificación del estudio

- Título:** «Nivel de satisfacción de los clientes por el servicio que ofrece la empresa de cable Telecom en la Ciudad de Arequipa, 2014»
- Problema de la investigación:** ¿Cuál es el nivel de satisfacción de los clientes por el servicio que ofrece la empresa de cable Telecom?
- Objetivo de investigación:** determinar nivel de satisfacción de los clientes por el servicio que ofrece la empresa de cable Telecom
- Hipótesis:** los clientes están muy insatisfechos por el servicio que ofrece la empresa de cable Telecom.
- Población:** total de clientes afiliados a la empresa de cable Telecom en la ciudad de Arequipa, 2014.
- Muestra:** n=400 clientes del empresa de cable Telecom elegidos al azar.
- Antecedentes:** Paredes Uceda J. En el año 2012 publicó el trabajo de investigación «Nivel de Satisfacción de los clientes externos en las empresas de telefonía móvil en la ciudad de Chiclayo»



<https://commons.wikimedia.org/wiki/>

B. RECOLECCIÓN DE LA INFORMACIÓN.

Es la búsqueda de la información, que se realiza mediante instrumentos de medición como encuestas (que utiliza cuestionarios), entrevistas, censos, registros, fichas. Las cuales se elaboran de manera muy minuciosa. Elaboración del instrumento de medición y su aplicación, análisis de validez y confiabilidad del instrumento de medición.

		Marqué con un X su nivel de Satisfacción, al frente de cada pregunta					
			No aplica	Muy Insatisfecho	Insatisfecho	Satisfecho	Muy Satisfecho
ELEMENTOS TANGIBLES	FIABILIDAD	P.1 Los equipos y la tecnología son de apariencia moderna ?	<input type="checkbox"/>				
		P.2 Los empleados tienen apariencia pulcra?	<input type="checkbox"/>				
CAPACIDAD DE RESPUESTA	SEGURIDAD	P.3 Los materiales ó elementos observados son visualmente atractivos?	<input type="checkbox"/>				
		P.4 Los vehículos en que transportar los productos son bien presentados?	<input type="checkbox"/>				
EMpatia		P.5 La empresa realiza los procesos en el tiempo prometido?	<input type="checkbox"/>				
		P.6 Al presentarse un problema hay un sincero interés en solucionarlo?	<input type="checkbox"/>				
EMpatia		P.7 Los empleados realizan bien el servicio a la primera vez?	<input type="checkbox"/>				
		P.8 Se atienden a tiempo las quejas y reclamos ?	<input type="checkbox"/>				
EMpatia		P.9 Se registran los tipos de errores cometidos?	<input type="checkbox"/>				
		P.10 Los empleados comunican a los clientes los horarios exactos del servicio?	<input type="checkbox"/>				
EMpatia		P.11 Los empleados ofrecen un servicio rápido a los clientes?	<input type="checkbox"/>				
		P.12 Los empleados están dispuesto a ayudar a los clientes?	<input type="checkbox"/>				
EMpatia		P.13 Los empleados responden oportunamente las preguntas a los clientes?	<input type="checkbox"/>				
		P.14 El comportamiento de los empleados transmite confianza a los clientes?	<input type="checkbox"/>				
EMpatia		P.15 Como cliente se siente seguro de las transacciones que hace con la compañía?	<input type="checkbox"/>				
		P.16 Los empleados son siempre amables con los clientes?	<input type="checkbox"/>				
EMpatia		P.17 Los empleados tienen conocimiento y responden las preguntas de los clientes?	<input type="checkbox"/>				
		P.18 La empresa da a los clientes servicios adicionales a los clientes?	<input type="checkbox"/>				
EMpatia		P.19 Se tienen horarios de trabajo convenientes para todos los clientes?	<input type="checkbox"/>				
		P.20 Los empleados ofrecen una atención personalizada a los clientes?	<input type="checkbox"/>				
EMpatia		P.21 La compañía se preocupa por los intereses de su clientes?	<input type="checkbox"/>				
		P.22 Los empleados comprenden las necesidades específicas de los clientes?	<input type="checkbox"/>				



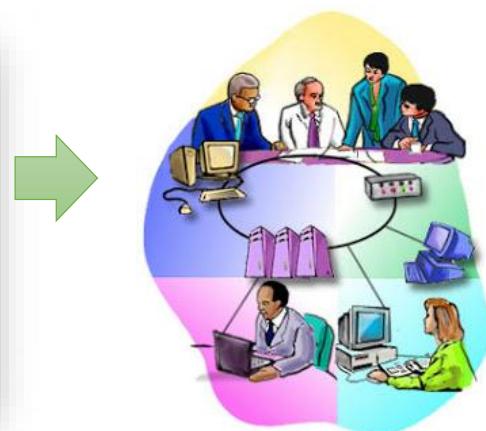
<https://gemmav58.files.wordpress.com/>

C. PROCESAMIENTO DE DATOS

- En esta etapa del método estadístico la información recogida es sometida a revisión, clasificación y cómputo numérico.
 - A veces el recuento puede realizarse de manera muy simple, por ejemplo con rayas o palotes y, en investigaciones con mucha información y muchos casos, puede requerirse el empleo de computadoras y programas especiales para el manejo de bases de datos.
- Por ejemplo:** el número de clientes de la empresa Telecom ordenados según su nivel de satisfacción: Muy satisfecho, Satisfecho, insatisfecho, Muy insatisfecho.



<http://bit.ly/1Qo1YY3>



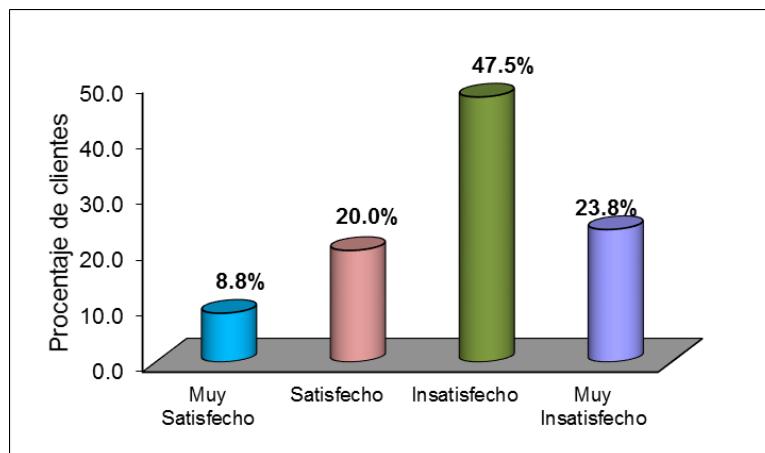
<http://bit.ly/20HVG4S>

D. PRESENTACIÓN Y CLASIFICACIÓN

- En esta etapa del método estadístico se elaboran los cuadros y los gráficos que permiten una inspección precisa y rápida de los datos.
- La elaboración de cuadros o tablas, tiene por propósito acomodar los datos para efectuar una revisión numérica precisa de los mismos.
- La elaboración de gráficos tiene por propósito facilitar la inspección visual rápida de la información.
- Casi siempre a cada cuadro con datos le puede corresponder una gráfica pertinente que represente la misma información.

Tabla N°1			
Distribución de clientes de la empresa Telecom según nivel de satisfacción con el servicio			
NIVELES	FRECUENCIAS	PORCENTAJE (%)	PORCENTAJES ACUMULADOS (%)
Muy Satisfecho	35	8.8	8.8
Satisfecho	80	20.0	28.8
Insatisfecho	190	47.5	76.3
Muy Insatisfecho	95	23.8	100.0
TOTAL	400	100.0	

Fuente: elaboración propia



Fuente: elaboración propia

Figura N°1: Distribución porcentual de clientes de la empresa Telecom según nivel de satisfacción con el servicio

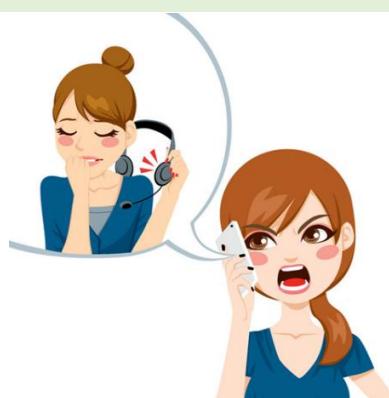
E. ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS.

En esta etapa la información es resumida en forma de medidas que tiene por propósito facilitar la comprensión global de las características fundamentales de los datos.

Entre las principales medidas de resumen para sintetizar a los datos cualitativos se encuentran las proporciones y las tasas. Y para los datos cuantitativos se encuentra el promedio.

Ejemplo: en la empresa Telecom.

- De la tabla anterior se tiene que del total de clientes encuestados, el 8.8% están muy satisfechos con el servicio de cable, 20% satisfechos, 47.5% insatisfechos y 23.8% Muy insatisfechos.
- La mayoría de clientes (47.5%) se sienten insatisfechos con el servicio de cable, también existe un alto porcentaje de muy insatisfechos.



<http://multihobbes.blogspot.pe/>

F. PUBLICACIÓN DE LOS RESULTADOS

- Es la divulgación de la investigación científica, mediante artículos, ponencias y conferencias, entre otros mecanismos. Es esencial, pues investigación que no se publica no existe.
- La investigación culmina al ser publicada en una revista científica; solo así será conocida por la comunidad académica y su contribución hará parte del conocimiento científico universal.



<http://guiasbus.us.es/ingenieria/articulosrevistas>

*Un artículo bien concebido debe transmitir la mayor parte de la información con sólo leer el **Abstract** y los resultados, siendo para ello vital que los datos, con las tablas y figuras correspondientes, estén bien presentados y organizados.*

GUÍA DE PRÁCTICA N°1

FUNDAMENTOS DE INVESTIGACIÓN EN ESTADÍSTICA

1. Del número de accidentes de tránsito por mes registrados por la policía nacional según causa, durante el periodo 1998 – 2003 se obtuvo una información de 50 accidentes, las causas más frecuentes fueron:

- | | |
|------------------------------|---------------------------|
| A. Exceso de velocidad | D. Imprudencia del peatón |
| B. Ebriedad del conductor | E. Desacato de señales |
| C. Imprudencia del conductor | F. Falla mecánica |



De lo anterior se tiene que 15 accidentes fueron por imprudencia del peatón siendo la causa más frecuente. Identifique:

- a) Población: _____
- b) Muestra: _____
- c) Unidad estadística: _____
- d) Variable analizada y su tipo: _____
- e) Estadígrafo y proporcione su valor _____
- f) Parámetro de interés e indique su valor _____

2. Se realiza un muestreo de opinión para determinar si los clientes de un Supermercado, prefieren un Celular con servicio de la compañía A con respecto a otras compañías. Con este fin se entrevistan a 1600 clientes y entre ellos 1200 prefieren la compañía A. Identifique:



- a) Población: _____
- b) Muestra: _____
- c) Unidad estadística: _____
- d) Variable analizada y su tipo: _____
- e) Estadígrafo y proporcione su valor _____
- f) Parámetro de interés e indique su valor _____

3. La secretaría de Salud Pública de un municipio decide realizar un estudio acerca de la posible influencia del aumento de la contaminación ambiental en el crecimiento y en la salud de la población infantil. Aunque el hospital municipal lleva un minucioso registro de los pacientes que atiende, la Secretaría considera que la población infantil que concurre no conforma una muestra representativa. Decide entonces, estudiar una muestra compuesta por 500 niños de 1 a 6 años de edad pertenecientes a familias de distintos niveles socioeconómicos, que habitan los diferentes barrios del municipio. Como parte de este estudio, se recolectan datos

referidos al número de varones y de mujeres, la distribución del peso y la altura por edades, y los trastornos de salud más frecuentes.

a) ¿Cuál es la población? _____

b) ¿Cuál es la muestra y que tipo de muestreo se utilizó? _____

c) ¿Qué tipo de estadística se aplicaría en este estudio? _____

d) Identifique todas las variables de estudio. _____

4. En los siguientes casos ¿Cuál probablemente exija solo el uso de la Estadística Descriptiva y cuál de la Estadística Inferencial?

a) Un gerente de personal desea conocer la aptitudes de cinco secretarias de una determinada oficina de la empresa, se aplica una prueba y las calificaciones son 85, 90, 93, 82 y 95 con promedio _____

b) Un médico investigador estudia la relación entre el consumo de cigarrillos y las enfermedades del corazón.

c) Una empresa de pernos desea conocer el porcentaje de unidades defectuosos de la producción para lo cual contabiliza el número de unidades defectuosas por lote tomando al azar 10 unidades por lote.

d) El año pasado, en la Universidad Autónoma el puntaje promedio en el curso de estadística fue 15.

e) El Dr. García, un ecólogo, informó que en cierto río de la selva la carne de los peces contienen un promedio de 300 unidades de mercurio.

f) Un Psicólogo estudia los efectos de las nuevas técnicas de automatización sobre el rendimiento de la producción.

5. En cada uno de los siguientes problemas identificar la variable y el tipo de variable (X):

	Problema	Variable	Variable Cualitativa	Variable Cuantitativa
1	El departamento de bomberos de una gran ciudad clasifica los incendios como grado 1,2,3,4,... etc			
2	La cantidad de dinero concedida por un tribunal en una demanda por alimentos.			
3	Un grupo de sociólogos clasificó a los internos de un penal de acuerdo a su peligrosidad			
4	Se registró el número de juicios que ha ganado cada mes un abogado, desde que empezó a ejercer su profesión.			
5	Durante el año 2002 en la ciudad de Lima se llevaron a cabo intervenciones policiales en los delitos de robo, homicidio, lesión, corrupción, aborto.			
6	La constitución física de un acusado es clasificada como: 1 Si es delgado, 2 si es regular y 3 si es obeso			
7	El tiempo de reacción de un conductor de automóvil cuando se enfrenta a un peligro inminente.			

6. De los siguientes enunciados identifique población, muestra, variable, estadígrafo o parámetro

- a) En una universidad se quiere saber cuál es el deporte más practicado por los alumnos para lo cual se entrevistan a 80 estudiantes cuyos resultados son: 40 prefieren futbol, 20 básquet, 8 natación y 12 Vóley. Identifique:

Población: _____

Muestra: _____

Variable: _____

Tipo de variable: _____

Estadígrafo: _____

- b) La enfermera de un centro de salud está interesada en realizar un estudio sobre el estado de la nutrición en niños de 5 años de edad de la comunidad San Román. La población está constituida por 900 niños de 5 años de edad. La enfermera está interesada, en particular, en conocer la proporción de niños que están desnutridos y la estatura promedio. Para tal efecto tomo el 10% de niños como muestra. Identifique:

Población: _____

Muestra: _____

Variable: _____

Tipo de variable: _____

Parámetro: _____

- c) El gerente general de una empresa de 460 empleados está interesado en determinar la proporción de empleados que tienen más de dos hijos, para lo cual analiza los datos personales de 90 trabajadores escogidos al azar en una base de datos de la empresa. Identifique:

Población: _____

Muestra: _____

Variable: _____

Tipo de variable: _____

Estadígrafo: _____

- d) Para realizar un pronóstico de turismo referido a la estimación de la demanda de turistas en Chiclayo, en el ítem “hospedaje”, se registró un tránsito de turistas en 179 hospederías (alojamientos, residencias, hoteles) de Chiclayo. Identifique:

Población: _____

Muestra: _____

Variable: _____

Tipo de variable: _____

Estadígrafo: _____

7. Aplica el método científico para resolver los siguientes problemas:

A. Problema: **Hoy me levante tarde para ir a la universidad**

Observación: _____

Hipótesis: _____

Experimentación: _____

Conclusión: _____

Resultado: _____

B. Problema: **el proyector multimedia no funciona correctamente para la exposición**

Observación: _____

Hipótesis: _____

Experimentación: _____

Conclusión: _____

Resultado: _____

8. Indica y justifica cuál de los métodos de muestreo explicados se aplicó en cada uno de estos casos:

Casos planteados	Tipo o tipos de muestreo aplicados
A. Se dispone de un directorio o lista de los 2000 bares y restaurantes de una gran ciudad, se elige uno al azar y a partir de este primer seleccionado y contando de 25 en 25 se ha ido seleccionando una muestra de 80 bares y restaurantes.	
B. Para investigar el impacto de la crisis en las empresas valencianas, tenemos una lista numerada con los nombres de las 169.000 empresas de la provincia de Valencia. El ordenador elige de forma aleatoria una muestra de 100 de esas empresas.	
C. Para seleccionar una muestra 100 de hogares que residen en municipios de menos de 1000 habitantes en la provincia de Valencia, se eligen al azar 10 municipios con menos de 1000 habitantes de la provincia de Valencia y en cada municipio seleccionado se selecciona una muestra aleatoria de 10 hogares.	
D. En una encuesta durante las elecciones, se elige al azar 2 mesas electorales y se analizan todos los votos emitidos de las mesas seleccionadas.	

9. Un investigador tiene un archivo de 80 casos de una enfermedad rara y está interesado en seleccionar muestra sistemática de 8 casos y dentro del proceso considera un arranque aleatorio de 6. Por consiguiente, la muestra queda constituida como:

- a. 6 12 18 24 30 36 42 48
- b. 6 17 28 39 50 61 72 80
- c. 6 15 21 30 39 58 67 78
- d. 6 16 26 36 46 56 66 76
- e. Ninguno de los anteriores

10. En una población estudiada, hay 2000 mujeres y 8000 hombres. Si queremos seleccionar una muestra de 250 individuos en dicha población. ¿Cuántos deberán ser mujeres para que la muestra sea considerada representativa?

	N (población)	n (muestra)
Mujeres		
Hombres		
Total		

11. En cierto barrio se quiere hacer un estudio para conocer mejor el tipo de actividades de ocio que gustan más a sus habitantes. Para ello van a ser encuestados 100 individuos elegidos al azar.

a) Identifica cual es la variable en estudio y su tipo

b) Como los gustos cambian con la edad y se sabe que en el barrio viven 2.500 niños, 7.000 adultos y 500 ancianos, posteriormente se decide elegir la muestra anterior utilizando un muestreo estratificado. Determinar el tamaño muestral correspondiente a cada estrato.

Estratos	N (población)	n (muestra)
Total		

12. El departamento de control de calidad de una empresa productora de latas de conserva, utiliza de forma periódica muestreo sistemático para estimar el peso medio de las latas en gramos. Un día concreto se produjeron 40 latas en una hora elegida al azar, cuyos pesos son:

12.1	11.97	12.01	12.03	12.01	11.8	11.91	11.98	12.03	11.98
12.3	11.83	11.87	12.01	11.98	11.97	12.05	11.03	11.12	12.6
11.9	11.94	11.65	11.19	12.02	11.45	12.01	11.18	11.9	12.6
12.3	11.9	11.65	11.84	12.6	12.35	11.88	12.05	11.6	12.09

Extraer una muestra sistemática de tamaño 10 y estimar el peso medio de las latas producidas

13. El Ministerio de Justicia, deseando mejorar el nivel de sus funcionarios en cargos de jefatura, dio un curso experimental para un grupo de 25 funcionarios. Luego se tomó una evaluación calificada en una escala del 1 al 5 (5 = Excelente 4 = Bueno 3 = Regular 2 = Malo 1 = Pésimo) y estos fueron los resultados:

1	3	3	4	1
2	2	2	5	1
4	5	1	5	3
5	1	4	1	2
2	1	2	3	5

- a) Identifique la variable en estudio _____
- b) Plantee el problema de investigación _____
- c) Defina el objetivo del estudio _____
- d) Formule una hipótesis _____
- e) ¿Qué Instrumento se utilizó para la recolección? _____
- f) Organiza y clasifica la información en el siguiente cuadro

Desempeño	Nº	%
Excelente		
Bueno		
Regular		
Malo		
Pésimo		
Total		



- g) Analiza e interpreta los resultados:

http://narcopolicias.blogspot.pe/2011_09_01_archive.html

14. Suponga que estamos investigando sobre el porcentaje de alumnos que trabajan de una población de alumnos de la Universidad de Talca. La Base de datos de la población es:

Alumno	Sexo	¿Trabaja?
Juan		SÍ
Alicia		NO
Pedro		NO
Marcos		NO
Alberto		SÍ
Jorge		SÍ
José		NO
Carlos		NO
Miguel		NO
Victoria		NO
Ricardo		SÍ
Silvia		NO

Alumno	Sexo	¿Trabaja?
María		NO
Fernanda		NO
Julio		SÍ
Rosa		NO
Fabián		NO
Ana		NO
Laura		NO
Enrique		NO
Carmen		SÍ
Marcelo		SÍ
Elena		NO
Xavier		NO

Alumno	Sexo	¿Trabaja?
Lizet		SÍ
Leo		SÍ
Joanna		SÍ
Amanda		NO
Roque		SÍ
Sergio		SÍ
Vanesa		NO
Elvira		SÍ
Ronaldo		NO
Mariana		NO
Daniela		SÍ
Roberta		NO

- a. Elija una muestra aleatoria simple de tamaño n=18 de esta población y calcule el porcentaje de alumnos que trabajan y además el % de sexo masculino. (Compare con el parámetro)
- b. Usando el muestreo sistemático obtener una muestra de tamaño 18 y calcular el porcentaje de alumnos de sexo masculino y además calcular el porcentaje de alumnos que trabajan.
- c. Elija una muestra estratificada de tamaño n=18 de esta población. Use el muestreo aleatorio simple para elegir la muestra dentro de cada estrato, Indique los pasos para elegir la muestra.
15. En una empresa textil se entrevistaron a 45 empleados sobre el clima organizacional de su entorno de trabajo. En la siguiente tabla fueron clasificadas sus opiniones (B=Bueno, R= Regular, M=Malo) según su sexo.

Nº	Sexo	Clima	Nº	Sexo	Clima	Nº	Sexo	Clima
1	M	M	16	F	M	31	M	M
2	F	B	17	M	B	32	F	B
3	M	B	18	F	R	33	M	B
4	F	M	19	F	M	34	F	M
5	M	R	20	F	B	35	F	B
6	F	M	21	M	R	36	F	R
7	F	R	22	F	R	37	M	B
8	F	B	23	M	B	38	F	R
9	M	M	24	M	M	39	F	B
10	F	B	25	F	R	40	F	B
11	F	R	26	F	B	41	M	R
12	M	M	27	F	B	42	M	B
13	F	B	28	F	M	43	M	B
14	M	B	29	M	B	44	F	M
15	M	B	30	F	M	45	F	M

- a) Usando el muestreo sistemático obtener una muestra de tamaño 15 y calcular el porcentaje de personas del sexo femenino y además calcular el porcentaje de empleados que respondieron “regular” (comparar con el parámetro)
- b) Usando la tabla de números aleatorios obtener una muestra de tamaño similar a la pregunta anterior y calcular los mismos porcentajes. (comparar con el parámetro)
- c) Estratificar la población según el sexo y obtener una muestra de tamaño 15 utilizando números aleatorios y calcular los mismos porcentajes. (comparar con el parámetro)

1.2. MARCO METODOLÓGICO DE INVESTIGACIÓN

1.2.1. MARCO METODOLÓGICO

Es el conjunto de acciones destinadas a describir y analizar el fondo del problema planteado, a través de procedimientos específicos que incluye las técnicas de observación y recolección de datos, determinando el “cómo” se realizará el estudio.

Es de gran importancia en la investigación, pues el planteamiento de una metodología adecuada nos permitirá lograr de manera precisa el objetivo planteado y obtener resultados con el máximo grado de exactitud.

Abarca los siguientes aspectos:

- a) Operacionalización de variables
- b) Tipo de estudio
- c) Diseño de estudio
- d) Población, muestra y muestreo
- e) Técnicas e instrumentos de recolección de datos



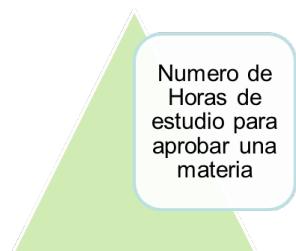
<http://es.123rf.com/imagenes-de-archivo/estadisticos.html>

1.2.2. VARIABLE ESTADÍSTICA

Una variable es una característica de la población que interesa al investigador. Son observables y susceptibles de tomar distintos valores o ser expresados en diferentes categorías. A las variables se les denota con las letra: x, y, z

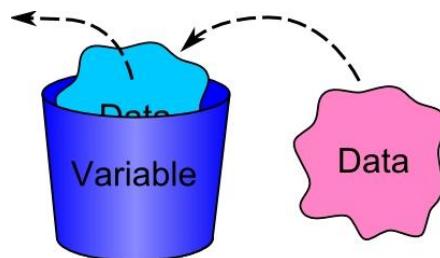
Ejemplos:

- a) La profesión de un grupo de docentes: contador, abogado, administrador, educador.
- b) Las temperaturas de las ciudades de Lima al medio día: 12°C, 15°C, 18°C, 20°C
- c) El medio de transporte utilizado para viajar a una ciudad: avión, bus, automóvil, barco.
- d) Los pesos de un grupo de mujeres que asisten a un gimnasio: 58 kg, 62 kg, 49 kg,



<https://debocaenbocacentre.com/blog/>

Los **datos** son los valores que adoptan las variables en cada caso particular, las **variables** no son más que aquello que estudiamos en cada individuo de la muestra.



<http://resources.esri.com/help/9.3/>

CLASIFICACION DE LAS VARIABLES



A. SEGÚN SU NATURALEZA:

A1. Variables cualitativas. Llamadas también no numéricas, son variables cuyos valores consisten en categorías de clasificación y responde a una calidad o atributo. No es posible realizar operaciones algebraicas con ellas.

Ejemplo: sexo, afiliación religiosa, tipo de automóvil que se posee, lugar de nacimiento, tipo de publicidad, calidad de un producto. etc.

Las variables cualitativas a su vez pueden clasificarse en:

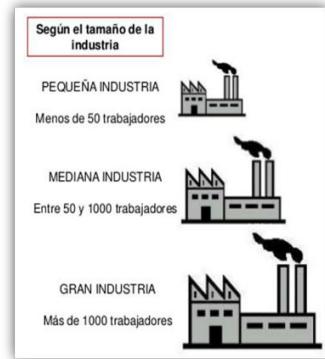
- ❖ **Variables cualitativas nominales.** Son aquellas que comprenden solo categorías de clasificación y no llevan ninguna ordenación. **Ejemplo:** estado civil, marca de productos, color preferido, procedencia, las profesiones, etc.



¿Una Marca de Autos?

http://www.vectorizados.com/vector/9086_coches-clasicos

- ❖ **Variables cualitativas ordinales.** Cuando las características no solo comprenden categorías de clasificación sino que llevan alguna ordenación. **Ejemplo:** grado de instrucción, calidad de un material determinado (excelente, bueno, regular o malo), grado académico, rango militar, jerarquía gerencial, clase social, el nivel de desempeño, etc.



<http://es.slideshare.net/pflores88/6-el-espacio-industrial>

A2. Variables Cuantitativas.

Son características cuyos valores pueden ser obtenidos por medición o por conteo y a su vez pueden ser clasificadas en:

- ❖ **Variables cuantitativas discretas.** Cuando las variables toman valores enteros y son obtenidos por conteo.
Ejemplo: número de hijos por familia, cantidad de libros que vende una editorial, el número de veces que un alumno lleva un curso, número de viajes, etc.



<http://aeaps.edu.pe/cms/upload/gallery/images/35//working>

- ❖ **Variables cuantitativas continuas.** Son aquellas que pueden tomar cualquier valor (entero, fraccionario o irracional) dentro un intervalo. Sus valores se obtienen principalmente a través de mediciones y están sujetos a la precisión de los instrumentos de medición.
Ejemplo: el peso, la estatura, la temperatura, el salario de los gerentes, el tiempo que un corredor tarda en cubrir una cierta distancia, la cantidad de hemoglobina en la sangre, colesterol, etc.



<http://www.ehowenespanol.com/relacion-altura-tamano-pies>

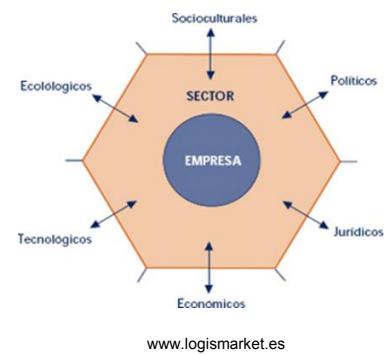
B. SEGÚN SU POSICIÓN EN UNA RELACIÓN CAUSAL

B1. Variable Independiente. Los valores de este tipo de variables no dependen de otras y cumple el papel de causa de algún efecto. Son las características controladas por el investigador y que afecta o influye a la variable dependiente.

B2. Variable Dependiente. También se las llama variables respuesta y cumple el papel de efecto causa. Son aquellas cuyo comportamiento es explicado por una o más variables independientes. Recibe la influencia o efecto de la variable independiente.

Ejemplos:

- La experiencia laboral alcanzada por las personas influye en su salario percibido mensualmente.
- El clima organizacional influye en el rendimiento y desarrollo del talento humano en una entidad.
- La calidad y los gastos invertidos en publicidad incrementan las ventas de un producto determinado.



Más ejemplos

- El grado de instrucción alcanzado por las personas influye en su salario percibido mensualmente.

Variable independiente: grado de instrucción alcanzado

Variable dependiente: salario percibido

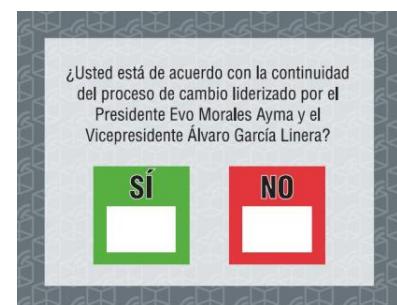
- El aumento en el precio de un producto disminuye el número de unidades vendidas de dicho producto. Identifica:

Variable independiente: _____

Variable dependiente: _____

C. SEGÚN LA CANTIDAD DE VALORES QUE CONTIENEN

C1. Variables Dicotómicas. Es aquella que solo puede tomar dos valores. Por ejemplo, sexo: masculino y femenino, tener o no una enfermedad (positivo y negativo), las notas de un curso pueden reducirse a dos grandes valores aprobados y desaprobados, la asistencia a una reunión (presente o ausente), etc.



<http://coocobo.bloasoot.pe/2008/07/referndum-revocatorio.html>

C2. Variables Polítómicas. Son aquellas que se pueden expresar con más de dos valores. El ejemplo propuesto de las notas de un curso tiene más de dos valores; igualmente, se suele considerar varios valores a la condición socioeconómica, a los niveles de escolaridad, la edad, etc.

¿En cuál de las siguientes situaciones se encontraba la mayor parte de la SEMANA PASADA?	
Trabajando	1 <input type="checkbox"/>
Sin trabajar, pero tiene empleo	2 <input type="checkbox"/>
Buscando trabajo	3 <input type="checkbox"/>
Quehaceres de su hogar	4 <input type="checkbox"/>
Estudiando	5 <input type="checkbox"/>

<http://www.fundacionsol.cl/2012/04/>

1.2.3. OPERACIONALIZACIÓN DE VARIABLES

- Es el procedimiento por el cual el investigador especifica las variables contenidas en el estudio. Desagrega las variables en dimensiones e indicadores a fin de que las variables puedan medirse empíricamente.
- Con la operacionalización de las variables el investigador pasa de lo abstracto de la hipótesis a lo concreto de los hechos

Tabla de operacionalización de variables

Variable	Dimensiones	Indicadores	Tipo de variable	Escala de Medición	Instrumento

Ejemplo:

Tabla N°1 Operacionalización de la variable Gestión empresarial

Variable	Dimensiones	Indicadores	Tipo de variable	Escala de Medición	Instrumento
Gestión empresarial	Organización	Grado de organización, clima organizacional, capacitación, promoción de la participación y relación con la comunidad.	Cualitativa ordinal	Escala Ordinal	Cuestionario Likert
	Dirección	Grado de Liderazgo, motivación, trabajo en equipo, delegación de funciones	Cualitativa ordinal	Escala Ordinal	Cuestionario Likert

1.2.4. TIPO DE ESTUDIO

Se entiende por tipo o clase de estudio a la precisión de la ubicación del estudio en una clasificación de las investigaciones y estos son:

A. Investigación básica: es conocida como pura o fundamental. Está destinada a aportar un cuerpo organizado de conocimientos científicos y no produce necesariamente resultados de utilidad práctica inmediata. Son de nivel exploratorio y descriptivo.



Ejemplo: la monografía sobre un tema en particular y las tesinas de pregrado en las que solo se recolecta información ya producida por otros.

B. Investigación aplicada: se le denomina también activa o dinámica. Busca conocer para hacer, para actuar, para construir, para modificar; le preocupa la aplicación inmediata sobre una realidad concreta. Son de nivel experimental y algunos descriptivos.



Ejemplo: estudio socioeconómico de los docentes de universidades nacionales para efectuar cambios en el presupuesto. Estudio comparativo para

1.2.5. DISEÑO DE ESTUDIO

El término diseño se refiere a plan o estrategia concebida para responder a las preguntas de investigación. El diseño señala al investigador lo que debe hacer para alcanzar sus objetivos de estudio y analizar la certeza de las hipótesis formuladas.

Clasificación: dividiremos en dos grupos:

A. Diseños descriptivos, los más usados son:

- Diseño descriptivo observacional
- Diseño descriptivo correlacional
- Diseño descriptivo transversal

B. Diseños Experimentales, los más usados son:

- Diseño experimental clásico
- Diseños pre experimentales
- Diseños cuasi experimentales

1.2.6. TÉCNICAS DE RECOLECCIÓN DE DATOS

- Un instrumento de recolección de datos es cualquier recurso de que pueda valerse el investigador para acercarse a los fenómenos y extraer de ellos información.

- Las técnicas están referidas a la manera como se van a obtener los datos y los instrumentos son los medios materiales, a través de los cuales se hace posible la obtención y archivo de la información requerida para la investigación. Algunos de ellos:

Técnica	Instrumento
Observación	Guía de observación Lista de chequeo Registro anecdótico Matriz de análisis
Entrevista	Guía de entrevista
Encuesta	Cuestionario Escala Test Prueba de conocimiento

1.2.7. ESCALAS DE MEDICIÓN

Medir en el campo de las ciencias exactas es comparar una magnitud con otra, tomada de manera arbitraria como referencia, denominada patrón y expresar cuántas veces la contiene. Al resultado de medir lo se le llama medida.

Las escalas de medición son el conjunto de los posibles valores que determinada variable puede tomar. Por tal razón, los tipos de escala de medición están íntimamente ligadas con los tipos de variables a estudiar. Las escalas de medición se clasifican de la siguiente forma:



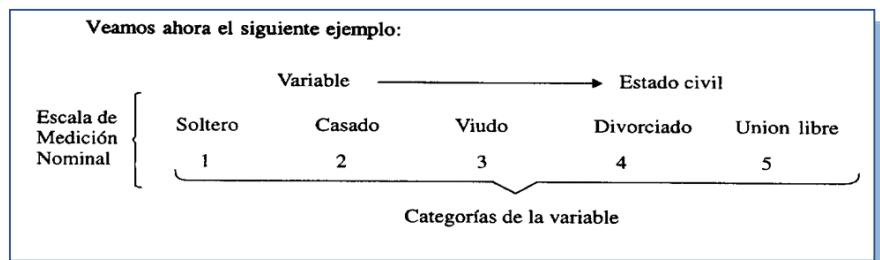
<https://cipe.uclm.es/noticias/nuevas-ofertas-para-ingeneros-en->

Al elaborar estadísticas con datos y su característica es necesario contarlas, jerarquizarlas y medirlas, es por ello que, se utilizan las escalas de medición como el proceso de asignar números o establecer una correspondencia uno a uno entre objetivos y observaciones. Las escalas de medición se clasifican de la siguiente forma:

Escala Nominal
Escala Ordinal
Escala de Intervalos
Escalas de razón

A. ESCALA NOMINAL

Consiste en clasificar a los elementos, personas, animales, etc, asignándoles símbolos o nombres. Los datos que se obtienen para una variable cualitativa se miden en una escala nominal y simplemente se clasifican en distintas categorías que no implican orden.



<http://es.slideshare.net/leamotova/variables-medicion>

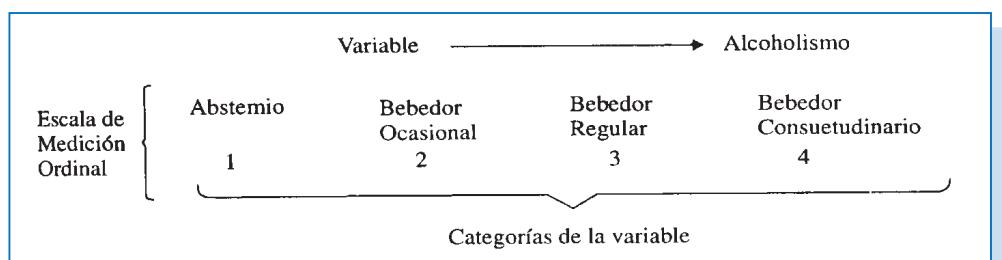
El estado civil tiene cinco categorías mutuamente excluyentes, cuyo orden de colocación es indistinto, ya que pudimos haber puesto primero viudo o casado y terminar en soltero. Además, si a "soltero" le llamamos 1, a "casado" 2, etc., estas cifras carecen de propiedades numéricas, ya que solo sirven para distinguir un estado civil de otro.

Propiedades de la escala nominal

1. No intervienen mediciones, ni escala, en vez de esto solo hay cuentas o conteos.
2. No existe un orden específico para esta categoría.
3. No se basa en diferencia cuantitativa.

B. ESCALA ORDINAL

Establece una relación de orden entre los elementos de una característica, sin que reflejen distancia entre ellos. La diferencia entre dos números ordinales no tiene significado cuantitativo, solo expresan, por ejemplo, que una situación es mejor que otra, pero no dice cuánto es uno que el otro.



<http://es.slideshare.net/leamotova/variables-medicion>

La medición de "alcoholismo" tiene categorías de dos extremos entre los cuales hay niveles. Estas categorías aunque se les llame por su nombre o por medio de cifras carecerían de propiedades numéricas: la cifra 3 indicaría un grado de dependencia menor que la 4 y mayor que la 2, pero nada más.

Propiedades de la escala ordinal

1. Las observaciones o elementos se ordenan en categorías diferentes.
2. Las categorías son mayores o menores que otras categorías, es decir, que existe una jerarquía.
3. Las categorías son mutuamente excluyentes y exhaustivas.

C. ESCALA DE INTERVALO

La medición de intervalo posee las características de la medición nominal y ordinal. Establece la distancia entre una medida y otra. La escala de intervalo se aplica a variables continuas pero carece de un punto cero absoluto.

Propiedades de la escala interválica

1. En estas medidas se utilizan unidades constantes de medición) los cuales producen intervalos iguales entre puntos de la escala.
2. En esta escala de intervalos el punto cero (0) y la unidad de medida es arbitrario.



El ejemplo más representativo de este tipo de medición es un termómetro, cuando registra cero grados centígrados de temperatura indica el nivel de congelación del agua y cuando registra 100 grados centígrados indica el nivel de ebullición, el punto cero es arbitrario no real, lo que significa que en este punto no hay ausencia de temperatura.

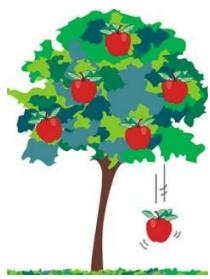
<http://es.slideshare.net/leamotova/variables-medicion>

D. ESCALA DE RAZÓN

Esta constituye el nivel más alto de medición, posee todas las características de las escalas nominales, ordinales y de intervalos; además tiene un cero absoluto o natural que tiene significado físico. Si en ella la medición es cero, significa ausencia o inexistencia total de la propiedad considerada.

Propiedades de la escala de razón

1. Los datos tienen un punto cero significativo y son posibles todas las operaciones aritméticas.
2. Permite hacer comparaciones entre los números verdaderos con un cero aritmético.



Los ingresos monetarios y gastos directos, la medición del peso o altura de un grupo de personas, el ingreso familiar, la intensidad de corriente eléctrica de un cable, la edad de un grupo de personas, **son ejemplos** de medidas con una escala de razón.

<https://consejonutricion.wordpress.com/taq/fruta/>

1.2.8. SUMATORIAS

La suma de los valores de la variable X : x_1, x_2, \dots, x_n se define mediante la notación

\sum que indica que han de sumarse los elementos de la sucesión desde el subíndice $i=1$ hasta el subíndice $i=n$, esto es:

$$\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$$

- i. La sumatoria de la suma de dos o más términos, es igual a la suma de las sumatorias separadas de los términos.

$$\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i$$

- ii. La sumatoria de la diferencia de dos o más términos es igual a la diferencia de las sumatorias separadas de los términos.

$$\sum_{i=1}^n (x_i - y_i) = \sum_{i=1}^n x_i - \sum_{i=1}^n y_i$$

- iii. La sumatoria de una constante multiplicada por una variable es igual a la constante multiplicada por la sumatoria de la variable.

$$\sum_{i=1}^n a \cdot x_i = a \cdot \sum_{i=1}^n x_i$$

- iv. La sumatoria de una constante es igual a la constante multiplicada por el número que indique los límites de la sumatoria.

$$\sum_{x=1}^n a = n.a$$

En la práctica frecuentemente se comenten algunos errores, los cuales los cuales mencionaremos para que no se incurra en ellos.

- Es falso el tomar a $\sum_{i=1}^n x^2 = \left(\sum_{i=1}^n x\right)^2$ ya que son valores completamente diferentes
- Otro error se comete es decir que $\sum_{i=1}^n x_i y_i = \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i$ ya que son términos diferentes.

Ejemplos:

$$\sum_{i=1}^5 x_i = x_1 + x_2 + x_3 + x_4 + x_5$$

$$\sum_{i=1}^n x_i = x_1 + x_2 + x_3 + x_4 + \dots + x_n$$

$$\sum_{i=1}^4 (2i+3) = (2(1)+3) + (2(2)+3) + (2(3)+3) + (2(4)+3) = 5 + 7 + 9 + 11 = 32$$

GUÍA DE PRÁCTICA N°2

MARCO METODOLOGICO EN ESTADÍSTICA

1. Clasificar adecuadamente las diversas variables con un aspa (X):

Nº	VARIABLE	CUALITATIVA		CUANTITATIVA	
		NOMINAL	ORDINAL	CONTINUA	DISCRETA
1	Situación laboral de una persona				
2	Superficie dedicada a cierto cultivo por hectáreas				
3	Opinión de los peruanos sobre el terrorismo				
4	Cantidad de Triglicéridos en la sangre				
5	Número de habitantes por kilómetro cuadrado				
6	Tipo de bebedor (Abstemio, leve, crónico)				
7	Volumen de agua de un reservorio				
8	Nivel de colesterol				
9	Tipo de municipio (rural, urbano, capital)				
10	Frecuencia de asistencia a actividades deportivas				
11	Período de duración de un automóvil.				
12	Número de acciones vendidas en la Bolsa.				

2. En los siguientes problemas identifica cual es la variable Independiente y cual la variable dependiente.

- a) Supongamos que estamos haciendo un estudio para comprobar la relación entre el clima laboral de la empresa y el rendimiento de los trabajadores.
V.Indep: _____ V.Dep: _____
- b) Al comprar un producto ¿La marca del producto está relacionado con su calidad?
V.Indep: _____ V.Dep: _____
- c) Las estafas a través de la web se debe a la falta de seguridad y medidas de control
V.Indep: _____ V.Dep: _____
- d) Un investigador quiere conocer si existe relación entre el sexo y la severidad del daño renal en los pacientes diabéticos.
V.Indep: _____ V.Dep: _____
- e) En una empresa se quiere determinar si la rápida atención de un trabajador influye en la satisfacción del cliente.
V.Indep: _____ V.Dep: _____
- f) ¿Puede la buena alimentación aumentar la capacidad mental?
V.Indep: _____ V.Dep: _____

3. Identifique cada una de las siguientes variables escribiendo su tipo:

Variable	Tipo de Variable
a) El uso más frecuente de su horno de microondas (recalentar, descongelar, calentar, otro)	
b) El número de consumidores que se rehusan a contestar una encuesta telefónica	
c) La puerta elegida por un ratón en un experimento con laberinto (A, B ó C)	
d) El tiempo ganador de un caballo que corre en el Derby de Monterrico	
e) El número de niños en una clase de quinto grado cuyo nivel de lectura está al nivel escolar superior.	

4. Determíñese qué tipo de escala de medida es la más adecuada para cada una de las siguientes variables:

Variable	Escala
a) Nuestro sistema de numeración cronológico de los años, por ejemplo: 1492, 1650, 1949, 1985, 1991	
b) La edad de los sujetos (entendiendo por edad el tiempo de vida extramaterna)	
c) La escala de dureza de los minerales	
d) Los diferentes números de las camisetas de los jugadores de equipos de fútbol	
e) La lista de éxitos discográficos del verano	
f) El tiempo empleado por los pilotos de automóviles en recorrer diez veces un circuito	
g) Las marcas de paquetes de cigarrillos	
h) La temperatura medida en grados kelvin	

5. Escribe el tipo de variable que corresponda según los siguientes enunciados:

- a) No se expresa mediante un número. _____
- b) Se expresa mediante un número. _____
- c) Solo admite valores aislados _____
- d) Puede admitir cualquier valor dentro de un intervalo _____
- e) Sus categorías son mutuamente excluyentes _____
- f) Admite jerarquías en sus categorías _____

6. La siguiente figura muestra una de las páginas de un cuestionario a clientes de una tienda por departamentos:

1-LE GUSTA VESTIR PRENDAS DEPORTIVAS? SI <input type="checkbox"/> NO <input type="checkbox"/>	PERSONAS ENCUESTADAS FEM MAS
3-COMPRO USTED PRENDAS DEPORTIVAS? SI <input type="checkbox"/> NO <input type="checkbox"/>	2-QUE PREnda DEPORTIVAS COMpra Remeras <input type="checkbox"/> Buzos <input type="checkbox"/> shorts <input type="checkbox"/> Otros - Especificar
5-COMO ACOSTUMBRA A PAGAR SU COMPRA? CONTADO <input type="checkbox"/> TARJETA <input type="checkbox"/> ASOCIAc.	4-QUE MARCA ACOSTUMBRA A COMPRAR? PUMA <input type="checkbox"/> Body Club KAPPA <input type="checkbox"/> Caro Cuore Hering <input type="checkbox"/> OTROS Especificificar
6-CUANDO ELIGE LA PREnda LO HACE EN BASE A: COMODIDAD <input type="checkbox"/> MODELO <input type="checkbox"/> MARCA <input type="checkbox"/> CALIDAD <input type="checkbox"/> RESISTENCIA <input type="checkbox"/> MODA <input type="checkbox"/> EXCLUSIVIDAD <input type="checkbox"/> PRECIO <input type="checkbox"/> STATUS <input type="checkbox"/>	5-DE DONDE GENERALMENTE COMpra SU PREnda? DISTRIBUIDORES <input type="checkbox"/> CASAS <input type="checkbox"/> SHOPPING <input type="checkbox"/> REPRESENTANTE TIENDAS <input type="checkbox"/> FABRICA <input type="checkbox"/>
7-CUAL ES EL TAMAÑO DE ROPA QUE USTED ELIGE? PEQUEÑO <input type="checkbox"/> MEDIANO <input type="checkbox"/> GRANDE <input type="checkbox"/> EXTRAGRANDE <input type="checkbox"/>	OTROS: Especificificar
8-CALIFIQUE SEGUN SU ESCALA DE IMPORTANCIA DEL 0 AL 10. SUMANDO EN TOTAL 10 MARA RENOMBRADA <input type="checkbox"/> COMODIDAD AL VESTIR <input type="checkbox"/> MODELO <input type="checkbox"/> CALIDAD/RESISTENCIA <input type="checkbox"/> MODA ACTUAL <input type="checkbox"/> TOTAL <input type="checkbox"/> 10	13 QUIEN NORMALMENTE DECIDE LA COMPRA? USTED <input type="checkbox"/> MAMA <input type="checkbox"/> PAPA <input type="checkbox"/> ESPOSO/A <input type="checkbox"/> HIOS <input type="checkbox"/> OTROS <input type="checkbox"/>

Identifica las variables que aparecen, así como sus sistemas de categorías y niveles de medición.

7. Operacionaliza las siguientes variables completando las tablas de operacionalización:

Variable	Dimensiones	Indicadores	Tipo de variable	Escala de medición	Instrumento
Calidad de un producto	Diseño de producto				
	Durabilidad				

Variable	Dimensiones	Indicadores	Tipo de variable	Escala de medición	Instrumento
Infraestructura de una empresa	Instalaciones de trabajo				
	Equipamiento y maquinaria				

8. Lea el siguiente caso

El Sr. Jesús Ramírez Obregón es el promotor del Consorcio Educativo "Mi Perú", la cual tiene dos sedes en la ciudad de Lima durante el 2014, estas se encuentran ubicadas en los distritos de Villa María y La Molina. Las instituciones educativas brindan servicio de Educación Básica Regular en los niveles de: inicial, primaria y secundaria. Además, cuenta aproximadamente con un población estudiantil que varía entre 1000 – 1500 estudiantes por sede, de los cuales 45% son mujeres y 55% son varones, la edad de los estudiantes fluctúa entre 12-15 años siendo la de mayor número la del distrito de La Molina y la de menor número la de Villa María. En los últimos tres años luego de revisar los datos estadísticos del consorcio, se ha dado cuenta que el porcentaje de "deserción estudiantil" se ha incrementado notoriamente especialmente en la sede de Jesús María. Consultando con los directivos de cada una de las sedes, estos explican que la deserción escolar se debe a diversos factores: cambio de domicilio, costo de pensiones, horario de clases e insatisfacción de los padres de familia por el servicio educativo prestado. El Sr. Ramírez luego de escuchar a sus directivos, reflexionó en que lo más conveniente era realizar una investigación que le permita mejorar el servicio educativo de su consorcio y elevar el nivel de satisfacción de los usuarios, tomando en cuenta: infraestructura, plana docente, currículo de estudios, servicios administrativos, tecnología educativa, clima institucional.

Luego de leer el caso, analice y elabore un listado de ocho (08) variables que encuentre en el contenido indicando su clasificación según naturaleza, número, tipo y valores que puede tomar.

Variable	Número: Dicotómica – Polítómica	Naturaleza: Cualitativa – Cuantitativa	Tipo: Ordinal, nominal, discreta, continua	Valores

9. Dada la siguiente tabla de datos

i	1	2	3	4	5	6	7	8
y_i	2	4	3	6	8	10	5	12
x_i	12	10	8	11	4	3	9	8

Hallar las siguientes sumatorias:

a) $\sum_{i=1}^4 (4y_i + 2)$

b) $\sum_{i=1}^6 (y_i + 3x_i - i)$

c) $\left(\sum_{i=4}^7 x_i \right) \left(\sum_{i=4}^7 y_i \right)$

d) $\sum_{i=3}^6 (y_i x_i)^2$

10. Dado $\sum_{i=1}^4 X_i = 7$ $\sum_{i=1}^4 Y_i = -3$ y $\sum_{i=1}^4 X_i Y_i = 5$ Halle:

a) $\sum_{i=1}^4 (2X_i + 5Y_i)$

b) $\sum_{i=1}^4 (X_i - 3)(2Y_i + 1)$

11. Calcula el porcentaje correspondiente en cada categoría

Nº DE HIJOS	n	%
	147	
1 hijo	59	
2 hijos	49	
3 hijos	21	
4 a mas	18	
Total		

12. Para cada uno de los siguientes problemas, identificar la variable y tipo de variable, y calcular e interpretar la proporción correspondiente:

- a) En el año 2002 se reportaron 1250 crímenes, de los cuales 300 fueron clasificados como muertes sin premeditación y no negligentes:

Variable: _____

Tipo de variables: _____

Proporción P= _____

- b) Un Director del Área de Marketing de una empresa dirigió una investigación y encontró que de un total de 950 compras que consideraba las siguientes escalas: menores a S/.1000, entre S/.1000 a S/.2000 y mayores a 2000. 118 fueron menores a S/.1000

Variable: _____

Tipo de variables: _____

Proporción P= _____

1.3. TABLAS DE DISTRIBUCIÓN DE FRECUENCIAS

1.3.1. TABLA DE DISTRIBUCIÓN DE FRECUENCIAS

Una vez recopilados los datos estos deben ser ordenados o clasificados en tablas. Cuando se dispone de un gran número de datos se debe distribuirlos en clases o categorías y determinar el número de observaciones pertenecientes a cada clase, que es la frecuencia de clase. Una distribución de frecuencias es un arreglo de los valores observados x_1, x_2, \dots, x_k de la variable x con sus respectivas frecuencias, en una tabla de la forma:

Tabla 1

Valor de x	n_i	N_i	h_i	H_i
x_1	n_1	N_1	h_1	H_1
x_2	n_2	N_2	h_2	H_2
\vdots	\vdots	\vdots	\vdots	\vdots
x_k	n_k	$N_k = n$	h_k	H_k
TOTAL	$\sum_{i=1}^k f_i = n$		$\sum_{i=1}^k h_i = 1$	

- **Frecuencia absoluta simple** (n_i), También llamado simplemente frecuencia absoluta, es el número de veces que aparece repetido el valor o cualidad x_i , y se cumple que:

$$n_1 + n_2 + n_3 + \dots + n_k = n$$

En notación sigma: $\sum_{i=1}^k n_i = n$

- **Frecuencia absoluta acumulada** (N_i), es la que resulta de sumar sucesivamente las frecuencias absolutas, así tenemos:

$$N_1 = n_1$$

$$N_2 = n_1 + n_2$$

$$N_3 = n_1 + n_2 + n_3$$

\vdots

$$N_i = n_1 + n_2 + n_3 + \dots + n_i, \quad i = 1, 2, 3, \dots, n$$

- **Frecuencia relativa simple** (h_i), es el cociente entre la frecuencia absoluta y el número de observaciones realizadas (n). Sus valores son números reales que oscilan entre 0 y 1. La suma de todas las frecuencias relativas es igual a 1. Se denota por:

$$h_i = \frac{f_i}{n} \quad ; \quad 0 \leq h_i \leq 1$$

- **Frecuencia relativa acumulada (H_i)**, es igual al cociente entre la frecuencia absoluta acumulada y el número de observaciones realizadas (n), o también es la que resulta de sumar sucesivamente las frecuencias relativas. Se denota por:

$$\begin{aligned} H_1 &= h_1 \\ H_2 &= h_1 + h_2 \\ H_3 &= h_1 + h_2 + h_3 \\ &\vdots \\ H_i &= h_1 + h_2 + h_3 + \dots + h_i \end{aligned} \quad \text{ó} \quad H_i = \frac{F_i}{n} = \frac{\sum_{j=1}^i f_j}{n}$$

1.3.2. TABLA DE DISTRIBUCIÓN DE FRECUENCIAS DE UNA VARIABLE CUALITATIVA

Se usa para clasificar variables cualitativas.

Ejemplo: se estudia cuál fue el impacto que produjo las nuevas políticas de marketing en la venta del jabón ZOTE, para esto se realizó una consulta a un grupo de amas de casa de la ciudad de Arequipa por su preferencia del jabón, respondiendo Sí en caso que siempre utilice el detergente, AV si a veces lo utiliza y NO si nunca utiliza el jabón, los resultados fueron:



<http://www.zote.com.mx/acerca.htm>

<u>Utiliza:</u> Si (siempre) A veces No (Nunca)	Sí	AV	NO	AV	Sí	NO	AV	Sí	Sí	Sí
	NO	NO	Sí	Sí	Sí	AV	Sí	Sí	Sí	AV
	NO	NO	AV	NO	NO	Sí	Sí	NO	AV	NO
	Sí	Sí	Sí	Sí	Sí	Sí	NO	NO	Sí	Sí
	AV	NO	NO	NO	Sí	NO	AV	Sí	Sí	NO

Organice los datos en una tabla de distribución de frecuencias:

Tabla 2

TABLA DE FRECUENCIAS SOBRE LA PREFERENCIA DE LAS AMAS DE CASA POR EL JABÓN ZOTE EN LA CIUDAD DE AREQUIPA

Preferencia	n_i	N_i	h_i	H_i
Siempre	24	24	0.48	0.48
A veces	9	33	0.18	0.66
Nunca	17	50	0.34	1
Total	50		1	

Interpretación:

n_2 : 9 amas de casa a veces utilizan el jabón

N_2 : 33 amas de casa siempre utilizan el jabón o a veces

h_3 : 34% de amas de casa nunca utilizan el jabón

H_2 : 66% amas de casa siempre utilizan el jabón o a veces

1.3.3. TABLA DE FRECUENCIAS DE VARIABLES CUANTITATIVAS DISCRETAS

Son aquellas que se utilizan para agrupar datos cuantitativos de acuerdo a los diferentes valores que toman las variables.

Ejemplo: los siguientes corresponden al número de viajes que realizan por día un grupo de vendedores entrevistados a azar en la empresa SPORTX

4	2	6	3	5	5	3	4	3	3
3	4	4	4	3	5	4	5	5	4
3	2	2	3	5	3	6	5	2	4
3	3	6	4	3	3	2	2	4	3
4	3	3	4	5	4	2	4	3	5



<http://elportaldemendoza.com/blog/>

Tabla 3

TABLA DE FRECUENCIAS SOBRE EL NÚMERO DE VIAJES DE LOS VENDEDORES

Nº Viajes	n_i	N_i	h_i	H_i
2 viajes	7	7	0.14	0.14
3 viajes	17	24	0.34	0.48
4 viajes	14	38	0.28	0.76
5 viajes	9	47	0.18	0.94
6 viajes	3	50	0.06	1.00
Total	50		1.00	

Interpretación:

n₃: 14 vendedores de SPORTX realizaron 4 viajes

N₂: 24 vendedores realizaron menos 4 viajes (máximo 3)

H₃: 76% de vendedores realizaron menos 5 viajes (máximo 4)

1.3.4. TABLA DE FRECUENCIAS DE VARIABLES CUANTITATIVAS CONTINUAS

Son aquellas que se utilizan para agrupar datos cuantitativos continuos mediante intervalos de frecuencias llamados intervalos de clase. Para construir la tabla con intervalos de clase se debe seguir los siguientes pasos:

Paso1: Rango (R). Llamado también “recorrido de los datos”, es la diferencia entre el valor máximo y el valor mínimo de la variable. Consideremos las siguientes variables $y_1, y_2, y_3, \dots, y_n$, entonces: $R = Y_{\max} - Y_{\min}$

Paso2: Número de intervalos de clase (m). Consiste en dividir el rango en un número conveniente de intervalos de clase. El número de intervalos depende principalmente del número de observaciones, sin embargo es recomendable que no sea menor que 5 ni mayor de 15 intervalos. Para determinar el número de intervalos usaremos la fórmula de **Sturges**:

$$m = 1 + 3.32 \log n$$

Cuando los resultados para obtener m son números decimales, entonces se redondeará al entero inmediato.

Ejemplo: Si $n = 40$ entonces $m = 1 + 3.32 \log 40 = 6.32$, redondeando $m = 7$

Si $n = 100$ entonces $m = 1 + 3.32 \log 100 = 7.64$, redondeando
 $m = 8$

Paso3: Amplitud de Clase (c). Es el tamaño o longitud que deben tener los intervalos; se recomienda tener intervalos del mismo tamaño. Se calcula mediante la fórmula:

$$c = \frac{R}{m}$$

Marca de clase (y_i). Es una medida ponderativa que va a representar al intervalo de datos. Es la semisuma entre el límite superior y el límite inferior del intervalo de clase. Sea el intervalo $[L_I - L_S]$ entonces $y_i = \frac{L_I + L_S}{2}$

Ejemplo: los siguientes datos indican el número de minutos que ocuparon sus asientos 50 clientes en una cafetería:

73	65	82	70	45	50	70	54	32	75
75	67	65	60	75	87	83	40	72	64
58	75	89	70	73	55	61	78	89	93
43	51	59	38	65	71	75	85	65	85
49	47	55	60	76	75	69	35	45	63

Construye una tabla de frecuencias adecuada para esta información e interprete.

Pasos:

$$V_{\min} = 32 \quad V_{\max} = 93$$

$$\text{Rango} : R = 93 - 32 = 61$$

$$N^{\circ} \text{ de intervalos} : m = 1 + 3.32 \cdot \log(50) = 6.64 = 7$$

$$\text{Amplitud de intervalo: } c = 61/7 = 8.71 = 9$$



<http://www.tenvinilo.com/vinilos-decorativos/>

Se construirá a continuación una tabla de frecuencias con 7 intervalos y amplitud constante de 9.

Tabla 3
TABLA DE FRECUENCIAS SOBRE EL NÚMERO DE MINUTOS EN LA CAFETERIA

L_i	L_s	Y_i	n_i	N_i	h_i	H_i
[32	41)	36.5	4	4	0.08	0.08
[41	50)	45.5	5	9	0.1	0.18
[50	59)	54.5	6	15	0.12	0.3
[59	68)	63.5	11	26	0.22	0.52
[68	77)	72.5	15	41	0.3	0.82
[77	86)	81.5	5	46	0.1	0.92
[86	95]	90.5	4	50	0.08	1
			50		1	

Interpretación:

n_3 : 4 clientes ocuparon sus asientos como mínimo 32 minutos pero menos de 41

N_4 : 26 clientes estuvieron sentados menos de 68 minutos

h_5 : 30% clientes se sentaron al menos 68 minutos pero menos de 77

H_3 : _____

GUÍA PRÁCTICA N° 3

TABLAS DE DISTRIBUCIÓN DE FRECUENCIAS

1. Un grupo de gerentes valora el desempeño del personal de su empresa como: Excelente (E), Bueno (B), Regular (R) o Malo (M). Los resultados obtenidos son:

E	B	B	R	E	B	R	M
M	B	E	B	R	R	E	E
R	R	M	B	B	E	B	B
E	M	E	R	R	B	B	B
B	B	E	R	R	E	R	M

- a) Construye la tabla de distribución de frecuencias

DESEMPEÑO	n_i	N_i	h_i	H_i
Excelente				
Bueno				
Regular				
Malo				
Total				

- b) Interpretar: $n_3 =$ _____
 $N_2 =$ _____
 $h_2 =$ _____
 $H_3 =$ _____

2. Los siguientes datos corresponden al número de viajes a provincia que realizan por mes un grupo de comerciantes entrevistados al azar en Gamarra Center.

2	5	1	4	2	1	2	4	2	3
5	1	3	1	2	2	3	2	4	3
1	1	2	1	2	3	1	2	1	3

- a) Construye la tabla de distribución de frecuencias

Nº de viajes	n_i	N_i	h_i	H_i
Total				

- b) Interpretar: $n_2 =$ _____
 $N_4 =$ _____
 $h_1 =$ _____

$$H_3 = \underline{\hspace{10cm}}$$

- c) ¿Cuántos comerciantes realizan 2 viajes o menos? _____
- d) ¿Cuántos comerciantes realizan al menos 3 viajes? _____
- e) ¿Cuántos viajes se han realizado entre todos los comerciantes? _____

3. En una clase de estadística hemos medido la altura de los 25 alumnos. Sus medidas en cm son:

167	159	164	170	164
151	168	174	158	163
160	178	150	172	169
159	158	153	157	156
175	165	164	158	163

- a) Elaborar una tabla de distribución de frecuencias adecuada

$$V_{\min} = \underline{\hspace{2cm}} \quad V_{\max} = \underline{\hspace{2cm}} \quad R = \underline{\hspace{2cm}}$$

$$m = \underline{\hspace{2cm}} \quad C = \underline{\hspace{2cm}}$$

Estatura	Y_i	n_i	N_i	h_i	H_i
[– >					
[– >					
[– >					
[– >					
[– >					
[–]					
Total					

a) Interpretar: $n_3 = \underline{\hspace{2cm}}$

$$N_4 = \underline{\hspace{2cm}}$$

$$h_6 = \underline{\hspace{2cm}}$$

$$H_5 = \underline{\hspace{2cm}}$$

b) ¿Cuántos alumnos miden de 155 hasta menos 170 cm? _____

c) ¿Qué porcentaje de alumnos miden al menos 165 cm? _____

4. Completar la siguiente tabla de frecuencia sobre edades:

Edad	n_i	N_i	h_i	H_i
20 años	2		0,05	0,05
21 años				0,15
24 años				0,4
30 años	15			0,775
40 años				1

5. Completar la siguiente tabla de frecuencia sobre edades:

Edad	n_i	N_i	h_i	H_i
20 años	2		0,05	0,05
21 años				0,15
24 años				0,4
30 años	15			0,775
40 años				1

6. Los siguientes datos pertenecen a la distribución de la producción de papas (en toneladas) en 40 zonas del país

$$y_1 = 20; \quad n_2 - n_5 = 2 \quad y_5 = 100 \quad n_1 = 4 \quad n_3 = 20$$

Si se sabe que la distribución es simétrica y presenta 5 intervalos de clase con amplitud constante. Construya una tabla de frecuencias.

$x = \underline{\hspace{2cm}}$	y_i	n_i	N_i	h_i	H_i
[-]					
[-]					
[-]					
[-]					
[-]					
TOTAL				1.00	

7. El cociente intelectual de los 120 alumnos de un centro se da en la tabla adjunta:

Cociente Intelectual $(L_I - L_S)$	y_i	n_i	N_i	h_i	H_i
[-]			12		
[-]	94	22			
[-]	100				
[-]				0.25	
[-]		18			
TOTAL				1.00	

a) Completar la tabla de frecuencias

b) Interpretar: $n_3 = \underline{\hspace{2cm}}$

$h_4 = \underline{\hspace{2cm}}$

$H_3 = \underline{\hspace{2cm}}$

8. Se tiene una tabla de distribución de frecuencias simétrica con 7 intervalos de amplitud constante e igual a 10 y la siguiente información acerca del número de artículos vendidos por un grupo de empleados:

$n_1 = 8$; $n_3 Y_3 = 1260$; $n_2 + n_5 = 62$; $h_3 = 0.21$ y $H_6 = 0.96$. Reconstruir la tabla.

X=_____	y_i	n_i	N_i	h_i	H_i
[-]					
[-]					
[-]					
[-]					
[-]					
[-]					
TOTAL				1.00	

9. El Monto vendido por los empleados de una empresa se da en la tabla adjunta:

Ventas (en miles) $[L_I - L_S)$	y_i	n_i	N_i	h_i	H_i
[82 -]			10		
[90 -]		32		0.20	
[-]					
[-]		30			
[-]		19			
TOTAL				1.00	

Completar la tabla de frecuencias

10. La tabla siguiente se refiere a los tiempos (en minutos) que permanecieron en la cafetería 90 alumnos. Completa la tabla y responde las siguientes preguntas:

Tiempo de permanencia en la cafetería (minutos)	y_i	n_i	N_i	h_i	H_i
[10 - 20]		5			
[20 - 30]		14			
[30 - 40]		29			
[40 - 50]					
[50 - 60]		12			
[60 - 70]		8			

- a) ¿Cuántos alumnos permanecieron en la cafetería al menos 20 minutos pero no más de 40 minutos?
- b) ¿Qué porcentaje de alumnos permaneció en la cafetería menos de 40 minutos?
- c) ¿Qué porcentaje de alumnos permaneció en la cafetería como mínimo de 50 minutos?
11. De una tabla de distribución de frecuencias de 6 intervalos de amplitud constante, se tiene la siguiente información sobre el precio de un conjunto de 300 productos de limpieza.

$$Y_2 = 550 \quad Y_5 = 850 \quad H_1 = h_6 = 0.04 \quad h_2 = 0.14$$

$$H_4 = 0.76 \quad h_5 = 0.20 \quad H_3 - h_2 = 0.26.$$

Reconstruir la tabla de distribución de frecuencias.

$X = \underline{\hspace{2cm}}$	y_i	n_i	N_i	h_i	H_i
[—]					
[—]					
[—]					
[—]					
[—]					
TOTAL				1.00	

12. Construye la tabla de frecuencias en base a la información dada a continuación:

$$\begin{array}{lll} X_i = \text{Nº de asaltos} & n_3 = 12 & H_2 = 0.40 \\ X_2 = 10.5 & h_1 = 0.08 & \sum n_i = 50 \\ X_4 = 18.5 & h_4 = 0.36 & m = 4 \end{array}$$

$X = \underline{\hspace{2cm}}$	y_i	n_i	N_i	h_i	H_i
[—]					
[—]					
[—]					
[—]					
TOTAL				1.00	

13. En un hospital se desea hacer un estudio sobre los pesos de los recién nacidos. Para ello, se recogen los datos de 40 bebés y se tiene:

3,2	3,7	4,2	4,6	3,7	3,0	2,9	3,1	3,0	4,5
4,1	3,8	3,9	3,6	3,2	3,5	3,0	2,5	2,7	2,8
3,0	4,0	4,5	3,5	3,5	3,6	2,9	3,2	4,2	4,3
4,1	4,6	4,2	4,5	4,3	3,2	3,7	2,9	3,1	3,5

- a. Construir la tabla de frecuencias
 - b. Interpretar: n_5 ; N_3 ; h_2 ; H_4
 - c. Si sabemos que los bebés que pesan menos de 3 kilos nacen prematuramente ¿Qué porcentaje de niños prematuros han nacido entre estos 40?
 - d. Normalmente los niños que pesan más de 3 kilos y medio no necesitan estar en la incubadora ¿Puedes decirme que porcentaje de niños están en esta situación?
14. Una empresa que se dedica a preparar dietas, proyecta lanzar al mercado una dieta rigurosa. Los empleados de una compañía se presentaron como voluntarios para dicha promoción. Se realizó un muestreo con 80 dichos empleados elegidos aleatoriamente. Los resultados del chequeo de los pesos (en kg) fueron los siguientes:

80.6	65.8	49.6	79.1	84.4	66.2	79.3	59.4	72.9	73.6
53.2	60.2	91.2	74.8	78.6	81.4	58.6	68.2	67.4	55.6
76.9	77.4	67.9	63.7	49.9	46.4	68.8	67.3	72.3	75.8
88.3	94.6	57.3	87.3	74.3	73.2	90.4	76.3	57.2	71.7
75.6	41.8	73.6	71.4	83.2	67.4	99.3	62.3	89.2	86.8
65.2	62.1	44.8	82.9	81.7	70.4	74.6	76.9	85.7	40.9
54.2	75.3	50.1	61.1	42.3	68.6	56.2	70.8	47.3	66.9
80.2	60.2	71.6	77.1	94.9	61.4	82.1	78.3	51.2	79.3

- a) Elaborar una tabla de distribución de frecuencias adecuada
- b) Interpretar: n_3 ; N_4 ; h_6 ; H_5
- c) ¿Cuántos empleados tienen pesos entre 45 y 60 kg?
- d) ¿Qué porcentaje de empleados tienen pesos mayores que 75.5 kg?

1.4. REPRESENTACIONES GRÁFICAS

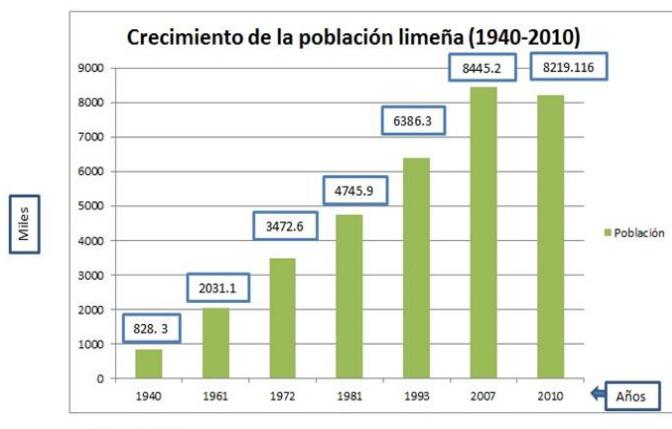
1.4.1. GRÁFICOS ESTADÍSTICOS

El gráfico es la representación de la información estadística, con el fin de obtener una impresión visual global del material presentado, que facilite su rápida comprensión. La utilidad de los gráficos es doble, ya que pueden servir no sólo como sustituto a las tablas, sino que también constituyen el medio más efectivo no sólo para describir y resumir la información, sino también para analizarla.

1.4.2. PARTES DE UN GRÁFICO

Al igual que las tablas estadísticas, los gráficos estadísticos deben tener un título y una explicación de QUÉ, DÓNDE y CUÁNDO se obtuvo la información. Tiene las siguientes partes:

-



Fuente: INEI

Interpretación: Según el gráfico, en el año de 1940 la población de Lima era de 828300 habitantes y para el año 2010 8 219116 habitantes, es decir existe un aumento de la población de Lima a lo largo de los años.

1.4.3. CLASES DE GRÁFICOS

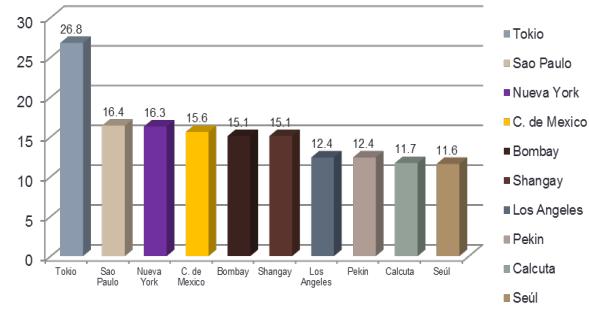
- Gráfico de barras
- Gráfico circular
- Gráfico de líneas
- Histogramas
- Polígono de frecuencias
- Gráfico de ojivas

A. GRÁFICO DE BARRAS. Se utilizan para representar la distribución de frecuencias de variables cualitativas y discretas. Cada categoría de la variable se representa por un rectángulo, cuya altura es proporcional a su frecuencia. Todos los rectángulos tienen la misma base, deben ser de igual ancho y estar igualmente espaciadas.

Ejemplo: Representa mediante un gráfico de barras las ciudades más pobladas (en 1995):

País	Ciudad	Habitantes (millones)
Japon	Tokio	26.8
Brasil	Sao Paulo	16.4
EE.UU	Nueva York	16.3
Mexico	C. de Mexico	15.6
India	Bombay	15.1
China	Shangay	15.1
EE.UU	Los Angeles	12.4
China	Pekin	12.4
India	Calcuta	11.7
Corea Sur	Seúl	11.6
(Fuente: Naciones Unidas)		

Numero Habitantes (en millones) de las Diez Primeras Ciudades mas pobladas del Mundo - 1995

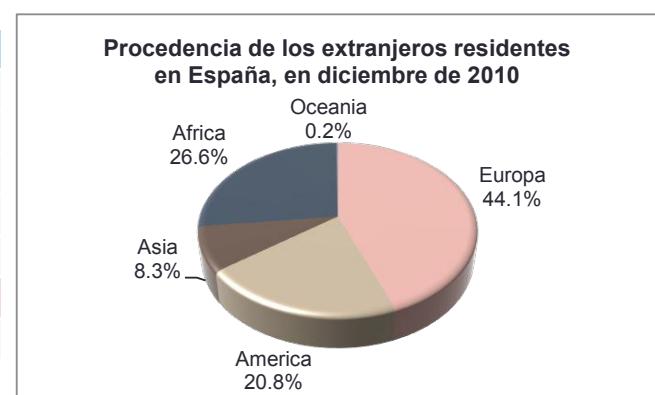


Fuente: elaboración propia

Interpretación: La ciudad de Tokio es la ciudad más poblada con 26.8 millones de habitantes aproximadamente.

B. GRÁFICO CIRCULAR. Sirve para representar, en términos de porcentaje, las distintas partes de un todo. El área de cada sector circular representa el porcentaje sobre el total de cada categoría. **Ejemplo:** Haz un diagrama de sectores para la siguiente tabla:

Procedencia	ni	hi%
Europa	353556	44.12%
América	166709	20.80%
Asia	66340	8.28%
África	213012	26.58%
Oceanía	1712	0.21%
Total	801329	100.00%
(Fuente: INE)		



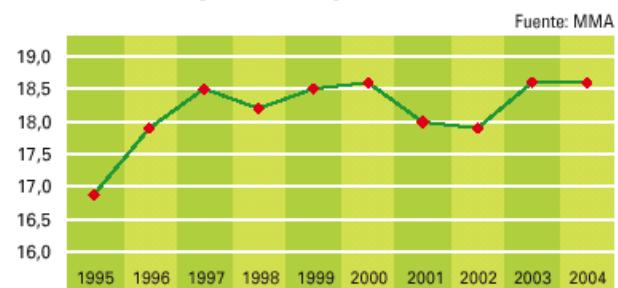
Fuente: elaboración propia

Interpretación: La mayoría de residentes (44.12%) proceden de Europa, el 20.8% proceden de América,....

C. GRÁFICO DE LÍNEAS. Normalmente usados para estudiar la evolución de uno o varios fenómenos a lo largo del tiempo. Esta última variable tiempo se representa en el eje horizontal, mientras que los datos estudiados se miden con referencia al eje vertical.

Año	(%)
1995	16.9
1996	18.0
1997	18.5
1998	18.3
1999	18.5
2000	18.6
2001	18.0
2002	17.9
2003	18.6
2004	18.6

Evolución de la superficie de regadío frente a la superficie agraria total (%)

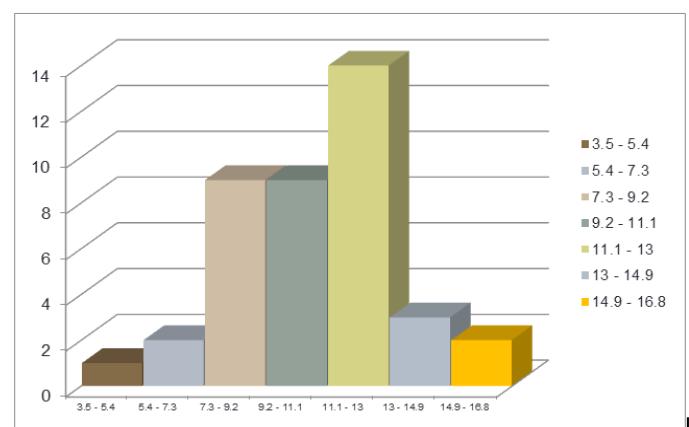


Fuente: Cifras INE. 3/2007

D. HISTOGRAMAS. Es una representación gráfica de una distribución de frecuencias agrupadas en intervalos de clase, mediante una serie de rectángulos contiguos. Su gráfica se realiza entre: (Intervalos vs n_i) e (Intervalos vs h_i)

Ejemplo: preguntando a una muestra de 40 ingenieros sobre sus ingresos mensuales (miles de soles), se tiene la siguiente tabla. Construye un histograma

L_i	L_s	n_i	h_i
3.5	5.4	1	0.025
5.4	7.3	2	0.050
7.3	9.2	9	0.225
9.2	11.1	9	0.225
11.1	13	14	0.350
13	14.9	3	0.075
14.9	16.8	2	0.050
		40	1

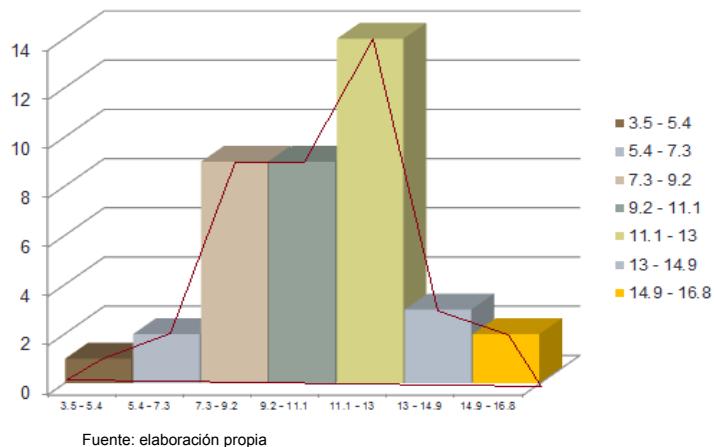


Fuente: elaboración propia

Interpretación: 14 ingenieros ganan como mínimo S/.11100 pero menos de S/.13000.

E. POLÍGONOS DE FRECUENCIAS. Cuando la variable esta agrupada en intervalos de clase se grafica sobre un histograma, el polígono de frecuencias se obtiene uniendo los puntos medios de las bases superiores de cada rectángulo en el histograma.

Ejemplo: En el ejemplo anterior construir el polígono de frecuencias:

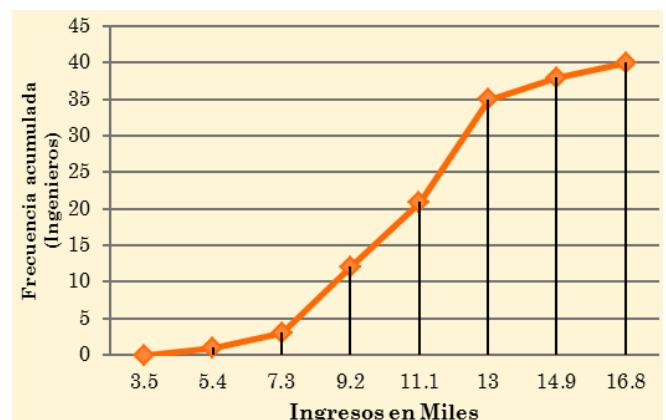


F. POLÍGONO DE FRECUENCIAS ACUMULADAS U OJIVAS.

Esta representación es válida para variables agrupadas en intervalos de clase. Su gráfica se realiza entre: (**Intervalos vs N_i**) ó (**Intervalos vs H_i**)

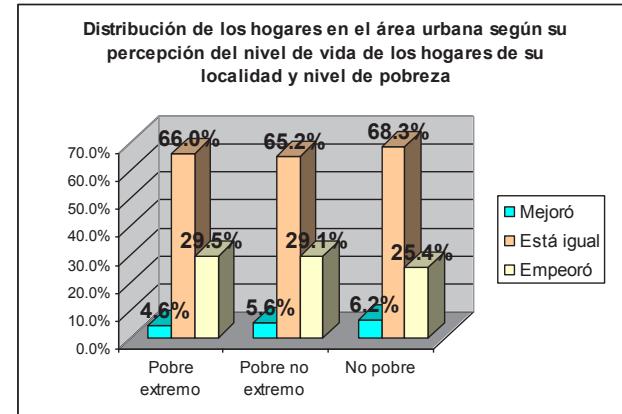
Ejemplo: en el ejemplo anterior sobre los ingresos de 40 ingenieros, construir un gráfico de ojivas.

L_i	L_s	n_i	N_i
3.5	5.4	1	1
5.4	7.3	2	3
7.3	9.2	9	12
9.2	11.1	9	21
11.1	13	14	35
13	14.9	3	38
14.9	16.8	2	40
		40	

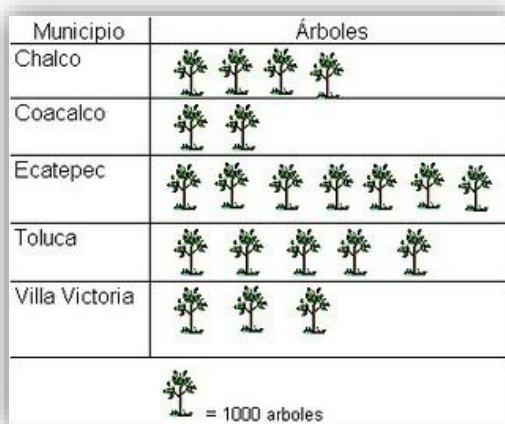


Fuente: elaboración propia

G. GRÁFICO BIDIMENSIONAL. Permiten representar las series de datos en dos dimensiones o sea los valores se representan alineados en dos ejes perpendiculares: el eje horizontal X y el eje vertical Y. Son aplicaciones estadísticas al estudio en conjunto de dos variables cualitativas.



<http://www.aularagon.org/files/espa/>



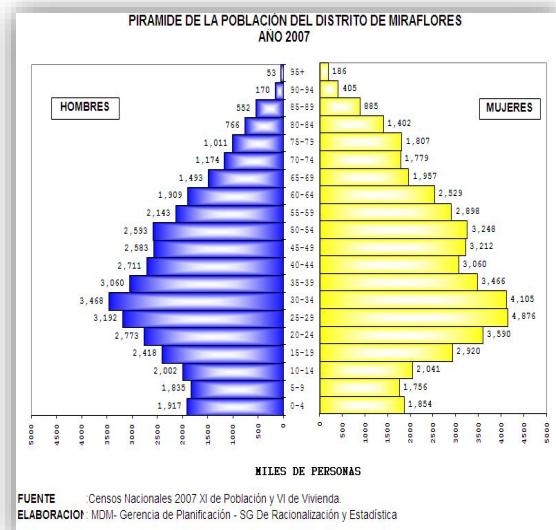
<http://www.aularagon.org/files/espa/>

I. PIRÁMIDES DE POBLACIÓN

Cuando se realizan representaciones correspondientes a edades de población, cambiamos el eje Y por el eje X para obtener las llamadas pirámides de población, que no son más que 2 histogramas a izquierda y derecha, para hombres y mujeres. La pirámide de población nos permite estudiar la estructura demográfica de la población en un momento concreto. Cada barra representa a los distintos grupos de población y nos indica la frecuencia con que podemos encontrar personas de esas características concretas.

H. PICTOGRAMAS

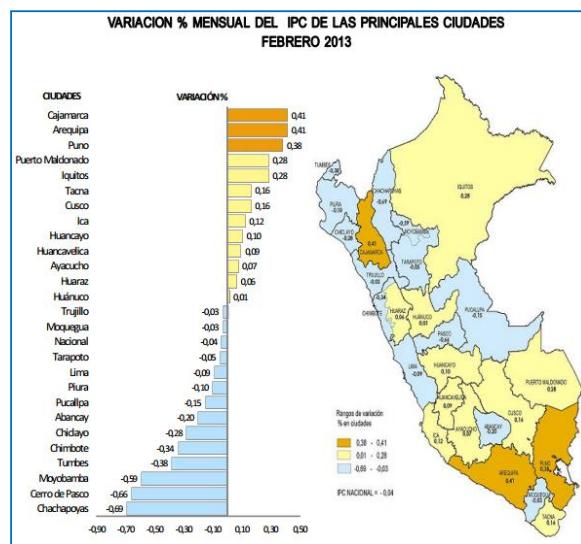
Son gráficos similares a los gráficos de barras. Son gráficos con dibujos alusivos al carácter que se está estudiando y cuyo tamaño es proporcional a la frecuencia que representan; dicha frecuencia se suele representar. Se usan para lograr el interés masivo del público.



<http://estadisticadef09.blogspot.pe/>

J. CARTOGRAMAS

Son gráficos realizados sobre mapas, en los que aparecen indicados sobre las distintas zonas cantidades o colores de acuerdo con el carácter que representan.



<http://www.aularagon.org/files/esp/>

K. DIAGRAMA DE TALLOS Y HOJAS

Un diagrama de tallo-hoja (Tukey, 1977) es un histograma que conserva información numérica. De manera similar al histograma permite ver el lote como un todo y advertir aspectos como:

- ❖ Cuán aproximadamente simétricos son los datos.
- ❖ Cuán dispersos están los valores.
- ❖ La aparición de valores inesperadamente más frecuentes.
- ❖ Si algunos valores están alejados del resto.
- ❖ Si hay concentraciones de valores.
- ❖ Si hay grupos separados.

Ventajas:

- ❖ Es más fácil de construir a mano.
- ❖ Facilita el ordenamiento de los datos.
- ❖ Permite ver la distribución de los datos dentro de cada intervalo como patrones dentro de los datos.

Construcción de tallos y hojas

El diagrama de tallos y hojas es otra forma de representación visual de una serie de valores, conformados por lo menos de dos dígitos.

Para su construcción los números se dividen en dos partes, una llamada de tallo, formada por uno o más dígitos principales y la otra la hoja, que contiene el resto de los dígitos

Ejemplo1: Trace un diagrama de tallo y hoja para los siguientes datos.

70, 72, 75, 64, 58, 83, 80, 82, 76, 75, 68, 65, 57, 78, 85, 72

Pasos:

- Ordenar en forma ascendente
- Definir los tallos
- Incorporar hojas

TALLO		HOJA						
5		7	8					
6		4	5	8				
7		0	2	2	5	5	6	8
8		0	2	3	5			

Fuente: elaboración propia

Ejemplo2: En la tabla N°1 se presenta el puntaje obtenido por 50 estudiantes en un examen de contabilidad financiera de 100 puntos. Construye su diagrama de Tallos y hojas respectivo e interpreta:

Tabla N°1

58	88	65	96	85
74	69	63	88	65
85	91	81	80	90
65	66	81	92	71
82	98	86	100	82
72	94	72	84	73
76	78	78	77	74
83	82	66	76	63
62	62	59	87	97
100	75	84	96	99

Prof.	NºHojas	Tallo	Hojas
2	2	5	8 9
12	10	6	2 2 3 5 5 5 5 6 6 9
24	12	7	1 2 2 3 4 4 5 6 6 7 8 8
26	15	8	0 1 1 2 2 2 3 4 4 5 5 6 7 8 8
11	9	9	0 1 2 4 6 6 7 8 9
2	2	10	0 0

Interpretaciones:

- Los datos son aproximadamente simétricos
- La mayor frecuencia se encuentra en el tallo 8
- 24 alumnos tienen puntajes entre 58 a 78 puntos
- 26 alumnos obtuvieron puntajes entre 80 a 100 puntos

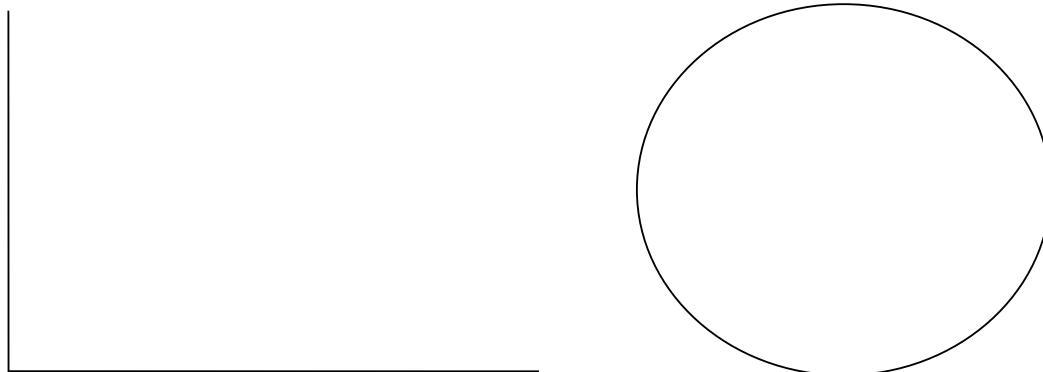
GUÍA DE PRÁCTICA N°4

REPRESENTACIONES GRÁFICAS

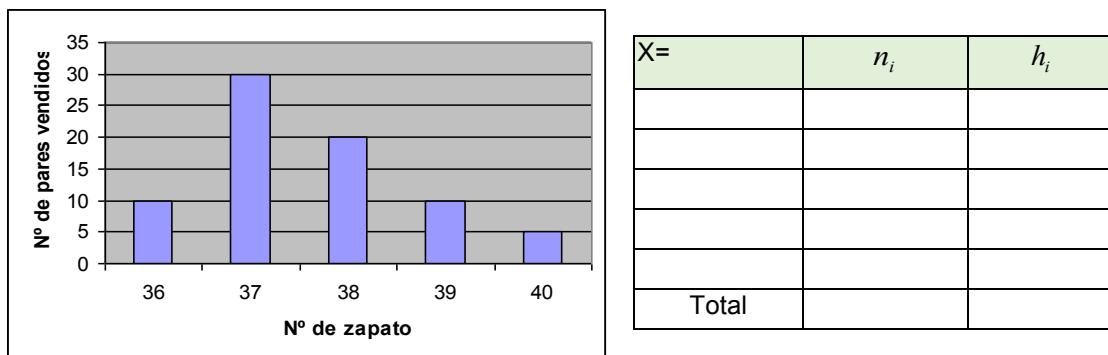
1. A continuación se presentan los datos de accidentes en atletismo durante los años 1994-2000, siendo la distribución según el tipo de lesión, la siguiente:

Lesión	n_i	N_i	h_i	H_i
Herida		6		
Contusión	12			
Esguince	19			
Luxación				
Fractura	5		0.10	
TOTAL				

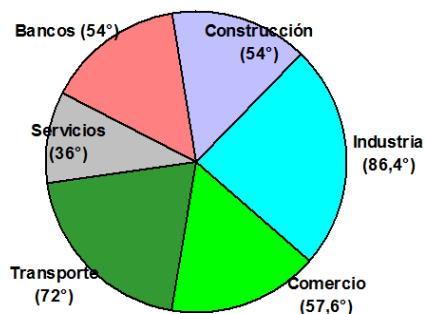
- a) Construye un gráfico de barras e interpreta
- b) Construye un gráfico circular e interpreta



2. La siguiente gráfica recoge la cantidad de parejas de zapatos de mujer vendidas en una tienda a lo largo del día, construye la tabla de frecuencias.



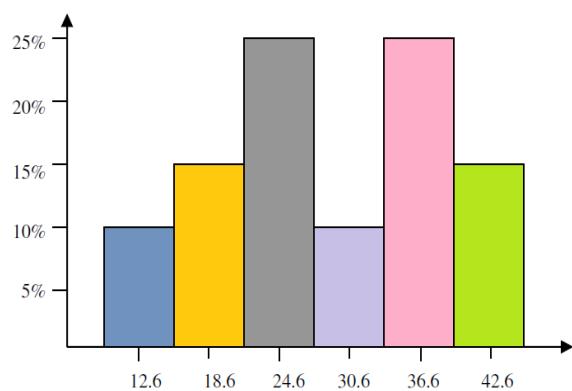
3. El siguiente gráfico circular muestra los datos extraídos de una encuesta sobre 500 empresas de determinada ciudad. Se agregó el valor en grados de cada uno de los ángulos para reconstruir la tabla de frecuencias correspondiente.



X=	n_i	$h_i \%$
Total		

<http://www.aularagon.org/files/espaa/>

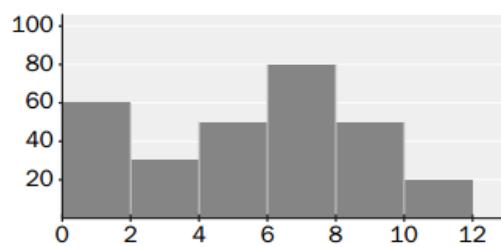
4. Los gastos diarios de una muestra 300 alumnos de una universidad está representado en el siguiente gráfico.



X=	y_i	n_i	N_i	h_i
Total				

<http://www.aularagon.org/files/espaa/>

5. Este histograma representa el número de artículos vendidos en una tienda en una semana, clasificados según su precio en euros...

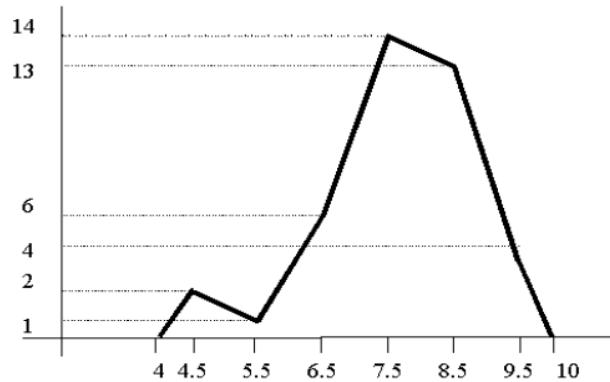


X=	y_i	n_i	N_i	h_i
Total				

<http://www.aularagon.org/files/espaa/>

- a) Construye la tabla de frecuencias.
b) ¿Qué grupo de artículos tuvo mayor venta?

6. Se ha analizado el tiempo que permanece conectado a Internet, a lo largo de un día, un determinado equipo informático, obteniéndose el siguiente gráfico como resultado del estudio realizado.



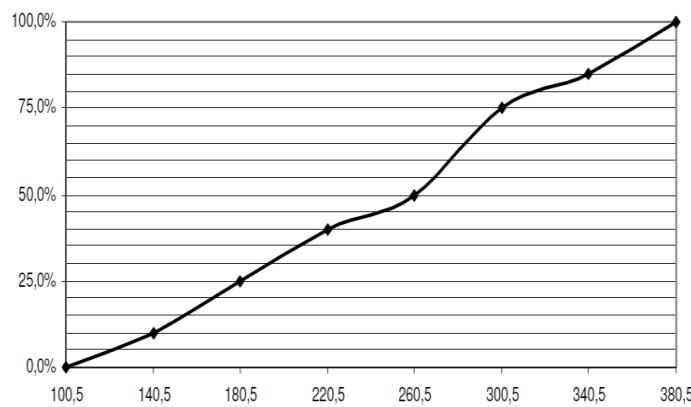
X=	y_i	n_i	N_i	h_i
—				
—				
—				
—				
—				
—				
Total				

<http://www.aularagon.org/files/espa/>

7. Interpreta el siguiente gráfico



8. A partir del gráfico de ojiva, responda las siguientes preguntas (Tamaño de la muestra es 500)
Diseñe la tabla de frecuencia respectiva



X=	y_i	n_i	h_i	H_i
—				
—				
—				
—				
—				
—				
Total				

<http://www.aularagon.org/files/espa/>

9. Un psicólogo prestigioso hizo un estudio sobre el número de ataques de epilepsia que tienen en el año un grupo de personas esquizofrénicas, como resultado se obtuvo la siguiente tabla:

Nº Ataques epilepsia [$L_I - L_S$)	y_i	n_i	N_i	h_i	H_i
[- >]					0.1
[- >]	8		20		
[- >]		15			
[- >]				0.20	
[- >]					
[-]		5	60		
TOTAL					

- a) Completar la tabla de distribución si $C= 4$
- b) Interpretar H_4 _____
- c) Construye un Histograma y un polígono de frecuencias
- d) Construye un gráfico de ojivas

10. Construye la tabla de frecuencias del siguiente pictograma e interprete

Meses	Días de lluvia en los meses de marzo a junio
Marzo	 
Abril	   
Mayo	 
Junio	
Clave	Una  representa 2 días

11. Dada la siguiente información correspondiente a la cantidad de vitamina administrada (en mm³) mensualmente a una muestra de 63 animales:

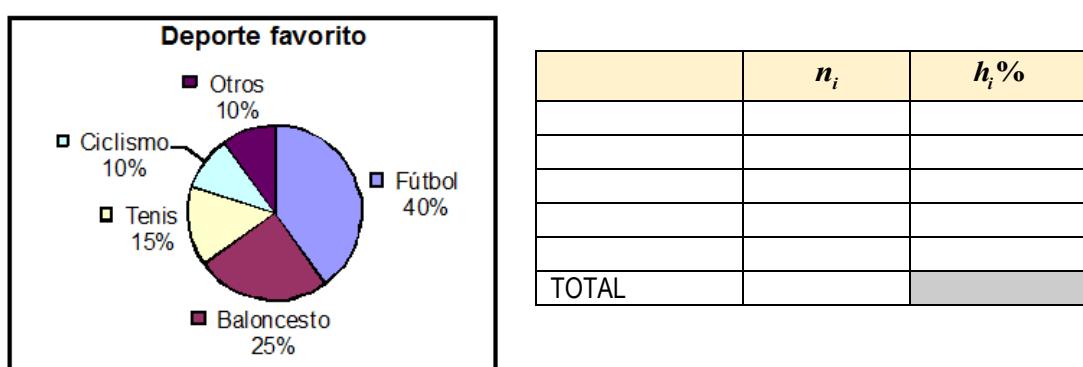
7,0	7,2	6,8	4,4	4,0	5,5	5,8	5,6	2,2	6,2	1,6
4,5	7,1	2,5	5,1	6,0	6,5	5,2	6,3	8,2	5,8	5,7
1,3	5,1	5,7	5,8	8,3	8,7	6,2	5,2	6,3	5,8	5,1
7,7	9,0	6,5	8,3	6,4	7,6	6,6	4,5	6,3	7,7	5,7
8,9	3,8	5,2	9,0	1,3	9,8	7,3	2,8	5,4	6,1	9,2
3,1	4,4	5,9	4,3	3,9	2,7	1,5	6,1			

- a) Identifica la variable y su tipo.
- b) Construye un diagrama de Tallos y hojas
- c) Interpreta el grafico. ¿Será simétrica la distribución?

12. A partir de una muestra de 26 observaciones de la variable X –que toma valores entre 320 y 430, se obtuvo el siguiente diagrama de tallos y hojas: (Unidad=0.1)

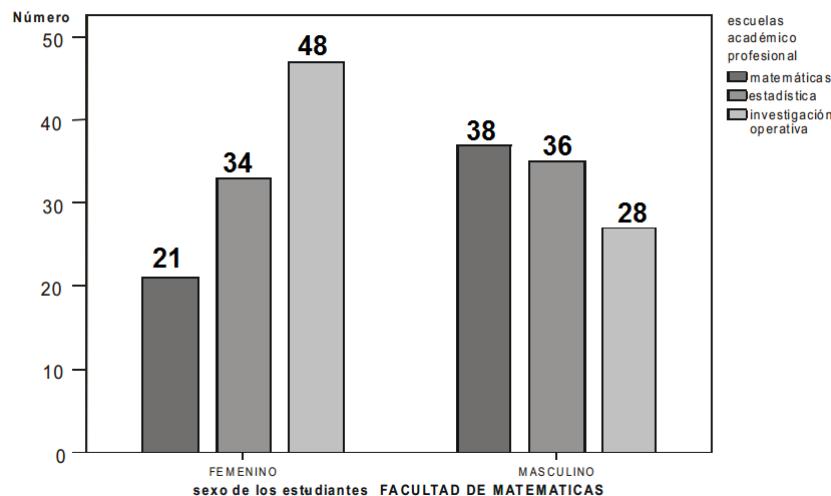
- | | | |
|----|-------|--|
| 32 | 55 | a) Complete con N° de hojas y profundidad |
| 33 | 49 | b) Reproduzca las 26 observaciones (en la ordenación de menor a mayor). |
| 34 | | c) ¿Cuál es la menor observación? |
| 35 | 6699 | d) ¿Cuál es la mayor observación? |
| 36 | 34469 | e) Entre que valores se encuentra agrupados la mayoría de observaciones. |
| 37 | 03345 | f) ¿Existen datos alejados de la distribución? |
| 38 | 9 | |
| 39 | 2347 | |
| 40 | 23 | |
| 41 | | |
| 42 | 4 | g) Escriba alguna interpretación |

13. Dado el siguiente diagrama de sectores sobre gustos en el deporte realizado gracias a una encuesta a 2500 individuos, realiza una tabla de frecuencia que organice los resultados:



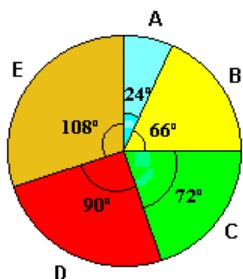
14. Interpreta y construye la tabla porcentual correspondiente del siguiente gráfico

Número de estudiantes según sexo por escuelas profesionales



<https://content.meteoblue.com/es/ayuda/standards>

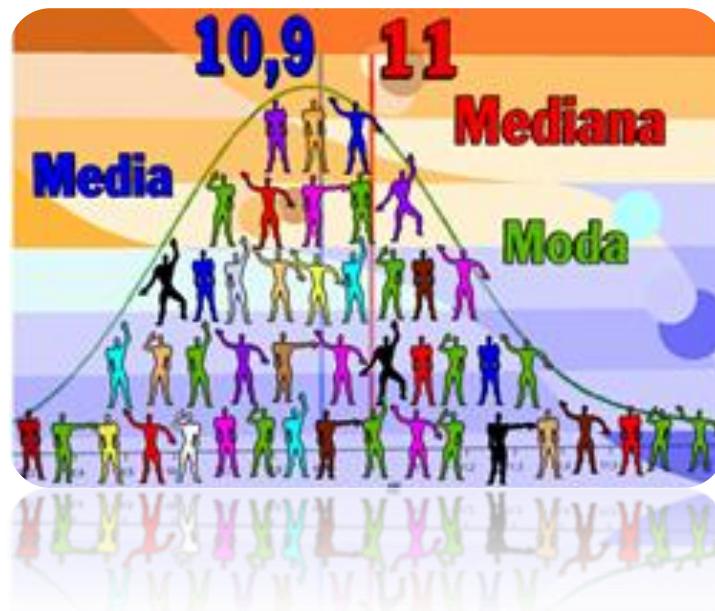
15. El siguiente gráfico representa un total de 600 elementos. ¿Cuál es la frecuencia de cada categoría?



<https://content.meteoblue.com/es/ayuda/standards>

X=	n_i	$h_i \%$	
Total			

UNIDAD 2: MEDIDAS ESTADÍSTICAS DESCRIPTIVAS Y ANÁLISIS DE REGRESIÓN - CORRELACIÓN



<http://estadisticas-ugma-faces-guayana.blogspot.pe/2013/>

Contenido Temático

- Medidas de Tendencia Central
- Medidas de Dispersion
- Análisis de Correlación y Regresión Lineal Simple

2.1. MEDIDAS DE TENDENCIA CENTRAL

2.1.1. DEFINICIÓN

Las medidas de tendencia central o medidas de posición son valores representativos de un conjunto de datos es decir describen con un solo valor un conjunto de observaciones o serie de datos. Dichos valores tienden a situarse en el centro del conjunto de datos ordenados según su magnitud. Las más comunes son:

- Media Aritmética
- Mediana
- Moda

2.1.2. MEDIA ARITMÉTICA

La media aritmética o simplemente media es el estadígrafo de tendencia central más importante y comúnmente se le conoce como **promedio**. La media aritmética se define como el cociente de la suma de los valores de una variable entre el número de observaciones o valores. Simbólicamente:

$$\bar{X} = \frac{\sum_{i=1}^N X_i}{N} = \frac{X_1 + X_2 + \dots + X_N}{N}$$

A. CALCULO DE LA MEDIA PARA DATOS NO AGRUPADOS

Se calculará cuando no está elaborada una tabla de frecuencias.

Ejemplo1: sea las edades en años de 5 niños son 8, 3, 5, 12 y 10. Entonces la media aritmética de las edades de éstos niños es:

$$\bar{X} = \frac{8+3+5+12+10}{5} = \frac{38}{5} = 7.6$$

Ejemplo2: calcule la media aritmética del número de televisores vendidos por 10 empleados durante una campaña. 73, 68, 59, 40, 81, 72, 40, 70, 59 y 72

$$\bar{x} = \frac{\sum x_i}{n} = \frac{73+68+59+\dots+59+72}{10} = 63$$

Interpretación: los empleados vendieron un promedio de 63 televisores.

B. CALCULO DE LA MEDIA PARA DATOS AGRUPADOS

Se utilizará cuando los datos están distribuidos en una tabla de frecuencias. Luego se calcula la media aritmética aplicando la fórmula:

$$\bar{x} = \frac{\sum_{i=1}^n n_i y_i}{n} ; \text{ donde } n \text{ es igual al número total de datos.}$$

Ejemplo: Calcule la media de la siguiente distribución de frecuencia correspondiente la inversión anual (en miles de dólares) de 40 empresas.

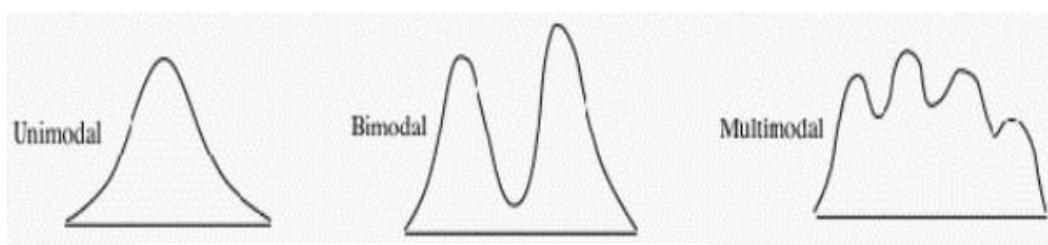
Inversiones [L _I – L _S)	y _i	n _i	n _i × y _i
[4 – 10)	7	1	7
[10 – 16)	13	3	39
[16 – 22)	19	6	114
[22 – 28)	25	12	300
[28 – 34)	31	11	341
[34 – 40)	37	5	185
[40 – 46]	43	2	86
TOTAL		n = 40	1072

Aplicando la formula se tiene: $\bar{x} = \frac{\sum n_i y_i}{n} = \frac{1072}{40} = 26.8$

Interpretación: La Inversión media anual del grupo de empresas fue de \$26800.

2.1.3. MODA

La moda de un conjunto de observaciones es el valor que se presenta con más frecuencia o el que más se repite. Puede ser:



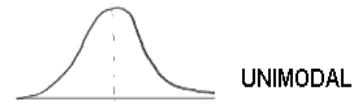
<http://temasdeenfermeria.com.ar/tag/estadistica-descriptiva/>

A. CALCULO DE LA MODA PARA DATOS NO AGRUPADOS

Se ordenan los datos en forma creciente o decreciente y se ubica los valores que más se repiten.

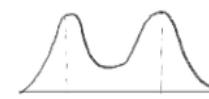
Ejemplo1: 18, 23, 25, 20, 25, 21, 20, 25

Ordenando: 18, 20, 20, 21, 23, 25, 25, 25 ; **Mo= 25**



Ejemplo2: 18, 23, 25, 20, 23, 25, 21, 22

Ordenando: 18, 20, 21, 22, 23, 23, 25, 25; **Mo= 23 ó Mo= 25**



Ejemplo3: 18, 19, 20, 21, 22, 23; Mo = No tiene



B. CALCULO DE LA MODA PARA DATOS AGRUPADOS

Se utilizará cuando los datos están distribuidos en una tabla de frecuencias. Luego se calcula la **moda** aplicando la fórmula:

$$M_o = L_I + c_j \left(\frac{\Delta_1}{\Delta_2 + \Delta_1} \right)$$

Donde:

L_I : Límite inferior de la clase modal

c_j : Amplitud del intervalo de la clase modal

$\Delta_1 = n_j - n_{j-1}$

$\Delta_2 = n_j - n_{j+1}$

n_j : Frecuencia absoluta modal

n_{j-1} : Frecuencia absoluta anterior a la frecuencia modal

n_{j+1} : Frecuencia absoluta posterior a la frecuencia modal

Observaciones:

- La moda se puede determinar en todos los dos tipos de variables cualitativas y cuantitativas.
- La moda tiene la ventaja de no ser afectada por valores extremos
- En muchas series de datos no hay moda porque ningún valor aparece más de una vez.

Ejemplo: Calcule la moda de la siguiente distribución de frecuencia correspondiente la inversión anual (en miles de dólares) de 40 empresas.

Inversiones $[L_I - L_S]$	y_i	n_i
[4 – 10]	7	1
[10 – 16]	13	3
[16 – 22]	19	6
[22 – 28]	25	12
[28 – 34]	31	11
[34 – 40]	37	5
[40 – 46]	43	2
TOTAL		n = 40

Pasos:

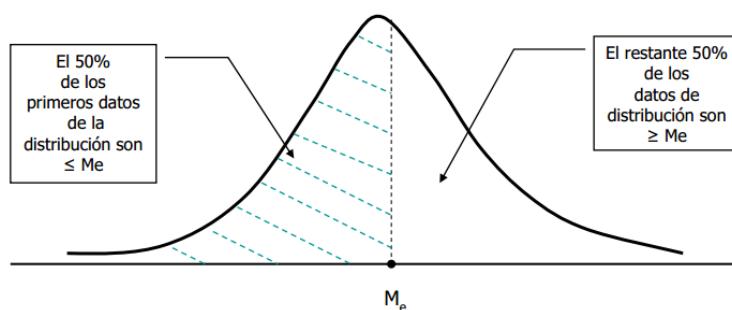
- Ubicamos primero la mayor frecuencia: $n_j = 12$
- Luego la moda se encuentra en la clase 22 - 28, por lo tanto
- $L_I = 22 ; n_j = 12 ; n_{j-1} = 6 ; n_{j+1} = 11 ; c_j = 6$
- $\Delta_1 = 12 - 6 = 6 ; \Delta_2 = 12 - 11 = 1$

$$Mo = 22 + 6 \left(\frac{6}{6+1} \right) = 28.857$$

Interpretación: El monto de inversión que más se repite es \$28857 lo que significa que la mayoría de las empresas invierte esa cantidad.

2.1.4. MEDIANA

La mediana es valor que se encuentra en el centro luego de ordenar los datos y divide el conjunto de datos en dos partes iguales.



A. MEDIANA DE DATOS NO AGRUPADOS

Para determinar la mediana de n observaciones x_1, x_2, \dots, x_n primero se ordenan dichas observaciones descendente o ascendente luego se ubica o se calcula la mediana dependiendo de la cantidad de datos “n” si es par o impar:

Ejemplo1. (Cuando el nº de datos es impar)

17, 24, 20, 18, 22, 21, 24; Ordenando: 17, 18, 20, **21**, 22, 24, 24 (n=7 impar)

$$\text{Posicion} = \frac{7+1}{2} = 4 \quad Me = 21$$

Ejemplo2. (Cuando el nº de datos es par)

13, 14, 7, 11, 15, 16, 12, 9; ordenando: 7, 9, 11, **12, 13**, 14, 15, 16 (n=8 par)

$$Me = \frac{12+13}{2} = 12.5$$

B. CALCULO DE LA MEDIANA PARA DATOS AGRUPADOS

Cuando se trabajan con tablas de frecuencias de intervalos, la fórmula para calcular la mediana es:

$$Me = L_I + c_j \left(\frac{\frac{n}{2} - N_{j-1}}{N_j - N_{j-1}} \right)$$

Donde:

L_I : Límite inferior de la clase mediana

c_j : Amplitud del intervalo de la clase mediana

n : Número total de observaciones o datos

N_j : Frecuencia acumulada de la clase mediana

N_{j-1} : Frecuencia acumulada anterior de la clase mediana.

Observación:

Se denomina clase mediana al intervalo de clase que contiene a la mediana en una tabla de distribución de frecuencias

Ejemplo: Calcule la mediana de la siguiente distribución de frecuencia correspondiente la inversión anual (en miles de dólares) de 40 empresas.

Inversiones $[L_I - L_s)$	y_i	n_i	N_i
[4 – 10)	7	1	1
[10 – 16)	13	3	4
[16 – 22)	19	6	10
[22 – 28)	25	12	22
[28 – 34)	31	11	33
[34 – 40)	37	5	38
[40 – 46]	43	2	40
TOTAL		n = 40	

Pasos:

i) Calcular $\frac{n}{2} = \frac{40}{2} = 20$ y ubicar en los N_i

ii) Luego la mediana se encuentra en la clase [22 – 28], por lo tanto:

$$L_I = 22 ; N_j = 22 ; N_{j-1} = 10 ; c_j = 6$$

$$Me = 22 + 6 \left[\frac{20-10}{22-10} \right] = 22 + 6 \left[\frac{10}{12} \right] = 27$$

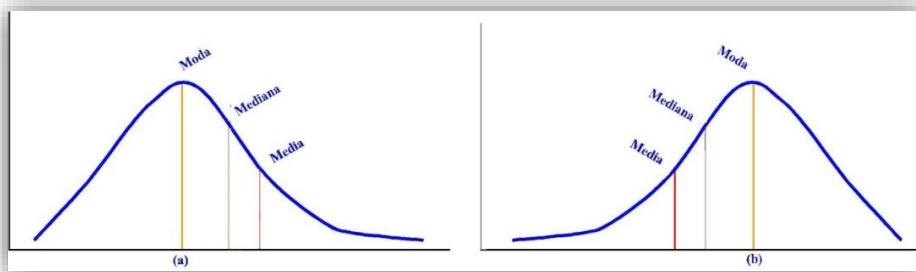
Interpretación: El 50% de las empresas invierten anualmente un monto menor o igual a \$27000.

2.1.5. COMPARACIÓN ENTRE LA MEDIA, MODA Y MEDIANA

- Las distribuciones simétricas tienen el mismo valor para la media, la mediana y la moda.
- En una distribución con sesgo positivo, la moda se halla en el punto más alto de la distribución, la mediana está hacia la derecha de la moda y la media más a la derecha. Es decir

$$Mo < Me < \bar{x}$$
 (a)
- En una distribución con sesgo negativo, la moda es el punto más alto, la mediana está a la izquierda de la moda y la media está a la izquierda de la mediana. Es decir

$$\bar{x} < Me < Mo$$
 (b)



Cuando la población tiene una distribución sesgada, con frecuencia la mediana resulta ser la mejor medida de posición. La mediana no se ve influida por la frecuencia de aparición de un solo valor como es el caso de la moda, ni se distorsiona con la presencia de valores extremos como la media.

2.1.6. MEDIA PONDERADA

Hay ocasiones en que se quiere expresar en una sola cifra los resultados de varios grupos de datos, cada uno de los cuales ha sido resumido previamente mediante un promedio, teniendo cada grupo diferente número de observaciones. Para hallar un promedio general de estos grupos hacemos uso de la media ponderada.

Definición: Sean $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ las medias aritméticas de k subconjuntos menores, cada uno con n_1, n_2, \dots, n_k observaciones respectivamente. La media aritmética del conjunto formado por los términos de k subconjuntos es dado por la fórmula:

$$\bar{x}_p = \frac{\sum_{j=1}^k n_j \bar{x}_j}{n} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \dots + n_k \bar{x}_k}{n_1 + n_2 + \dots + n_r}; \text{ donde } n = \sum_{j=1}^k n_j$$

Ejemplo: durante el mes de octubre de 2008 el promedio de salarios en 3 empresas fueron:

Empresa	Promedio de salarios	Nº de obreros
A	200	10
B	220	15
C	300	20

Hallar el salario medio ponderado durante ese mes.

$$\bar{x} = \frac{200 \times 10 + 220 \times 15 + 300 \times 20}{10 + 15 + 20} = 251.11$$

Interpretación: el salario promedio en las empresas fue de 251.11 para el mes de Octubre del 2008.

2.1.7. CUARTILES

Son medidas de resumen que dividen en cuatro partes iguales al conjunto de valores ordenados de una distribución de frecuencias. Trabajando con tablas de frecuencias con intervalos, la fórmula para calcular cuartiles es:

$$Q_k = L_I + c_j \left(\frac{\frac{nk}{4} - N_{j-1}}{N_j - N_{j-1}} \right) ; \quad k=1,2,3$$

Donde:

L_I : Límite inferior de la clase cuartil

c_j : Amplitud del intervalo de la clase cuartil

n : Número total de observaciones o datos

N_j : Frecuencia acumulada de la clase cuartil

N_{j-1} : Frecuencia acumulada anterior de la clase cuartil

k : k -ésimo cuartil

Así por ejemplo:

$$Q_1 = L_I + c_j \left(\frac{\frac{n}{4} - N_{j-1}}{N_j - N_{j-1}} \right) \quad Y \quad Q_3 = L_I + c_j \left(\frac{\frac{3n}{4} - N_{j-1}}{N_j - N_{j-1}} \right)$$

Ejemplo: dada la siguiente distribución, determinar los cuartiles Q_1 y Q_3

EDADES	N_i	N_i
[6–16)	8	8
[16–26)	20	28
[26–36)	25	53
[36–46)	10	63
[46–56]	5	68
TOTAL	n =68	

← Clase que contiene a Q_1
 ← Clase que contiene a Q_3

Solución:

Determinación de Q_1 : $\frac{n}{4} = \frac{68}{4} = 17$

Determinación de Q_3 : $\frac{3n}{4} = \frac{3(68)}{4} = 51$

Usando las formulas dadas para calcular:

$$Q_1 = 16 + 10 \left(\frac{17-8}{28-8} \right) = 16 + 4.5 = 20.5$$

$$Q_3 = 26 + 10 \left(\frac{51-28}{53-28} \right) = 26 + 9.2 = 35.2$$

2.1.8. DECILES

Son medidas de posición que dividen el total de observaciones en 10 partes iguales, la fórmula para calcular deciles es:

$$D_k = L_I + c_j \left(\frac{\frac{nk}{10} - N_{j-1}}{N_j - N_{j-1}} \right) ; \quad k = 1, 2, 3, \dots, 10$$

Donde:

L_I : Límite inferior de la clase decil

c_j : Amplitud del intervalo de la clase decil

n : Número total de observaciones o datos

N_j : Frecuencia acumulada de la clase decil

N_{j-1} : Frecuencia acumulada anterior de la clase decil

k : k-ésimo decil

2.1.9. PERCENTILES

Son medidas de posición que dividen el total de observaciones en 100 partes iguales, la fórmula para calcular percentiles es:

$$P_k = L_I + c_j \left(\frac{\frac{nk}{100} - N_{j-1}}{N_j - N_{j-1}} \right) ; \quad k = 1, 2, 3, \dots, 100$$

Donde:

L_I : Límite inferior de la clase percentil

c_j : Amplitud del intervalo de la clase percentil

n : Número total de observaciones o datos

N_j : Frecuencia acumulada de la clase percentil

N_{j-1} : Frecuencia acumulada anterior de la clase percentil

k : k-ésimo percentil

Así:

$$P_{10} = L_I + c_j \left(\frac{\frac{10n}{100} - N_{j-1}}{N_j - N_{j-1}} \right) \quad y \quad P_{90} = L_I + c_j \left(\frac{\frac{90n}{100} - N_{j-1}}{N_j - N_{j-1}} \right)$$

Ejemplo: dada la siguiente distribución, determinar los percentiles P_{65} y P_{85}

EDADES (x_i)	n_i	N_i
$[10 - 30]$	94	94
$[30 - 50]$	140	234
$[50 - 70]$	160	394
$[70 - 90]$	98	492
$[90 - 100]$	8	500
TOTAL	n = 500	

← Clase de P_{65}
 ← Clase de P_{85}

Determinación de P_{65} : $\frac{nk}{100} = \frac{500(65)}{100} = 325$

Determinación de P_{85} : $\frac{nk}{100} = \frac{500(85)}{100} = 425$

Usando las formulas dadas para calcular:

$$P_{65} = 50 + \left(\frac{\frac{500(65)}{100} - 234}{\frac{100}{394} - 234} \right) 20 = 50 + 11.37 = 61.37$$

$$P_{85} = 70 + \left(\frac{\frac{500(85)}{100} - 394}{\frac{100}{492} - 394} \right) 20 = 70 + 6.33 = 76.33$$

GUÍA PRÁCTICA N° 5

MEDIDAS DE TENDENCIA CENTRAL

1. Sean los precios en dólares de 7 libros de metodología que se venden en una librería:

65	45	47	51	36	65	65
----	----	----	----	----	----	----

Calcula las medidas de tendencia central e interpreta:

Medida	Valor	Interpretación
Media =		
Moda =		
Mediana =		

2. Se preguntó a 10 abogados de una consultoría seleccionados aleatoriamente cuantas horas semanales dedicaban a leer periódico. Los resultados fueron:

12	8	10	12	12	8	12	9	11	13
----	---	----	----	----	---	----	---	----	----

Calcula las medidas de tendencia central e interpreta:

Medida	Valor	Interpretación
\bar{x} =		
M_o =		
M_e =		

3. El administrador de una empresa textil desea conocer los años que vienen trabajando los técnicos en la compañía, para lo cual revisa los registros de 90 técnicos donde figura su tiempo en años de permanencia y luego obtiene las siguientes medidas:

MEDIDA		INTERPRETACION
\bar{x} =	13.5	
M_o =	12.6	
M_e =	13.9	
Q_1 =	5.2	
D_6 =	14.3	
P_{37} =	8.7	

4. Se tienen los gastos en movilidad por día de 40 empleados de la empresa SYS.COM cuyos resultados están distribuidos en la siguiente tabla de frecuencias:

Gastos (S/.)		y_i	n_i	$n_i \times y_i$
Li	Ls			
8	12		4	
12	16		12	
16	20		14	
20	24		7	
24	28		3	

$$\bar{x} = \frac{\sum n_i y_i}{n} =$$

$$M_o = L_I + C \left(\frac{\Delta_1}{\Delta_2 + \Delta_1} \right)$$

$$\Delta_1 = n_i - n_{i-1}, \quad \Delta_2 = n_i - n_{i+1}$$

Calcula la media aritmética y moda e interpreta:

$\bar{x} =$		
$M_o =$		

5. En el ejemplo anterior calcular e interpretar la mediana e interpreta:

Gastos (S/.)		y_i	n_i	N_i
Li	Ls			
8	12		4	
12	16		12	
16	20		14	
20	24		7	
24	28		3	

$$Me = L_I + C \left(\frac{\frac{n}{2} - N_{i-1}}{N_i - N_{i-1}} \right)$$

$Me =$		
--------	--	--

6. Un empresario desea repartir unas bonificaciones entre sus empleados en base a la categoría y productividad de los mismos. Dicha distribución quedó de la siguiente forma:

Bonificación (U\$)		y_i	n_i	N_i	$n_i \times y_i$
Li	Ls				
15			3		
21			8		
			12		
			15		
			7		
			5		
	Total				

Calcula la media, moda, mediana e interpreta

7. Las ventas de cinco vendedores de una empresa son: \$8000, \$9000, \$10500, \$9800 y \$55000.
 Calcula el sueldo medio, la moda y la mediana e indica cuál representa mejor a los datos.

Medida	Valor
$\bar{x} =$	
$Mo =$	
$Me =$	

8. En una empresa hay 3 directivos, 50 operarios y 8 vendedores. Los sueldos mensuales, en euros, de cada categoría son los siguientes: directivos, 4.000; operarios, 1.400; vendedores, 2.000.
- Halla la moda, la mediana y la media de los sueldos.
 - ¿Qué medida es más representativa del promedio?

9. Los Sueldos de una empresa están distribuidos de la siguiente manera:

Sueldos	y_i	n_i	N_i	$y_i n_i$
[– >				
[– >	300	25	36	
[– >	500	26		
[– >		35		
[– >		8		
[–]			120	
Total				

Calcular las medidas de tendencia central e interprete los resultados

Medida	Valor	Interpretación
$\bar{x} =$		
$Mo =$		
$Me =$		

10. En una empresa donde el sueldo medio es de \$400 se incrementa un personal igual al 25% del ya existente con un sueldo medio igual al 60% de los antiguos. Si 3 meses más tarde se incrementan cada sueldo en 20%, más 30\$, ¿cuánto es el nuevo salario medio?
11. Los sueldos en una empresa varían de \$300 a \$800 distribuidos en forma simétrica en 5 intervalos de igual amplitud, con el 15%, 20%, y 30% de casos en el primer, segundo y tercer intervalo respectivamente. Calcule los diferentes indicadores de tendencia central.

12. Si la media de horas de estudio diarios de los 2130 alumnos de cierta universidad es de 2,58 horas ¿cuál es la media de horas de estudio diario de los alumnos de las facultades de letras? teniendo en cuenta los datos de la tabla siguiente:

	Medicina	Derecho	Ciencias	Letras	Total
Media	2,50	3,00	4,00		2,58
n	580	250	350		2130

13. Se desea analizar el precio de las viviendas en el municipio de Villa el Salvador para ello tras un minucioso estudio en el registro de la propiedad y una agrupación de los datos se obtiene que durante los primeros 5 meses del presente año los precios vienen reflejados en la tabla siguiente y las unidades vienen dadas en miles de euros:

Precio	y_i	n_i	N_i	$y_i n_i$
[12 – 16 >		3		
[– >		14		
[– >		33		
[– >		21		
[–]		9		

- a) ¿Cuál es el precio medio de las viviendas?
- b) ¿Cuál es el precio más común de las viviendas?
- c) La mitad de las viviendas cuánto cuestan como máximo?

14. Una compañía tiene cuatro departamentos, en el primero trabajan 200 personas cuyo sueldo promedio es \$1200, en el segundo, 100 personas con un sueldo medio de \$800, en el tercero, 150 personas con un sueldo medio de \$1000 y en el cuarto, 250 personas con un sueldo medio de \$600.

- a) Hallar el sueldo medio de los trabajadores de la compañía.
- b) Para el mes próximo, la compañía incrementará los sueldos de los trabajadores del primer departamento en el 20%, a los del segundo departamento en el 15%, a los del tercer departamento en el 10% más \$50 y a los del cuarto departamento en \$100. ¿Cuál es el nuevo sueldo promedio de los trabajadores?

15. Los costos de fabricación, en soles, de diez objetos son los siguientes: 9.35, 9.46, 9.20, 9.80, 9.77, 9.00, 9.99, 9.36, 9.50, 9.60, si el precio de venta de cada objeto es 3 veces su costo de fabricación menos 5 soles, calcular la utilidad media por objeto.

16. Dado el siguiente cuadro estadístico con ancho de clase constante igual a 20. Determine la media de los datos.

$[L_i - L_s)$	y_i	n_i	N_i	$y_i n_i$
				880
				1950
			35	1800
		13		
200)				
		4	70	
Total				

17. Dada la siguiente distribución de frecuencias, calcular el valor de "n" sabiendo que la moda es 60 y pertenece al tercer intervalo.

$[L_i - L_s)$	n_i
[16 – 32[6
[32 – 48[n
[48 – 64[8
[64 – 80[3n
[80 – 96]	3

18. Calcular la frecuencia correspondiente al tercer intervalo de la siguiente distribución, sabiendo que la media aritmética es igual a 11,50.

$L_i - L_s$	n_i
4-6	4
6-10	5
10-16	X
16-20	3
20-30	1

19. En una encuesta sobre los ingresos anuales en miles de soles de un grupo de familias se obtuvo la siguiente información:

$[L_i - L_s)$	y_i	n_i
10 – 30		20
30 – 50		
50 – 70		
70 - 90		20

Además, $\bar{x} = 54$ y $n_2 / n_3 = 1/5$, calcular el número de familias con ingreso no menos de 50 mil soles. Construya además un gráfico circular e interprete.

20. Los siguientes datos corresponden a la altura de alumnos de un grupo de curso.

Altura en metros	Nº alumnos
1.50	1
1.55	5
1.60	10
1.65	15
1.70	5
1.75	1
1.80	2

Calcular las medidas de tendencia central e interpretar los resultados

Medida	Valor	Interpretación
$\bar{x} =$		
$M_o =$		
$M_e =$		

21. La siguiente información corresponde al rendimiento que tuvieron en la asignatura de religión los alumnos de segundo de secundaria de cierto establecimiento educacional:

RENDIMIENTO	ni
DEFICIENTE	5
REGULAR	10
BUENO	11
MUY BUENO	7
EXCELENTE	9
TOTAL	42

a) ¿Qué medidas de tendencia central tienen sentido calcular en este caso? Justifique.

b) Calcule e interprete la(s) medida(s) de tendencia central que tenga(n) sentido.

Medida	Valor	Interpretación
$\bar{x} =$		
$M_o =$		
$M_e =$		

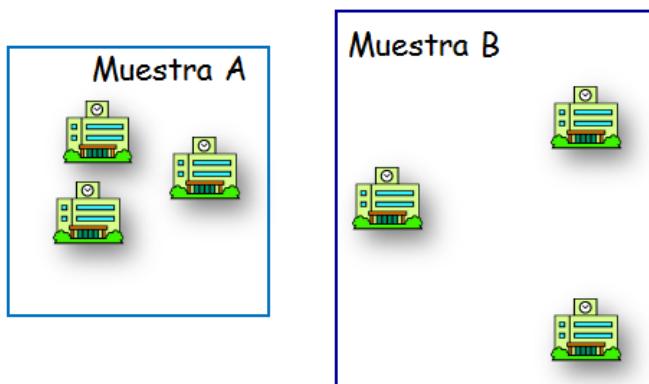
2.2. MEDIDAS DE DISPERSIÓN

2.2.1. DEFINICIÓN

Son medidas que nos dan el grado de concentración o dispersión de las observaciones alrededor de un valor central o de posición. Las más comunes son:

- Recorrido o rango
- Varianza
- Desviación Estándar o Típica
- Coeficiente de Variación

Ejemplo: sean las distancias de tres viviendas en dos muestras ¿en qué muestra las viviendas están más separadas o dispersas?



2.2.2. RANGO O RECORRIDO

Es la diferencia entre el valor mayor y menor de una variable. Su fórmula es:

$$R = X_{\max} - X_{\min}$$

Ejemplo1: para una serie de datos de carácter cuantitativo como es la estatura tal y como:

$$x_1 = 185, \quad x_2 = 165, \quad x_3 = 170, \quad x_4 = 182, \quad x_5 = 155$$

De este modo, el rango sería la diferencia entre el valor máximo y el mínimo, o lo que es lo mismo: $R = 185 - 155 = 30$.

Ejemplo2: Observando las edades de un grupo de personas: 6, 10, 16, 22, 36, 48 ,56

Hallar el rango de los datos. $R = 56 - 6 = 50$ años.

2.2.3. VARIANZA Y DESVIACIÓN ESTÁNDAR

La varianza y la desviación estándar son las medidas de dispersión más importantes y expresan el grado de dispersión de las observaciones respecto a la media aritmética. Si se trabaja con la población se calcula la varianza poblacional (σ^2) y si se toma una muestra se calcula la varianza muestral (s^2)

A. PARA DATOS NO AGRUPADOS

Tipo	Varianza	Desviación Estándar
Poblacional	$\sigma^2 = V(X) = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N} = \frac{\sum X_i^2 - N\bar{X}^2}{N}$	$\sigma = \sqrt{V(X)}$
Muestral	$s^2 = V(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$	$s = \sqrt{V(x)}$

B. PARA DATOS AGRUPADOS

Tipo	Varianza	Desviación Estándar
Poblacional	$\sigma^2 = V(X) = \frac{\sum_{i=1}^N n_i (X_i - \bar{X})^2}{N} = \frac{\sum n_i X_i^2 - N\bar{X}^2}{N}$	$\sigma = \sqrt{V(X)}$
Muestral	$s^2 = V(x) = \frac{\sum_{i=1}^n n_i (x_i - \bar{x})^2}{n-1} = \frac{\sum n_i x_i^2 - n\bar{x}^2}{n-1}$	$s = \sqrt{V(x)}$

OBSERVACIONES:

- A la desviación estándar también se la llama desviación típica y es la raíz cuadrada de la varianza. La varianza nunca es negativa.
- Cuando la variable toma un único valor; es decir cuando es constante entonces la varianza es cero.
- Mientras más se approxima a cero, más concentrados están los valores de la serie alrededor de la media. Por el contrario, mientras mayor sea la varianza, más dispersos están.

Ejemplo1: (cálculo de la varianza y desviación estándar para datos no agrupados)

El número de días necesarios por 10 equipos de trabajadores para terminar 10 instalaciones de iguales características han sido: 21, 32, 15, 59, 60, 61, 64, 60, 71, y 80 días. Calcule e interprete la Varianza y Desviación Típica.

- Hallamos primero la media: $\bar{x} = 52.3$

- Varianza :

$$s^2 = \frac{(21-52.3)^2 + (32-52.3)^2 + \dots + (80-52.3)^2}{10-1} = 475.12$$

- Desviación estándar : $s = \sqrt{475.12} = 21.8$

Interpretación: el número de días necesarios para terminar el trabajo de los 10 equipos se alejan del promedio en 21.8 días.

Ejemplo2: (Calculo de la varianza y desviación estándar para datos agrupados)

Sean los impuestos pagados anualmente por 46 empresas organizadas en la siguiente tabla:

Impuestos (miles)	x_i	n_i	$n_i x_i$	$n_i x_i^2$
$[35-41]$	38	4	152	5776
$[41-47]$	44	9	396	17424
$[47-53]$	50	14	700	35000
$[53-59]$	56	11	616	34496
$[59-65]$	62	8	496	30752
TOTAL		n=46	2360	123448

- Determinando la media: $\bar{x} = \sum \frac{n_i x_i}{n} = \frac{2360}{46} = 51.3$

- La varianza es: $s^2 = \frac{123448 - 46(51.3)^2}{46-1} = 53.12$

- La desviación estándar es $s = \sqrt{s^2} = \sqrt{53.12} = 7.29$

Interpretación: En promedio el monto de impuestos pagado por los contribuyentes se desvían de su media aritmética en S/. 7290

2.2.4. COEFICIENTE DE VARIACIÓN

Es una medida relativa porque no tiene unidades de medida, se utiliza para comparar la variabilidad de series de datos que tengan unidades diferentes o de dos distribuciones distintas (diferente tamaño de muestra o unidades). Se calcula mediante la fórmula:

$$C.V\% = \frac{S}{\bar{X}} \times 100 \quad (\text{Coeficiente de variación muestral})$$

OBSERVACIÓN:

- Al realizar comparaciones entre dos variables, el coeficiente de variación más pequeño será el que tenga menor dispersión relativa.
- Un coeficiente de variación mayor a 30% indica un alto grado de dispersión y pequeña representatividad de la media, pero cuanto menor sea a 30% la media será más representativa.

Ejemplo: se ha evaluado a 3 obreros para saber su rapidez en la fabricación de un artículo, sus resultados en segundos se presentan en la siguiente tabla. Mediante el C.V. indicar que obrero tiene el rendimiento más uniforme:

Obrero1	Obrero2	Obrero3
61.8	60.8	61.4
61.9	60.7	61.2
63.2	62.9	63.3
63.8	62.8	61.0
61.4	62.9	63.2
61.2	62.5	60.8
63.3	62.1	61.4
61.0	61.9	61.2
62.5	62.8	63.2
60.8	63.8	61.8

Medidas	Obrero1	Obrero2	Obrero3
\bar{x}	62.09	62.32	61.85
S	1.06	0.97	0.99
$C.V$	0.017	0.0156	0.016
$C.V\%$	1.70%	1.56%	1.60%

Conclusión

Comparando los tres coeficientes de variación se observa que el segundo obrero tiene el menor coeficiente, por lo tanto dicho obrero tiene rendimiento más homogéneo o uniforme.

GUÍA PRÁCTICA N°6

MEDIDAS DE DISPERSIÓN

1. Sean los datos referentes a los años de antigüedad que tienen en una empresa sus empleados:

19	23		25	19	19	15
----	----	--	----	----	----	----

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Calcula las siguientes medidas e interpreta:

$$S^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$$

Medida	Valor	Interpretación
$\bar{x} =$		
$Mo =$		
$Me =$		
$S^2 =$		
$S =$		
$C.V. = S / \bar{x}$		

2. Se hizo un estudio en 7 viviendas de Villa el salvador sobre el monto ahorrado de los jefes de hogar hasta la fecha. Los resultados en miles de soles fueron:

13	10	9	13	8	12	13
----	----	---	----	---	----	----

Calcula las siguientes medidas e interpreta:

Medida	Valor	Interpretación
$\bar{x} =$		
$Mo =$		
$Me =$		
$S^2 =$		
$S =$		
$C.V. = S / \bar{x}$		

3. Se tiene a continuación la atención mensual de usuarios en 3 consultorios psicológicos:

Medidas	Consultorio A	Consultorio B	Consultorio C
Promedio	144	142	158
Desviación estándar	20	19	21
C.V.			

¿En qué consultorio la cantidad de usuarios atendidos es más homogénea?

4. Se hizo un estudio en una muestra de 60 vendedores de seguros sobre el nivel de ansiedad al realizar su trabajo. Las calificaciones en una prueba de ansiedad fueron:

Calificación en la prueba de ansiedad		y_i	n_i	$n_i \times y_i$	$n_i \times y_i^2$
Li	Ls				
			6		
60			12		
80			15		
			18		
			7		
			2		
Total					

$$\bar{x} = \frac{\sum n_i y_i}{n} =$$

$$s^2 = \frac{\sum_{i=1}^m n_i y_i^2 - n \bar{x}^2}{n-1}$$

Calcula las siguientes medidas e interpreta:

$\bar{x} =$		
$S^2 =$		
$S =$		
$C.V. =$		

5. Se ha realizado una investigación en un grupo de médicos para determinar el número de horas al año que dedican a actualizarse y capacitarse, se obtuvieron los siguientes resultados:

Horas de estudio mensuales		y_i	n_i	N_i	$n_i \times y_i$	$n_i \times y_i^2$
Li	Ls					
					112	
		36		12		
			11			
			13			
		60	6	42		
TOTAL						

Calcula las siguientes medidas e interpreta:

$\bar{x} =$		
$Mo =$		
$Me =$		
$S^2 =$		
$S =$		
$C.V. =$		

6. En una fábrica, el personal de planta está dividido en supervisores, operarios calificados y asistentes. Se cuenta con la siguiente información:

	Supervisores	Operarios calificados	Asistentes
Salario promedio	2000	1200	800
Desviación estándar	300	2400	200

- a) ¿Qué grupo de trabajadores tiene los salarios más homogéneos?
 b) Se decide incrementar los salarios del personal de la siguiente manera:

- Supervisores : 30% de su salario actual
- Operarios calificados : S/. 650
- Asistentes : 20% de sus salario actual más S/. 200

Después de los incrementos ¿Qué grupo de trabajadores tiene los salarios más homogéneos?

7. La distribución de sueldos en una empresa es tal que el sueldo promedio es S/. 1200 y la varianza S/.400. Si el sueldo de cada empleado se incrementa en 10% y adicionalmente se otorga una bonificación de S/. 50 a cada empleado ¿Cómo cambia la varianza de la distribución de sueldos?
 8. Se investigaron 8 muestras de un producto alimenticio en cuanto al contenido de vitaminas A. Las cantidades x_i de vitamina A medidos en mg mostraron los siguientes resultados:

$$\sum x_i = 187; \quad \sum x_i^2 = 5009;$$

Calcule el Coeficiente de Variación e interprete.

9. Un jugador de baloncesto anota, cada domingo, el número de puntos que encesta en el partido de la liga. Las anotaciones de los 10 últimos encuentros, jugados por su equipo, se muestran en el siguiente cuadro.

Encuentro	1	2	3	4	5	6	7	8	9	10
Anotaciones	10	18	17	8	10	9	19	10	7	10

- a) Calcular la media y la moda de las anotaciones.
 b) Calcular el coeficiente de variación y representar el diagrama de barras, utilizando las frecuencias relativas.
10. El salario promedio en una ciudad es de 11 000 u.m. con una variancia de 2 000 u.m. ¿Cuáles serán la nueva media y la nueva variancia si se efectúan los siguientes cambios:

- a) Se aumenta 810 u.m a todos
- b) Se aumenta el 15 % de su salario a cada trabajador
- c) Si se duplican los sueldos

11. Se sabe que los tiempos que tardan los trabajadores de la empresa automotriz AVF, que construye autos compactos, para colocar la llanta de refacción debajo del chasis, es una variable cuyos tiempos tienen una gran variación. El superintendente tomó 13 tiempos al azar, durante todo un día de ese ensamblaje, de cada uno de los cuatro trabajadores que realizan la operación: Juan, José, Virgilio y Mario. Esos tiempos se muestran enseguida, dados en minutos.

JUAN	4.5	4.3	4.5	4.6	4.3	4.7	4.5	4.4	4.5	4.7	4.5	4.5	4.5
JOSÉ	4.5	5.0	5.0	5.3	4.8	4.3	5.4	4.2	5.0	5.0	4.6	5.4	5.0
VIRGILIO	5.1	5.2	5.0	5.2	5.1	5.0	5.2	5.0	5.2	5.1	5.0	5.2	5.2
MARIO	4.1	4.2	4.0	4.0	4.4	4.3	4.3	4.4	4.3	4.2	4.3	4.2	4.2

- a. Para cada uno de los trabajadores, calcula los estadísticos: media aritmética, mediana, moda, rango, varianza y desviación estándar.

	X	Me	Mo	R	S ²	S
Juan						
José						
Virgilio						
Mario						

- b. ¿Por qué el rango y la desviación estándar son estadísticos?
 c. ¿Los datos son los de una muestra o los de una población? ¿Por qué?
 d. De acuerdo con los resultados obtenidos en las diferentes estadísticas, ¿cuál trabajador parece ser el mejor de todos? ¿cuál trabajador parece ser el peor de ellos? ¿Por qué?

12. Los 16 edificios más altos de Montreal tiene 47, 43, 42, 40, 38, 36, 33, 33, 33, 32, 32, 32, 27, 27, 26, y 22 pisos.

- a) Calcule la desviación estándar de la muestra de edificios.
 b) Vuelva a determinar la desviación luego de eliminar los cuatro edificios más altos.

13. Dos caballos de resistencia, Reo y Petrarca, compiten frecuentemente uno contra otro. Sus respectivos dueños han registrado los siguientes datos para los tiempos que han hecho al correr cuatro millas en competencia entre ellos.

VARIABLES	ESTADÍSTICAS DESCRIPTIVAS					
	n	MEDIA	MEDIANA	DESVIACIÓN ESTÁNDAR	DATO MENOR	DATO MAYOR
Reo	40	4.55 min	4.50 min	0.20 min	4.45 min	5.40 min
Tetrarca	40	4.57 min	4.58 min	0.15 min	4.40 min	5.00 min

- a. ¿Cuál distribución de tiempos es más simétrica? ¿Alguna es normal? ¿Por qué?
 b. ¿Hacia dónde se da el sesgo de cada distribución? ¿Qué significa esto? Dibuja una sobre otra las distribuciones de frecuencias respectivas aproximadas para cada caballo.
 c. ¿Siempre gana Reo a Petrarca? ¿Por qué? Explica con base en los datos.
 d. ¿La desviación estándar de Reo se ve afectada por el dato mayor? Explica por qué.

14. En un estudio comparativo de los rendimientos de ciertos bonos, se elaboró la siguiente distribución de los rendimientos al vencimiento de una muestra de 50 bonos.

Resumen porcentual	y_i	Números de bonos	$n_i \times y_i$	$n_i \times y_i^2$
6.0-7.9		1		
8.0-9.9		5		
10.0-11.9		11		
12.0-13.9		21		
14.0-15.9		9		
16.0-17.9		3		

- a) Determina la media y la desviación típica
- b) Se puede concluir que los rendimientos son uniformes
15. Una muestra de siete automóviles de alquiler, de una flota grande, utilizó las siguientes cantidades de gasolina en un día: 19.9, 19.3, 14.7, 13.8, 15.3, 11.4 y 12.6 galones. Calcule la desviación estándar del número de galones de gasolina que utilizaron.
16. La siguiente tabla de frecuencias corresponde a los jornales, en soles de los obreros de una fábrica que cuenta con 500 obreros:

Jornal(Soles)	y_i	n_i	N_i	h_i	H_i
$[80 - 90]$				0.10	0.10
$[90 - 100]$		110		0.22	0.32
$[100 - 110]$		120			
$[110 - 120]$		80			
$[120 - 130]$		70			
$[130 - 140]$		60	490		
$[140 - 150]$					
TOTAL					

- a) Calcule e interprete la media, moda, mediana y desviación estándar
- b) Calcule e interprete el coeficiente de variación
- c) Se decidió otorgar un aumento a todos los obreros del 30% de su jornal más una bonificación por concepto de pasajes de 25 soles ¿Cuál es la nueva varianza y el nuevo coeficiente de variación de los salarios?

2.3. ANÁLISIS DE CORRELACIÓN Y REGRESIÓN LINEAL SIMPLE

2.3.1. INTRODUCCIÓN

Es frecuente que estudiemos sobre una misma población los valores de dos variables estadísticas X e Y distintas, con el fin de ver si existe alguna relación entre ellas. El análisis de este relacionamiento presenta dos aspectos diferentes.

Unas veces nuestro interés está en conocer si las dos variables están asociadas y medir hasta qué punto los cambios en una pueden explicarse por los cambios que ocurren en la otra. En tal caso tenemos un problema de correlación y la medida que cuantifica es el llamado **Coeficiente de Correlación**.

Otras veces, cuando estamos seguros que existe un alto grado de asociación entre las dos variables, el análisis se encamina a cuantificar la relación existente con el fin de predecir cuáles serán los valores de la variable respuesta, en este caso tenemos un problema de **Regresión**.

2.3.2. ANÁLISIS DE CORRELACIÓN

La correlación es la teoría que se encarga de estudiar las posibles relaciones existentes entre dos variables X e Y estadísticas. En lo sucesivo, cuando queramos hablar de la dependencia entre dos variables, hablaremos de correlación entre ambas variables. La correlación es el método empleado para determinar el grado de asociación o relacionamiento entre las variables que se estudian.

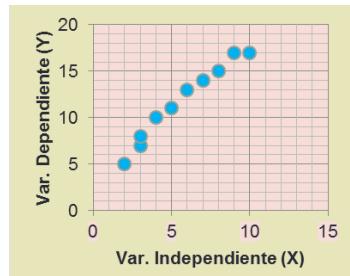
2.3.3. TIPOS DE CORRELACIÓN

Atendiendo al relacionamiento entre las variables X e Y podemos tener:

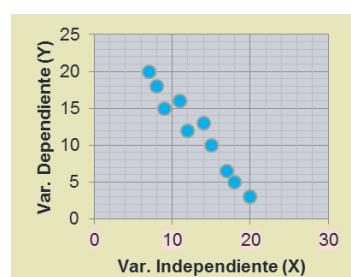
- A. **Correlación directa o positiva.** Cuando las variables X e Y presentan variaciones en un mismo sentido, esto es, para mayores valores de X corresponde mayores valores de Y.
- B. **Correlación inversa o negativa.** Cuando las variaciones de X e Y son en sentidos contrarios, esto es, para mayores valores de X corresponden menores valores de Y.
- C. **Sin correlación.** Cuando no existe ningún tipo de relacionamiento entre las variables.

2.3.4. DIAGRAMA DE DISPERSIÓN

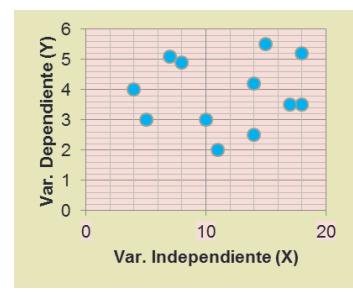
Es la representación gráfica más útil para describir el comportamiento conjunto de dos variables X e Y, consiste en una nube de puntos que indicará si existe o no correlación dependiendo de la tendencia de la tendencia de los puntos.



Correlación Positiva
(Directa)



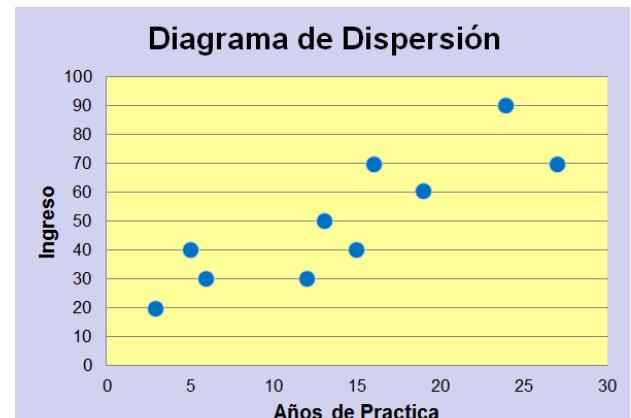
Correlación Negativa
(Inversa)



Sin correlación

Ejemplo 1: los siguientes datos representan los años de práctica profesional y el ingreso anual (en miles de soles) para un conjunto de servidores públicos. Construir el diagrama de dispersión e indicar el tipo de correlación.

Años de Práctica (X)	Ingreso (Y)
5	40
15	40
24	90
16	70
19	60
3	20
6	30
12	30
27	70
13	50



Interpretación: observando el gráfico podemos decir que los años de práctica y el ingreso anual tienen una correlación Directa o Positiva.

Observación: si se desea investigar la relación existente entre dos variables el primer paso será trazar el diagrama de dispersión, el cual proporcionará una idea del tipo de relación existente entre ambas variables.

2.3.5. COEFICIENTE DE CORRELACIÓN

Es un número que indica el grado de asociación entre las variables y se define del siguiente modo:

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right) \left(\sum_{i=1}^n y_i^2 - n \bar{y}^2 \right)}}$$

Características:

- i) El coeficiente r es un número comprendido entre -1 y +1

$$-1 \leq r \leq +1$$

- ii) Si $r > 0$ Existe una correlación directa o positiva
 iii) Si $r < 0$ Existe una correlación inversa o negativa
 iv) Si $r = +1$ Existe una perfecta asociación positiva entre las dos variables
 v) Si $r = -1$ Existe una perfecta asociación negativa entre las dos variables
 vi) Si $r = 0$ No existe asociación entre las dos variables, no existe asociación lineal

Ejemplo2: del ejemplo1 sobre los años de práctica profesional y el ingreso anual hallar el coeficiente de correlación:

Nº	Años de Práctica(X)	Ingreso(Y)	XY	X ²	Y ²
1	5	40	200	25	1600
2	15	40	600	225	1600
3	24	90	2160	576	8100
4	16	70	1120	256	4900
5	19	60	1140	361	3600
6	3	20	60	9	400
7	6	30	180	36	900
8	12	30	360	144	900
9	27	70	1890	729	4900
10	13	50	650	169	2500
Total	140	500	8360	2530	29400
	ΣX	ΣY	ΣXY	ΣX^2	ΣY^2

$$n = 10 \quad ; \quad \bar{x} = \frac{\sum x}{n} = \frac{140}{10} = 14 ; \quad \bar{y} = \frac{500}{10} = 50$$

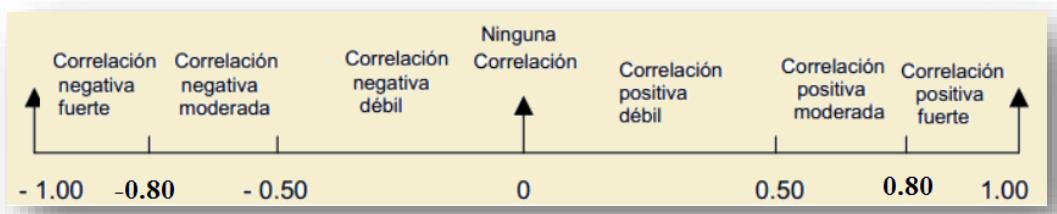
Entonces:

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right) \left(\sum_{i=1}^n y_i^2 - n \bar{y}^2 \right)}} = \frac{8360 - 10(14)(50)}{\sqrt{(2530 - 10(14)^2)(29400 - 10(50)^2)}} = 0.859$$

Interpretación: como $r=0.859$, existe una correlación positiva fuerte entre los años de práctica y el ingreso.

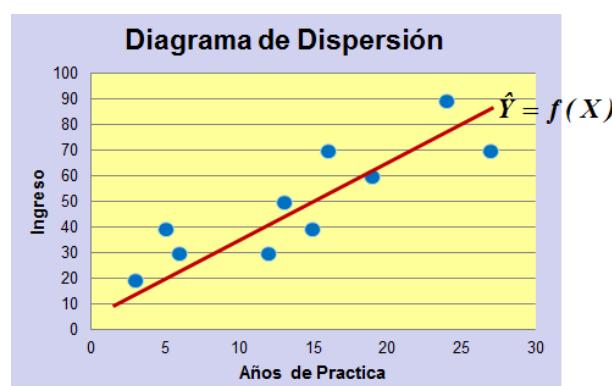
NIVELES DE CORRELACIÓN r

Como se observa en los diagramas anteriores, el valor de r se aproxima a +1 cuando la correlación tiende a ser lineal directa (mayores valores de X significan mayores valores de Y), y se aproxima a -1 cuando la correlación tiende a ser lineal inversa. El siguiente diagrama resume el análisis del coeficiente de correlación entre dos variables:



2.3.6. ANÁLISIS DE REGRESIÓN

Es un método que se emplea para encontrar una función que se ajusta a una nube de puntos o diagrama de dispersión, con la finalidad de obtener una predicción aproximada de una de las variables a partir de la otra. Es decir la función o ecuación nos permitirá hacer pronósticos.



2.3.7. REGRESIÓN LINEAL SIMPLE

Este tipo de regresión se utiliza cuando existe solo una variable independiente X para una variable dependiente Y. Está definida por la siguiente ecuación lineal en su forma general:

$$Y = b_0 + b_1 X + e$$

Donde:

Y Es la variable respuesta o valor de la predicción de la variable Y dado un valor X

b_0, b_1 Coeficientes de la regresión lineal

b_0 Es el valor de Y cuando X = 0, es decir, es el valor de Y cuando la línea de regresión cruza el eje de las Y

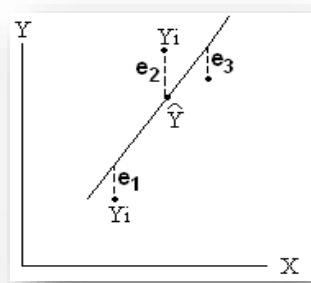
b_1 Es la pendiente de la línea, o la variación promedio en Y por cada variación de una unidad en X

X Es cualquier valor seleccionado de la variable independiente X

e Es el error de predicción

2.3.8. MÉTODO DE MÍNIMOS CUADRADOS PARA ESTIMAR LOS COEFICIENTES DE REGRESIÓN

Es un método para obtener la recta que se ajuste mejor a los datos, al graficar los datos sabemos que podemos trazar infinidad de rectas pero este método nos proporciona la de mejor ajuste.



A e_i se le llama desviación o error que puede ser negativo o positivo, si elevamos las desviaciones al cuadrado, para obtener la recta más representativa la suma de los errores al cuadrado $e_1^2 + e_2^2 + e_3^2 + \dots + e_n^2$ debe ser lo más pequeña posible.

2.3.9. RECTA DE MÍNIMOS CUADRADOS

Para poder obtener la recta de regresión Y en X utilizaremos la siguiente ecuación de estimación:

$$\hat{Y} = b_0 + b_1 X$$

Donde los valores de b_0 y b_1 en la ecuación de regresión son conocidos como coeficientes de regresión y las fórmulas para calcularlos son:

$$b_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

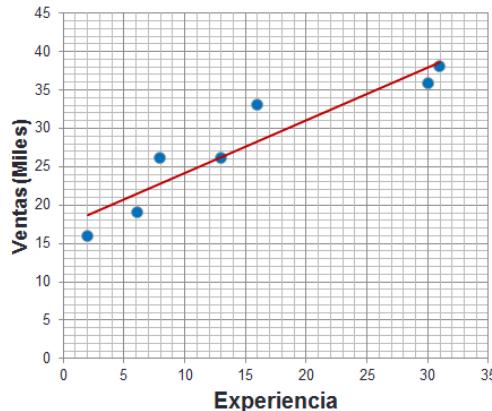
Ejemplo 1: Una empresa tiene 7 vendedores, y se quiere analizar las ventas mensuales (en miles de soles) frente a los años de experiencia. Los datos están en la siguiente tabla.

Nº	Experiencia (X)	Ventas Y (miles)
1	13	26
2	16	33
3	30	36
4	2	16
5	8	26
6	6	19
7	31	38

- a) ¿Se trata de una relación lineal? (analizar el diagrama de dispersión)
- b) Si es una ecuación lineal hallar la ecuación de regresión determinando los coeficientes de regresión.
- c) Predecir la venta de un empleado con 40 años de experiencia (Estimar Y para un valor X=40)

Solución:

- a) Diagrama de dispersión entre los años de experiencia y las ventas mensuales



Según el diagrama de dispersión existe una relación aproximadamente lineal.

- b) Hallando la ecuación de la recta de regresión:

Nº	Experiencia (X)	Ventas(miles) (Y)	XY	X ²	Y ²
1	13	26	338	169	676
2	16	33	528	256	1089
3	30	36	1080	900	1296
4	2	16	32	4	256
5	8	26	208	64	676
6	6	19	114	36	361
7	31	38	1178	961	1444
	$\sum X = 106$	$\sum Y = 194$	$\sum XY = 3478$	$\sum X^2 = 2390$	$\sum Y^2 = 5798$

$$n=7 \quad ; \quad \bar{x} = \frac{\sum x}{n} = \frac{106}{7} = 15.14 ; \quad \bar{y} = \frac{\sum y}{n} = \frac{194}{7} = 27.71$$

$$b_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{3478 - 7(15.14)(27.71)}{2390 - 7(15.14)^2} = 0.688$$

$$b_0 = \bar{y} - b_1 \bar{x} = 27.71 - (0.688)15.14 = 17.29$$

Entonces la ecuación de regresión estimada será:

$$\hat{Y} = 17.29 + 0.688X$$

Interpretación:

$b_0 = 17.29$, Son las ventas de un empleado cuando X es cero o cuando no tienen años de experiencia. En otras palabras un vendedor tendrá en promedio ventas de S/. 17290 si no tiene años de experiencia.

$b_1 = 0.688$, Es el incremento en 0.688 cuando X aumenta en una unidad. En otras palabras es el aumento en S/. 688 por cada año de experiencia que adquiera un empleado.

c) Predicción para X=40

$$\hat{Y} = 17.29 + 0.688(40) = 44.826$$

Un empleado con 40 años de experiencia en ventas, venderá aproximadamente S/.44826

2.3.10. COEFICIENTE DE DETERMINACIÓN (R^2)

Mide el porcentaje de variabilidad en Y que puede explicarse a través del conocimiento de la variable independiente X. Se calcula con la siguiente fórmula:

$$R^2 = \frac{\left[\sum (x - \bar{x})(y - \bar{y}) \right]^2}{\left(\sum x_i^2 - n\bar{x}^2 \right) \left(\sum y_i^2 - n\bar{y}^2 \right)} = r^2$$

Características:

- i) Es un valor no negativo ya que se encuentra entre 0 y 1 $0 \leq R^2 \leq 1$
- ii) Es un valor muy importante en cualquier análisis de regresión, ya que muestra el grado hasta el cual están relacionadas la variabilidad de X e Y

Ejemplo2: Del ejemplo anterior sobre la regresión entre los años experiencia y las ventas:

$$r = 0.939, \text{ por lo cual } R^2 = (0.939)^2 = 0.883$$

Interpretación: El 88.3% de la variación de las ventas (Y) son explicados por los años de experiencia. Existe además un $(100 - 88.3)\% = 11.7\%$ que no es explicado por los años de experiencia.

GUÍA DE PRÁCTICA N°7

ANÁLISIS DE CORRELACIÓN Y REGRESIÓN

1. El Psicólogo de una empresa realiza un estudio para determinar la relación entre los años trabajados de un empleado (X) y su estrés laboral (Y) y hemos obtenido los siguientes resultados:

Años trabajados	Estrés laboral (puntaje)	XY	X ²	Y ²
12	82			
10	74			
06	55			
08	58			
09	62			
04	40			

A) Identifique cual es la variable dependiente y cual la independiente:

X= _____

Y= _____

B) Interpreta el diagrama de dispersión:

C) Calcula e Interpreta los siguientes coeficientes:

$r =$ _____

$R^2 =$ _____

COEFICIENTE DE CORRELACIÓN

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right) \left(\sum_{i=1}^n y_i^2 - n \bar{y}^2 \right)}}$$

COEFICIENTE DE DETERMINACIÓN

$$R^2 = r^2$$

COEFICIENTES DE REGRESION

$$\hat{Y} = b_0 + b_1 X$$

$$b_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

2. Si se tienen la ecuación de regresión: X: Gasto de una familia
Y: Ahorro de una familia

$$y = 500 - 1,7x \quad r = -0,81$$

- a. Interpreta el coeficiente de correlación:
-

- b. Interpreta el coeficiente de determinación
-

- c. Interpreta los coeficientes de la ecuación regresión:

$$b_0 = \underline{\hspace{10cm}}$$

$$b_1 = \underline{\hspace{10cm}}$$

- d. ¿Cuánto se espera que ahorre una familia cuyo gasto es S/. 280?
-

3. El número de horas semanales dedicadas al estudio de una asignatura y la calificación obtenida en el examen correspondiente de 6 alumnos es:

Horas ()	Calificación ()	XY	X ²	Y ²
4	09			
8	15			
6	12			
11	18			
8	13			
2	08			

A) Identifique cual es la variable dependiente y cual la independiente:

X=

Y=

B) Calcula e interpreta los siguientes coeficientes:

r =

R^2 =

C) Estime la ecuación lineal e Interprete los coeficientes de la ecuación de regresión lineal

$$\hat{Y} = b_0 + b_1 X = \underline{\hspace{10cm}}$$

$$b_0 = \underline{\hspace{10cm}}$$

$$b_1 = \underline{\hspace{10cm}}$$

D) Estimar la calificación para una persona que hubiese estudiado 13 horas.

4. Se hace un estudio para determinar la relación entre las edades de un grupo de trabajadores de una fábrica y su rendimiento en el trabajo. Los datos se dan en la siguiente tabla:

N°		Edad (X)	Eficiencia%(Y)	XY	X ²	Y ²
1		22	82			
2		24	76			
3		51	40			
4		39	60			
5		37	64			
6		30	78			
7		46	54			
Total						

A) Calcula e Interpreta los siguientes coeficientes:

$$r = \underline{\hspace{10cm}}$$

$$R^2 = \underline{\hspace{10cm}}$$

B) Estime la ecuación lineal e interprete los coeficientes de la ecuación de regresión lineal

$$\hat{Y} = b_0 + b_1 X = \underline{\hspace{10cm}}$$

$$b_0 = \underline{\hspace{10cm}}$$

$$b_1 = \underline{\hspace{10cm}}$$

C) Estimar la eficiencia de un trabajador con 40 años de edad

5. Una casa de empanadas realiza promociones a través de publicidad repartiendo volantes publicitarios. Se cree que existe relación entre "x" cantidad de volantes repartidos por semana en cientos e "y" ventas en cientos de pesos. Se tomaron 8 semanas al azar que arrojaron la siguiente información:

$$\sum x = 31 \quad \sum y = 81.1 \quad \sum xy = 329.7 \quad \sum x^2 = 135 \quad \sum y^2 = 839.17$$

- a) Determine en qué medida la variable cantidad de volantes explica las variaciones de las ventas.
 b) Determine la recta de ajuste e interprete los coeficientes.
 c) Pronostique el valor las ventas, cuando se reparten 350 volantes.

6. Los siguientes datos corresponden al número de policías destacados por día (durante una semana) y al número de asaltos que se produjeron en cierto distrito.

N° policías en servicio	N° de asaltos	XY	X ²	Y ²
20	54			
40	25			
45	20			
60	12			
32	27			
25	48			
28	42			

A. Identifique cual es la variable dependiente y cual la independiente:

X= _____

Y= _____

B. Interpreta el diagrama de dispersión:

- C. Calcula e Interpreta los siguientes coeficientes:

$$r = \text{_____}$$

$$R^2 = \text{_____}$$

- D. Estime la ecuación lineal e Interprete los coeficientes de la ecuación de regresión lineal

$$\hat{Y} = b_0 + b_1 X = \text{_____}$$

$$b_0 = \text{_____}$$

$$b_1 = \text{_____}$$

- E. Si se destacan 15 policías para protección policial ¿cuántos asaltos se espera que se produzcan?
- _____

7. Pamela Salas, gerente de personal del Textiles Peruvian, está interesada en pronosticar si un aspirante en particular se convertirá en un buen vendedor. Pamela decide emplear las ventas en miles dólares del primer mes y la calificación en una prueba de aptitud. El Gerente de personal reúne todos los datos que se muestran en la siguiente tabla.

Calificación en aptitud	Ventas	XY	X ²	Y ²
10	44			
19	47			
27	58			
31	62			
64	65			

B. Identifique cual es la variable dependiente y cual la independiente:

X= _____

Y= _____

C. Interpreta el diagrama de dispersión:

- D. Calcula e Interpreta los siguientes coeficientes:

$$r = \text{_____}$$

$$R^2 = \text{_____}$$

- E. Estime la ecuación lineal e Interprete los coeficientes de la ecuación de regresión lineal

$$\hat{Y} = b_0 + b_1 X = \text{_____}$$

$$b_0 = \text{_____}$$

$$b_1 = \text{_____}$$

- F. Predecir las ventas de un vendedor con 100 puntos de calificación en la prueba de aptitud.
- _____

8. Se ha estudiado las calificaciones de 100 alumnos en dos asignaturas: Matemática I y Estadística Básica, obteniéndose los siguientes resultados:

$$\sum x_i = 1066 ; \quad \sum y_i = 1230 ; \quad S_x^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n} = 4 ; \quad S_y^2 = 2.25 \quad r = 0.9$$

- a) Hallar la ecuación de recta de regresión de Y sobre X ; y X sobre Y
 - b) Predecir la nota en estadística cuando el alumno saca 17 en matemática I
9. El índice de mortalidad de siete grupos que consumían diariamente cigarrillos aparece en la tabla adjunta:

Nº Cigarrillos(X)	3	5	6	15	20	30	40	45
Tasa mortalidad(Y)	0.2	0.3	0.3	0.5	0.7	0.8	1.4	1.5

- a) Calcule el coeficiente de correlación lineal entre X e Y. Interprete el resultado.
 - b) Obtenga la recta de regresión que explique la tasa de mortalidad en función del número de cigarrillos consumidos. Interpreta los parámetros estimados
 - c) Hallar el coeficiente de determinación e interpretar.
 - d) ¿Qué mortalidad se puede predecir para un consumidor de 60 cigarrillos diarios?
10. En un país europeo se han obtenido estadísticas que relacionan el número de vehículos matriculados y el número de accidentes habidos en un período determinado. Los datos recogidos son los siguientes:

periodo	nº de accidentes	nº de vehículos matriculados
1	166	352
2	153	373
3	177	411
4	201	441
5	216	462
6	208	490
7	227	529
8	238	577

Se pide:

- a) Un modelo de regresión que nos explique el nº de accidentes en función de los vehículos matriculados.
- b) Coeficiente de correlación lineal.
- c) Porcentaje de las causas ajena a la regresión que influyen en la variable dependiente.
- d) Deducir cuál sería el nº de accidentes si se matriculan 800 vehículos.
- e) Estimar el parque de vehículos matriculados para reducir el número de accidentes hasta 175.

11. La información estadística obtenida de una muestra de tamaño 12 sobre la relación existente entre la inversión realizada y el rendimiento obtenido en cientos de miles de euros para explotaciones agrícolas, se muestra en el siguiente cuadro:

Inversión (X)	11	14	16	15	16	18	20	21	14
Rendimiento (Y)	2	3	5	6	5	3	7	10	6

- a) Construya un diagrama de dispersión e interprete
- b) Calcule los coeficientes de correlación y determinación e interprete.
- c) Estima la recta de regresión del rendimiento respecto de la inversión e interprete sus coeficientes.
- d) La previsión del rendimiento que se obtendrá con un inversión de 1 250 000 €.

12. Se realizó un estudio para determinar los efectos de no dormir en la capacidad de las personas para resolver problemas sencillos. Un grupo de personas participó en el estudio que consistió en dar a cada persona, después de un periodo específico sin dormir, un conjunto de problemas sencillos de sumar y se registró el número de errores. Se obtuvieron los siguientes resultados:

Nº de errores	10	13	16	22	25	12
Nº de Horas sin dormir	4	8	13	15	18	10

- a) Identifique cual es la variable dependiente y cual la independiente.
- b) Construya el diagrama de dispersión y calcule el coeficiente de correlación e interprete.
- c) Estima la ecuación de regresión lineal que se ajuste a los datos e Interpreta los coeficientes.
- d) Calcule el coeficiente de determinación e interprete
- e) Si una persona deja de dormir 48 horas ¿cuántos errores se espera que cometa?

13. Se supone que se puede establecer cierta relación lineal entre las exportaciones de un país y la producción interna de dicho país. En el caso de España, tenemos los datos anuales (expresados en miles de millones de pesetas) para tales variables correspondientes al quinquenio 1992-97 en la siguiente tabla:

Años	Producción	Exportaciones
1992	52.6	10.4
1993	53.9	11.8
1994	57.3	14.4
1995	61.8	16.7
1996	65.3	18.7
1997	78.5	23.9

A partir de tal información, y considerando como válida dicha relación lineal, se pide:

- a) Identifique cual es la variable dependiente y cual la independiente.

- b) Construya el diagrama de dispersión y calcule el coeficiente de correlación e interprete.
 - c) Estima la ecuación de regresión lineal que se ajuste a los datos e Interpreta los coeficientes estimados.
 - d) Calcule el coeficiente de determinación e interprete
 - e) Si la producción para el año 1997 fue de 210610 millones de pesetas, ¿cuál sería la predicción de las exportaciones para este año? ¿Qué grado de precisión tendría dicha predicción?
14. La administración de una cadena de farmacias quiere estudiar la rentabilidad de su inversión en publicidad. Para ello ha recogido datos del volumen de ventas de sus medicinas y del gasto en publicidad referidos a los años 2000 - 2005 expresados en millones de soles.

Gasto en publicidad	Ventas de medicinas	XY	X^2	Y^2
3	15			
5	21			
4	20			
6	30			
8	36			
7	32			

- a) Identifique cual es la variable dependiente y cual la independiente.
 - b) Construya un diagrama de dispersión e interprete
 - c) Calcule el coeficiente de correlación e interprete.
 - d) Calcule el coeficiente de determinación e interprete
 - e) Estima la ecuación de regresión lineal e Interpreta los coeficientes estimados.
 - f) Predecir la venta de medicinas si se gasta S/.10 millones en publicidad
15. En el servicio central de turismo de un país se ha observado que el número de plazas hoteleras ocupadas es diferente según sea el precio de la habitación. Sobre el total de plazas ocupadas en un año se tiene:

Precio (dólares/noche)	25	65	100	140	210
Nº habitaciones ocupadas	472	261	187	94	45

- a) Representa gráficamente para comprobar que existe cierta dependencia lineal entre las variables.
- b) Halla la ecuación de la recta de regresión e interpreta los coeficientes.
- c) ¿Cuántas habitaciones se llenarían a 1500 dólares?
- d) ¿En qué medida podemos considerar que el nivel de ocupación depende de la estructura de precios?

UNIDAD 3: PROBABILIDADES Y DISTRIBUCIONES MUESTRALES



<http://bit.ly/1oM8N9p>

Contenido Temático

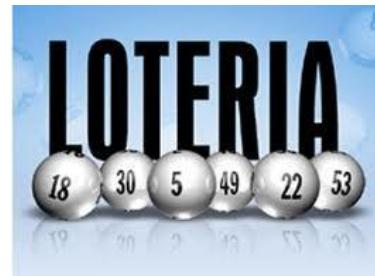
- ❑ Probabilidad básica
- ❑ Probabilidad condicional
- ❑ Distribución de probabilidad continua
- ❑ Estimación de parámetros y tamaño de muestra

3.1. PROBABILIDAD BÁSICA

3.1.1. PROBABILIDAD Y EVENTOS

El concepto de probabilidad es manejado por mucha gente. Frecuentemente se escuchan preguntas como las que se mencionan a continuación:

- ¿Cuál es la probabilidad de que me saque la Lotería?
- ¿Qué posibilidad hay de que me pase un accidente automovilístico?
- ¿Qué posibilidad hay de que hoy llueva? para llevar mi paraguas o no.
- ¿Existe alguna probabilidad de que repreube el primer parcial?



<http://bit.ly/1osZUll>

Las preguntas anteriores esperan como respuesta una medida de confianza para conocer un evento futuro, o bien de una forma sencilla interpretar la probabilidad. El estudio de las probabilidades proporciona las reglas para el estudio de los experimentos aleatorios o de azar, que constituyen la base para la estadística inferencial.

3.1.2. EXPERIMENTO

Es un proceso mediante el cual se obtiene un resultado de una observación. **Ejemplos:**

- Si un producto lanzado al mercado tiene aceptación o no
- Observar artículos defectuosos en un producción
- El lanzamiento de un dado y observar su cara superior.



<http://bit.ly/1PSMjdV>



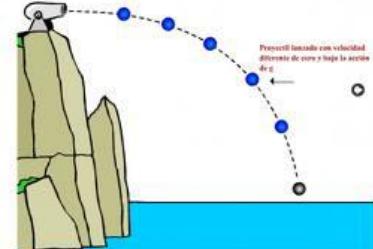
<http://bit.ly/210FKjR>

Un experimento puede ser de dos tipos:

A. EXPERIMENTO DETERMINÍSTICO. Cuando el resultado de la observación se puede predecir con exactitud antes de realizar el experimento.

Ejemplos:

- De cierta altura se deja caer una piedra sin que hay obstáculo alguno entre ella y el suelo. Lógicamente la piedra caerá al suelo por la ley de la gravedad y se puede predecir la distancia que recorrerá en un tiempo dado.
- La hora en que se despierta una persona utilizando un reloj despertador.
- Observar el color de una bola extraída de una urna que contiene solo bolas negras.
- Observar la suma de dos números naturales pares.



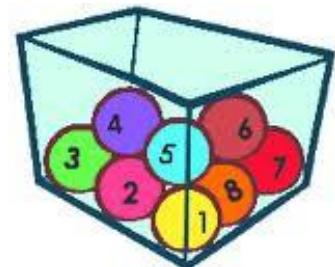
<http://bit.ly/1PSMy8D>

B. EXPERIMENTO NO DETERMINÍSTICO.

Cuando los resultados del experimento no pueden predecirse con exactitud antes de realizar el experimento.

Ejemplos:

- Lanzar un dado y observar el número que aparece en la cara superior
- Lanzar una moneda 8 veces y observar la sucesión de caras y sellos.
- Extraer bolas de una urna que contienen 3 bolas blancas, 2 negras y 4 rojas.
- Elegir un presidente de un grupo de 50 personas.



3.1.3. EXPERIMENTO ALEATORIO.

Llamado también No determinístico y tiene las siguientes características:

- a) Cada experimento podrá ser repetido indefinidamente sin cambiar esencialmente las condiciones.
- b) No se puede determinar un valor “A priori”, sin embargo es posible describir de antemano todos sus resultados posibles.



Ejemplo: todos los juegos de azar

3.1.4. ESPACIO MUESTRAL Y EVENTOS

A. ESPACIO MUESTRAL

Es denotado por Ω , es un conjunto formado por todos los posibles resultados de un experimento.

Ejemplo1: consideremos el experimento aleatorio de lanzar un dado y observar el número que aparece en la cara superior, entonces su espacio muestral será:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Ejemplo2: sea el experimento lanzar dos monedas y observar el resultado, entonces su espacio muestral será: $\Omega = \{cc, cs, sc, ss\}$

B. EVENTO O SUceso

Un evento es un subconjunto del espacio muestral Ω y se denota con las letras mayúsculas A, B, C, etc.

Ejemplo3: sea el experimento de lanzar un dado y observar el número que aparece en la cara superior. El espacio muestral asociado a este experimento como ya sabemos es:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Para este experimento se pueden definir los siguientes eventos:

- A: Observar un número impar. $A = \{1, 3, 5\}$
- B: Observar un número menor que 4. $B = \{1, 2, 3\}$
- C: Observar un número mayor que 2. $C = \{3, 4, 5, 6\}$

Ejemplo4: sea el experimento lanzar tres monedas y observar el resultado, entonces su espacio muestral será: $\Omega = \{ccc, ccs, csc, scc, css, scs, ssc, sss\}$

Se definirán los siguientes eventos:

- D: Observar que las tres monedas salgan sello. $D = \{sss\}$
- E: Observar que al menos 2 monedas sean caras. $E = \{ccc, ccs, csc, scc\}$
- F: Observar que salgan solo 1 sello en cualquiera de las tres. $F = \{ccs, csc, scc\}$

3.1.5. TÉCNICAS DE CONTEO



En muchas situaciones solo nos interesarán el número de elementos que tiene un espacio muestral o un evento particular, en tales situaciones acudiremos a las técnicas de conteo. Las técnicas de conteo son procedimientos o arreglos de enumeración para determinar el tamaño del espacio muestral. Es necesario desarrollar algunas técnicas de enumeración entre las cuales está: el análisis combinatorio.

A. ANÁLISIS COMBINATORIO

Es un procedimiento más sencillo para determinar el número total de resultados. Con este fin, nos apoyaremos en los conceptos permutaciones y combinaciones, los cuales tienen como base el principio fundamental del conteo.

A1. PERMUTACIONES

Una permutación de un conjunto de elementos, es un ordenamiento específico de todos o algunos elementos del conjunto, facilita el recuento de las ordenaciones diferentes que pueden hacerse con los elementos del conjunto. En una permutación el orden en que se disponen los elementos del conjunto es importante.

A2. PERMUTACIONES DE n ELEMENTOS

Por el principio fundamental del conteo podemos enunciar que el número de permutaciones de n objetos distintos tomados de n en n , es: $P_n = n!$

El símbolo $!$ se lee factorial y es el producto resultante de todos los enteros positivos de 1 a n ; es decir, sea n un número entero positivo, el producto $n(n-1)(n-2)\dots 3 \times 2 \times 1$ se llama factorial de n .

- $n! = n(n-1)(n-2)\dots 3 \times 2 \times 1$
- $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$
- Por definición $0! = 1$



Ejemplo1: se quiere conocer el conjunto de todas las disposiciones posibles de tres personas sentados en una banca. $P_3 = 3! = 6$ entonces $\Omega = \{abc, acb, bac, bca, cab, cba\}$

Ejemplo2: cinco personas desean nombrar un Comité Directivo compuesto de un presidente, un vicepresidente, un secretario, un tesorero y un vocal. ¿Cuántas maneras hay de constituir el comité? $P_5 = 5! = 120$

Ejemplo3: hay seis banderas de distintos colores. ¿Cuántas señales diferentes se pueden enviar usando las seis banderas al mismo tiempo? $P_6 = 6! = 720$

A3. PERMUTACIONES CIRCULARES.

Cuando los elementos se encuentran dispuestos en forma circular tenemos:

$${}_n P_c = (n - 1)!$$



Ejemplo 4: ¿De cuántas maneras se pueden sentar 6 personas alrededor de una mesa circular?

$${}_6 P_c = (6 - 1)! = 5! = 120$$

B. COMBINACIONES

Una combinación es un subconjunto o una disposición de todos los elementos de un conjunto, sin tener en cuenta el orden de ellos. El número de combinaciones o subconjuntos no ordenados, cada uno formado por r elementos, que pueden obtenerse de un conjunto de n elementos es:

$$nCr = \frac{n!}{(n-r)!r!} \circ \binom{n}{r} = \frac{n!}{(n-r)!r!}$$

Ejemplo5: se tienen cinco obreros para un trabajo especial que requiere de tres de ellos.
¿De cuántas maneras diferentes se puede seleccionar un equipo de tres?

$$5C_3 = \frac{5!}{(5-3)!3!} = \frac{5!}{2!3!} = \frac{120}{12} = 10$$

Ejemplo6: de un club de 20 socios, se van a seleccionar 3 para formar la mesa directiva.
¿De cuántas formas puede constituirse?

$$20C_3 = \frac{20!}{(20-3)!3!} = \frac{20!}{17!3!} = \frac{6840}{6} = 1140$$

3.1.6. CONCEPTO CLÁSICO DE PROBABILIDAD

Dado un evento A , asociado a un experimento aleatorio, se llama probabilidad de A , y se representa por el símbolo $P(A)$, al cociente que se obtiene dividiendo el número de resultados favorables para la ocurrencia del evento, entre el número total de posibilidades o número de elementos del espacio muestral (Ω). Y se denota por:

$$P(A) = \frac{\#(A)}{\#(\Omega)}$$

Ejemplo7: en el experimento de lanzar un dado,

- Determinar la probabilidad de que en la cara superior aparezca el número 5
 $\Omega = \{1, 2, 3, 4, 5, 6\}$, 6 resultados posibles

Sea el evento $A = \{5\}$, un resultado favorable $P(A) = \frac{1}{6} = 0,167$

- Determinar la probabilidad de que se obtenga un número par

Sea el evento $B = \{2, 4, 6\}$, 3 resultados favorables $P(B) = \frac{3}{6} = \frac{1}{2} = 0,5$

A. PROPIEDADES DE LAS PROBABILIDADES

Sean los eventos A y B asociados al espacio muestral Ω . Entonces se cumplen las siguientes propiedades.

- i. $0 \leq P(A) \leq 1$
- ii. $P(\Omega) = 1$ es decir, la probabilidad del suceso seguro Ω , es igual a la unidad.
- iii. $P(\emptyset) = 0$
- iv. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- v. $P(A \cup B) = P(A) + P(B)$ para todo A y B eventos disjuntos ($A \cap B = \emptyset$)

Ejemplo 8: en cierta ciudad, la probabilidad que una familia tenga televisor es 0.75, un refrigerador es 0.60 y que tengan ambos es 0.50 ¿Cuál es la probabilidad que una familia tenga un refrigerador o un televisor?

Solución: $P(T) = 0.75$ $P(R) = 0.60$ $P(T \cap R) = 0.50$

$$P(T \cup R) = 0.75 + 0.60 - 0.50 = 0.85$$

GUÍA DE PRÁCTICA N°8

PROBABILIDAD BÁSICA

I. PROBABILIDAD DE UN EVENTO

1. Los accidentes en una empresa, que se dedica a la fabricación de correas para damas, se clasificaron de acuerdo con la zona del daño en: 1=Dedos 2=Ojos 3=Brazos 4=Piernas; A continuación se tiene una muestra con los siguientes resultados:

Zona de lesión	n_i	$P(A)$
Dedos	6	
Ojos	10	
Brazos	16	
Piernas	8	
Total		

$$P(A) = \frac{\#(A)}{\#(\Omega)}$$

- a) ¿Cuál es la probabilidad que un trabajador tenga una lesión en los ojos? _____
- b) ¿Cuál es la probabilidad que tenga una lesión en los dedos? _____
- c) ¿Cuál es la probabilidad que tenga una lesión en brazos o piernas? _____
- d) ¿Cuál es la probabilidad que tenga una lesión en dedos y brazos? _____
2. En una tienda de electrodomésticos se realiza una encuesta a un grupo de empleados para determinar el número de televisores que han vendido:

Número de televisores	n_i	$P(A)$
1	4	
2	10	
3	16	
4	14	
5	6	
Total		

- a) ¿Cuál es la probabilidad que un empleado venda 4 televisores? _____
- b) ¿Cuál es la probabilidad que venda 2 ó 3 televisores? _____
- c) ¿Cuál es la probabilidad que venda menos de 4 televisores? _____
- d) ¿Cuál es la probabilidad que venda al menos 3 televisores? _____
3. En una capacitación empresarial, hubo 99 asistentes enfermos de gripe (E) entre las 158 personas que acudieron a la capacitación. Calcule la probabilidad de que una persona seleccionada al azar:
- a) se encuentre enferma. $P(E) = \text{_____} =$
- b) se encuentre sana $P(S) = \text{_____} =$

II. PROPIEDAD DE LA ADICIÓN

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

4. En cierta hospital la probabilidad de que un paciente esté enfermo del corazón es de 0.35, la probabilidad que se fumador 0.75 y la probabilidad de que este enfermo del corazón y sea fumador es de 0.45 ¿Cuál es la probabilidad de que un paciente seleccionado aleatoriamente de esta población, sea fumador ó este enfermo del corazón?

$$P(C) = \underline{\hspace{2cm}} \quad P(F) = \underline{\hspace{2cm}} \quad P(C \cap F) = \underline{\hspace{2cm}} \quad P(C \cup F) = \underline{\hspace{2cm}}$$

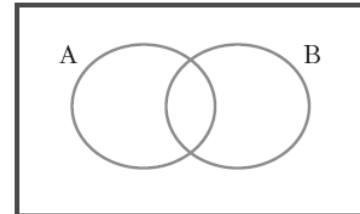
5. En una entidad bancaria, la probabilidad que un directivo tenga título de economista es 0.46, que tenga título contable es 0.34 y que tengan alguno de los dos 0.62 ¿Cuál es la probabilidad que un directivo elegido al azar tenga ambos títulos?

$$P(E) = \underline{\hspace{2cm}} \quad P(C) = \underline{\hspace{2cm}} \quad P(\dots\dots\dots\dots) = \underline{\hspace{2cm}} \quad P(\dots\dots\dots\dots) = \underline{\hspace{2cm}}$$

III. OPERACIONES CON EVENTOS

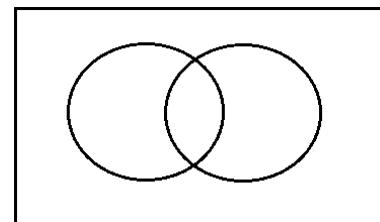
6. El 60% de la población de una determinada ciudad lee el periódico A, el 35% el B y un 15% ambos. Elegido un ciudadano al azar, calcular la probabilidad de:

- a) Ser lector de algún periódico
- b) No leer ninguno
- c) Leer solo el periódico A
- d) Leer solo uno de los dos periódicos



7. De 120 estudiantes de una escuela 24 estudian biología, 80 estudian física y 12 estudian ambas materias. Si se selecciona un estudiante al azar, encontrar las siguientes probabilidades:

- a) No estudie biología
- b) No estudie física
- c) Estudie biología o física
- d) Estudie biología pero no física
- e) No estudie ni biología ni física



8. Una pareja al planificar una familia está interesada en tener 4 hijos, de acuerdo a esto determine los siguientes eventos:

A: Todos los hijos del mismo sexo $A = \{ \dots \}$

B: Exactamente un varón $B = \{ \dots \}$

C: Por lo menos dos varones $C = \{ \dots \}$

9. Suponga que el siguiente espacio muestral $S = \{a, b, c, d\}$ es un espacio equiprobable. Se definen los siguientes eventos: $A = \{a, b\}$, $B = \{a, c\}$, $C = \{a, d\}$. Calcula las probabilidades:

$$a. P(A \cup B) = \underline{\hspace{2cm}} = b. P(B \cap C) = \underline{\hspace{2cm}} =$$

IV. PROBABILIDADES CON COMBINATORIAS

10. De una baraja de 52 naipes bien mezclada se sacan 5 naipes. Hallar la probabilidad de que:

- a) 4 sean ases:
c) 3 sean dieces y 2 Jotas

- b) 4 sean ases y 1 rey;
d) 2 sean nueves y 2 ases

11. De una baraja de 52 cartas se extraen al azar 6 cartas. Determinar la probabilidad que 3 de ellas sean espadas y dos tréboles.

12. De 12 personas que contraen influenza al mismo tiempo, 9 se recuperan en 5 días. Suponga que pasados los 5 días se escogen 3 personas al azar de las 12.

Calcular la probabilidad de que:

- a) Las tres se hayan recuperado.
b) Exactamente dos se hayan recuperado.
c) Ninguna se haya recuperado.

13. Se elige aleatoriamente de una baraja de 52 cartas y se pide:

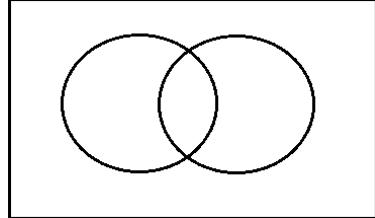
- a. ¿Cuál es la probabilidad que sea una carta negra? _____
b. ¿Cuál es la probabilidad que sea un diez? _____
c. ¿Cuál es la probabilidad que sea un cuatro o menos? _____

14. Una caja contiene 12 bolas negras y 8 rojas, ¿qué probabilidad hay de **no** sacar una bola negra?

15. Se lanzó un dado honesto –no cargado- dos veces, obteniéndose 4 en ambas oportunidades
¿Cuál es la probabilidad de que en un tercer lanzamiento se obtenga nuevamente el cuatro?

16. Hay 16 monedas de \$ 100; 22 monedas de \$ 50 y 12 monedas de \$ 10. Al sacar una moneda, ¿cuál es la probabilidad de sacar una de \$ 100 o de \$ 50?

17. Se tiran dos dados al mismo tiempo: a) ¿Cuál es la probabilidad de que salgan dos cincos? b)
¿Cuál es la probabilidad de que la suma de los números sea 10?

18. De un grupo de personas, el 30% practica futbol y el 40% ajedrez. De los futbolistas el 50% juega ajedrez. Si se elige aleatoriamente una persona ¿Cuál es la probabilidad que
- Juegue futbol o ajedrez? _____
 - Practica solo uno de estos deportes? _____
 - No practica ni futbol ni ajedrez? _____
- 
19. En un almacén hay 12 artículos de los cuales 5 son de la marca A y 7 de la marca B; si se extraen 4 artículos al azar, calcule la probabilidad de que:
- Todos los artículos sean de la Marca A b) 2 artículos sean de Marca A y 2 de B
 - 3 artículos sean de la Marca B y uno de A
20. De 20 personas que contrajeron cierta enfermedad al mismo tiempo y que fueron llevados a una misma sala del hospital, 15 se recuperan completamente en 15 días; al cabo del cual, se escogen aleatoriamente 5 personas para un chequeo
- ¿Cuál es la probabilidad que los 5 sean dados de alta?
 - ¿Cuál es la probabilidad de que exactamente 4 sean dados de alta?
 - ¿Cuál es la probabilidad que ninguno sea dado de alta?
21. En una urna son mezcladas diez bolas numeradas del 1 al 10. Dos bolas (a, b) son retiradas sin reposición ¿Cuál es la probabilidad de que $a + b = 10$?
22. Se reúne el comité directivo de un club de fútbol para decidir si despiden o no al entrenador. Cinco quieren despedirlo y tres no quieren. Viene un reportero e interroga a dos de ellos al azar. ¿Cuál es la probabilidad de que los dos opinen que hay que despedirlo? ¿Cuál es la probabilidad de que los dos opinen que hay que renovarle el contrato?
23. En cierta ciudad, la probabilidad que una familia tenga televisor es 0.85, un refrigerador es 0.60 y que tengan ambos es 0.50 ¿Cuál es la probabilidad que una familia tenga un refrigerador o un televisor?
24. Para obtener licencia para conducir es necesario aprobar tanto el examen teórico como el práctico. Se sabe que la probabilidad que un alumno apruebe la parte teórica es 0,68, la de que apruebe la parte práctica es 0,72 y la de que haya aprobado alguna de las dos partes es 0,82. Si se elige un alumno al azar, ¿cuál es la probabilidad de que apruebe el examen para obtener licencia?
25. Jorge se presenta a dos universidades A y B. El estima la probabilidad que sea admitido en la universidad A en 0.8; a la universidad B en 0.75, en al menos una de ellas en 0.95 ¿Cuál es la probabilidad que ingrese a ambas universidades?

3.2. PROBABILIDAD CONDICIONAL

3.2.1. CONCEPTO

Ocurre cuando dos eventos se relacionan de manera tal que la probabilidad de ocurrencia de uno depende de la ocurrencia del otro.

Definición: sean A y B dos eventos tal que $P(B) > 0$, la probabilidad condicional de que ocurra el evento A dado que ha ocurrido B, se denota:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Ejemplo 1: en una ciudad el 55% de los habitantes consume arroz, el 30% consume trigo y el 20% consume ambos. Se pide: A= Arroz T= Trigo

a) ¿Cuál es la probabilidad de que coma trigo dado que consumió arroz?

$$P(T/A) = \frac{P(T \cap A)}{P(A)} = \frac{0.20}{0.55} = 0.364$$

b) Sabiendo que un habitante consume trigo, ¿cuál es la probabilidad de que consuma arroz?

$$P(A/T) = \frac{P(T \cap A)}{P(T)} = \frac{0.20}{0.30} = 0.667$$

Ejemplo 2: En una universidad el 70% de los estudiantes son de ciencias y el 30% de letras; de los estudiantes de ciencias el 60% son varones y los de letras son varones el 40%. Si se elige aleatoriamente un estudiante, calcular la probabilidad que:

- a) Sea un estudiante varón
- b) Sea un estudiante varón si es de ciencias
- c) Sea un estudiante de ciencias si es varón
- d) Sea un estudiante de ciencias y varón

Solución: Ordenando los datos en una tabla:

Esp/Sexo	Varones	Mujeres	Total
Ciencias	0.42	0.28	0.70
Letras	0.12	0.18	0.30
Total	0.54	0.46	1.00

a) $P(V) = 0.54$

c) $P(C/V) = \frac{P(C \cap V)}{P(V)} = \frac{0.42}{0.54} = 0.778$

b) $P(V/C) = \frac{P(V \cap C)}{P(C)} = \frac{0.42}{0.70} = 0.6$ d) $P(V \cap C) = 0.42$

3.2.2. REGLA DE MULTIPLICACIÓN

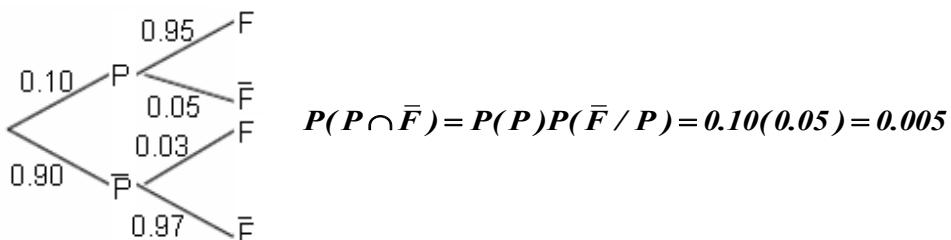
Se define a partir de la probabilidad condicional:

Definición: sean A y B dos eventos cualesquiera, entonces se tiene:

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

Ejemplo 3: en un sistema de alarma, la probabilidad de que se produzca un peligro es 0.10. Si este se produce, la probabilidad de que la alarma funcione es de 0.95. La probabilidad que funcione la alarma sin haber habido peligro es 0.03. Determinar la probabilidad que haya un peligro y la alarma no funcione.

Solución:

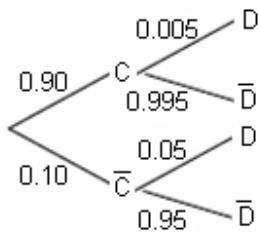


3.2.3. TEOREMA DE LA PROBABILIDAD TOTAL

Definición: Sea A_1, A_2, \dots, A_n una partición del espacio muestral Ω tal que: $P(A_i) > 0$, entonces para cualquier evento B de Ω se tiene:

$$P(B) = \sum_{i=1}^n P(A_i)P(B/A_i) = P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + \dots + P(A_n)P(B/A_n)$$

Ejemplo 4: se conoce que cierta máquina que produce tornillos trabaja correctamente el 90% del tiempo. Si la máquina no está trabajando correctamente, el 5% de los tornillos producidos son defectuosos. Cuando está trabajando bien solamente el 0.5% de tornillos son defectuosos. Si se escoge un tornillo aleatoriamente ¿Cuál es la probabilidad que sea defectuoso?



$$P(D) = P(C)P(D/C) + P(\bar{C})P(D/\bar{C}) \\ = 0.90(0.005) + 0.10(0.05) = 0.0095$$

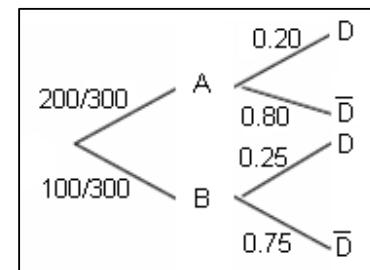
3.2.4. TEOREMA DE BAYES

Si los eventos A_1, A_2, \dots, A_n forman una partición del espacio muestral Ω , y B un evento cualquiera de Ω , entonces:

$$P(A_i / B) = \frac{P(A_i)P(B / A_i)}{\sum_{i=1}^n P(A_i)P(B / A_i)}$$

Ejemplo 1: en una línea de producción hay dos procesos A y B. En el proceso A hay un 20% de defectuosos y en B 25%. En una muestra de 300 productos hay 200 de A y 100 del B.

- a) Si se extrae un producto al azar, hallar la probabilidad que sea defectuoso.
- b) Si al extraer el producto resultó defectuoso, halle la probabilidad de que sea del proceso A.



Solución: sean los siguientes eventos:

- | | |
|-----------------------------------|--|
| A: "El producto es del proceso A" | B: "El producto es del proceso B" |
| D: "El Producto es defectuoso" | \bar{D} : "El Producto es no defectuoso" |

- a) Aplicaremos Teorema de la Probabilidad Total:

$$P(D) = P(A)P(D/A) + P(B)P(D/B) = \frac{200}{300}(0.20) + \frac{100}{300}(0.25) = 0.217$$

- b) Aplicando Teorema de Bayes:

$$P(A/D) = \frac{P(A)P(D/A)}{P(A)P(D/A) + P(B)P(D/B)} = \frac{(2/3)(0.2)}{(2/3)(0.2) + (1/3)(0.25)} = 0.615$$

GUÍA DE PRÁCTICA N°9

PROBABILIDAD CONDICIONAL

I. METODO TABULAR

1. En la siguiente tabla de contingencia se muestra La calificación final de los estudiantes de un curso de estadística según las carreras de Psicología e Ingeniería

Condición/Carrera	Psicología	Ingeniería	Total
Sobresaliente	8	12	
Aprobado	30	25	
Desaprobado	10	6	
Total			

Si se escoge un estudiante al azar:

- a. ¿Cuál es la probabilidad que el estudiante este aprobado? $P(A)=$ _____
- b. ¿Cuál es la probabilidad que estudie Psicología? $P(P)=$ _____
- c. ¿Cuál es la probabilidad que sea sobresaliente y este en Ingeniería?
- d. ¿Cuál es la probabilidad que este desaprobado o estudie para ser psicólogo?
- e. ¿Cuál es la probabilidad que un estudiante este aprobado si es de Ingeniería?
- f. ¿Cuál es la probabilidad que el estudiante sea de Psicología dado que es sobresaliente?

2. Una cafetería quiere conocer la opinión de sus clientes con respecto al servicio prestado y la calidad de sus productos. A cada cliente se le entrega un cuestionario para que lo conteste. De este cuestionario se seleccionó en forma aleatoria la pregunta de calidad del servicio. Los resultados obtenidos de la primera semana se muestran en el cuadro adjunto.

Cliente	Calidad del servicio			Total
	Buena	Regular	Mala	
Adulto (26-45 años)	25	21	8	
Joven (18-25 años)	20	18	6	
Total				

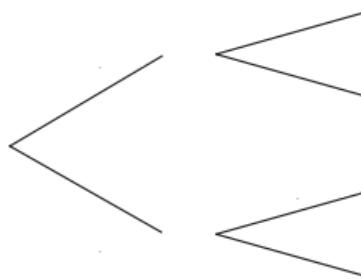
Si se escoge un empleado al azar:

- a. ¿Cuál es la probabilidad que el cliente sea joven? $P(J)=$ _____
- b. ¿Cuál es la probabilidad que el cliente indique un regular servicio? $P(R)=$ _____

- c. ¿Cuál es la probabilidad que el cliente sea joven y opine que el servicio es malo?
- d. ¿Cuál es la probabilidad que el cliente sea Adulto o haya opinado que el servicio es bueno?
- e. ¿Cuál es la probabilidad de que indique que el servicio es bueno si el cliente es joven?
- f. ¿Cuál es la probabilidad que sea Adulto, dado que indica que el servicio es malo?

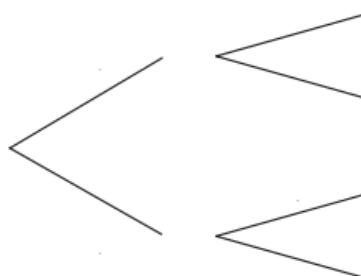
II. DIAGRAMA DEL ARBOL (TEOREMA DE BAYES)

3. Una empresa tiene a dos recepcionistas para atender a sus clientes: María y Carmen. En cierto día el 44% del total de clientes fueron atendidos por María y el resto por Carmen. Además se dispone de un registro de quejas por la atención recibida: 10% se quejaron de María y 12% de Carmen. Si se elige un cliente atendido al azar:



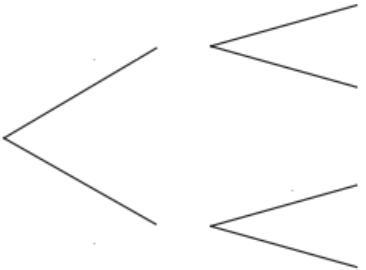
- a) ¿Cuál es la probabilidad de que el cliente se queje por la atención recibida?
 $P(Q) =$
- b) ¿Cuál es la probabilidad de que el cliente no se queje por la atención?
 $P(\bar{Q}) =$
- c) ¿Cuál es la probabilidad que el cliente fue atendido por María si es que hubo queja?
- d) ¿Cuál es la probabilidad que el cliente fuera atendido por Carmen si es que no hubo queja?

4. La probabilidad de que haya un accidente en una fábrica que dispone de alarma es 0.10. La probabilidad de que esta funcione sí se ha producido algún incidente es de 0.88 y la probabilidad de que funcione si no ha sucedido ningún incidente es 0.05.



- a) ¿Cuál es la probabilidad de que funcione la alarma?
 $P(F) =$
- b) ¿Cuál es la probabilidad de que no funcione la alarma?
 $P(\bar{F}) =$
- c) ¿Cuál es la probabilidad que haya un accidente en la fábrica si no funciona la alarma?
- d) En el supuesto de que haya funcionado la alarma, ¿cuál es la probabilidad de que no haya habido ningún incidente?

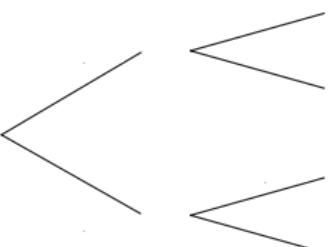
5. En una industria de acero se estima que hay un 60% de probabilidad de que las inversiones de capital aumenten en el próximo año. Si el próximo año hay un aumento en las inversiones de capital, la probabilidad de que el acero para construcciones suba de precios es de 90%. Si no hay incremento en dichas inversiones, la probabilidad de un aumento en los precios del acero es de 40%.

- 
- a) ¿Cuál es la probabilidad de que no suban los precios del acero?
 - b) ¿Cuál es la probabilidad de que suban los precios?
 - c) ¿Cuál es la probabilidad de que haya un aumento de la inversión de capital si es que no suben los precios del acero para construcciones?
 - d) Si sabemos que subieron los precios del acero ¿Cuál es la probabilidad que no aumenten las inversiones?

6. De una población de alumnos de primaria y secundaria se seleccionó una muestra de 660 estudiantes para medir su nivel de conocimientos en cuidados del medio ambiente. Los resultados se muestran a continuación:

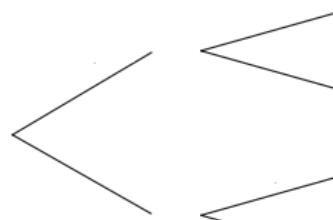
<u>Grado</u>	Nivel de Conocimientos			Total
	<u>Alto</u>	<u>Medio</u>	<u>Bajo</u>	
Primaria	120	170	86	
Secundaria	34	115	135	
Total				

Si se escoge un alumno al azar:

- a. ¿Cuál es la probabilidad que tenga un nivel medio de conocimientos? $P(M)=$
 - b. ¿Cuál es la probabilidad que sea de secundaria y tenga bajo conocimiento?
 - c. ¿Cuál es la probabilidad que sea de primaria o tenga alto conocimiento?
 - d. ¿Cuál es la probabilidad que tenga bajo conocimiento si es de secundaria?
 - e. ¿Cuál es la probabilidad que sea de primaria dado que tiene alto conocimiento?
7. Un monedero contiene 2 monedas de plata y 3 de cobre, y otro contiene 4 de plata y 3 de cobre. Si se elige un monedero al azar y se extrae una moneda ¿cuál es la probabilidad de que sea de plata?
- 

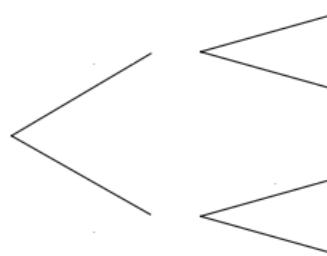
8. Un hombre toma un microbús o un ómnibus para ir a su trabajo, con probabilidades 0.3 y 0.7 respectivamente. 30% de las veces que toma el microbús llega tarde al trabajo, mientras que 20% de las veces que toma el ómnibus llega tarde a su trabajo.

- a) ¿Cuál es la probabilidad que llegue tarde a su trabajo?
- b) Si el hombre llega tarde a su trabajo un día particular, ¿Cuál es la probabilidad de que haya tomado el microbús?
- c) ¿Cuál es la probabilidad que llegue temprano?
- d) Dado que el hombre llega temprano a su trabajo ¿Cuál es la probabilidad de que haya tomado el Ómnibus?

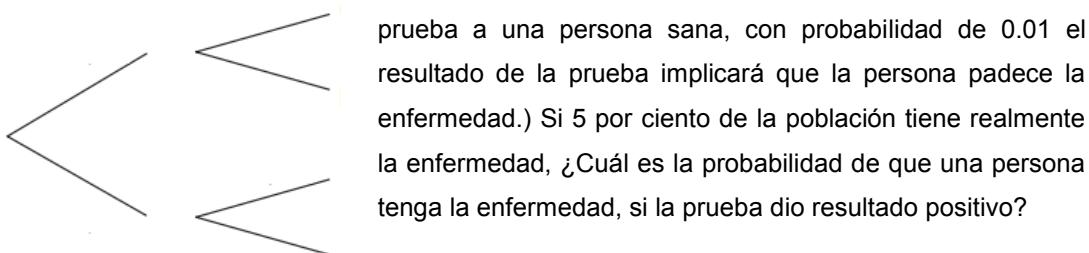


9. En un supermercado el 70% de las compras las realizan las mujeres; de las compras realizadas por estas, el 80% supera los S/. 2000, mientras que de las compras realizadas por hombres sólo el 30% supera esa cantidad.

- a) Elegido un ticket de compra al azar, ¿cuál es la probabilidad de que supere los S/. 2000?
- b) Si se sabe que el ticket de compra no supera los S/. 2000. ¿cuál es la probabilidad de que la compra haya sido hecha por una mujer?



10. Una prueba de sangre de laboratorio es 99 por ciento efectiva para detectar una cierta enfermedad cuando ocurre realmente. Sin embargo, la prueba también da un resultado “positivo falso” en 1 por ciento de las personas sanas a las que se les aplica. (Es decir, si se le hace la prueba a una persona sana, con probabilidad de 0.01 el resultado de la prueba implicará que la persona padece la enfermedad.) Si 5 por ciento de la población tiene realmente la enfermedad, ¿Cuál es la probabilidad de que una persona tenga la enfermedad, si la prueba dio resultado positivo?



11. En un colegio hay dos grupos de 25 alumnos de quinto curso y dos grupos de 20 alumnos de sexto curso. El 50 % de los alumnos de quinto no tienen faltas de ortografía, porcentaje que sube a 70% en los alumnos de sexto. En un concurso de redacción entre alumnos de quinto y sexto se Elije una redacción al azar.

- a) ¿Qué probabilidad hay de que sea de un alumno de quinto?
- b) Si tiene faltas de ortografía, ¿qué probabilidad hay de que sea de un alumno de quinto?

12. En un curso integrado por 16 damas y 14 varones, se sabe que 10 damas y 12 varones prefieren Coca Cola y el resto Sprite. Si elegimos un estudiante al azar, ¿Cuál es la probabilidad de que ese estudiante sea varón y prefiera Sprite? (use diagrama del árbol)

- A) 2/30 B) 6/30 C) 12/30 D) 2/14 E) 12/14

13. Se sabe que en determinado periodo invernal el 30% de la población escolar contrae gripe. Una campaña de vacunación alcanza una cobertura del 70% de esta población. Si de los vacunados, solo el 10% contrae gripe, ¿Cuál es la probabilidad que un escolar contraiga gripe? (use diagrama del árbol)

- A) 28% B) 21% C) 16% D) 30% F) 63%

14. En un viaje organizado por Europa para 120 personas, 48 de los que van saben hablar inglés, 36 saben hablar francés, y 12 de ellos hablan los dos idiomas. Escogemos uno de los viajeros al azar.

	HABLAN FRANCÉS	NO HABLAN FRANCÉS	
HABLAN INGLÉS	12	36	48
NO HABLAN INGLES	24	48	72
	36	84	120

- a. ¿Cuál es la probabilidad de que hable alguno de los dos idiomas?
 b. ¿Cuál es la probabilidad de que hable francés, sabiendo que habla inglés?
 c. ¿Cuál es la probabilidad de que solo hable francés?

15. En una cadena de televisión se hizo una encuesta a 2 500 personas para saber la audiencia de un debate y de una película que se emitieron en horas distintas: 2 100 vieron la película, 1 500 vieron el debate y 350 no vieron ninguno de los dos programas. Si elegimos al azar a uno de los encuestados:

	DEBATE	NO DEBATE	
PELÍCULA	1 450	650	2 100
NO PELÍCULA	50	350	400
	1 500	1 000	2 500

- a. ¿Cuál es la probabilidad de que viera la película y el debate?
 b. ¿Cuál es la probabilidad de que viera la película, sabiendo que no vio el debate?
 c. Sabiendo que vio la película, ¿cuál es la probabilidad de que viera el debate?

16. El 20% de los empleados de una empresa son ingenieros y otro 20% son economistas. El 75% de los ingenieros ocupan un puesto directivo y el 50% de los economistas también, mientras que los no ingenieros y los no economistas solamente el 20% ocupa un puesto directivo. ¿Cuál es la probabilidad de que un empleado directivo elegido al azar sea ingeniero?

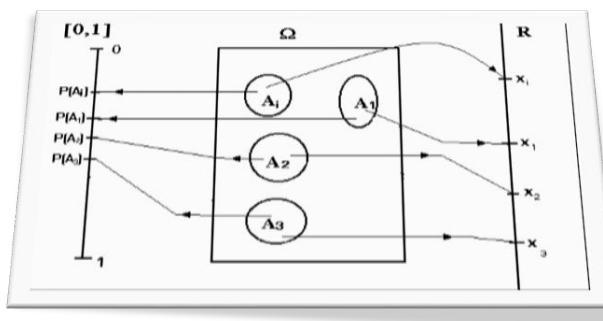
17. En cierta universidad 20% de los hombres y 1% de las mujeres miden más de 1.80m de altura. Asimismo 40% de los estudiantes son mujeres. Si se selecciona un estudiante al azar y se observa que mide más de 1.80m ¿Cuál es la probabilidad de que sea mujer?

18. En una ciudad el 30% de las personas son conservadores, el 50% son liberales y el 20% son independientes. Los registros muestran que en las últimas elecciones votaron el 65% de los conservadores, el 82% de los liberales y el 50% de los independientes. Si se selecciona al azar una persona de la ciudad y se sabe que no votó en las elecciones pasadas, ¿cuál es la probabilidad de que sea un liberal?

3.3. DISTRIBUCIÓN DE PROBABILIDAD CONTINUA

3.3.1. VARIABLE ALEATORIA

Muchas veces se desea resumir con un número el resultado de un experimento aleatorio. En muchos de los ejemplos relativos a experimentos aleatorios que han sido considerados hasta ahora, el espacio muestral es solo una descripción de los posibles resultados. En algunos casos tales descripciones son suficientes, pero en otros se hace útil asociar un número con cada resultado del espacio muestral. Es así como se llega a la definición de **variable aleatoria**.



<http://bit.ly/1RRKuCl>

Definición: una **variable aleatoria X** es una función que asigna un número real a cada resultado en el espacio muestral Ω de un experimento aleatorio. El conjunto de los posibles valores de la variable aleatoria X se denomina rango.

Ejemplo1: Sea el experimento aleatorio “Lanzar una moneda 2 veces y observar su resultado”, entonces $\Omega = \{ss, cs, sc, cc\}$ y sea X la variable “Nº de caras obtenidas” donde X es una función definida sobre Ω .

$$\begin{array}{cccc} \Omega = \{ss, & cs, & sc, & cc\} \\ \downarrow & \downarrow & \downarrow & \downarrow \\ x(ss) & x(cs) & x(sc) & x(cc) \\ 0 & 1 & 1 & 2 \end{array}$$

Entonces $X(w)$ tiene como dominio el espacio muestral Ω y como rango

$$\begin{aligned} R_x = \{x / x = 0, 1, 2\} ; \text{ en símbolos } X: \Omega &\rightarrow \{0, 1, 2\} \\ w \rightarrow X(w) \end{aligned}$$

3.3.2. TIPOS DE VARIABLES ALEATORIAS

- **Variable aleatoria discreta:** una variable aleatoria es discreta si su recorrido es un conjunto discreto. Es decir, un conjunto que se puede enumerar.
- **Variable aleatoria continua:** una variable aleatoria es continua si su recorrido no es un conjunto numerable. Intuitivamente esto significa que el conjunto de posibles valores de la variable abarca todo un intervalo de números reales.

3.3.3. VARIABLE ALEATORIA DISCRETA

Se llama variable aleatoria discreta si el rango de la variable aleatoria X es un conjunto finito o infinito numerable.

$$R_x = \{x_1, x_2, x_3, \dots\}$$

Ejemplo2: en 100 días de trabajo, los records de los empleados se marcan cada día que ellos están ausentes del trabajo. Se selecciona aleatoriamente un record y se observa los días marcados. Hallar R_x

$$R_x = \{0, 1, 2, 3, \dots, 100\}$$

i. FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD

La distribución de probabilidad de X será la descripción del conjunto de valores posibles de X (rango de X), junto con la probabilidad asociada con cada uno de estos valores. Se denota por:

$$p(x) = p[X=x] = \sum P[\{w\}]$$

A esta función de probabilidad también se le conoce como “**función de cuantía**” de X . Diremos que la función $p(x)=P(X=x)$ que va del conjunto de valores posibles de la variable aleatoria X al intervalo $[0, 1]$ es la **función distribución de probabilidad** para X si y solo si se satisfacen las siguientes propiedades:

- $0 \leq p(x) \leq 1 \quad \forall x \in R_x$
- $\sum_{x \in R_x} p(x) = 1$
- $p(x) = 0 \quad \text{si } x \neq x_j, \quad j=1, 2, 3, \dots$

ii. REPRESENTACIÓN TABULAR DE LA FUNCIÓN DE PROBABILIDAD

Se denomina distribución de probabilidad al conjunto de pares $(x; p(x)) \quad \forall x \in R_x$

X	x_1	x_2	x_k
$p(x)$	$p(x_1)$	$p(x_2)$	$p(x_k)$

Ejemplo 3: en el experimento anterior “Lanzar una moneda 2 veces y observar su resultado” y siendo X una variable aleatoria “Nº de caras obtenidas”.

a) Describe el dominio de X : $\Omega = \{ss, cs, sc, cc\}$

b) Hallar el rango de X : $R_x = \{0, 1, 2\}$

c) Hallar la distribución de probabilidad en forma tabular

X	0	1	2
$p(x)$	1/4	2/4	1/4

iii. FUNCIÓN DE DISTRIBUCIÓN ACUMULADA DE UNA VARIABLE ALEATORIA DISCRETA

Se denota por $F(x)$, $x \in \mathbb{R}$ y es dado por:

$$F(x) = P[X \leq x] = \sum_{x_i \leq x} p(x_i)$$

Ejemplo 4: En el ejemplo 3 hallar la función de Distribución Acumulada $F(x)$:

Solución: su función de distribución acumulada será:

X	0	1	2
$p(x)$	1/4	2/4	1/4
$F(x)$	1/4	3/4	1

PROPIEDADES: Para todo $a, b \in \mathbb{R}$ se tiene

- a) $P[X > b] = 1 - P[X \leq b]$
- b) $P[X < b] = P[X \leq b] - P[X = b]$
- c) $P[a < X \leq b] = P[X \leq b] - P[X \leq a]$
- d) $P[a \leq X \leq b] = P[X \leq b] - P[X < a]$
- e) $P[a < X < b] = P[X < b] - P[X \leq a]$

iv. ESPERANZA MATEMÁTICA DE UNA VARIABLE ALEATORIA DISCRETA

También se le conoce como promedio o media aritmética y está definida por:

$$\mu = E[X] = \sum_{i=1}^{\infty} x_i p(x_i)$$

Propiedades: a. $E[c] = c$; donde c es una constante

b. $E[ax \pm b] = aE[x] \pm b$

v. VARIANZA DE UNA VARIABLE ALEATORIA DISCRETA

$$\sigma^2 = Var[x] = E[(x - \mu)^2] = E[x^2] - [E(x)]^2 ; \text{ donde } E[x^2] = \sum_{i=1}^n x_i^2 p(x_i)$$

Propiedades: a) $Var[c] = 0$; donde c es una constante

b) $Var[ax \pm b] = a^2 Var[x]$

Ejemplo 5: una determinada marca de coches, ante la competencia existente en el mercado para la venta de coches nuevos, ha decidido rebajar sus precios con el fin de aumentar las ventas y disminuir sus existencias. El director comercial ha estimado la siguiente distribución de probabilidad del número total X de coches, que se venderán el próximo mes después de rebajar los precios.

X	0	1	2	3	4
P(x)	a	0.15	0.35	5a	0.20

- a) Calcular el valor de "a"
- b) Hallar su función de distribución acumulada
- c) Obtener el promedio y desviación estándar del número de coches que espera vender.
- d) Calcular $P[1 \leq X < 4]$; $P[X < 3]$; $P[X = 2]$

Solución:

a) Aplicando la propiedad de que la suma de todas las probabilidades es 1 tenemos que:

$$6a + 0.70 = 1 \text{ entonces } a = 0.05$$

b) Hallando F(X)

X	0	1	2	3	4
P(x)	0.05	0.15	0.35	0.25	0.20
F(X)	0.05	0.20	0.55	0.80	1

- c) • Número promedio de coches vendidos

$$E[x] = 0 \times 0.05 + 1 \times 0.15 + 2 \times 0.35 + 3 \times 0.25 + 4 \times 0.20 = 2.4$$

- Varianza y desviación estándar

$$E[x^2] = 0^2 \times 0.05 + 1^2 \times 0.15 + 2^2 \times 0.35 + 3^2 \times 0.25 + 4^2 \times 0.20 = 7$$

$$Var[x] = 7 - 2.4^2 = 1.24$$

Entonces la desviación estándar $S_x = \sqrt{1.24} = 1.11$

- d) Calculando según la tabla:

$$P[1 \leq X < 4] = 0.15 + 0.25 + 0.35 = 0.75$$

$$P[X < 3] = 0.05 + 0.15 + 0.35 = 0.55$$

$$P[X = 2] = 0.35$$

3.3.4. VARIABLE ALEATORIA CONTINUA

Una variable aleatoria $X: \Omega \rightarrow \mathbb{R}$ es llamada continua cuando el rango R_x , es un intervalo o una colección de intervalos sobre la recta real.

Ejemplo 1: sea X la variable aleatoria “Nº de kilogramos que pierde una persona” al seguir una dieta durante cierto periodo. Su R_x pueden ser todos los puntos de un intervalo como $[1;3]$.

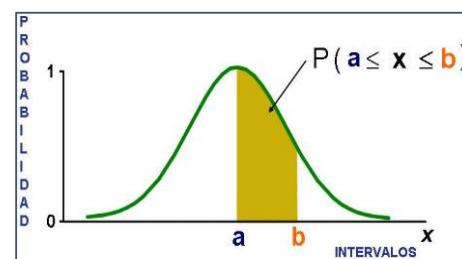
Ejemplo 2: sea X : “Tiempo de vida de un transistor”. Es variable aleatoria continua por que R_x puede ser $[0;+\infty)$

i. FUNCIÓN DE DENSIDAD DE PROBABILIDAD

Sea X una variable aleatoria continua con rango $R_x \subset \mathbb{R}$. La función de densidad asociado a la variable aleatoria x es una función $f(x)$ integrable que se satisface las siguientes condiciones:

- $f(x) \geq 0 ; \forall x \in R_x \subset \mathbb{R}$
- $\int_{R_x} f(x) dx = 1$
- Para todo $a, b \in R_x$ con $a \leq b$ se cumple:

$$P[a \leq x \leq b] = \int_a^b f(x) dx$$



<https://www.google.com.p>

Ejemplo 3: sea X una variable aleatoria con función de densidad:

$$f(x) = \begin{cases} a(3x - x^2), & \text{Si } 0 \leq x \leq 3 \\ 0, & \text{En otros casos} \end{cases}$$

- a) Encuentre el valor de la constante a
- b) Calcular la probabilidad que X esté en el intervalo $[1, 2]$

Solución: todos los $R_x \in [0, 3]$ entonces:

$$\begin{aligned} a) \int_0^3 a(3x - x^2) dx &= a \left[\frac{3x^2}{2} - \frac{x^3}{3} \right]_0^3 = a \left(\frac{27}{2} - 9 \right) = 1 \Rightarrow a = \frac{2}{9} \\ b) P[1 \leq x \leq 2] &= \int_1^2 \frac{2}{9} (3x - x^2) dx = \left[\frac{x^2}{3} - \frac{2x^3}{27} \right]_1^2 = \frac{13}{17} \end{aligned}$$

ii. ESPERANZA MATEMÁTICA DE UNA VARIABLE ALEATORIA CONTINUA

$$E[X] = \int_{R_x} x \cdot f(x) dx$$

iii. VARIANZA DE UNA VARIABLE ALEATORIA CONTINUA

$$Var[x] = E[x^2] - [E(x)]^2$$

$$\text{Donde: } E[x^2] = \int_{R_x} x^2 \cdot f(x) dx$$

Ejemplo 4: hallar la esperanza matemática y varianza en el ejemplo 3

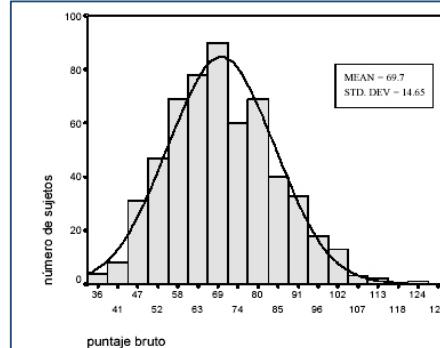
$$a) E[X] = \int_0^3 \frac{2x}{9} (3x - x^2) dx = \int_0^3 \left(\frac{6x^2}{9} - \frac{2x^3}{9} \right) dx = \left[\frac{2x^3}{9} - \frac{x^4}{18} \right]_0^3 = 1.5$$

$$b) Var[x] = E[x^2] - [E(x)]^2 = 2.7 - (1.5)^2 = 0.45$$

$$E[X^2] = \int_0^3 \frac{2x^2}{9} (3x - x^2) dx = \int_0^3 \left(\frac{6x^3}{9} - \frac{2x^4}{9} \right) dx = \left[\frac{x^4}{6} - \frac{2x^5}{45} \right]_0^3 = 2.7$$

3.3.5. DISTRIBUCIÓN NORMAL

La distribución normal es llamada también campana de Gauss por su forma acampanada. Es la distribución de probabilidad continua más importante en todo el campo de la estadística ya que muchos datos que ocurren en la naturaleza, la industria, la economía y la investigación describen una distribución aproximadamente normal o gaussiana. Por ejemplo, la distribución de las alturas y pesos de individuos en poblaciones homogéneas de personas tienen distribución aproximadamente normal.

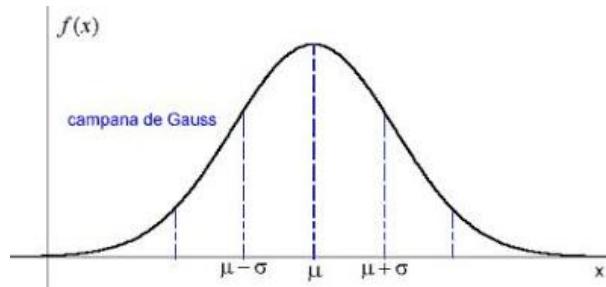


<http://bit.ly/1Tt8n4p>

i. DEFINICIÓN

Se dice que una variable aleatoria continua sigue una distribución normal de media μ y varianza σ^2 , y se escribe $X \sim N(\mu, \sigma^2)$, cuando tiene la función de densidad:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



<http://bit.ly/1mKGyX>

i. PROPIEDADES DE LA DISTRIBUCIÓN NORMAL

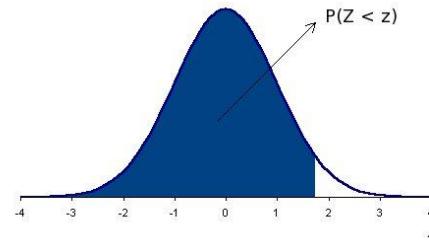
- La distribución normal tiene forma de campana con distribución de probabilidad que tiene media $\mu = 0$ y desviación estándar $\sigma = 1$.
- El área bajo la curva o la probabilidad desde menos infinito a más infinito vale 1.
- La distribución normal es simétrica, es decir cada mitad de curva tiene un área de 0.5.
- La escala horizontal de la curva se mide en desviaciones estándar.
- La forma y la posición de una distribución normal dependen de los parámetros μ y σ , en consecuencia hay un número infinito de distribuciones normales.

ii. DISTRIBUCIÓN NORMAL ESTANDAR

Es la Distribución Normal con media $\mu = 0$ y varianza $\sigma^2 = 1$. La función de densidad de la distribución normal estándar usualmente se denota por el símbolo Φ .

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

$$\Phi(z) = P[Z \leq z] = \int_{-\infty}^z \Phi(u) du , \quad \forall z \in \mathbb{R}$$



<http://bit.ly/1orsWBg>

iii. ESTANDARIZACIÓN DE UNA VARIABLE ALEATORIA X (Para buscar en tablas)

Si $X \sim N(\mu, \sigma^2)$, la estandarización o tipificación consiste en pasar de la variable aleatoria X con distribución $N(\mu, \sigma^2)$ a una nueva variable aleatoria Z con distribución $N(0, 1)$ mediante la expresión:

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

La cual tiene distribución $N(0, 1)$, esto es la variable aleatoria Z tiene distribución normal con media cero y varianza 1. Este tipo de procedimiento se denomina estandarización de la variable aleatoria X .

Ejemplo 1. En un examen de matemáticas, la calificación media fue 72 y varianza 225. Determinar en unidades estándar las puntuaciones de los alumnos que obtuvieron a) 60, b) 93 y c) 72

Solución:

$$\text{a)} \quad Z = \frac{X - \mu}{\sigma} = \frac{60 - 72}{15} = -0.8 \quad \text{b)} \quad Z = \frac{93 - 72}{15} = 1.4 \quad \text{c)} \quad Z = \frac{72 - 72}{15} = 0$$

Ejemplo 2. El gerente de personal de una gran compañía requiere que los solicitantes a un puesto efectúen cierta prueba y alcancen una calificación de 500. Si las calificaciones de la prueba se distribuyen normalmente con media $\mu = 485$ y $\sigma = 30$

a) ¿Cuál es la probabilidad que los solicitantes pasen la prueba?

$$\begin{aligned} P(X \geq 500) &= 1 - P(X < 500) = 1 - P\left(\frac{X - 485}{30} < \frac{500 - 485}{30}\right) = 1 - P(Z < 0.5) \\ &= 1 - 0.69146 = 0.30854 \end{aligned}$$

b) ¿Cuál es la probabilidad que un solicitante obtenga menos de 550 en la prueba?

$$P(x < 520) = P\left(Z < \frac{520 - 485}{30}\right) = P(Z < 1.17) = 0.878$$

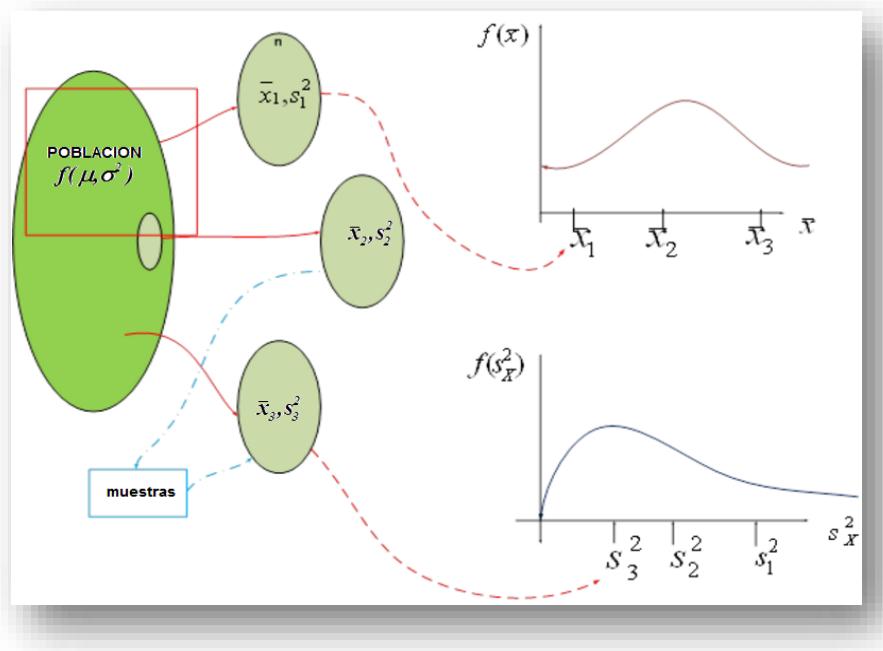
c) ¿Cuál es la probabilidad que un solicitante obtenga entre 450 y 510 en la prueba?

$$P(450 < x < 510) = P(x < 510) - P(x < 450)$$

$$\begin{aligned} P\left(Z < \frac{510 - 485}{30}\right) - P\left(Z < \frac{450 - 485}{30}\right) &= P(Z < 0.83) - P(Z < -1.17) \\ &= 0.79767 - 0.12167 = 0.676 \end{aligned}$$

3.3.6. DISTRIBUCIONES MUESTRALES

DEFINICIÓN. Una distribución muestral es la distribución de todos los posibles valores del estadístico de la muestra, que se pueden obtener de la población para un determinado tamaño de muestra.

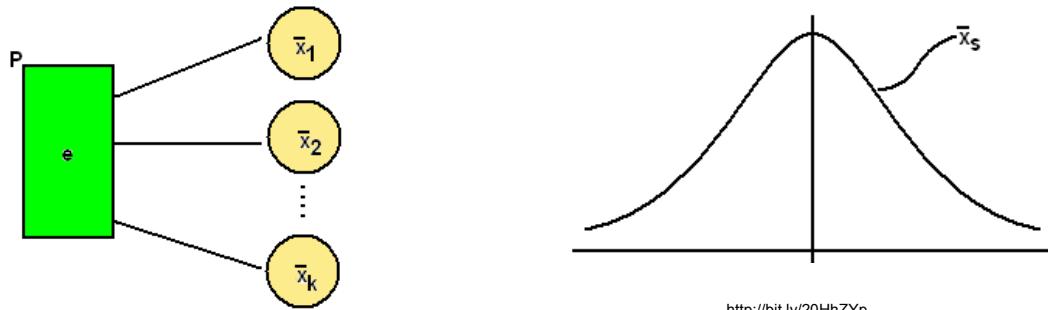


<http://bit.ly/1OizWFA>

Por ejemplo, se podría tomar de una población una muestra aleatoria de 100 personas y pesarlas para calcular después la media de su peso. Se puede pensar en esta media muestral como si se hubieran extraído de la distribución de todas las medias muestrales posibles, para muestras de tamaño 100 que pudieran tomarse de la población.

3.3.7. DISTRIBUCIÓN MUESTRAL DE LA MEDIA

Consiste en tomar de una población todas las muestras posibles de tamaño n . Luego se calcula las medias de cada muestra. Obteniéndose así la distribución de todas las medias muestrales posibles:



i. MUESTREO CON REEMPLAZO Y SIN REEMPLAZO

a) **Muestreo con reemplazo.** Es cuando de una población N podemos extraer n elementos, de tal manera que cada elemento extraído es devuelto al total poblacional para al siguiente extracción, de esta forma un elemento puede ser extraído varias veces. El total de muestras de tamaño n extraídas con reemplazamiento de una población N está dado por la formula N^n .

b) **Muestreo sin reemplazo.** Es cuando un elemento extraído de la población para conformar la muestra ya no es devuelto para una siguiente extracción. Lo que significa que cada elemento se escogerá por única vez sin importar incluso el orden. Esta dado por la fórmula:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

ii. MEDIA Y VARIANZA PARA LA DISTRIBUCIÓN MUESTRAL DE LA MEDIA

Observación1: cuando el muestreo es **con reemplazo**:

a) La media de todas las medias muestrales $\mu_{\bar{x}}$ es exactamente igual a la media de la población μ , esto es:

$$\mu_{\bar{x}} = E[\bar{X}] = \mu$$

- b) También la varianza de las medias muestrales es igual a la varianza poblacional dividido por el tamaño de la muestra:

$$\sigma_{\bar{x}}^2 = Var[\bar{X}] = \frac{\sigma^2}{n}$$

Observación2: cuando el muestreo es **sin reemplazo**:

a) $\mu_{\bar{x}} = E[\bar{X}] = \mu$

- b) La varianza de las medias muestrales es igual a la varianza poblacional dividido por el tamaño de la muestra pero multiplicado por el factor de corrección poblacional

$$\frac{N-n}{N-1}$$

$$\sigma_{\bar{x}}^2 = Var[\bar{X}] = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)$$

iii. PROBABILIDAD DE LA MEDIA MUESTRAL

Si recordamos a la distribución normal, esta es una distribución continua, en forma de campana simétrica. Con esta distribución podíamos calcular la probabilidad de algún evento relacionado con la variable aleatoria, mediante la siguiente fórmula:

$$z = \frac{x - \mu}{\sigma} \sim N(0,1)$$

En donde z es una variable estandarizada con media igual a cero y varianza igual a uno. Con esta fórmula se pueden hacer los cálculos de probabilidad para cualquier ejercicio, utilizando la tabla de la distribución z .

Entonces la fórmula para calcular la probabilidad del comportamiento del estadístico, en este caso la media de la muestra, quedaría de la siguiente manera:

Para muestras **con reemplazo**:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

y para muestras **sin reemplazo**:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}} \sim N(0,1)$$

Importante: para el desarrollo de los ejercicios se trabajará con la fórmula para **muestras con reemplazo**.

Ejemplo: En cierta población de alcohólicos, la duración promedio del abuso del alcohol es de 12 años y la desviación estándar de 5 años si es escoge una muestra aleatoria de 36 individuos de esta población:

- a) ¿Cuál es la probabilidad que la muestra de alcohólicos tengan una duración promedio de abuso del alcohol menor a 11 años?

$$\text{Identificando: } \mu = 12 \quad \sigma = 5 \quad \text{entonces} \quad \bar{x} \sim N(12, 5/\sqrt{36})$$

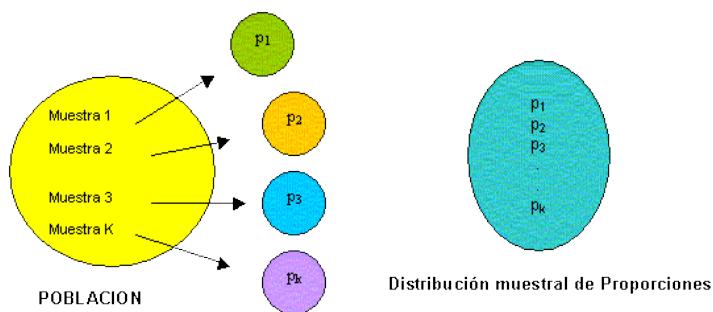
$$P(\bar{x} < 11) = P\left(z < \frac{11 - 12}{\frac{5}{\sqrt{36}}}\right) = P(z < -1.20) = 0.11507 \quad (\text{valor buscado en tabla normal})$$

- b) ¿Cuál es la probabilidad que la muestra de alcohólicos tengan una duración promedio de abuso del alcohol entre 10 y 13 años?

$$\begin{aligned} P(10 < \bar{x} < 13) &= P(\bar{x} < 13) - P(\bar{x} < 10) \\ &= P\left(z < \frac{13 - 12}{\frac{5}{\sqrt{36}}}\right) - P\left(z < \frac{10 - 12}{\frac{5}{\sqrt{36}}}\right) \\ &= P(z < 1.20) - P(z < -2.40) \\ &= 0.88493 - 0.00820 = 0.87673 \end{aligned}$$

3.3.8. DISTRIBUCIÓN MUESTRAL DE UNA PROPORCIÓN

Existen ocasiones en las cuales no estamos interesados en la media de la muestra, sino que queremos investigar la proporción de artículos defectuosos o la proporción de alumnos reprobados en la muestra. La distribución muestral de proporciones es la adecuada para dar respuesta a estas situaciones.



PROBABILIDAD DE UNA PROPORCIÓN MUESTRAL

Para hallar dicha probabilidad se usará la fórmula de la normal estandarizada en los siguientes casos:

- A) Si el muestreo se hace **con reemplazo**, entonces:

$$\hat{p} \sim N\left(p, \frac{pq}{n}\right) \text{ o equivalente } z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \sim N(0,1)$$

- B) Si el muestreo se hace **sin reemplazo** entonces:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n} \cdot \frac{N-n}{N-1}}} \sim N(0,1)$$

Ejemplo1: Se ha determinado que 60% de los estudiantes de una universidad grande fuman cigarrillos. Se toma una muestra aleatoria de 800 estudiantes. Calcule la probabilidad de que la proporción de la muestra de la gente que fuma cigarrillos sea menor que 0.55.

Solución

$$n = 800 \quad p = 0.6 \quad q = 0.4 \quad \hat{p} = 0.55$$

$$P(\hat{p} < 0.55) = P\left(z < \frac{0.55 - 0.60}{\sqrt{\frac{0.6 \times 0.4}{800}}}\right) = P(z < -2.89) = 0.00193$$

Ejemplo2: se sabe que la verdadera proporción de los componentes defectuosos fabricados por una firma es de 4%, y encuentre la probabilidad de que una muestra aleatoria de tamaño 60 tenga:

- a) Menos del 3% de los componentes defectuosos. $P(\hat{p} < 0.03)$
 b) Más del 1% pero menos del 5% de partes defectuosas. $P(0.01 < \hat{p} < 0.05)$

GUÍA DE PRÁCTICA N°10

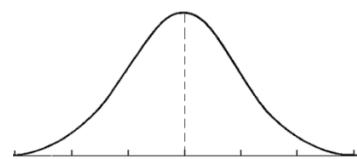
DISTRIBUCIÓN DE PROBABILIDAD CONTINUA

I. DISTRIBUCIÓN NORMAL

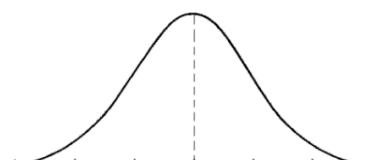
1. Los pesos de 2000 pacientes de un hospital presentan una distribución normal con un peso medio de 65 kg y desviación estándar 6kg. Si un paciente es elegido al azar, calcula:

* Identifica: $\mu = \underline{\hspace{2cm}}$ $\sigma = \underline{\hspace{2cm}}$ entonces $X \sim N(\underline{\hspace{2cm}}, \underline{\hspace{2cm}})$

- a) La probabilidad que pese menos de 75 kg



- b) La probabilidad que pese más de 60 kg

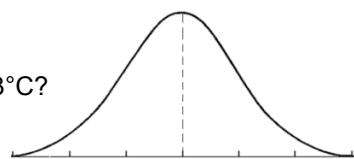


- c) La probabilidad que pese entre 58 y 70 kg

2. En una ciudad, las temperaturas máximas diarias durante el mes de enero se distribuyen normalmente con una media de 26°C y una varianza de 16°C^2 . Si se escoge un día cualquiera de ese mes:

* Identifica: $\mu = \underline{\hspace{2cm}}$ $\sigma = \underline{\hspace{2cm}}$ entonces $X \sim N(\underline{\hspace{2cm}}, \underline{\hspace{2cm}})$

- a) ¿Cuál es la probabilidad que tenga una temperatura menor a 20°C?



- c) ¿Cuál es la probabilidad que tenga una temperatura entre 21°C y 29°C?

3. Las ventas diarias, en euros, en un determinado comercio siguen una distribución con media poblacional de 950 euros y desviación estándar de 150 euros es decir $N(950, 150^2)$. Calcula la probabilidad de que las ventas diarias en ese comercio:

- a) Sean menores a 750 euros.

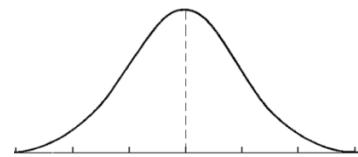
- c) Sean exactamente de 1000 euros

- b) Estén entre 700 y 1200 euros.

- d) Sean al menos 1100 euros

4. Un estudio reciente de los sueldos por hora del personal de mantenimiento para aerolíneas importantes mostró que el salario medio por hora era de \$16,50 (dólares), con una desviación estándar de \$3,50. Si se selecciona al azar un empleado, ¿cuál es la probabilidad de que gane:
- entre \$16,50 y \$22,00 por hora?

b) más de \$21,00 por hora?



c) menos de \$14,00 por hora?

5. Se sabe que la talla media de una población en edad escolar es de 165cm con una desviación típica de 12 cm. Un centro tiene 1400 alumnos matriculados, se pide:

- ¿Cuántos alumnos es esperable que midan más de 155cm?
- ¿Qué proporción (%) de alumnos miden entre 150 y 178 cm?
- Determina la probabilidad de que un cierto alumno mida entre 170 y 186 cm.
- ¿Qué talla permite asegurar que, elegido un alumno al azar, el 67% de sus compañeros son más bajos que él?

6. Se han utilizado dos tipos de pruebas, A y B, para medir los conocimientos sobre cierta materia en una misma población. Los resultados en ambas tienen distribución Normal. La prueba A tiene como media 78,3 y como desviación típica 4,2. La prueba B tiene 85,1 de media y 3,2 de desviación típica. Una persona ha obtenido 83,1 en la prueba A y otra ha obtenido 87,5 en la prueba B. ¿Cuál de las dos se encuentra en mejor posición? ¿Por qué?

7. Gensa es una asociación internacional de personas con alto coeficiente intelectual. Para pertenecer a ella, una persona debe tener un coeficiente intelectual de 132 o más alto (USA today, 13 de febrero de 1992). Si las calificaciones del coeficiente de inteligencia se distribuyen normalmente con promedio de 100 y desviación estándar de 15, ¿qué porcentaje de personas califican para ser miembros de Gensa?

II. DISTRIBUCIÓN MUESTRAL DE LA MEDIA

$$\text{Estandarización: } Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

8. El número de clientes por semana en cada tienda de una cadena de autoservicios tiene una media poblacional $\mu=5000$ clientes y una desviación estándar $\sigma=500$. Si se selecciona una muestra aleatoria de 25 tiendas. * Identifica: $\mu = \underline{\hspace{2cm}}$ $\sigma = \underline{\hspace{2cm}}$ entonces $\bar{x} \sim N(\underline{\hspace{2cm}}, \underline{\hspace{2cm}})$

¿Cuál es la probabilidad de que la media muestral sea inferior a 5075 clientes por semana?

9. El contenido promedio de cereal en un paquete es de 450 gramos. Si se tomó una muestra de 23 paquetes con una desviación estándar de 13 gramos.

* Identifica: $\mu = \underline{\hspace{2cm}}$ $\sigma = \underline{\hspace{2cm}}$ entonces $\bar{x} \sim N(\dots, \dots)$

a) ¿Cuál es la probabilidad de que el promedio de esta muestra sea mayor a 455 gramos?

b) ¿Cuál es la probabilidad de que el promedio de ésta muestra se encuentre entre 445 y 458 gramos?

10. El tiempo de atención por cliente de un cajero de un banco es normal con media 6 minutos y desviación estándar 2.5 minutos.* Identifica: $\mu = \underline{\hspace{2cm}}$ $\sigma = \underline{\hspace{2cm}}$ entonces

$\bar{x} \sim N(\dots, \dots)$

a) ¿Cuál es la probabilidad de que el tiempo promedio de atención para una muestra de 15 clientes sea menor de 7 minutos?

c) Si el tiempo promedio en que el cajero atiende a un grupo de 15 clientes excede los 10 minutos entonces éste es despedido. ¿Cuál es la probabilidad de que esto ocurra?

11. En cierta ciudad americana hay 400 agentes que se dedican al negocio de venta de propiedades. El valor medio de las propiedades vendidas por estos agentes en un año es de \$800000 y su desviación típica de \$300000. Se selecciona una muestra de 100 agentes y se anota el valor de las propiedades que han vendido en un año. ¿Cuál es la probabilidad de que la media muestral sea mayor que \$825000?

* Identifica: $\mu = \underline{\hspace{2cm}}$ $\sigma = \underline{\hspace{2cm}}$ entonces $\bar{x} \sim N(\dots, \dots)$

12. Un inspector de pesos y medidas visita una planta de empacado para verificar que el peso neto de las cajas sea el indicado en éstas. El gerente de la planta asegura al inspector que el peso promedio de cada caja es de 750 gr. con una desviación estándar de 5 gr. El inspector selecciona, al azar, 100 cajas y encuentra que el peso promedio es de 748 gr. Bajo estas condiciones, ¿qué tan probable es tener un peso de 748 o menos? ¿Qué actitud debe tomar el inspector?

13. El tiempo que se usa el correo electrónico por sesión tiene una distribución normal con media de 8 minutos y varianza de 4 minutos². Si se seleccionan muestras de 25 correos:

a) ¿Cuál es la probabilidad de que la media de una muestra cualquiera esté entre 7,8 y 8,2 minutos inclusive?

b) ¿Qué es más probable que ocurra, una media muestral mayor de 9 minutos en una muestra de 25 sesiones, o una media muestral mayor de 8,6 minutos en una muestra de 100 sesiones?

III. DISTRIBUCIÓN MUESTRAL DE LA PROPORCIÓN

Estandarización:
$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \sim N(0,1)$$

14. En una investigación realizada entre los habitantes de la delegación Miraflores, se ha encontrado que el 18% de ellos, ha tenido problemas de tránsito. Se seleccionó una muestra aleatoria de 100 personas

* Identifica: $p = \underline{\hspace{2cm}}$ $q = \underline{\hspace{2cm}}$ $n = \underline{\hspace{2cm}}$

a) ¿Cuál es la probabilidad de que, menos de 35% haya tenido problemas de tránsito?

b) ¿Cuál es la probabilidad de que, entre el 15% y el 25% haya tenido contacto con policías de tránsito?

15. Un proceso para llenar botellas de cerveza presenta un producción promedio de en la que el 10% de las botellas no están completamente llenas. Si mediante este proceso se selecciona al azar una muestra aleatoria de 225 botellas de un lote de 625 envases llenos ¿Cuál es la probabilidad de que la proporción muestral de botellas parcialmente llenas se encuentre en el intervalo que va del 9 al 11%?

* Identifica: $p = \underline{\hspace{2cm}}$ $q = \underline{\hspace{2cm}}$ $n = \underline{\hspace{2cm}}$

16. Un partido político cree que el 60% del electorado está a favor de su programa. Como su líder encuentra que esta predicción es demasiado optimista decide hacer un sondeo con una muestra de 90 personas.

* Identifica: $p = \underline{\hspace{2cm}}$ $q = \underline{\hspace{2cm}}$ $n = \underline{\hspace{2cm}}$

a) ¿Cuál será la probabilidad que como máximo 60 personas estén a favor de su partido

b) ¿Cuál será la probabilidad que al menos 50 personas estén a favor de su partido?

17. Supongamos que con una terapia para tratar “el miedo a volar en avión” se recupera el 80% de los pacientes. Si seleccionamos al azar 16 pacientes que han acudido a la consulta de un psicólogo clínico con este tipo de fobia ¿Cuál es la probabilidad de que al menos 12 se hayan recuperado y puedan tomar aviones?

18. Se toma una muestra de 100 trabajadores de una gran empresa para estudiar su actitud frente a un cambio en el método de trabajo. Si el 60% de todos los trabajadores de la empresa están a favor del cambio. ¿Cuál es la probabilidad de que menos de 50 de los miembros de la muestra estén a favor?

19. Una encuesta citó a los distribuidores de los automóviles Chevrolet y Toyota como los dos mejores en lo que respecta a servicio al cliente. Sólo el 4% de sus clientes mostró cierta inconformidad con la agencia. Si se toma una muestra de 250 clientes
- ¿Cuál es la probabilidad de que 12 clientes o menos tengan cierta inconformidad con la agencia?
 - ¿Cuál es la probabilidad de que 5 o más clientes estén descontentos con la agencia?
 - ¿Cuál es la probabilidad de que entre 6 y 10 clientes (ambas inclusive) estén descontentos con la agencia?

IV. VARIABLE ALEATORIA CONTINUA

20. Sea X una variable aleatoria continua con función de densidad

$$f(x) = \begin{cases} K(1+x^2), & 0 < x \leq 3 \\ 0, & \text{en otro caso} \end{cases}$$

- Calcular la constante K y la función de distribución de X .
- Calcular $F(X)$
- Obtener las probabilidades $P(X \geq 2)$, $P(-\infty < X < 1)$ y $P(1 \leq X \leq 2)$.
- Sabiendo que $X > 1$, calcular la probabilidad de que $X \leq 2$.

21. Suponga que el tiempo de atención de cada cliente en una estación de servicio es una variable aleatoria continua X : duración en horas, con la función de densidad de probabilidad:

$$f(x) = \begin{cases} \frac{2}{5}(x+2) & , 0 \leq x \leq 1 \\ 0 & , \text{ otro } x \end{cases}$$

- Calcule la probabilidad que el tiempo de atención esté entre 15 y 30 minutos
- Suponga que el costo de atención a cada cliente está dado por la siguiente variable aleatoria: $G(X) = 10 + 5X$ (dólares), Calcule su media y Varianza.

22. Sea X una variable aleatoria continua con función de densidad dado por

$$f(x) = \begin{cases} C(6x - 2x^2), & \text{Si } 0 < x < 2 \\ 0 & \text{En otro caso} \end{cases}$$

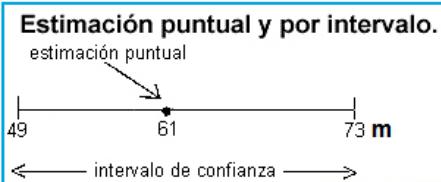
- Encuentre el valor de la constante C
- Hallar $E(X)$ y $V(X)$
- Calcular $P[X > 1]$

3.3. ESTIMACIÓN DE PARÁMETROS Y TAMAÑO DE MUESTRA

3.4.1. ESTIMACIÓN DE PARÁMETROS

Una estimación puntual (de punto) sabemos que está dado por un solo número, pero una estimación de un parámetro dada por dos números entre los cuales se encuentra el parámetro se llama una estimación de intervalos del parámetro:

Ejemplo: si decimos que una distancia mide 61m estamos dando una estimación puntual, pero si decimos que la distancia está entre 49 m y 73 m estamos dando una estimación de intervalo.



Observación: una estimación puntual no proporciona información de la precisión o error debido al muestreo y esto se logrará mediante estimación de intervalos junto con una medida de seguridad que tal intervalo contenga al parámetro desconocido.

3.4.2. INTERVALO DE CONFIANZA

DEFINICIÓN

Sea X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n extraída de una población. Y sean L_i y L_s dos límites tales que $L_i < L_s$, para los cuales se cumple:

$$P(L_i < \theta < L_s) = \gamma = 1 - \alpha$$

Donde:

- i. El intervalo $\langle L_i; L_s \rangle$ se llama intervalo de confianza para θ
- ii. L_i = Límite inferior de confianza para θ
- iii. L_s = Límite superior de confianza para θ
- iv. $\gamma = 1 - \alpha$ es Nivel de confianza
- v. α = nivel de significación

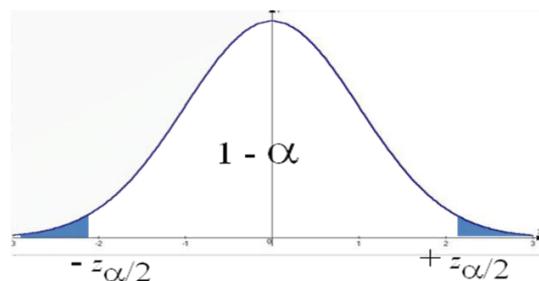
La elección del nivel de confianza depende del investigador y que sea alto (cercano a 1), los más utilizados son:

$$\gamma = 0.90 ; \quad \gamma = 0.95 ; \quad \gamma = 0.99$$

Ejemplo: Si $P(10,5 < \mu < 12,5) = 0.95$, significa que la probabilidad que el intervalo incluya a μ es del 95%. Otra interpretación sería que del total de muestras, el 95% incluirá al verdadero valor y 5% no lo incluirá.

3.4.3. INTERVALO DE CONFIANZA PARA LA MEDIA POBLACIONAL μ DE UNA POBLACIÓN NORMAL SI LA VARIANZA POBLACIONAL σ^2 ES CONOCIDA

Sea X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n extraída de una población Normal $N(\mu; \sigma^2)$ con varianza poblacional σ^2 conocida.



$$P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha$$

Despejando de la formula anterior tenemos que los límites inferiores y superiores para la media poblacional son:

$$\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

En Resumen:

El intervalo de confianza con coeficiente de confianza $\gamma = 1 - \alpha$ de la media poblacional μ con varianza poblacional σ^2 conocida es expresado como:

$$\mu \in \left(\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

Donde

\bar{X} es la media muestral .

$Z_{\alpha/2}$ es el valor crítico de la distribución Normal Estándar correspondiente a una

confianza del $100.\gamma\% = 100(1 - \alpha)\%$, es decir $P(Z \leq -Z_{\alpha/2}) = \frac{\alpha}{2}$

3.4.4. INTERVALO DE CONFIANZA PARA UNA PROPORCIÓN POBLACIONAL

Sea X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n extraída de una población Bernoulli $B(1; p)$ de parámetro p .

Entonces $Q = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \sim N(0; 1)$ es una cantidad pivotal, donde $q = 1 - p$

Luego el intervalo de Confianza para la proporción poblacional está dado por

$$p \in \left(\hat{p} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} ; \hat{p} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \right), \text{ donde } \hat{q} = 1 - \hat{p}$$

3.4.5. TAMAÑO DE MUESTRA

La muestra debe reproducir las características de la población, por lo tanto surgen entonces dos preguntas, sobre la cantidad de elementos que debe incluir la muestra y hasta qué punto pueden generalizarse a la población. Ambas preguntas convergen en un problema de exactitud o precisión cuya finalidad es no incurrir en errores a la hora de obtener los resultados, no obstante los errores son inevitables, lo importante entonces es minimizarlos. Existen dos tipos de errores:



<http://bit.ly/1KcAt17>

- a) Los sistemáticos o distorsiones, que son causados por factores externos a la muestra y que se pueden producir en cualquier momento de la investigación.
- b) El error de muestreo, de azar o de estimación, inevitable, ya que siempre habrá diferencia entre los valores medios de la muestra y los valores medios de la población, la magnitud de este error depende del tamaño de la muestra (a mayor tamaño de muestra menor error) y de la dispersión o desviación (a mayor dispersión mayor error).

Por lo tanto, que para que una muestra sea representativa debe estar dentro de ciertos límites y proporciones establecidas por la estadística:

i. ELEMENTOS PARA EL CÁLCULO DEL TAMAÑO MUESTRAL

El cálculo del tamaño de la muestra depende de los siguientes elementos:

a) **Tamaño de la población.** Se considera finito cuando está constituido por 100,000 elementos o menos e infinito si excede esta cifra.

b) **Nivel de confianza adoptado.** El nivel más utilizado es el 95% de probabilidad de que los resultados obtenidos en la muestra sean válidos para la población (riesgo del 5%). A mayor intervalo de confianza mayor debe ser el tamaño de la muestra, por lo que se puede decir que el intervalo de confianza es propio de cada investigación.



c) **Error de estimación permitido.** Los resultados obtenidos de la muestra no son rigurosamente exactos con respecto al universo por lo que siempre existirá un margen de error mayor o menor. El error de estimación es siempre inversamente proporcional al tamaño de la muestra, a

mayor tamaño menor error.

<http://bit.ly/1QWKwXp>

ii. TAMAÑO DE MUESTRA PARA ESTIMAR UNA MEDIA μ

El procedimiento para seleccionar el tamaño de muestra es

a) Elija d el error máximo permisible y un coeficiente de confianza $\gamma = 1 - \alpha$

b) Resuelva la ecuación $d = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ para n . Es decir:

$$n = \left[\frac{Z_{\frac{\alpha}{2}} \sigma}{d} \right]^2$$

c) Si el tamaño de la población N es conocido ó $\frac{n}{N} \geq 0.05$ se hace uso de la

corrección por población finita y se debe resolver: $d = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$ para hallar n

$$n = \frac{N \cdot Z_{\alpha/2}^2 \cdot \sigma^2}{d^2(N-1) + Z_{\alpha/2}^2 \cdot \sigma^2}$$

iii. TAMAÑO DE MUESTRA PARA ESTIMAR UNA PROPORCIÓN

a) El tamaño n de la manera tal que el error en la estimación de p sea menor que un valor especificado d , por lo expresado en el párrafo anterior se debe escoger de tal manera que:

$$d = Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}, \quad \text{de donde despejando } n \text{ se tiene:}$$

$$n = \frac{Z_{\alpha/2}^2 \hat{p}\hat{q}}{d^2}$$

b) Si N es conocido ó $\frac{n}{N} \geq 0.05$ se hace uso de la corrección por población finita

y se obtiene la fórmula:

$$n = \frac{N \cdot Z_{\alpha/2}^2 \cdot pq}{d^2(N-1) + Z_{\alpha/2}^2 \cdot pq}$$

Donde:

N : Tamaño poblacional

p : Proporción estimada de característica principal de la variable en estudio

q : Proporción estimada de característica secundaria de la variable en estudio

d : Error máximo tolerable

$Z_{\alpha/2}$: Valor tabular correspondiente a la distribución normal estándar

considerando una confianza $100(1-\alpha)\%$

GUÍA DE PRÁCTICA N°11

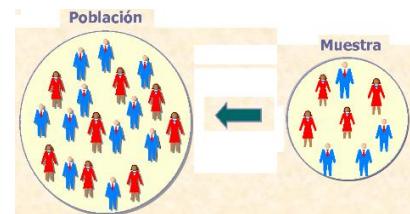
ESTIMACIÓN DE PARÁMETROS Y TAMAÑO DE MUESTRA

1. Un candidato se postula para un puesto ministerial y de acuerdo a una encuesta tiene una apoyo del $40\% \pm 5$ ¿Entre que valores estará su porcentaje de aceptación?
 - Su porcentaje de aceptación estará en el intervalo [____ , ____] es decir ____ $< p <$ ____

2. Un gerente luego de analizar las ventas de sus representantes de ventas, obtuvo que el promedio de ventas es de 9000 ± 500 ¿entre que valores estará la media?
 - El promedio de ventas estará en el intervalo [____ , ____] es decir ____ $< \mu <$ ____

PROCEDIMIENTO PARA HALLAR INTERVALOS DE CONFIANZA DE UN PARÁMETRO

Un parámetro es una medida de la población y para estimar el parámetro necesitamos obtener una muestra con cuyos resultados estimaremos el parámetro.



<http://bit.ly/1WuE37s>

II. INTERVALO DE CONFIANZA PARA LA MEDIA POBLACIONAL (μ) DE UNA POBLACIÓN NORMAL

A) SI LA VARIANZA POBLACIONAL ES CONOCIDA $\mu \in \left(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$

B) SI LA VARIANZA POBLACIONAL ES DESCONOCIDA $\mu \in \left(\bar{X} - Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} ; \bar{X} + Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right)$

CALCULO DE $Z_{\frac{\alpha}{2}}$: α =Nivel de significancia (0.05, 0.01, 0.10) ; γ =Nivel de confianza (0.95;0.99;0.90)

Si $\gamma =$ _____ entonces $\alpha =$ _____ para buscar en tabla: $1 - \alpha / 2$ _____ ; $Z_{\frac{\alpha}{2}} =$ _____

γ (Confianza)	90%	95%	99%
$Z_{\alpha/2}$ =			

3. Un fabricante afirma que el peso promedio de las latas de fruta en conserva que saca al mercado es 19 onzas. Para verificar esta afirmación se escogen al azar 35 latas de la fruta y se encuentra que el peso promedio es 18.5 onzas con una desviación estándar de 2 onzas. Utilizando un intervalo de confianza del 95% para μ , ¿se puede aceptar la afirmación del fabricante?

$$n = \underline{\hspace{2cm}} \quad \bar{X} = \underline{\hspace{2cm}} \quad \sigma = \underline{\hspace{2cm}} \quad \alpha = \underline{\hspace{2cm}} \quad Z_{\frac{\alpha}{2}} = \underline{\hspace{2cm}}$$

$$\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \underline{\hspace{2cm}} ; \quad \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \underline{\hspace{2cm}}$$

Entonces el intervalo de confianza para $\mu \in \langle \underline{\hspace{2cm}} ; \underline{\hspace{2cm}} \rangle$

Interpretación: _____

¿Se acepta la afirmación del fabricante? _____

4. Para estimar el gasto promedio de los clientes en la cafetería de la universidad, los estudiantes de una clase de estadística toman una muestra de 80 clientes y encuentran un gasto promedio de S/. 5.67, con una desviación estándar de S/. 1.10.

a) ¿Cuál es el valor del estimador puntual para la cantidad promedio de gastos? _____

b) ¿Cuál es el intervalo de confianza del 99% para los gastos promedios de todos los clientes?

Interprete sus resultados.

$$n = \underline{\hspace{2cm}} \quad \bar{X} = \underline{\hspace{2cm}} \quad \sigma = \underline{\hspace{2cm}} \quad \alpha = \underline{\hspace{2cm}} \quad Z_{\frac{\alpha}{2}} = \underline{\hspace{2cm}}$$

$$\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \underline{\hspace{2cm}} ; \quad \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \underline{\hspace{2cm}}$$

Entonces el intervalo de confianza para $\mu \in \langle \underline{\hspace{2cm}} ; \underline{\hspace{2cm}} \rangle$

Interpretación: _____

5. Una máquina de empaquetar bolsas de café, está regulada para embalar bolsas cuyos pesos se distribuyen normalmente con media 500 gr y desviación estándar 10 gr. Supongamos que la máquina está desregulada y deseamos saber el nuevo promedio μ . Una muestra aleatoria de 25 paquetes arroja una media de 485 gr. Hallar un intervalo de confianza de 95% de confianza para μ .

$$n = \underline{\hspace{2cm}} \quad \bar{X} = \underline{\hspace{2cm}} \quad \sigma = \underline{\hspace{2cm}} \quad \alpha = \underline{\hspace{2cm}} \quad Z_{\frac{\alpha}{2}} = \underline{\hspace{2cm}}$$

$$\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \underline{\hspace{2cm}} ; \quad \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \underline{\hspace{2cm}}$$

Entonces el intervalo de confianza para $\mu \in \langle \underline{\hspace{2cm}} ; \underline{\hspace{2cm}} \rangle$

Interpretación: _____

6. Los tiempos de reacción, en mili segundos, de 17 sujetos frente a una matriz de 15 estímulos fueron los siguientes: 448, 460, 514, 488, 592, 490, 507, 513, 492, 534, 523, 452, 464, 562, 584, 507, 461. Suponiendo que el tiempo de reacción se distribuye Normalmente, determine un intervalo de confianza para la media a un nivel de confianza del 95%.

$$n = \underline{\quad} \quad \bar{X} = \underline{\quad} \quad \sigma = \underline{\quad} \quad \alpha = \underline{\quad} \quad Z_{\frac{\alpha}{2}} = \underline{\quad}$$

$$\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \underline{\quad}; \quad \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \underline{\quad}$$

Entonces el intervalo de confianza para $\mu \in \langle \underline{\quad} ; \underline{\quad} \rangle$

Interpretación: _____

7. Fueron retiradas 25 piezas de la producción diaria de una máquina; se encontró para una cierta medida una media de 5,2 mm. Se sabe que las medidas tienen distribución normal con desviación estándar de 1,2 mm. Construir el intervalo de confianza para la media con coeficiente de confianza de 99%.

$$n = \underline{\quad} \quad \bar{X} = \underline{\quad} \quad \sigma = \underline{\quad} \quad \alpha = \underline{\quad} \quad Z_{\frac{\alpha}{2}} = \underline{\quad}$$

$$\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \underline{\quad}; \quad \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \underline{\quad}$$

Entonces el intervalo de confianza para $\mu \in \langle \underline{\quad} ; \underline{\quad} \rangle$

Interpretación: _____

8. Se tomó una muestra aleatoria de 62 estudiantes de marketing en cierta universidad y se les pidió que calificasen en una escala de uno (totalmente en desacuerdo) a siete (totalmente de acuerdo) la siguiente afirmación: "La mayoría de los anuncios publicitarios insultan la inteligencia del consumidor medio". La media y la desviación típica de las respuestas fue de 3,92 y 1,57 respectivamente. Calcular un intervalo de confianza del 95% para la calificación media poblacional.

$$n = \underline{\quad} \quad \bar{X} = \underline{\quad} \quad \sigma = \underline{\quad} \quad \alpha = \underline{\quad} \quad Z_{\frac{\alpha}{2}} = \underline{\quad}$$

$$\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \underline{\quad}; \quad \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \underline{\quad}$$

Entonces el intervalo de confianza para $\mu \in \langle \underline{\quad} ; \underline{\quad} \rangle$

Interpretación: _____

III. INTERVALO DE CONFIANZA PARA UNA PROPORCIÓN POBLACIONAL

$$p \in \left(\hat{p} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} ; \quad \hat{p} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \right) \quad , \text{ donde } \hat{q} = 1 - \hat{p}$$

9. En una encuesta realizada en el MINSA se desea verificar las actitudes de los médicos ante el boletín mensual, se les pidió a 500 médicos que indicaran con qué frecuencia leían el boletín de noticias. De los 500, 375 informaron que leían todas las ediciones. Construir el intervalo de confianza para la proporción real de los que leen todas las ediciones, usar $\alpha = 0.01$.

$$\alpha = \underline{\hspace{2cm}} \quad Z_{\frac{\alpha}{2}} = \underline{\hspace{2cm}}; \quad n = \underline{\hspace{2cm}} \quad \hat{p} = \underline{\hspace{2cm}}; \quad \hat{q} = \underline{\hspace{2cm}}$$

Entonces el intervalo de confianza para $p \in \langle \underline{\hspace{2cm}} ; \quad \underline{\hspace{2cm}} \rangle$

Interpretación: _____

10. Tomada, al azar, una muestra de 120 estudiantes de una Universidad, se encontró que 54 de ellos hablaban inglés. Halle, con un nivel de confianza del 90%, un intervalo de confianza para estimar la proporción de estudiantes que hablan el idioma inglés entre los estudiantes de esa Universidad.

$$\alpha = \underline{\hspace{2cm}} \quad Z_{\frac{\alpha}{2}} = \underline{\hspace{2cm}}; \quad n = \underline{\hspace{2cm}} \quad \hat{p} = \underline{\hspace{2cm}}; \quad \hat{q} = \underline{\hspace{2cm}}$$

Entonces el intervalo de confianza para $p \in \langle \underline{\hspace{2cm}} ; \quad \underline{\hspace{2cm}} \rangle$

Interpretación: _____

11. Se recibe un lote muy grande de artículos provenientes de un fabricante que asegura que el porcentaje de artículos defectuosos en la producción es del 1%. Al seleccionar una m.a. de 200 artículos y después de inspeccionarlos, se descubren 8 defectuosos. Obtener el intervalo de confianza aproximado del 99% para la verdadera proporción de artículos defectuosos en el proceso de manufactura del fabricante. Con base a estos resultados ¿Qué se puede concluir con respecto a la afirmación del fabricante?

$$\alpha = \underline{\hspace{2cm}} \quad Z_{\frac{\alpha}{2}} = \underline{\hspace{2cm}}; \quad n = \underline{\hspace{2cm}} \quad \hat{p} = \underline{\hspace{2cm}}; \quad \hat{q} = \underline{\hspace{2cm}}$$

Entonces el intervalo de confianza para $p \in \langle \underline{\hspace{2cm}} ; \quad \underline{\hspace{2cm}} \rangle$

Interpretación: _____

12. En una encuesta para verificar las actitudes de los empleados ante el boletín mensual, se les pidió a 500 empleados de una gran organización nacional que indicaran con qué frecuencia leían el boletín de noticias. De los 500, 375 informaron que leían todas las ediciones. Construir el intervalo de confianza para la proporción real de los que leen todas las ediciones, usar $\alpha = 0.01$.

$$\alpha = \text{_____} \quad Z_{\frac{\alpha}{2}} = \text{_____}; \quad n = \text{_____} \quad \hat{p} = \text{_____}; \quad \hat{q} = \text{_____}$$

Entonces el intervalo de confianza para $p \in \langle \text{_____} ; \text{_____} \rangle$

Interpretación: _____

IV. TAMAÑO DE MUESTRA PARA ESTIMAR UNA MEDIA

a) Si N no se conoce

$$n = \left[\frac{Z_{\frac{\alpha}{2}} \sigma}{d} \right]^2$$

b) Si N se conoce

$$n = \frac{N \cdot Z_{\frac{\alpha}{2}}^2 \cdot \sigma^2}{d^2(N-1) + Z_{\frac{\alpha}{2}}^2 \cdot \sigma^2}$$

13. Se desea estimar la edad media de los empleados que sufren de estrés cuando llega navidad. De estudios anteriores se sabe que la varianza de la población es 108.

- a) ¿Cuál es tamaño de la muestra que se requiere con una confianza del 90% y un error permisible de 2 años?

$$\alpha = \text{_____} \quad Z_{\alpha/2} = \text{_____}; \quad \sigma = \text{_____}; \quad d = \text{_____} \quad n = \text{_____}$$

- b) Si el tamaño de la población es **N=800** ¿Qué tamaño muestral se requiere con una confianza del 99% y un error permisible de 3 años?

$$\alpha = \text{_____} \quad Z_{\alpha/2} = \text{_____}; \quad \sigma^2 = \text{_____}; \quad d = \text{_____} \quad n = \text{_____}$$

14. Un fabricante de electrodomésticos sabe que la vida media de estos sigue una distribución normal con media $\mu = 100$ meses y desviación típica $\sigma = 12$ meses. Determíñese el mínimo tamaño muestral que garantiza, con una probabilidad de 0,99 que la vida media de los electrodomésticos en dicha muestra se encuentra entre 90 y 110 meses.

$$\alpha = \text{_____} \quad Z_{\alpha/2} = \text{_____}; \quad \sigma^2 = \text{_____}; \quad d = \text{_____} \quad n = \text{_____}$$

15. El director de un colegio desea usar la media de la muestra aleatoria para estimar el tiempo promedio que tardan los alumnos en ir de una clase a la siguiente y además quiere poder asegurar con una confianza del 99% que el error es a lo sumo 0.25 minutos. Si puede suponer por experiencia que $\sigma = 1.40$ minutos ¿qué tamaño debe tener la muestra?

$$\alpha = \underline{\quad} \quad Z_{\alpha/2} = \underline{\quad}; \quad \sigma^2 = \underline{\quad}; \quad d = \underline{\quad} \quad n = \underline{\quad}$$

16. Una firma constructora desea estimar la resistencia media de cierto material ¿Qué tamaño muestral se requiere para garantizar que haya un riesgo de solo 0.1% de sobrepasar un error de 5kg o más en la estimación?. La desviación típica de la resistencia de este material es 25 kg.

$$\alpha = \underline{\quad} \quad Z_{\alpha/2} = \underline{\quad}; \quad \sigma^2 = \underline{\quad}; \quad d = \underline{\quad} \quad n = \underline{\quad}$$

V. TAMAÑO DE MUESTRA PARA ESTIMAR UNA PROPORCIÓN

a) Si N no se conoce

$$n = \frac{Z_{\alpha/2}^2 \hat{p}\hat{q}}{d^2}$$

b) Si N se conoce

$$n = \frac{N \cdot Z_{\alpha/2}^2 \cdot pq}{d^2(N-1) + Z_{\alpha/2}^2 \cdot pq}$$

17. La oficina de planificación familiar de cierto distrito desea determinar la proporción de familias con un ingreso mensual inferior a S/. 3000, estudios previos indicaron que esta proporción era de 20%

a) ¿Qué tamaño muestral se requiere para asegurar con confianza 0.95 que el error en la estimación de esta proporción no sobrepasará a 0.05?

$$p = \underline{\quad} \quad q = \underline{\quad} \quad d = \underline{\quad} \quad \alpha = \underline{\quad} \quad Z_{\alpha/2} = \underline{\quad} \text{ entonces } n = \underline{\quad}$$

b) Si el tamaño de la población es **N=800** ¿Qué tamaño muestral se requiere con una confianza del 90% y un error permisible de 0.04?

$$p = \underline{\quad} \quad q = \underline{\quad} \quad d = \underline{\quad} \quad \alpha = \underline{\quad} \quad Z_{\alpha/2} = \underline{\quad} \text{ entonces } n = \underline{\quad}$$

18. El administrador de un market desea determinar la proporción de clientes que pagan con tarjeta de débito, estudios previos indicaron que esta proporción era de 28%. Si el tamaño de la población de una semana es **N=400** ¿Qué tamaño muestral se requiere con una confianza del 95% y un error permisible de 0.06?

$$p = \underline{\quad} \quad q = \underline{\quad} \quad d = \underline{\quad} \quad \alpha = \underline{\quad} \quad Z_{\alpha/2} = \underline{\quad} \text{ entonces } n = \underline{\quad}$$

19. En una muestra aleatoria de 500 familias que tienen televisores en la ciudad de Hamilton, Canadá, se encuentra que 340 están suscritas a HBO. ¿Qué tan grande se requiere que sea una muestra si se quiere tener 95% de confianza de que la estimación de P esté con un margen de 0.02?

$$p = \underline{\quad} \quad q = \underline{\quad} \quad d = \underline{\quad} \quad \alpha = \underline{\quad} \quad Z_{\alpha/2} = \underline{\quad} \text{ entonces } n = \underline{\quad}$$

UNIDAD 4: PRUEBAS DE HIPÓTESIS



<http://bit.ly/1PmOmsk>

Contenido Temático

- ❑ Prueba de Hipótesis en una muestra para media y proporción poblacional
- ❑ Prueba de Hipótesis en dos muestras relacionadas y dos muestras independientes para media y proporción poblacional
- ❑ Prueba de independencia Chi-cuadrado de dos variables cualitativas

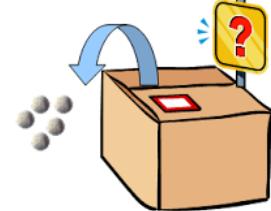
4.1. PRUEBA DE HIPÓTESIS EN UNA MUESTRA PARA MEDIA Y PROPORCIÓN POBLACIONAL

4.1.1. INTRODUCCIÓN

La estadística inferencial es el proceso de usar la información de una muestra para describir el estado de una población. Sin embargo es frecuente que usemos la información de una muestra para probar un reclamo o conjetura sobre la población. El reclamo o conjetura se refiere a una hipótesis. El proceso que corrobora si la información de una muestra sostiene o refuta el reclamo se llama prueba de hipótesis.

En la prueba de hipótesis se pone a prueba un reclamo hecho sobre la naturaleza de una población a base de la información de una muestra.

El reclamo se llama hipótesis estadística.



<http://bit.ly/1TfhPbz>

4.1.2. HIPÓTESIS ESTADÍSTICA

“Una hipótesis estadística es un reclamo hecho sobre la naturaleza de una población”.

Por ejemplo, la premisa formulada por un productor de baterías para autos de que su batería dura en promedio 48 meses, es una hipótesis estadística porque el manufacturero no inspecciona la vida de cada batería que él produce. Si surgieran quejas de parte de los clientes, entonces se pone a prueba el reclamo del manufacturero. La hipótesis estadística sometida a prueba se llama la hipótesis nula, y se denota como H_0 .



<http://bit.ly/1Xk8KgT>

En otras palabras, **una hipótesis estadística** es un supuesto acerca de la distribución de probabilidad de una o más variables aleatorias. En la práctica, La distribución de la población es a menudo implícitamente supuesta, especificándose una hipótesis con el valor o los valores del parámetro o los parámetros que la definen.

Ejemplo1: el promedio poblacional de la altura de los peruanos es 1.70m, esto es:

$$\mu = 1.70$$

Ejemplo 2: la proporción de unidades defectuosas por cierto proceso es menor o igual a 8%.

$$p \leq 0.08$$

4.1.3. PRUEBA DE HIPÓTESIS

Una prueba de hipótesis estadística, es una regla de decisión que permite rechazar o aceptar la hipótesis en base a la información dada por la muestra aleatoria extraída de una población en estudio.

Hipótesis nula. Es la que se formula con la finalidad de rechazarla y se denota por H_0 . Se construye artificialmente para que el investigador evalúe su hipótesis de investigación. Plantea que no existe relación entre los dos fenómenos.

Hipótesis alterna. Simplemente señala la existencia de un hecho o de un evento, o la relación entre dos o más fenómenos. Se denota por H_1

El rechazo de H_0 implicará la aceptación la aceptación de H_1 . Generalmente la hipótesis alternativa representa la suposición que el investigador quiere probar, siendo H_0 formulada con el propósito expreso de ser rechazada.



<http://bit.ly/1KIVgJl>

4.1.4. ERRORES TIPO I Y TIPO II

El procedimiento de decisión de aceptar o rechazar la hipótesis nula está sujeta a dos tipos de error. Estos errores son debidos a fluctuaciones al azar en el muestreo.

Hipótesis Nula	Se acepta H_0	Se rechaza H_0
H_0 es verdadera	Decisión correcta	Error tipo I, α
H_0 es falsa	Error tipo II, β	Decisión Correcta

Error tipo I: Es la probabilidad de rechazar la hipótesis nula, cuando H_0 es cierta. Esta probabilidad es comúnmente denotada por la letra α , conocido también como el nivel de significancia.

Error tipo II: Es la probabilidad de aceptar la hipótesis nula cuando H_0 es falsa. Esta probabilidad es comúnmente denotada por la letra β .

4.1.5. PASOS DE UNA PRUEBA DE HIPÓTESIS

Una prueba de hipótesis se realiza mediante un procedimiento sistemático de cinco paso:



<http://bit.ly/1Ub92H3>

PASO 1: PLANTEAR LA HIPÓTESIS NULA H_0 Y LA HIPÓTESIS ALTERNATIVA H_1

La hipótesis nula (H_0): es una afirmación que no se rechaza a menos que los datos muestrales proporcionen evidencia convincente de que es falsa. Por lo general hay un "no" en la hipótesis nula que indica que "no hay cambio". El planteamiento de la hipótesis nula siempre contiene un signo de igualdad con respecto al valor especificado del parámetro.

La hipótesis alternativa (H_1) es cualquier hipótesis que difiera de la hipótesis nula. Es una afirmación que se acepta si los datos muestrales proporcionan evidencia suficiente de que la hipótesis nula es falsa. Se le conoce también como la **hipótesis de investigación**. El planteamiento de la hipótesis alternativa nunca contiene un signo de igualdad con respecto al valor especificado del parámetro.



<http://bit.ly/1OjHXdF>

PASO 2: SELECCIONAR EL NIVEL DE SIGNIFICANCIA

El nivel de significancia es la probabilidad de rechazar la hipótesis nula cuando es verdadera. Se le denota mediante la letra griega α , también es denominada como error tipo I ó nivel de riesgo. En resumen, el estadístico controla el error tipo I y generalmente se toma el valor de α como 0.05 ó si se quiere ser más riguroso 0.01.



<http://bit.ly/1XmjtaH>

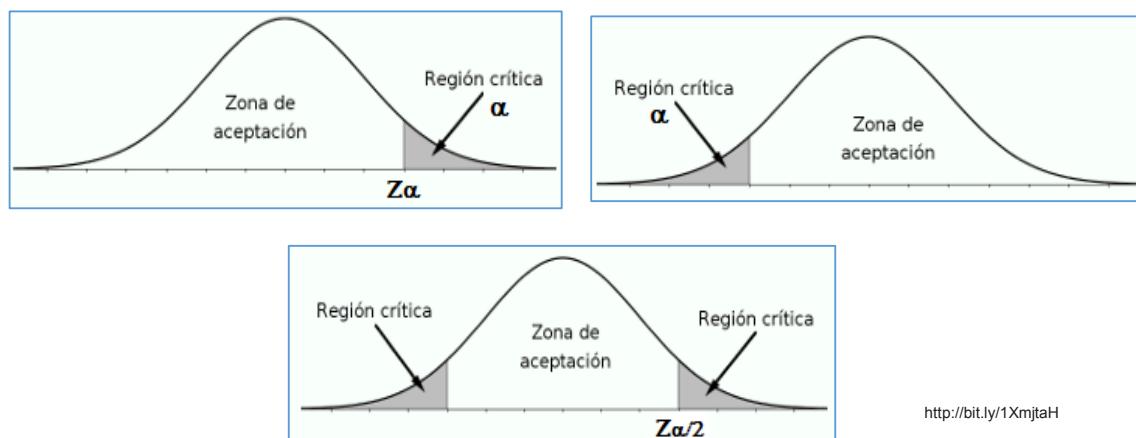
La distribución de muestreo de la estadística de prueba se divide en dos regiones, una región de rechazo (conocida como región crítica) y una región de no rechazo (aceptación). Si la estadística de prueba cae dentro de la región de aceptación, no se puede rechazar la hipótesis nula.

PASO 3: CALCULO DEL VALOR ESTADÍSTICO DE PRUEBA

Valor determinado a partir de la información muestral, que se utiliza para determinar si se rechaza la hipótesis nula., existen muchos estadísticos de prueba para nuestro caso utilizaremos los estadísticos z y t. La elección de uno de estos depende de la cantidad de muestras que se toman, si las muestras de la prueba son iguales a 30 o más se utiliza el estadístico z, en caso contrario se utiliza el estadístico t.

PASO 4: FORMULAR LA REGLA DE DECISIÓN

Se establece las condiciones específicas en la que se rechaza la hipótesis nula y las condiciones en que no se rechaza la hipótesis nula. La región de rechazo define la ubicación de todos los valores que son tan grandes o tan pequeños, que la probabilidad de que se presenten bajo la suposición de que la hipótesis nula es verdadera, es muy remota



<http://bit.ly/1XmjtaH>

Valor crítico: es el punto de división entre la región en la que se rechaza la hipótesis nula y la región en la que no se rechaza la hipótesis nula.

PASO 5: TOMAR UNA DECISIÓN

En este último paso de la prueba de hipótesis, se calcula el estadístico de prueba, se compara con el valor crítico y se toma la decisión de rechazar o no la hipótesis nula. Tenga presente que en una prueba de hipótesis solo se puede tomar una de dos decisiones: aceptar o rechazar la hipótesis nula. Debe subrayarse que siempre existe la posibilidad del error tipo I y el error de tipo II



<http://bit.ly/1mvfyf>

4.1.6. PRUEBA DE HIPÓTESIS PARA LA MEDIA CON VARIANZA POBLACIONAL σ^2 CONOCIDA

Sea X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n extraída de una población Normal $N(\mu; \sigma^2)$ con varianza poblacional σ^2 conocida.

Entonces tenemos.

$$\text{Hipótesis: } \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases} \quad \text{ó} \quad \begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases} \quad \text{ó} \quad \begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$$

$$\text{Estadístico de prueba: } Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

Reglas de decisiones

$$\text{Caso A: } \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases} \quad \text{Si } |Z| > Z_{\alpha/2}, \text{ se rechaza } H_0.$$

$$\text{Caso B: } \begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases} \quad \text{Si } Z > Z_\alpha, \text{ se rechaza } H_0.$$

$$\text{Caso C: } \begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases} \quad \text{Si } Z < -Z_\alpha, \text{ se rechaza } H_0.$$

Donde: $Z_{\alpha/2}$ y Z_α son los valores tabulares de la distribución normal estándar a un nivel de significancia α de dos colas y una cola respectivamente.

4.1.7. PRUEBA DE HIPÓTESIS PARA UNA PROPORCIÓN

Sea X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n extraída de una población Bernoulli $B(1; p)$ de parámetro p .

$$\text{Y sea la proporción muestral es } \hat{p} = \frac{\sum_{i=1}^n X_i}{n} = \frac{\text{Número de éxitos en la muestra}}{n}$$

Donde $X_i = \begin{cases} 1, & \text{éxito} \\ 0, & \text{fracaso} \end{cases}$ Entonces para $n > 30$ tenemos:

Hipótesis: $\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}$ ó $\begin{cases} H_0: p \leq p_0 \\ H_1: p > p_0 \end{cases}$ ó

$$\begin{cases} H_0: p \geq p_0 \\ H_1: p < p_0 \end{cases}$$

Estadístico de prueba: $Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

Reglas de decisiones

Caso A: $\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}$ Si $|Z| > Z_{\alpha/2}$, se rechaza H_0 .

Caso B: $\begin{cases} H_0: p \leq p_0 \\ H_1: p > p_0 \end{cases}$ Si $Z > Z_\alpha$ se rechaza H_0 .

Caso C: $\begin{cases} H_0: p \geq p_0 \\ H_1: p < p_0 \end{cases}$ Si $Z < -Z_\alpha$ se rechaza H_0 .

Donde: $Z_{\alpha/2}$ y Z_α son los valores tabulares de la distribución normal estándar a un nivel de significancia α de dos colas y una cola respectivamente.

GUÍA DE PRÁCTICA N°12

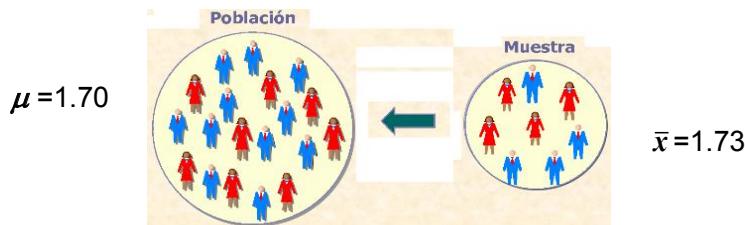
PRUEBA DE HIPÓTESIS EN UNA MUESTRA PARA MEDIA Y PROPORCIÓN POBLACIONAL

Ejemplo1: el promedio poblacional de la altura de los peruanos es 1.70m, esto es:

$$\text{Hipótesis: } \mu = 1.70$$

Ejemplo2: la temperatura de los pacientes del hospital es mayor a 38°C

$$\text{Hipótesis: } \mu \text{ _____}$$



<http://bit.ly/1WuE37s>

I. PRUEBA DE HIPÓTESIS PARA LA MEDIA POBLACIONAL CON VARIANZA CONOCIDA

i) Hipótesis: $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$ ó $\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$ ó $\begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$

ii) Nivel de Significancia α : es el Nivel de riesgo de la comprobación y debe ser bien pequeño. $\alpha = 0.01; 0.05$ y 0.10

iii) Estadístico de prueba: $Z_c = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$

iv) Reglas de decisiones

Caso A: $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$ Si $|Z_c| > Z_{\alpha/2}$, se rechaza H_0 .

Caso B: $\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$ Si $Z_c > Z_\alpha$ se rechaza H_0 .

Caso C: $\begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$ Si $Z_c < -Z_\alpha$ se rechaza H_0 .

II. VALORES DE Z_α y $Z_{\alpha/2}$

α	0.01	0.05	0.10
$Z_{\alpha/2}=$			
$Z_\alpha=$			

III. PRUEBA DE HIPOTESIS PARA UNA MEDIA

1. La estatura media de la población de varones de 15 años de la ciudad de Lima es de 172 cm con una desviación estándar de 8 cm. Una muestra aleatoria de 50 varones de 15 años de edad de un distrito dio una estatura media de 176 cm. ¿Es esto una indicación de que los varones de 15 años de este distrito son más altos que la estatura promedio de los varones de Lima? Pruebe la hipótesis con un nivel de significancia $\alpha = 0.05$.

a) Obtener los datos: $n =$ _____ $\bar{x} =$ _____ $\sigma =$ _____

H₀: _____ Para $\alpha =$ _____ entonces _____

H1: _____

d) Estadístico de prueba $z_c = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} =$ _____

e) Decisión: _____

2. En una pizzería se seleccionó una muestra de 31 repartidores. El tiempo medio (calculado a partir de los datos de la muestra) que requieren para llegar a su destino es 15 minutos con una desviación estándar de 5 minutos. Supóngase que la población de tiempos presenta una distribución normal. ¿Puede concluirse que la media de la población es igual que 17 minutos? con un nivel de significancia $\alpha = 0.01$.

a) Obtener los datos: $n =$ _____ $\bar{x} =$ _____ $\sigma =$ _____

H₀: _____ Para $\alpha =$ _____ entonces _____

H1: _____

d) Estadístico de prueba $z_c = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} =$ _____

e) Decisión: _____

3. De estudios anteriores se sabe que la vida media de los peruanos sigue una distribución normal de media 72 años y desviación típica 8,5 años. En una determinada región, se ha hallado que la vida media de una muestra aleatoria de 90 personas que han fallecido en los tres últimos meses,

es de 69 años, ¿Se puede afirmar que la vida media en esa región ha disminuido con respecto a la media general? Use $\alpha = 0.10$

a) Obtener los datos: $n =$ _____ $\bar{x} =$ _____ $\sigma =$ _____

H_0 : _____ Para $\alpha =$ _____ entonces _____

H1: _____

Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$

d) Estadístico de prueba $z_c = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} =$

e) Decisión: _____

4. En diversos anuncios se afirma que el contenido promedio de nicotina de cierto cigarrillo es de 0.3 miligramos. Al sospechar que esta cifra es demasiado baja, una agencia de protección al consumidor toma una muestra al azar de 35 cigarrillos de diversos lotes de producción y advierte que su contenido de nicotina fue en promedio 0.33 miligramos, con una desviación estándar de 0.018 miligramos. Utilice un nivel de significancia de 0.05 para probar la hipótesis de que el nivel medio de nicotina es mayor a 0.3 miligramos.

a) Obtener los datos: $n =$ _____ $\bar{x} =$ _____ $\sigma =$ _____

Ho: _____ Para $\alpha =$ _____ entonces _____

H1: _____

Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$

d) Estadístico de prueba $z_c = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} =$ _____

e) Decisión: _____

5. El balance promedio de las cuentas de ahorros durante 2013 en el banco Financiero fue de \$1300 con una desviación estándar de \$80. Una muestra aleatoria de 45 cuentas de ahorros promediaron \$1,350 durante 2013. Usando un nivel de significancia $\alpha = 0.10$ ¿podemos concluir que el balance promedio de las cuentas de ahorros durante 2013 difiere del balance de las cuentas de ahorros.

a) Obtener los datos: $n =$ _____ $\bar{x} =$ _____ $\sigma =$ _____

H₀: _____ Para $\alpha =$ _____ entonces _____

H1: _____

Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$

d) Estadístico de prueba

$$z_c = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} =$$

e) Decisión: _____

6. Una encuesta en 64 laboratorios reveló que el precio medio cobrado por realizar cierta prueba es de 50.00 nuevos soles con una desviación estándar de 15.00 nuevos soles ¿proveen estos datos la suficiente información para indicar que la media de la población es mayor que 53.00 nuevos soles? usar un nivel de significancia de $\alpha=0.05$

a) Obtener los datos: $n=$ _____ $\bar{x} =$ _____ $\sigma =$ _____

b) Plantee las hipótesis nula y alterna: c) Nivel de significancia

$H_0:$ _____

Para $\alpha =$ _____ entonces _____

$H_1:$ _____

d) Estadístico de prueba $z_c = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} =$ _____

e) Decisión: _____

7. El tiempo de vida de una olla arrocera sigue una distribución normal con media igual a 8000 horas con una desviación típica de 120 horas de duración. Se escoge al azar una muestra de 60 ollas arroceras de un lote de producción y, después de comprobarlas, se obtiene que su vida media es de 7750 horas. Con un nivel de significación de 0,01, ¿Se rechazaría el lote de producción?

a) Obtener los datos: $n=$ _____ $\bar{x} =$ _____ $\sigma =$ _____

b) Plantee las hipótesis nula y alterna: c) Nivel de significancia

$H_0:$ _____

Para $\alpha =$ _____ entonces _____

$H_1:$ _____

d) Estadístico de prueba $z_c = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} =$ _____

e) Decisión: _____

IV. PRUEBA DE HIPÓTESIS PARA UNA PROPORCIÓN

a) Tipos de Contrastes

Entonces para $n > 30$ tenemos.

$$\text{Hipótesis: } \begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases} \quad \text{ó} \quad \begin{cases} H_0: p \leq p_0 \\ H_1: p > p_0 \end{cases} \quad \text{ó} \quad \begin{cases} H_0: p \geq p_0 \\ H_1: p < p_0 \end{cases}$$

$$\text{Estadístico de prueba: } Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

b) Reglas de Decisión

$$\text{Caso A: } \begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases} \quad \text{Si } |Z| > Z_{\alpha/2}, \text{ se rechaza } H_0$$

$$\text{Caso B: } \begin{cases} H_0: p \leq p_0 \\ H_1: p > p_0 \end{cases} \quad \text{Si } Z > Z_\alpha \text{ se rechaza } H_0.$$

$$\text{Caso C: } \begin{cases} H_0: p \geq p_0 \\ H_1: p < p_0 \end{cases} \quad \text{Si } Z < -Z_\alpha \text{ se rechaza } H_0.$$

8. Se sabe que el 10% de los fumadores prefieren la marca de cigarrillo Malboro. Después de una campaña publicitaria del cigarrillo Malboro, se entrevistaron a 200 fumadores para determinar la eficiencia de la campaña publicitaria. El resultado de la muestra realizada detectó un total de 26 personas que fumaban Malboro. ¿Pueden considerarse que esos datos presentan evidencia suficiente para indicar que hubo un aumento en la aceptación del cigarrillo Malboro. Obtenga las conclusiones del planteamiento desarrollando un contraste de hipótesis con un nivel de significancia del 5 %.

a) Obtener los datos: $n =$ _____ $\hat{p} =$ _____

H₀: _____ Para $\alpha =$ _____ entonces _____

H1: _____

Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$

d) Estadístico de prueba $z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} =$ _____

e) Decisión: _____

9. Históricamente la proporción de clientes que compran con tarjeta de crédito en una determinada tienda es como máximo 25%, sin embargo la dueña de la tienda piensa que esta cifra ha disminuido significativamente. De los últimos 1122 clientes 242 compraron con tarjeta de crédito, si $\alpha = 10\%$. ¿Se está cumpliendo lo que piensa la dueña?

a) Obtener los datos: $n =$ _____ $\hat{p} =$ _____

H0: _____ Para $\alpha =$ _____ entonces _____

H1: _____

Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$

d) Estadístico de prueba $z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} =$ _____

e) Decisión: _____

10. Una agencia de empleos afirma que el 80% de todas las solicitudes hechas por mujeres con hijos prefieren trabajos a tiempo parcial. En una muestra aleatoria de 200 solicitantes mujeres con niños, se encontró que 110 prefirieron trabajos a tiempo parcial. Pruebe la hipótesis de la agencia con un nivel de significancia de 5%.

a) Obtener los datos: $n =$ _____ $\hat{p} =$ _____

H₀: _____ Para $\alpha =$ _____ entonces _____

H1: _____

Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$

d) Estadístico de prueba $z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} =$ _____

e) Decisión: _____

11. Suponga que, en el pasado, 10% de todos los adultos estaba de acuerdo con la pena de muerte. ¿Se tiene razón para creer que la proporción de adultos que apoya la pena de muerte hoy en día ha disminuido; si en una muestra aleatoria de 150 adultos, 20 la favorecen? Usar $\alpha=0.05$

a) Obtener los datos: $n =$ _____ $\hat{p} =$ _____

H₀: _____ Para $\alpha =$ _____ entonces _____

H1: _____ ^

d) Estadístico de prueba $z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} =$ _____

e) Decisión: _____

12. En una muestra de 1500 residentes de un barrio, quienes participaron en un programa de concurso 125 participantes fueron elegidos ¿proporcionan estos datos la evidencia suficientes para indicar que la proporción de individuos seleccionados en la población muestreada es mayor que 0.06? Sea $\alpha=0.05$.

a) Obtener los datos: $n =$ _____ $\hat{p} =$ _____

H₀: _____ Para $\alpha =$ _____ entonces _____

H1: _____

d) Estadístico de prueba $z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} =$ _____

e) Decisión: _____

13. Un sociólogo ha pronosticado, que en una determinada ciudad, el nivel de abstención en las próximas elecciones será del 40% como mínimo. Se elige al azar una muestra aleatoria de 200 individuos, con derecho a voto, 75 de los cuales estarían dispuestos a votar. Determinar con un nivel de significación del 0.10, si se puede admitir el pronóstico.

a) Obtener los datos: $n =$ _____ $\hat{p} =$ _____

H₀: _____ Para $\alpha =$ _____ entonces _____

H1: _____ ^

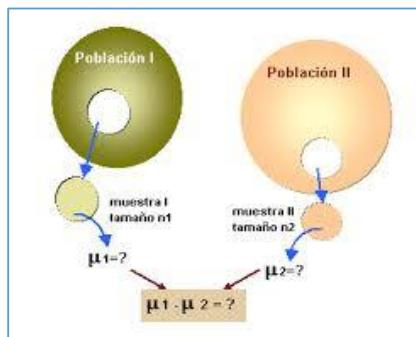
d) Estadístico de prueba $z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} =$ _____

e) Decisión: _____

4.2. PRUEBA DE HIPÓTESIS EN DOS MUESTRAS INDEPENDIENTES PARA MEDIA Y PROPORCIÓN POBLACIONAL

4.2.1. PRUEBA DE HIPÓTESIS PARA LA DIFERENCIA DE MEDIAS

La prueba de hipótesis para la diferencia de medias de dos poblaciones es un problema muy frecuente en todas las áreas que se sirven de la estadística como instrumento de trabajo. Así un administrador puede estar interesado en averiguar la diferencia entre dos técnicas de ventas, lo cual conseguirá contrastando la diferencia de los promedio de ventas obtenidos



con cada técnica; un docente puede estar interesado en la eficacia de un nuevo método de enseñanza, para lo cual ensayará la diferencia de las medias de las calificaciones obtenidas por un grupo de alumnos a los que ha aplicado el nuevo método y otro grupo de alumnos con lo que utilizó un método clásico de enseñanza.

<http://bit.ly/1WhbdYf>

En estos casos hay un modelo común de trabajo, que consiste en seleccionar dos muestras, una formada por individuos de la población en los que se va a ensayar la nueva experiencia y otra segunda muestra a la que se aplica el método clásico y que se utiliza para contrastar los resultados.

PRUEBA DE HIPÓTESIS PARA DIFERENCIAS DE MEDIAS DE POBLACIONES NORMALES INDEPENDIENTES CON VARIANZAS σ_1^2 y σ_2^2 CONOCIDAS

Sean X_1, X_2, \dots, X_{n_1} una muestra aleatoria extraída de una población Normal $N(\mu_1 ; \sigma_1^2)$ y Y_1, Y_2, \dots, Y_{n_2} otra m. a. extraída de una población Normal $N(\mu_2 ; \sigma_2^2)$ donde σ_1^2 y σ_2^2 son varianzas poblacionales conocidos. Supóngase también que las poblaciones son independientes.

Entonces tenemos.

Hipótesis: $\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$ ó $\begin{cases} H_0: \mu_1 \leq \mu_2 \\ H_1: \mu_1 > \mu_2 \end{cases}$ ó $\begin{cases} H_0: \mu_1 \geq \mu_2 \\ H_1: \mu_1 < \mu_2 \end{cases}$

Estadístico de prueba: $Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

Donde

$\overline{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$ es la media muestral de la primera variable X

$\overline{X}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$ es la media muestral de la segunda variable Y

n_1 es el tamaño muestral tomada de la primera Población X

n_2 es el tamaño muestral tomada de la segunda Población Y

Reglas de decisiones

Caso A: $\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$ Si $|Z| > Z_{\alpha/2}$, se rechaza H_0 .

Caso B: $\begin{cases} H_0: \mu_1 \leq \mu_2 \\ H_1: \mu_1 > \mu_2 \end{cases}$ Si $Z > Z_\alpha$ se rechaza H_0 .

Caso C: $\begin{cases} H_0: \mu_1 \geq \mu_2 \\ H_1: \mu_1 < \mu_2 \end{cases}$ Si $Z < -Z_\alpha$ se rechaza H_0 .

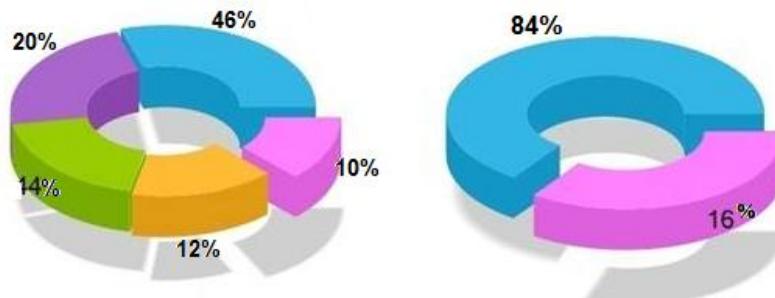
Donde

$Z_{\alpha/2}$ y Z_α son los valores tabulares de la Distribución Normal Estándar a un nivel de significancia α de dos colas y una cola respectivamente.

4.2.2. PRUEBA DE HIPÓTESIS PARA DIFERENCIA DE PROPORCIONES

En algunos diseños de investigación, el plan muestral requiere seleccionar dos muestras independientes, calcular las proporciones muestrales y usar la diferencia de las dos proporciones para estimar o probar una diferencia entre las mismas. Las aplicaciones son similares a la diferencia de medias.

Por ejemplo, si dos empresas ofrecen datos de proporciones de empleados que tienen maestría y al hacer dos estudios diferentes salen resultados ligeramente diferentes ¿pero qué tanta diferencia se requiere para que sea estadísticamente significativo? de eso se tratan las pruebas estadísticas de diferencias de proporciones.



<http://bit.ly/1QZ6JnL>

Definición: Sean X_1, X_2, \dots, X_{n_1} una muestra aleatoria extraída de una población Bernouilli $B(1; p_1)$ y Y_1, Y_2, \dots, Y_{n_2} otra muestra aleatoria extraída de una población Bernoulli $B(1; p_2)$. Supóngase también que las poblaciones son independientes. Sean:

$$\hat{p}_1 = \frac{\sum_{i=1}^{n_1} X_i}{n_1} = \frac{\text{Número de éxitos en la muestra}}{n_1}, \quad X_i = \begin{cases} 1, & \text{éxito} \\ 0, & \text{fracaso} \end{cases}$$

$$\hat{p}_2 = \frac{\sum_{i=1}^{n_2} Y_i}{n_2} = \frac{\text{Número de éxitos en la muestra}}{n_2}, \quad Y_i = \begin{cases} 1, & \text{éxito} \\ 0, & \text{fracaso} \end{cases}$$

Entonces tenemos:

Hipótesis: $\begin{cases} H_0: p_1 = p_2 \\ H_1: p_1 \neq p_2 \end{cases}$ ó $\begin{cases} H_0: p_1 \leq p_2 \\ H_1: p_1 > p_2 \end{cases}$ ó $\begin{cases} H_0: p_1 \geq p_2 \\ H_1: p_1 < p_2 \end{cases}$

Estadístico de prueba: $Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \quad \hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n}, \quad n = n_1 + n_2$

Reglas de decisiones

Caso A: $\begin{cases} H_0: p_1 = p_2 \\ H_1: p_1 \neq p_2 \end{cases}$ Si $|Z| > Z_{\alpha/2}$, se rechaza H_0 .

Caso B: $\begin{cases} H_0: p_1 \leq p_2 \\ H_1: p_1 > p_2 \end{cases}$ Si $Z > Z_\alpha$ se rechaza H_0 .

Caso C: $\begin{cases} H_0: p_1 \geq p_2 \\ H_1: p_1 < p_2 \end{cases}$ Si $Z < -Z_\alpha$ se rechaza H_0 .

Donde

$Z_{\alpha/2}$ y Z_α son los valores tabulares de la distribución normal estándar a un nivel de significancia α de dos colas y una cola respectivamente.

GUÍA DE PRÁCTICA N°13

PRUEBA DE HIPÓTESIS PARA DIFERENCIA DE MEDIAS Y PROPORCIONES

I. HIPÓTESIS PARA DIFERENCIA DE MEDIAS

1. Para comparar dos métodos de enseñanza de matemáticas, se aplicaron a 200 alumnos elegidos al azar el método tradicional y a otra muestra de 250 alumnos el método nuevo resultando las calificaciones promedio de 13 y 15 respectivamente. Suponga que las varianzas poblacionales respectivas son 9 y 16. Usando un nivel de significación del 5%, ¿podemos afirmar que el método nuevo es superior al método antiguo?

a) Obtener los datos: $n_1 = \underline{\hspace{2cm}}$ $\bar{x}_1 = \underline{\hspace{2cm}}$ $\sigma_1 = \underline{\hspace{2cm}}$

$n_2 = \underline{\hspace{2cm}}$ $\bar{x}_2 = \underline{\hspace{2cm}}$ $\sigma_2 = \underline{\hspace{2cm}}$

b) Plantee las hipótesis nula y alterna: c) Nivel de significancia

$H_0: \underline{\hspace{2cm}}$ Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$

$H_1: \underline{\hspace{2cm}}$

d) Estadístico de prueba

$$Z_c = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \underline{\hspace{2cm}}$$

e) Decisión: $\underline{\hspace{2cm}}$

2. Se selecciona una muestra de 40 habitaciones de un hotel, resultando una media de 102 turistas y una desviación estándar de 5. Otra muestra de 50 habitaciones se selecciona de un segundo hotel resultando una media de 99 turistas y una desviación estándar de 6. Utilizando un nivel de significación del 4 % ¿se puede indicar que el promedio de turistas en ambos hoteles son diferentes?

a) Obtener los datos: $n_1 = \underline{\hspace{2cm}}$ $\bar{x}_1 = \underline{\hspace{2cm}}$ $\sigma_1 = \underline{\hspace{2cm}}$

$n_2 = \underline{\hspace{2cm}}$ $\bar{x}_2 = \underline{\hspace{2cm}}$ $\sigma_2 = \underline{\hspace{2cm}}$

b) Plantee las hipótesis nula y alterna: c) Nivel de significancia

$H_0: \underline{\hspace{2cm}}$ Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$

$H_1: \underline{\hspace{2cm}}$

d) Estadístico de prueba

$$Z_c = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \underline{\hspace{2cm}}$$

e) Decisión: $\underline{\hspace{2cm}}$

3. Un fabricante quiere comparar los tiempos de proceso de dos marcas de máquinas A y B, para fabricar un tipo de artículo. Al observar dos muestras aleatorias de 60 artículos procesados por A y B respectivamente, encuentra que las medias respectivas son 1,230 y 1,190 segundos. Suponga $\sigma_A = 120$ y $\sigma_B = 90$ segundos. Al nivel de significación de 5%, ¿se puede inferir que la máquina A es más rápida que la máquina B?

a) Obtener los datos: $n_1 = \underline{\hspace{2cm}}$ $\bar{x}_1 = \underline{\hspace{2cm}}$ $\sigma_1 = \underline{\hspace{2cm}}$
 $n_2 = \underline{\hspace{2cm}}$ $\bar{x}_2 = \underline{\hspace{2cm}}$ $\sigma_2 = \underline{\hspace{2cm}}$

b) Plantee las hipótesis nula y alterna: $H_0: \underline{\hspace{2cm}}$ c) Nivel de significancia
 $H_1: \underline{\hspace{2cm}}$ Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$

d) Estadístico de prueba $Z_c = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \underline{\hspace{2cm}}$

e) Decisión: $\underline{\hspace{2cm}}$

4. Una muestra aleatoria de 100 profesores de universidades particulares mostró que en 9 meses de trabajo obtuvieron un salario promedio de 1600 mensuales con una desviación estándar de 150 mientras que una muestra de 200 profesores de universidades públicas mostró un salario de 1520 con una desviación estándar de 140. Prueba la hipótesis de que el salario promedio de los profesores de universidades públicas es menor que el salario de profesores de universidades particulares.

a) Obtener los datos: $n_1 = \underline{\hspace{2cm}}$ $\bar{x}_1 = \underline{\hspace{2cm}}$ $\sigma_1 = \underline{\hspace{2cm}}$
 $n_2 = \underline{\hspace{2cm}}$ $\bar{x}_2 = \underline{\hspace{2cm}}$ $\sigma_2 = \underline{\hspace{2cm}}$

b) Plantee las hipótesis nula y alterna: $H_0: \underline{\hspace{2cm}}$ c) Nivel de significancia
 $H_1: \underline{\hspace{2cm}}$ Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$

d) Estadístico de prueba $Z_c = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \underline{\hspace{2cm}}$

e) Decisión: $\underline{\hspace{2cm}}$

5. Se selecciona una muestra de 40 habitaciones de un hotel, resultando una media de 102 turistas y una desviación estándar de 5. Otra muestra de 50 habitaciones se selecciona de un segundo hotel resultando una media de 99 turistas y una desviación estándar de 6. Utilizando un nivel de significación del 4 % se puede indicar ¿que las medias de los turistas son diferentes?

a) Obtener los datos: $n_1 = \underline{\hspace{2cm}}$ $\bar{x}_1 = \underline{\hspace{2cm}}$ $\sigma_1 = \underline{\hspace{2cm}}$
 $n_2 = \underline{\hspace{2cm}}$ $\bar{x}_2 = \underline{\hspace{2cm}}$ $\sigma_2 = \underline{\hspace{2cm}}$

b) Plantee las hipótesis nula y alterna: $H_0: \underline{\hspace{2cm}}$ c) Nivel de significancia
 $H_1: \underline{\hspace{2cm}}$ Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$

d) Estadístico de prueba $Z_c = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \underline{\hspace{2cm}}$

e) Decisión: $\underline{\hspace{2cm}}$

6. Para medir la calidad del aire de cierta zona industrial, con relación a los óxidos de azufre, se sacaron dos muestras de tamaños 50 y 75, respectivamente. Los promedios fueron de 76 mg/L y de 82 mg/L, respectivamente. Asumir que las varianzas de estas poblaciones son conocidas e iguales a 16. Asumir un nivel de significancia de $\alpha = 0.05$. Usando el valor de p , probar que no hay diferencias entre las dos poblaciones muestreadas.

a) Obtener los datos: $n_1 = \underline{\hspace{2cm}}$ $\bar{x}_1 = \underline{\hspace{2cm}}$ $\sigma_1 = \underline{\hspace{2cm}}$
 $n_2 = \underline{\hspace{2cm}}$ $\bar{x}_2 = \underline{\hspace{2cm}}$ $\sigma_2 = \underline{\hspace{2cm}}$

b) Plantee las hipótesis nula y alterna: $H_0: \underline{\hspace{2cm}}$ c) Nivel de significancia
 $H_1: \underline{\hspace{2cm}}$ Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$

d) Estadístico de prueba $Z_c = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \underline{\hspace{2cm}}$

e) Decisión: $\underline{\hspace{2cm}}$

7. En la facultad de Administración se seleccionó una muestra aleatoria de 20 estudiantes (grupo A) de una población de estudiantes cuyos padres trabajan. Se seleccionó también una muestra aleatoria de 16 estudiantes (grupo B) entre aquellos estudiantes en que solamente el padre trabaja. El análisis de los puntajes de rendimiento académico de los dos grupos dio los siguientes resultados:

Grupo A	Grupo B
$\bar{X}_A = 14$	$\bar{X}_B = 17$

La experiencia muestra que las poblaciones de puntajes para ambos grupos están distribuidas en forma aproximadamente normal, con varianzas $\sigma_A^2 = 36$ y $\sigma_B^2 = 20$. ¿Se puede concluir, con estos datos, que la media de la población de la que se seleccionó el grupo A es inferior a la media de la población de la que se seleccionó el grupo B?

II. PRUEBA DE HIPÓTESIS PARA DIFERENCIA DE PROPORCIONES

8. Un analista realizó un estudio para comparar la efectividad de dos marcas de medicina A y B, que se aplica para curar cierta enfermedad. Para esto, durante 8 días se suministró a pacientes que sufren la enfermedad: la medicina A en una muestra de 300 y la medicina B a una muestra de 400 pacientes. Si resultó efectiva para 270 y 320 pacientes de cada muestra respectivamente, ¿Se puede inferir que la medicina A es más efectiva que la B? Utilice $\alpha=0.05$.

a) Obtener los datos: $n_1 = \underline{\hspace{2cm}}$ $\hat{p}_1 = \underline{\hspace{2cm}}$ $n_2 = \underline{\hspace{2cm}}$ $\hat{p}_2 = \underline{\hspace{2cm}}$ $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n} = \underline{\hspace{2cm}}$

b) Plantee las hipótesis nula y alterna: c) Nivel de significancia
 $H_0: \underline{\hspace{2cm}}$ Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$
 $H_1: \underline{\hspace{2cm}}$

d) Estadístico de prueba $Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \underline{\hspace{2cm}}$

e) Decisión: _____

9. Un grupo de investigadores del Ministerio de Educación afirman que en Lima, la proporción de hombres que recibieron educación primaria es igual a la de mujeres. Para probar su afirmación los investigadores tomaron una muestra aleatoria de 1722 hombres, de los cuales 411 recibieron educación primaria y una muestra aleatoria de 1572 mujeres, de las cuales 393 recibieron educación primaria. En base a los datos ¿Se puede decir que los investigadores tenían razón? Utilice $\alpha=0.01$.

a) Obtener los datos: $n_1 = \underline{\hspace{2cm}}$ $\hat{p}_1 = \underline{\hspace{2cm}}$ $n_2 = \underline{\hspace{2cm}}$ $\hat{p}_2 = \underline{\hspace{2cm}}$ $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n} = \underline{\hspace{2cm}}$

b) Plantee las hipótesis nula y alterna: c) Nivel de significancia
 $H_0: \underline{\hspace{2cm}}$ Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$
 $H_1: \underline{\hspace{2cm}}$

d) Estadístico de prueba $Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \underline{\hspace{2cm}}$

e) Decisión: _____

10. Una prueba de 100 jóvenes y 200 adultos muestra que 50 de los jóvenes 80 de los adultos son conductores descuidados de vehículos. Probar la hipótesis para la afirmación de que el porcentaje de jóvenes choferes descuidados es mayor que el correspondiente a los adultos. Utilice $\alpha=0.02$

a) Obtener los datos: $n_1 = \underline{\hspace{2cm}}$ $\hat{p}_1 = \underline{\hspace{2cm}}$
 $n_2 = \underline{\hspace{2cm}}$ $\hat{p}_2 = \underline{\hspace{2cm}}$ $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n} = \underline{\hspace{2cm}}$

b) Plantee las hipótesis nula y alterna:
 $H_0: \underline{\hspace{2cm}}$ c) Nivel de significancia
 $H_1: \underline{\hspace{2cm}}$ Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$

d) Estadístico de prueba $Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \underline{\hspace{2cm}}$

e) Decisión: $\underline{\hspace{2cm}}$

11. Se va a efectuar una encuesta sobre habitación en San Isidro y en Miraflores para determinar la proporción de unidades habitacionales ocupadas por familias de ingresos altos. Una muestra aleatoria de 600 unidades habitacionales en San Isidro reveló 150 unidades ocupadas por familias de ingresos altos. Una muestra de 300 unidades en Miraflores reveló 120 unidades ocupadas por familias de ingresos altos. ¿Existe alguna diferencia entre San Isidro y Miraflores en la proporción de unidades habitacionales ocupadas por familias de ingresos altos? Utilice $\alpha=0.10$

a) Obtener los datos: $n_1 = \underline{\hspace{2cm}}$ $\hat{p}_1 = \underline{\hspace{2cm}}$
 $n_2 = \underline{\hspace{2cm}}$ $\hat{p}_2 = \underline{\hspace{2cm}}$ $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n} = \underline{\hspace{2cm}}$

b) Plantee las hipótesis nula y alterna:
 $H_0: \underline{\hspace{2cm}}$ c) Nivel de significancia
 $H_1: \underline{\hspace{2cm}}$ Para $\alpha = \underline{\hspace{2cm}}$ entonces $\underline{\hspace{2cm}}$

d) Estadístico de prueba $Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \underline{\hspace{2cm}}$

e) Decisión: $\underline{\hspace{2cm}}$

12. Un médico ha sugerido que un ataque cardíaco es menos probable que ocurra en hombres que practican alguna clase de deporte. Se elige una muestra al azar de 300 hombres, de los cuales 100 practican alguna clase de deporte y de ellos sólo 10 han sufrido un ataque cardíaco. De los 200 que no practican deportes, 25 han sufrido ataques cardíacos. Probar si los resultados de las muestras apoyan lo sugerido por el médico.

a) Obtener los datos: $n_1 = \underline{\hspace{2cm}}$ $\hat{p}_1 = \underline{\hspace{2cm}}$
 $n_2 = \underline{\hspace{2cm}}$ $\hat{p}_2 = \underline{\hspace{2cm}}$ $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n} = \underline{\hspace{2cm}}$

b) Plantee las hipótesis nula y alterna: c) Nivel de significancia
H₀: _____ Para $\alpha = \underline{\hspace{2cm}}$ entonces _____
H₁: _____

d) Estadístico de prueba $Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \underline{\hspace{2cm}}$

e) Decisión: _____

13. Yahoo hizo una encuesta para determinar el porcentaje de personas que usaban Internet en el trabajo:

- En México se encontró que el 40% de los adultos usa Internet de una muestra de 240.
- En Monterrey el 32% de los adultos usaba Internet de una muestra de 250.

¿Es mayor la proporción que usa Internet en México que en Monterrey?

a) Obtener los datos: $n_1 = \underline{\hspace{2cm}}$ $\hat{p}_1 = \underline{\hspace{2cm}}$
 $n_2 = \underline{\hspace{2cm}}$ $\hat{p}_2 = \underline{\hspace{2cm}}$ $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n} = \underline{\hspace{2cm}}$

b) Plantee las hipótesis nula y alterna: c) Nivel de significancia
H₀: _____ Para $\alpha = \underline{\hspace{2cm}}$ entonces _____
H₁: _____

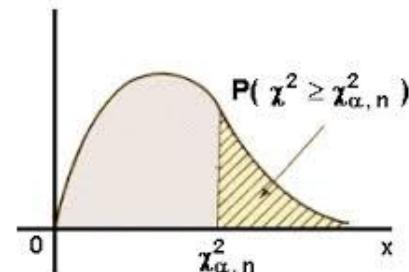
d) Estadístico de prueba $Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \underline{\hspace{2cm}}$

e) Decisión: _____

4.3. PRUEBA DE INDEPENDENCIA CHI-CUADRADO DE DOS VARIABLES CUALITATIVAS

4.3.1. PRUEBAS NO PARAMÉTRICAS

En este capítulo presentamos el uso de la distribución Chi-Cuadrado en contrastes no paramétricos, esto es, aquellos que no dependen de los parámetros poblacionales ni de sus respectivos estimadores. Este tipo de pruebas frecuentemente ocurre, cuando el análisis se basa en conteos o frecuencias y no en medidas tales como metros, kilogramos o puntajes de pruebas.



<http://bit.ly/1KJDEgA>

4.3.2. PRUEBA CHI-CUADRADO PARA LA INDEPENDENCIA DE DOS VARIABLES

La Prueba de Independencia consiste en comprobar si dos variables cualitativas están relacionadas entre sí (por ejemplo: ¿el color de ojos está relacionado con el color de los cabellos?). Este tipo de contrastes se aplica cuando deseamos comparar una variable en dos situaciones o poblaciones diferentes, i.e., deseamos estudiar si existen diferencias en las dos poblaciones respecto a la variable de estudio.

i. TABLA DE CONTINGENCIA

Es una tabla estadística en la que cada observación de la muestra es clasificada en dos o más niveles de categorías.

TABLA DE CONTINGENCIA DE LAS VARIABLES CUALITATIVAS X E Y					
Categorías de X	Categorías de Y				Total
	C_{Y1}	C_{Y2}	C_{Yk}	
C_{X1}	O_{11} (e_{11})	O_{12} (e_{12})	O_{1k} (e_{1k})	O_{X1}
C_{X2}	O_{21} (e_{21})	O_{22} (e_{22})	O_{2k} (e_{2k})	O_{X2}
.
C_{Xm}	O_{m1} (e_{m1})	O_{m2} (e_{m2})	O_{mk} (e_{mk})	O_{Xm}
Subtotales	O_{Y1}	O_{Y2}	O_{Yk}	n

Donde

O_{ij} son las Frecuencias observadas

e_{ij} son las frecuencias esperadas

$$O_{X_i} = \sum_{j=1}^k O_{ij} = \text{nº de veces que se presenta el nivel } x_i, i = 1, 2, \dots, m$$

$$O_{Y_j} = \sum_{i=1}^m O_{ij} = \text{nº de veces que se presenta el nivel } y_i, i = 1, 2, \dots, k$$

$$n = \sum_{i=1}^m O_{X_i} = \sum_{j=1}^k O_{Y_j} = \sum_{i=1}^m \sum_{j=1}^k O_{ij} = \text{tamaño de la muestra}$$

CONTRASTE DE INDEPENDENCIA

Con frecuencia un investigador está interesado en saber si dos métodos de clasificación son independientes o probablemente están relacionadas. Se dice que dos métodos de categorización son independientes si la distribución de un método no depende de la distribución del otro. Para el contraste de independencia seguiremos los siguientes pasos:

A. HIPÓTESIS

H₀: los dos métodos de clasificación son independientes

H₁: los dos métodos de clasificación no son independientes

Que también se puede expresar:

H₀: no existe relación entre los dos métodos de clasificación

H₁: existe relación entre los dos métodos de clasificación

B. ESTADÍSTICO DE PRUEBA A USAR

Dada la tabla de contingencia

TABLA DE CONTINGENCIA DE LAS VARIABLES CUALITATIVAS X E Y					
<i>Categorías de X</i>	<i>Categorías de Y</i>				<i>Total</i>
	<i>C_{Y1}</i>	<i>C_{Y2}</i>	<i>C_{Yk}</i>	
<i>C_{X1}</i>	<i>O₁₁</i> (<i>e₁₁</i>)	<i>O₁₂</i> (<i>e₁₂</i>)	<i>O_{1k}</i> (<i>e_{1k}</i>)	<i>O_{X1}</i>
<i>C_{X2}</i>	<i>O₂₁</i> (<i>e₂₁</i>)	<i>O₂₂</i> (<i>e₂₂</i>)	<i>O_{2k}</i> (<i>e_{2k}</i>)	<i>O_{X2}</i>
.
<i>C_{Xm}</i>	<i>O_{m1}</i> (<i>e_{m1}</i>)	<i>O_{m2}</i> (<i>e_{m2}</i>)	<i>O_{mk}</i> (<i>e_{mk}</i>)	<i>O_{Xm}</i>
<i>Subtotales</i>	<i>O_{Y1}</i>	<i>O_{Y2}</i>	<i>O_{Yk}</i>	<i>n</i>

Donde

O_{ij} son las Frecuencias observadas

e_{ij} son las frecuencias esperadas; donde $e_{ij} = \frac{O_{Xi} \cdot O_{Yj}}{n}$

Entonces el estadístico de prueba es

$$\chi_o^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(O_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^m \sum_{j=1}^k \frac{O_{ij}^2}{e_{ij}} - n$$

C. DECISIÓN: Si $\chi_o^2 > \chi_{(\alpha, (m-1)(k-1))}^2 = \chi_\alpha^2$ se rechaza H_0 ,

Donde :

χ_o^2 es denominado valor calculado

α es el nivel de significancia a considerar por ejemplo **0.05**

$\chi_{(\alpha, (m-1)(k-1))}^2 = \chi_\alpha^2$ es el valor tabular correspondiente a la distribución Chi-

Cuadrada con

(m - 1)(k - 1) grados de libertad con nivel de significancia α

m es el número de filas de la tabla de contingencia

k es el número de columnas de la tabla de contingencia

Observaciones finales:

Cuando empleamos la variable aleatoria χ^2 en un contraste de hipótesis debemos tener en cuenta las siguientes consideraciones:

- Para que la variable aleatoria χ^2 tenga una buena aproximación a la distribución Chi cuadrado es necesario que las frecuencias esperadas de las distintas categorías no sea inferior a 5 (es decir, debemos tener $e_{ij} \geq 5$)
- Si hay alguna categoría que tiene una frecuencia esperada menor que cinco se agrupan dos o más categorías en una sola hasta lograr que la nueva frecuencia esperada sea mayor o igual que cinco.
- Se puede aplicar la prueba Chi cuadrado en situaciones en las que deseamos decidir si una serie de observaciones se ajustan o no a una distribución teórica previamente determinada que puede ser binomial, Poisson, exponencial, normal o hipotética

Ejemplo1. Verificar si existe o no independencia entre el sexo y carrera profesional escogida por 400 alumnos de Economía y Administración de cierta universidad al nivel de 5%. Los datos se dan en la siguiente tabla de contingencia:

Carrera	Sexo		Total
	Masculino	Femenino	
Economía	36	11	47
Administración	14	19	33
Total	50	30	80

SOLUCIÓN:

i) **Hipótesis**

H_0 : El sexo y la carrera son independientes (no tienen relación)

H_1 : El sexo y la carrera no son independientes (están relacionados)

ii) **Calculando las frecuencias esperadas:**

$$e_{11} = \frac{47 \times 50}{80} = 29.38 \quad e_{12} = \frac{47 \times 30}{80} = 17.62$$

$$e_{21} = \frac{33 \times 50}{80} = 20.62 \quad e_{22} = \frac{33 \times 30}{80} = 12.38$$

iii) Con lo anterior podemos formar la siguiente tabla de contingencia 2x2

Carrera	Sexo		Total
	Masculino	Femenino	
Economía	36 (29.38)	11 (17.63)	47
Administración	14 (20.62)	19 (12.38)	33
Total	50	30	80

iv) Calculamos ahora el valor de χ^2 con nuestros datos según el estadístico de prueba:

$$\chi_o^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(O_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^m \sum_{j=1}^k \frac{o_{ij}^2}{e_{ij}} - n$$

$$\chi^2 = \frac{(36 - 29.38)^2}{29.38} + \frac{(14 - 20.62)^2}{20.62} + \frac{(11 - 17.63)^2}{17.63} + \frac{(19 - 12.38)^2}{12.38} = 9.659$$

v) Hallamos en la tabla el valor crítico $\chi^2_{(\alpha, (m-1)(k-1))} = \chi^2_{\alpha}$ para $\alpha = 0.05$. Para nuestro caso

$m=2$ filas y $k=2$ columnas, por lo cual los grados de libertad = $(2-1)(2-1)=1$, entonces buscaremos en la tabla con $\chi^2_{0.05,1} = 3.84$

vi) Decisión:

Como $\chi^2 = 9.659 > 3.84$ entonces rechazamos H_0 y decimos que El sexo y la carrera no son independientes (están relacionados)

Ejemplo 2: En un trabajo de Investigación se tiene que su Hipótesis central es:

i. Hipótesis

H_0 : No existe una relación entre la Gestión de los Directores y el Clima Laboral de los docentes en las Instituciones Educativas

H_1 : Si existe una relación entre la Gestión de los Directores y el Clima Laboral de los docentes en las Instituciones Educativas.

Ahora determinaremos el valor calculado valor del estadístico de prueba a partir de la siguiente tabla de contingencia

TABLA DE CONTINGENCIA DE LA GESTIÓN DE LOS DIRECTORES Y EL CLIMA LABORAL DE LOS DOCENTES EN LAS I. E. DE NIVEL SECUNDARIA			NIVEL DE CALIDAD DEL CLIMA LABORAL		Total
GESTIÓN DE DIRECTORES	NIVELES DE CALIDAD		Bueno	Regular o malo	Subtotales
	Bueno	Frecuencias Observadas: O_{ij} (Frecuencias Esperadas e_{ij})	21 (12.5)	8 (16.5)	29
	Regular o Malo	Frecuencias Observadas: O_{ij} (Frecuencias Esperadas e_{ij})	4 (12.5)	25 (16.5)	29
Subtotales		25	33	58	

ii. Estadístico de prueba

De la tabla de contingencia tenemos que

$$\chi^2_o = \sum_{i=1}^m \sum_{j=1}^k \frac{O_{ij}^2}{e_{ij}} - n = \frac{21^2}{12.5} + \frac{8^2}{16.5} + \frac{4^2}{12.5} + \frac{25^2}{6.5} - 58 = 20.318$$

Entonces

$$\chi_o^2 = 20.318.$$

El valor tabular de la Ch-Cuadrada con 1 grado de libertad y a un nivel de significancia

$$\text{de } \alpha = 0.05 \text{ es } \chi_{\alpha}^2 = \chi_{(0.05, 1)}^2 = 3.841$$

iii. Decisión

Como $\chi_o^2 = 20.318 > \chi_{\alpha}^2 = 3.841$, rechazamos H_0 , es decir que entre la gestión de los directores y el clima laboral de los docentes en las Instituciones Educativas de nivel secundario de la zona urbana de Andahuaylas existe una relación estadísticamente significativa.

GUÍA DE PRÁCTICA N°14

PRUEBA DE INDEPENDENCIA CHI-CUADRADO DE DOS VARIABLES

CUALITATIVAS

- ESSALUD desea verificar si existe relación entre el estado nutricional de los niños las ciudades del sur del Perú, para lo cual toma una muestra de niños de dos ciudades y los clasifica según su estado nutricional obteniendo la siguiente tabla. Pruebe la hipótesis correspondiente con un nivel de significancia de 0.01

ESTADO NUTRICIONAL	Ciudad		Total
	Arequipa	Puno	
Sobrepeso	93 ()	53 ()	
Normal	25 ()	21 ()	
Total			

i) Hipótesis

H₀: _____

H₁: _____

ii) Estadístico de prueba

$$\chi_o^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{o_{ij}^2}{e_{ij}} - n =$$

iii) Nivel de significancia: $\alpha =$ _____ entonces: $\chi^2_{(\dots, \dots)} =$ _____

iv) Decisión: como $\chi_o^2 =$ _____ “que $\chi^2_{\alpha} =$ _____ Entonces _____ H_o
y decimos que _____

- Una compañía de venta de libros quiere saber si el volumen de ventas de sus distribuidores es independiente del carácter, de los mismos. Para ello recoge los siguientes datos de una muestra de vendedores. Realiza un test de independencia a nivel de significación de $\alpha=0,01$

Carácter	Ventas			
	Bajo	Medio	Alto	Total
Antipáticos	38 ()	29 ()	9 ()	
Simpáticos	32 ()	59 ()	4 ()	
Total				

¿Presentan estos datos suficiente evidencia de que el salario mensual depende de la educación lograda? A un nivel de confianza del 90%

i) Hipótesis

H₀: _____

H₁: _____

ii) Estadístico de prueba

$$\chi_o^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{o_{ij}^2}{e_{ij}} - n =$$

iii) Nivel de significancia: $\alpha =$ _____ entonces: $\chi^2_{(.....;)} =$ _____

iv) Decisión: como $\chi_o^2 =$ _____ “que $\chi_\alpha^2 =$ _____ Entonces _____ H_o

y decimos que _____

3. La calificación final de los estudiantes de un curso de estadística se clasifica por carreras. ¿Se podría concluir que existe una asociación entre la carrera y la calificación final con un nivel de significación del 0,01?

Condición	Carrera			Total
	Psicología	Administración	Economía	
Sobresaliente	20 ()	34 ()	30 ()	
Aprobado	22 ()	8 ()	13 ()	
Desaprobado	6 ()	4 ()	9 ()	
Total				

i) Hipótesis

H₀: _____

H₁: _____

ii) Estadístico de prueba

$$\chi_o^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{o_{ij}^2}{e_{ij}} - n =$$

iii) Nivel de significancia: $\alpha =$ _____ entonces: $\chi^2_{(.....;)} =$ _____

iv) Decisión: como $\chi_o^2 =$ _____ “que $\chi_\alpha^2 =$ _____ Entonces _____ H_o

y decimos que _____

4. Al final de un semestre, las calificaciones de matemáticas fueron tabuladas en la siguiente tabla de contingencia de 3×2 para estudiar la relación entre la asistencia a clase y la calificación obtenida.

Ausencias	Condición		Total
	Aprobado	No aprobado	
0 - 3	135	110	
4 - 6	36	4	
7 - 45	9	6	
Total			

Con $\alpha = 0.05$, ¿indican los datos que son independientes la asistencia a clase y la calificación obtenida?

i) **Hipótesis**

$H_0:$ _____

$H_1:$ _____

ii) **Estadístico de prueba**

$$\chi_o^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{o_{ij}^2}{e_{ij}} - n =$$

iii) **Nivel de significancia:** $\alpha =$ _____ entonces: $\chi^2_{(\dots, \dots)} =$ _____

iv) **Decisión:** como $\chi_o^2 =$ _____ “que $\chi_\alpha^2 =$ _____ Entonces _____ H_o

y decimos que _____

5. Un investigador clasificó en forma cruzada a 355 niños de una escuela primaria de acuerdo con su grupo socioeconómico y la presencia o ausencia de un defecto congénito. Con base en estos datos, ¿Podríamos concluir que los efectos congénitos están relacionados con la posición económica a un nivel de confianza del 99%?

Defecto Congénito	Grupo Socioeconómico			Total
	Alto	Medio	Bajo	
Presente	4 ()	32 ()	35 ()	
Ausente	46 ()	138 ()	100 ()	
Total				

i) Hipótesis

H₀: _____

H₁: _____

ii) Estadístico de prueba

$$\chi_o^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{O_{ij}^2}{e_{ij}} - n =$$

iii) Nivel de significancia: $\alpha =$ _____ entonces: $\chi^2_{(\dots, \dots)} =$ _____

iv) Decisión: como $\chi_o^2 =$ _____ “que $\chi_\alpha^2 =$ _____ Entonces _____ H_o

y decimos que _____

6. La siguiente tabla muestra la distribución de una muestra aleatoria de 400 truchas cafés de un gran río., según la longitud y el sector donde fueron extraídas. Pruebe la hipótesis de que existe alguna relación entre la longitud de las truchas y el sector del río donde fueron extraídas, usando $\alpha = 0.05$.

Longitud	Sector del río			Total
	Alto	Centro	Bajo	
Bajo el promedio	67 ()	64 ()	25 ()	
Promedio	42 ()	76 ()	56 ()	
Sobre el promedio	10 ()	23 ()	37 ()	
Total				

i) Hipótesis

H₀: _____

H₁: _____

ii) Estadístico de prueba

$$\chi_o^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{O_{ij}^2}{e_{ij}} - n =$$

iii) Nivel de significancia: $\alpha =$ _____ entonces: $\chi^2_{(\dots, \dots)} =$ _____

iv) Decisión: como $\chi_o^2 =$ _____ “que $\chi_\alpha^2 =$ _____ Entonces _____ H_o

y decimos que _____

7. Un estudio que se realizó con 81 personas referente a la relación entre la cantidad de violencia vista en la televisión y la edad del televíidente produjo los siguientes resultados.

	16 - 34	34 - 55	55 o mas	Total
Poca violencia	8	12	21	41
Mucha violencia	18	15	7	40
Total	26	27	28	81

¿Indican los datos que ver violencia en la televisión depende de la edad del televíidente, a un nivel de significación del 5%?

i) **Hipótesis**

$H_0:$ _____

$H_1:$ _____

ii) **Estadístico de prueba**

$$\chi_o^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{O_{ij}^2}{e_{ij}} - n =$$

iii) **Nivel de significancia:** $\alpha =$ _____ entonces: $\chi^2_{(\dots, \dots)} =$ _____

iv) **Decisión:** como $\chi_o^2 =$ _____ “que $\chi^2_{\alpha} =$ _____ Entonces _____ H_o

y decimos que _____

8. Una empresa minera hizo un estudio para verificar si el tipo de trabajo se relaciona con el grado de estrés de los trabajadores. Para lo cual se elige una muestra aleatoria de 300 trabajadores y se clasifican en la tabla siguiente:

Tipo de Trabajo	Grado de Estrés			
	I	II	III	Total
Oficina	42	24	30	96
Terreno	54	78	72	204
Total	96	102	102	300

Probar la hipótesis de que el tipo de trabajo afecta el grado de estrés del trabajador con un nivel de significación de 5%

i) **Hipótesis**

$H_0:$ _____

$H_1:$ _____

ii) **Estadístico de prueba**

$$\chi^2_o = \sum_{i=1}^m \sum_{j=1}^k \frac{O_{ij}^2}{e_{ij}} - n =$$

iii) **Nivel de significancia:** $\alpha =$ _____ entonces: $\chi^2_{(\dots, \dots)} =$ _____

iv) **Decisión:** como $\chi_o^2 =$ _____ “que $\chi^2_{\alpha} =$ _____ Entonces _____ H_o

y decimos que _____

ANEXOS

ANEXO 1

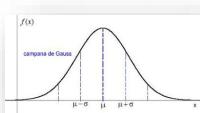
TABLA DE NÚMEROS ALEATORIOS

3690 2492 7171 7720 6509 7549 2330 5733 4730 0813 6790 6858 1489 2669 3743 1901
4971 8280 6477 5289 4092 4223 6454 7632 7577 2816 9002 0772 2160 7236 0812 4195
5589 0830 8261 9232 5692 9870 3583 8997 1533 6466 8830 7271 3809 2080 3828 7880
0586 8482 7811 6807 3309 2729 1039 3382 7600 1077 4455 8806 1822 1669 7501 7227
0104 4141 1521 9104 5563 1392 8238 4882 8506 6348 4612 8252 1062 1757 0964 2983
2244 5086 0303 7423 3298 3979 2831 2257 1508 7642 0092 1629 0377 3590 2209 4839
6332 1490 3092 0935 5565 2315 8030 7651 5189 0075 9353 1921 2605 3973 8204 4143
2677 0034 8601 3340 8383 7277 9889 0390 5579 4620 5650 0210 2082 4664 5484 3900
3485 0741 9069 5920 4326 7704 6525 6905 7127 5933 1137 7583 6450 5658 7678 3444
8387 5323 3753 1859 6043 0294 5110 6340 9137 4094 4957 0163 9717 4118 4276 9465
8820 4127 4951 3781 5101 1815 7068 6379 7252 1086 8919 9047 0199 5068 7447 1664
9278 1708 3625 2864 7274 9512 0074 6677 8676 0222 3335 1976 1645 9192 4011 0255
5458 6942 8043 6201 1587 0972 0554 1690 6333 1931 9433 2661 8690 2313 6999 9231
5627 1815 7171 8036 1832 2031 6298 6073 3995 9677 7765 3194 3222 4191 2734 4469
8617 2402 6250 9362 7373 4757 1716 1942 0417 5921 5295 7385 5474 2123 7035 9983
5192 1840 6176 5177 1191 2106 3351 5057 0967 4538 1246 3374 7315 3365 7203 1231
0546 6612 1038 1425 2709 5775 7517 8974 3961 2183 5295 3096 8536 9442 5500 2276
6307 2346 1285 7000 5306 0414 3383 3251 8902 8843 2112 8567 8131 8116 5270 5994
4675 1435 2192 0874 2897 0262 5092 5541 4014 3543 6130 4247 4859 2660 7852 9096
0578 0097 3521 8772 6612 0721 3899 2999 1263 7017 8057 5573 9396 3464 1702 9204
3389 5678 2589 0288 7478 7569 7551 3380 2152 5411 2647 7242 2800 3339 2854 9691
9562 3252 9848 6030 8472 2266 5505 8474 3167 8552 5409 1556 4247 4652 2953 6381
2086 5457 7703 2758 2963 8167 6712 9823 9362 4373 4757 1716 3042 0222 2335 1916

Donald B. Owen, *Handbook of Statistical Tables*, Reading Mass:Addison-Wesley, 1.962.

ANEXO 2

TABLA DE DISTRIBUCIÓN NORMAL ESTANDAR A1



z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99118	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999
3.1	0.99903	0.99906	0.9991	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.9994	0.99942	0.99944	0.99946	0.99948	0.9995
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.9996	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.9997	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.9998	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.9999	0.9999	0.9999	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998

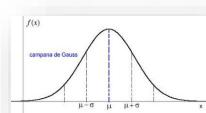


TABLA DE DISTRIBUCIÓN NORMAL ESTANDAR A2

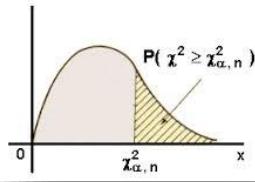
z	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0
-4	0.00002	0.00002	0.00002	0.00002	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003
-3.9	0.00003	0.00003	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00005	0.00005
-3.8	0.00005	0.00005	0.00005	0.00006	0.00006	0.00006	0.00006	0.00007	0.00007	0.00007
-3.7	0.00008	0.00008	0.00008	0.00008	0.00009	0.00009	0.0001	0.0001	0.0001	0.00011
-3.6	0.00011	0.00012	0.00012	0.00013	0.00013	0.00014	0.00014	0.00015	0.00015	0.00016
-3.5	0.00017	0.00017	0.00018	0.00019	0.00019	0.0002	0.00021	0.00022	0.00022	0.00023
-3.4	0.00024	0.00025	0.00026	0.00027	0.00028	0.00029	0.0003	0.00031	0.00032	0.00034
-3.3	0.00035	0.00036	0.00038	0.00039	0.0004	0.00042	0.00043	0.00045	0.00047	0.00048
-3.2	0.0005	0.00052	0.00054	0.00056	0.00058	0.0006	0.00062	0.00064	0.00066	0.00069
-3.1	0.00071	0.00074	0.00076	0.00079	0.00082	0.00084	0.00087	0.0009	0.00094	0.00097
-3	0.001	0.00104	0.00107	0.00111	0.00114	0.00118	0.00122	0.00126	0.00131	0.00135
-2.9	0.00139	0.00144	0.00149	0.00154	0.00159	0.00164	0.00169	0.00175	0.00181	0.00187
-2.8	0.00193	0.00199	0.00205	0.00212	0.00219	0.00226	0.00233	0.0024	0.00248	0.00256
-2.7	0.00264	0.00272	0.0028	0.00289	0.00298	0.00307	0.00317	0.00326	0.00336	0.00347
-2.6	0.00357	0.00368	0.00379	0.00391	0.00402	0.00415	0.00427	0.0044	0.00453	0.00466
-2.5	0.0048	0.00494	0.00508	0.00523	0.00539	0.00554	0.0057	0.00587	0.00604	0.00621
-2.4	0.00639	0.00657	0.00676	0.00695	0.00714	0.00734	0.00755	0.00776	0.00798	0.0082
-2.3	0.00842	0.00866	0.00889	0.00914	0.00939	0.00964	0.0099	0.01017	0.01044	0.01072
-2.2	0.01101	0.0113	0.0116	0.01191	0.01222	0.01255	0.01287	0.01321	0.01355	0.0139
-2.1	0.01426	0.01463	0.015	0.01539	0.01578	0.01618	0.01659	0.017	0.01743	0.01786
-2	0.01831	0.01876	0.01923	0.0197	0.02018	0.02068	0.02118	0.02169	0.02222	0.02275
-1.9	0.0233	0.02385	0.02442	0.025	0.02559	0.02619	0.0268	0.02743	0.02807	0.02872
-1.8	0.02938	0.03005	0.03074	0.03144	0.03216	0.03288	0.03362	0.03438	0.03515	0.03593
-1.7	0.03673	0.03754	0.03836	0.0392	0.04006	0.04093	0.04182	0.04272	0.04363	0.04457
-1.6	0.04551	0.04648	0.04746	0.04846	0.04947	0.0505	0.05155	0.05262	0.0537	0.0548
-1.5	0.05592	0.05705	0.05821	0.05938	0.06057	0.06178	0.06301	0.06426	0.06552	0.06681
-1.4	0.06811	0.06944	0.07078	0.07215	0.07353	0.07493	0.07636	0.0778	0.07927	0.08076
-1.3	0.08226	0.08379	0.08534	0.08692	0.08851	0.09012	0.09176	0.09342	0.0951	0.0968
-1.2	0.09853	0.10027	0.10204	0.10383	0.10565	0.10749	0.10935	0.11123	0.11314	0.11507
-1.1	0.11702	0.119	0.121	0.12302	0.12507	0.12714	0.12924	0.13136	0.1335	0.13567
-1	0.13786	0.14007	0.14231	0.14457	0.14686	0.14917	0.15151	0.15386	0.15625	0.15866
-0.9	0.16109	0.16354	0.16602	0.16853	0.17106	0.17361	0.17619	0.17879	0.18141	0.18406
-0.8	0.18673	0.18943	0.19215	0.19489	0.19766	0.20045	0.20327	0.20611	0.20897	0.21186
-0.7	0.21476	0.2177	0.22065	0.22363	0.22663	0.22965	0.2327	0.23576	0.23885	0.24196
-0.6	0.2451	0.24825	0.25143	0.25463	0.25785	0.26109	0.26435	0.26763	0.27093	0.27425
-0.5	0.2776	0.28096	0.28434	0.28774	0.29116	0.2946	0.29806	0.30153	0.30503	0.30854
-0.4	0.31207	0.31561	0.31918	0.32276	0.32636	0.32997	0.3336	0.33724	0.3409	0.34458
-0.3	0.34827	0.35197	0.35569	0.35942	0.36317	0.36693	0.3707	0.37448	0.37828	0.38209
-0.2	0.38591	0.38974	0.39358	0.39743	0.40129	0.40517	0.40905	0.41294	0.41683	0.42074
-0.1	0.42465	0.42858	0.43251	0.43644	0.44038	0.44433	0.44828	0.45224	0.4562	0.46017
0	0.46414	0.46812	0.4721	0.47608	0.48006	0.48405	0.48803	0.49202	0.49601	0.5

ANEXO 3

TABLA DE DISTRIBUCIÓN CHI CUADRADO χ^2

α = Probabilidad de encontrar un valor mayor o igual que el chi cuadrado tabulado,

g.l. = Grados de Libertad



g.l	Valores de α								
	0,01	0,025	0,05	0,10	0,15	0,20	0,25	0,30	0,35
1	6.634	5.023	3.841	2.705	2.07	1.642	1.323	1.074	0.873
2	9.210	7.377	5.991	4.605	3.79	3.218	2.772	2.407	2.099
3	11.34	9.348	7.814	6.251	5.31	4.641	4.108	3.664	3.283
4	13.27	11.14	9.487	7.779	6.74	5.988	5.385	4.878	4.437
5	15.08	12.83	11.07	9.236	8.11	7.289	6.625	6.064	5.573
6	16.81	14.44	12.59	10.64	9.44	8.558	7.840	7.231	6.694
7	18.47	16.01	14.06	12.01	10.74	9.803	9.037	8.383	7.806
8	20.09	17.53	15.50	13.36	12.02	11.03	10.21	9.524	8.909
9	21.66	19.02	16.91	14.68	13.28	12.24	11.38	10.65	10.00
10	23.20	20.48	18.30	15.98	14.53	13.44	12.54	11.78	11.09
11	24.72	21.92	19.67	17.27	15.76	14.63	13.70	12.89	12.18
12	26.21	23.33	21.02	18.54	16.98	15.81	14.84	14.01	13.26
13	27.68	24.73	22.36	19.81	18.20	16.98	15.98	15.11	14.34
14	29.14	26.11	23.68	21.06	19.40	18.15	17.11	16.22	15.42
15	30.57	27.48	24.99	22.30	20.60	19.31	18.24	17.32	16.49
16	31.99	28.84	26.29	23.54	21.79	20.46	19.36	18.41	17.56
17	33.40	30.19	27.58	24.76	22.97	21.61	20.48	19.51	18.63
18	34.80	31.52	28.86	25.98	24.15	22.75	21.60	20.60	19.69
19	36.19	32.85	30.14	27.20	25.32	23.90	22.71	21.68	20.76
20	37.56	34.16	31.41	28.41	26.49	25.03	23.82	22.77	21.82
21	38.93	35.47	32.67	29.61	27.66	26.17	24.93	23.85	22.88
22	40.28	36.78	33.92	30.81	28.82	27.30	26.03	24.93	23.94
23	41.63	38.07	35.17	32.00	29.97	28.42	27.14	26.01	25.00
24	42.97	39.36	36.41	33.19	31.13	29.55	28.24	27.09	26.06
25	44.31	40.64	37.65	34.38	32.28	30.67	29.33	28.17	27.11
26	45.64	41.92	38.88	35.56	33.42	31.79	30.43	29.24	28.17
27	46.96	43.19	40.11	36.74	34.57	32.91	31.52	30.31	29.22
28	48.27	44.46	41.33	37.91	35.71	34.02	32.62	31.39	30.27
29	49.58	45.72	42.55	39.08	36.85	35.13	33.71	32.46	31.33

REFERENCIAS BIBLIOGRÁFICAS

BÁSICA

Hernández, R., Fernández, C., Baptista, P. (2010). *Metodología de la Investigación*. 5^a ed. México: Mc Graw-Hill.

Tamayo, M. (2012). *El proceso de la investigación científica*. 5^a ed. México: Limusa

Castillo, I., y Guijarro, M. (2009). *Estadística Descriptiva y Cálculo de Probabilidades*. España: Editorial Pearson.

Cordova Zamora, M. (2010). *Estadística Aplicada Básica*. Lima, Perú: Editorial San Marcos.

Hernandez, A. (2008). *Curso Elemental de Estadística Descriptiva*. España: Ed. Pirámide.

Mitacc Meza, M. (2011). *Tópicos de Estadística y Probabilidad*. Lima, Perú: Editorial Thales.

Moya, R., y Saravia,G. (2012). *Probabilidades e Inferencia Estadística*. Lima, Perú: Editorial San Marcos.

Quispe, Q. (2008). *Fundamentos de estadística básica*. México: Trillas.

Visauta, B. (2010). *Análisis Estadístico con SPSS 17(Estadística Básica)*. España: Editorial McGraw-Hill.

COMPLEMENTARIA

Bunge, M. (2004). *La investigación científica*. 3^a ed. Barcelona: Ariel

Anderson, D. (2009). *Estadística para administración y economía*. México: Pearson education S.A.

Castro, A., y Villacampa, Y. (2000). *Estadística Aplicada a la Ingeniería Civil*. Barcelona, España: Editorial Club Universitario

Díaz, A. (2010). *Estadística Aplicada a la Administración y la Economía*. México: Editorial McGraw-Hill

Romero, R., y Zunica, L. (2005). *Métodos Estadísticos en Ingeniería*. Valencia, España: Editorial Universidad Politécnica de Valencia.

Sarriá, A., Guardia, J., y Montserrat, F. (1999). *Introducción a la Estadística en Psicología*. España: Edicions Universitat Barcelona

Wackerly, D., Scheffe, R., y Mendenhal, W. (2010). *Estadística para Administradores*. México: Thomson Editores.

PAGINAS WEB

[http://www.academia.edu/6399195/Metodologia de la investigacion 5ta Edicion Sampieri](http://www.academia.edu/6399195/Metodologia_de_la_investigacion_5ta_Edicion_Sampieri)

<http://www.gestiopolis.com/metodos-y-tecnicas-de-investigacion/>

<http://www4.ujaen.es/~ajsaez/recursos/EstadisticaIngenieros.pdf>

<http://www.rosaweb.org/descargas/temasei.pdf>

<http://www.estadisticaparaadministracion.blogspot.com/>

<http://www.dm.uba.ar/materias/estadistica>

<http://www.aulafacil.com/CursoEstadistica/CursoEstadistica>

<https://www.inei.gob.pe/>