EDUCACIÓN

SECRETARÍA DE EDUCACIÓN PÚBLICA

DGETI Guanajuato CBTis 75

Unidad III

PENSAMIENTO MATEMÁTICO

Semestre

Agosto´23 – Enero´24

Mate maycko

EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA
DIRECCIÓN GENERAL DE EDUCACIÓN TECNOLÓGICA INDUSTRIAL Y DE SERVICIOS
2023 Francisco VILLA
DGETI Guanajuato CBTis 75

PENSAMIENTO MATEMÁTICO
PROGRESIÓN No. 11

Identifica ante la imposibilidad de estudiar la totalidad de una población, la opción de extraer informacion de ésta a través de l empleo de técnicas de muestro en particular ,valora la importancia de la aleatoriedad al momento de tomar una muestra.

METAS	CATEGORÍAS	SUBCATEGORÍAS
M1 Observa y obtiene información de una situación o fenómeno para establecer estrategias o formas de visualización que ayuden a entenderlo	C2 Procesos de intuición y razonamiento	S1 Capacidad para observar y conjeturar S2 Pensamiento intuitivo S3 Pensamioento Formal
M2 Socializa con sus pares sus conjeturas descubrimientos o procesos en la solución de un problema tanto teorico como de su entorno	C3 Solución de Problemas y Modelación	S2 Pensamiento intuitivo S3 Pensamioento Formal

Mate mayeko

MUESTREO PROBABILISTICO Y NO PROBABILISTICO

La necesidad de conocer diversas características de una población, hace indispensable el recurrir a procedimientos que eviten registrar la información correspondiente a todos los elementos que componen esa población, lo que implica describir está a partir del estudio de solo una porción proveniente de ella. Estos procedimientos, conocidos bajo el nombre genérico de *muestreo*, son de diversos tipos, pero todos ellos pueden clasificarse en dos grandes grupos: *no probabilísticos y probabilísticos.*

Independientemente del procedimiento de selección de muestra el conocimiento que se adquiere respecto a determinada descripciones de una población a partir de dicha muestra es siempre una estimación.

En estadística, se entiende por *estimación*, el procedimiento que conduce a una descripción numérica aproximada de una característica poblacional, llamada por algunos autores *la estimada*, a partir de la correspondiente descripción numérica de los datos muestrales.

Volviendo al muestreo, en el *no probabilístico*, la muestra se elige aplicando criterio; es decir, juega un papel decisivo en su selección el punto de vista del investigador. La muestra puede ser *circunstancial o intencional*.

La circunstancial, llamada también *de conveniencia*, surge atendiendo razones de comodidad; en un universo, de estudiantes, por ejemplo; se puede elegir como muestra los propios condiscípulos o a cualquier subgrupo de estudiantes del mismo universo con que se tope uno fácilmente.

En *la intencional*, se elige la muestra en función de los objetivos y de la experiencia del investigador; es este el que decide cuántos y cuáles serán los elementos que la integran; en una muestra de automóviles, se puede decidir cuantos, de que marca y de qué modelo la compondrán. Las muestras circunstanciales o intencionales pueden ser representativas, pero no permiten calcular el error de muestreo no el nivel de confianza de las estimaciones.

En los noticieros televisivos abundan ambos tipos de muestras, sobre todo, las circunstanciales cuyos resultados son siempre de dudosa representatividad no solo por la escasez de sus elementos, sino también y sobre todo, por la manipulación que suele hacerse de la información. Dicho esto el muestreo circunstancial tiene gran utilidad en los procesos de diseño de investigaciones fuertemente estructuradas; permite de manera rápida efectuar sondeos que ayudan, entre otras cosas, a la afinación de los instrumentos de medición.

En el muestreo probabilístico la muestra se elige mediante sorteo y ya no depende en lo absoluto del criterio del investigador.

Una muestra probabilística suele garantizar altos niveles de representatividad y permite siempre calcular el error de muestreo y el nivel de confianza de las estimaciones. Sólo en este tipo de muestras tienen aplicación justificada los métodos de inferencia estadística que se verán adelante.

Aunque hay muchas técnicas que intentan asegurar que la muestra sea suficientemente representativa de la población, existe una en que se basan la mayoría de los métodos de la estadística inferencial: LA MUESTRA ALEATORIA SIMPLE.

Una muestra aleatoria, o muestra al azar es aquella en que cada elemento de la población tiene la misma probabilidad de ser seleccionado.

El término azar o aleatorio no significa “de cualquier modo” o “como sea”, significa, más bien, adhesión a ciertas probabilidades. Así, cuando hablemos de muestra al azar o muestra aleatoria, deberá entenderse que se trata de una porción extraída de un universo en el cual los elementos de este tuvieron la misma probabilidad de pertenecer a la muestra.

El procedimiento de muestreo aleatorio simple exige, como requisito indispensable, la identificación previa y precisa de cada elemento o unidad esencial de la población, seguido del uso de un dispositivo de selección de elementos de muestra que este al margen de prejuicio y que garantice la exigencia de equiprobabilidad. Este dispositivo se conoce como tabla de números aleatorios.

TABLA DE NÚMEROS ALEATORIOS

Formadas por regla general de cientos o miles de dígitos, cada uno de los cuales genera independientemente, estas tablas poseen la característica de que cualquiera de las cifras del cero al nueve tiene la misma probabilidad de aparecer.

Ejemplos de tablas de números aleatorios (ver cuadros anexos)al final del material

Debemos señalar que los procedimientos modernos de construcción son por vía de computadora.

Considerando que el universo del cual se pretende extraer una muestra aleatoria, debe ser manipulable física o simbólicamente.

La manipulación física es muy difícil si no imposible; la simbólica, en cambio, es con frecuencia accesible mediante un MARCO:

MARCO

Es todo que hace posible el manejo simbólico del universo que se intenta investigar por muestreo.

Una lista de nombres, el mapa actualizado de una ciudad, son ejemplos comunes que caen bajo la denominación de marco. Este instrumento debe representar fielmente o con un mínimo de error a todos y cada uno de los elementos que conforman la población sometida a estudio. Una de las primeras tareas con relación al marco que se va a utilizar, es verificar que esté libre de errores; el más común entre estos suele ser la repetición o la omisión de elementos. Si hay repetición, algunos elementos tendrán al menos doble probabilidad de pertenecer a la muestra; si hay omisión, habrá algunos cuya probabilidad de pertenecer a la muestra sea cero.

En cualquier caso, no se cumplirá el requisito fundamental de la muestra aleatoria simple.

USO DE LAS TABLAS DE NUMEROS ALEATORIOS.

Ejemplo 1.

Supongamos que deseamos elegir a dos números entre 1 y 50.

Para ello podemos utilizar pares de números en la tabla No. 1, y estos pueden ser adyacentes. Iniciemos con los primeros dos del segundo grupo de números (66) y avancemos hacia la derecha (06), (57). El 66 no está comprendido entre 1 y 50 de manera que se descarta, el 06, si está comprendido se toma el número, el 57 no está comprendido se deja, si se elige el 47 también; estos números constituyen nuestra selección, (06), (57). Se queda

Ahora elijamos a cinco números aleatorios comprendidos entre 1 y 195, para ello podemos usar agrupaciones de tres dígitos. Si comenzamos arriba y a la derecha de la tabla con el número 494 y continuamos hacia abajo, la muestra queda dada por:

494
166
3600
765
683

Ejemplo No. 2

Se decide seleccionar aleatoriamente a 5 estudiantes (n = 5) de un grupo de 35 (N = 35)

Para ello se cuenta con una lista de los nombres de todos ellos **(marco)**

Tomando de la tabla No. 2 el sentido de lectura vertical y en sentido ascendente, comenzando por la primera columna.

Los números que leeríamos serían el 25, el 59, el 07, que serían los primeros elementos de la muestra; el **25**, el **07** y así sucesivamente. Los estudiantes que resultarían elegidos al azar serían aquellos cuyos números asociados fuesen **07 25,05,24 y 38**

*Si consideramos de nuevo el ejemplo que acabamos de desarrollar, veremos que de los 35 elementos solo **1/7, ó 14%**, forman parte de la muestra; pues bien, a la proporción que guarda el tamaño de la muestra con respecto al del universo se le da el nombre de **FRACCIÓN DE MUESTREO**.*

Si el universo constara de 5,000 elementos y de el se extrajera una muestra de 250, la fracción de muestreo sería del 5%.

En símbolos, la fracción de muestreo se representa así:

$$f = \frac{n}{N}$$

Con frecuencia, el investigador fija el valor de la fracción de muestreo, a partir de la cual, conocido el tamaño de la población, se determina el de la muestra. Si N = 800 y f = 15% entonces n = 120.

NOTA: La relación que guarda el tamaño de la muestra con respecto al del universo suele no ser tan arbitrario; depende básicamente, del nivel de confianza y de la precisión que se quiera alcanzar en las estimaciones de los parámetros que se busca conocer.

De los ejemplos anteriores podemos establecer un orden para el uso de las tablas de números aleatorios.

Definido un universo y depurado el marco correspondiente tenemos:

Primero, se asocia un número progresivo a todas y cada una de las unidades esenciales, de modo que cada uno tenga tantas cifras como cifras tiene el tamaño del universo.

Segundo, conocido el tamaño de muestra, se procede a la selección de los elementos empleando una tabla apropiada, como se indica a continuación:

a) Se elige arbitrariamente un punto de partida en la tabla y un sentido de lectura que deberá ser respetado hasta el fin de la fase de selección.
b) Se leen los números en grupos formados por tantas cifras como cifras tiene N.

Pertenecerán a la muestra todos los elementos del universo $\leq$ N, que se irán anotando aparte. Si durante el proceso se repite un número admisible, se le pasa por alto.

Si por acaso se agotara la tabla y aún faltasen elementos para completar el tamaño de muestra, bastará con elegir otro sentido de lectura para generar nuevos números aleatorios. En cualquier caso, la fase toca a su fin tan pronto se alcanza el tamaño de muestra, tercero, obtenidos los componentes de esta última, se agruparán de suerte que facilite su localización con fines de observación o encuesta.

Por grupos académicos si son estudiantes; por colonias, barrios o manzanas si son familias; por orden cronológico si son editoriales de periódicos.

DISTRIBUCIONES DE MUESTREO

Sabías que….

La necesidad de conocer diversas características de una población, hace indispensable

el recurrir a procedimientos que eviten registrar la información correspondiente a todos los elementos que componen esa población, lo que implica describir está a partir del estudio de solo una porción proveniente de ella. Estos procedimientos, conocidos bajo el nombre genérico de muestreo, son de diversos tipos, pero todos ellos pueden clasificarse en dos grandes grupos: no probabilísticos y probabilísticos.

Independientemente del procedimiento de selección de muestra el conocimiento que se adquiere respecto a determinada descripciones de una población a partir de dicha muestra es siempre una estimación

En estadística, se entiende por ***estimación*,** el procedimiento que conduce a una descripción numérica aproximada de una característica poblacional, llamada por algunos autores ***la estimada*,** a partir de la correspondiente descripción numérica de los datos muestrales.

El muestreo, **no probabilístico**, la muestra se elige aplicando criterio; es decir, juega un papel decisivo en su selección el punto de vista del investigador**. La muestra puede ser *circunstancial o intencional*.**

La circunstancial, llamada también **de conveniencia**, surge atendiendo razones de **comodidad;** en un universo, de estudiantes, por ejemplo; se puede elegir como muestra los propios condiscípulos o a cualquier subgrupo de estudiantes del mismo universo con que se tope uno fácilmente.

En *la intencional*, se elige la muestra en función de los objetivos y de la experiencia del investigador; es este el que decide cuántos y cuáles serán los elementos que la integran; en una muestra de automóviles, se puede decidir cuantos, de que marca y de qué modelo la compondrán. Las muestras circunstanciales o intencionales pueden ser representativas, pero no permiten calcular el error de muestreo no el nivel de confianza de las estimaciones.

En el muestreo probabilístico la muestra se elige mediante sorteo y ya no depende en lo absoluto del criterio del investigador.

Una muestra probabilística suele garantizar altos niveles de representatividad y permite siempre calcular el error de muestreo y el nivel de confianza de las estimaciones.

Aunque hay muchas técnicas que intentan asegurar que la muestra sea suficientemente representativa de la población, existe una en que se basan la mayoría de los métodos de la estadística inferencial: **LA MUESTRA ALEATORIA SIMPLE**.

Una muestra aleatoria, o muestra al azar es aquella en que cada elemento de la población tiene la misma probabilidad de ser seleccionado.

Sabías que....

El muestreo aleatorio simple es una técnica de muestreo probabilístico en la que todos los sujetos tienen una probabilidad conocida, distinta de cero, de ser seleccionados, es una técnica que utiliza métodos aleatorios para la selección.

Tabla 1: Números aleatorios

10	09	73	25	33	76	52	01	35	86	34	67	35	48	76	80	95	90	91	17	39	29	27	49	45
37	54	20	48	05	64	89	47	42	96	24	80	52	40	37	20	63	61	04	02	00	82	29	16	65
08	42	26	89	53	19	64	50	93	03	23	20	90	25	60	15	95	33	47	64	35	08	03	36	06
99	01	90	25	29	09	37	67	07	15	38	31	13	11	65	88	67	67	43	97	04	43	62	76	59
12	80	79	99	70	80	15	73	61	47	64	03	23	66	53	98	95	11	68	77	12	17	17	68	33
66	06	57	47	17	34	07	27	68	50	36	69	73	61	70	65	81	33	98	85	11	19	92	91	70
31	06	01	08	05	45	57	18	24	06	35	30	34	26	14	86	79	90	74	39	23	40	30	97	32
85	26	97	76	02	02	05	16	56	92	68	66	57	48	18	73	05	38	52	47	18	62	38	85	79
63	57	33	21	35	05	32	54	70	48	90	55	35	75	48	28	46	82	87	09	83	49	12	56	24
73	79	64	57	53	03	52	96	47	78	35	80	83	42	82	60	93	52	03	44	35	27	38	84	35
98	52	01	77	67	14	90	56	86	07	22	10	94	05	58	60	97	09	34	33	50	50	07	39	98
11	80	50	54	31	39	80	82	77	32	50	72	56	82	48	29	40	52	42	01	52	77	56	78	51
83	45	29	96	34	06	28	89	80	83	13	74	67	00	78	18	47	54	06	10	68	71	17	78	17
88	68	54	02	00	86	50	75	84	01	36	76	66	79	51	90	36	47	64	93	29	60	91	10	62
99	59	46	73	48	87	51	76	49	69	91	82	60	89	28	93	78	56	13	68	23	47	83	41	13
65	48	11	76	74	17	46	85	09	50	58	04	77	69	74	73	03	95	71	86	40	21	81	65	44
80	12	43	56	35	17	72	70	80	15	45	31	82	23	74	21	11	57	82	53	14	38	55	37	63
74	35	09	98	17	77	40	27	72	14	43	23	60	02	10	45	52	16	42	37	96	28	60	26	55
69	91	62	68	03	66	25	22	91	48	36	93	68	72	03	76	62	11	39	90	94	40	05	64	18
09	89	32	05	05	14	22	56	85	14	46	42	75	67	88	96	29	77	88	22	54	38	21	45	98
91	49	91	45	23	68	47	92	76	86	46	16	28	35	54	94	75	08	99	23	37	08	92	00	48
80	33	69	45	98	26	94	03	68	58	70	29	73	41	35	53	14	03	33	40	42	05	08	23	41
44	10	48	19	49	85	15	74	79	54	32	97	92	65	75	57	60	04	08	81	22	22	20	64	13
12	55	07	37	42	11	10	00	20	40	12	86	07	46	97	96	64	48	94	39	28	70	72	58	15
63	60	64	93	29	16	50	53	44	84	40	21	95	25	63	43	65	17	70	82	07	20	73	17	90
61	19	69	04	46	26	45	74	77	74	51	92	43	37	29	65	39	45	95	93	42	58	26	05	27
15	47	44	52	66	95	27	07	99	53	59	36	78	38	48	82	39	61	01	18	33	21	15	94	66
94	55	72	85	73	67	89	75	43	87	54	62	24	44	31	91	19	04	25	92	92	92	74	59	73
42	48	11	62	13	97	34	40	87	21	16	86	84	87	67	03	07	11	20	59	25	70	14	66	70
23	52	37	83	17	73	20	88	98	37	68	93	59	14	16	26	25	22	96	63	05	52	28	25	62
04	49	35	24	94	75	24	63	38	24	45	86	25	10	25	61	96	27	93	35	65	33	71	24	72
00	54	99	76	54	64	05	18	81	59	96	11	96	38	96	54	69	28	23	91	23	28	72	95	29
35	96	31	53	07	26	89	80	93	54	33	35	13	54	62	77	97	45	00	24	90	10	33	93	33
59	80	80	83	91	45	42	72	68	42	83	60	94	97	00	13	02	12	48	92	78	56	52	01	06
46	05	88	52	36	01	39	09	22	86	77	28	14	40	77	93	91	08	36	47	70	61	74	29	41
32	17	90	05	97	87	37	92	52	41	05	56	70	70	07	86	74	31	71	57	85	39	41	18	38
69	23	46	14	06	20	11	74	52	04	15	95	66	00	00	18	74	39	24	23	97	11	89	63	38
19	56	54	14	30	01	75	87	53	79	40	41	92	15	85	66	67	43	68	06	84	96	28	52	07
45	15	51	49	38	19	47	60	72	46	43	66	79	45	43	59	04	79	00	33	20	82	66	95	41
94	86	43	19	94	36	16	81	08	51	34	88	88	15	53	01	54	03	54	56	05	01	45	11	76
98	08	62	48	26	45	24	02	84	04	44	99	90	88	96	39	09	47	34	07	35	44	13	18	80
33	18	51	62	32	41	94	15	09	49	89	43	54	85	81	88	69	54	19	94	37	54	87	30	43
80	95	10	04	06	96	38	27	07	74	20	15	12	33	87	25	01	62	52	98	94	62	46	11	71
79	75	24	91	40	71	96	12	82	96	69	86	10	25	91	74	85	22	05	39	00	38	75	95	79
18	63	33	25	37	98	14	50	65	71	31	01	02	46	74	05	45	56	14	27	77	93	89	19	36
74	02	94	39	02	77	55	73	22	70	97	79	01	71	19	52	52	75	80	21	80	81	45	17	48
54	17	84	56	11	80	99	33	71	43	05	33	51	29	69	56	12	71	92	55	36	04	09	03	24
11	66	44	98	83	52	07	98	48	27	59	38	17	15	39	09	97	33	34	40	88	46	12	33	56
48	32	47	79	28	31	24	96	47	10	02	29	53	68	70	32	30	75	75	46	15	02	00	99	94
69	07	49	41	38	87	63	79	19	76	35	58	40	44	01	10	51	82	16	15	01	84	87	69	38

Tomada de las tablas de la RAND Corporation. Se reimprimen de la obra de Wilfred J. Dixon y Frank Massey, Jr., *Introduction to Statistical Analysis*, 3ª ed., McGraw-Hill Book Company, Nueva York, 1969, pág. 446.

2.- **Centro de Ciencias Básicas**
Departamento de Estadística

Tabla de Número Aleatorios

Renglón	Columna 1	2	3	4	5	6	7	8	9	10
1	34600	19108	69812	93480	65191	57359	24408	36527	60414	94913
2	79151	13078	01872	84469	83906	06881	22936	49856	97607	04230
3	92494	97825	58734	08516	37704	20133	70505	06395	54808	57036
4	44852	06858	81140	89296	54813	56856	24316	70468	90027	08372
5	97467	69926	51148	73026	43306	89484	33330	19093	80101	48435
6	96207	18877	70523	29690	44458	99242	35456	39595	87653	32716
7	60337	14292	12704	08359	36120	29596	67888	93498	74984	72836
8	04812	88937	96641	22579	73721	31921	35923	14615	40883	03776
9	30697	44518	57792	97046	99380	17005	30846	55406	22689	88659
10	60331	18044	08728	03094	03465	49651	90558	38744	11275	83301
11	18237	87670	02435	72480	99308	66631	17864	56993	98537	72231
12	98035	63712	25899	61025	35983	46596	59199	36711	03279	15780
13	67961	65714	61082	75324	85711	68100	91197	62429	68027	21201
14	70218	24572	67326	26462	87248	17841	87067	78185	42740	57149
15	83363	17664	88351	55077	07062	17763	60613	60318	05146	02800
16	68761	46051	17313	89765	00076	37890	69373	83061	32370	43278
17	67671	08649	76236	27897	17142	49988	96564	96447	51142	19597
18	95378	01544	76192	69697	29253	70416	17232	38553	21685	22376
19	84149	79121	41425	91820	04102	66022	43084	52345	42530	13834
20	95722	26655	74689	06488	39904	89072	54856	41955	54177	23443
21	19752	28685	28588	43556	66010	50637	37566	74944	20588	98308
22	45683	63873	88430	66485	06903	21488	50694	63228	23797	55052
23	99371	57461	20036	00612	19257	63458	57497	08098	74158	72297
24	72580	53039	43441	98578	54184	45921	65127	01318	68949	48418
25	34315	22973	71948	22061	65262	45078	31623	68896	05562	69511
26	68713	40962	66760	59066	51208	26809	54870	28032	73369	85440
27	44238	95669	78727	90871	08582	59089	73503	97694	68497	94423
28	95424	98332	30624	05323	17194	75596	56225	48613	19599	79610
29	49124	66002	32001	79866	31301	48747	93177	34517	05604	90547
30	74466	36981	62140	54336	98307	84174	31450	67320	24019	93067
31	62705	87371	27786	60655	04768	28167	89910	73654	39125	08345
32	79052	64426	81519	48547	52989	10767	44335	37239	39975	01336
33	77042	31204	27133	25775	26464	74715	90253	64489	09105	40317
34	92857	14367	89222	76064	42903	39980	98344	59800	29486	71233
35	38651	89649	14561	27064	20533	91217	63644	53458	40351	58349
36	09378	90060	82564	32916	71102	81222	74533	11621	10693	96972
37	58472	38563	24415	34840	58615	33807	12195	90969	73059	29497
38	42007	40047	43323	68349	91582	39506	77927	89534	87133	02449
39	53059	66560	80929	47058	64544	75657	68385	92748	31013	97111
40	82725	70760	74764	97880	46162	58002	62728	78882	77898	23641

Fuente: R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

SIMBOLOS		
INDICADOR	**EN LA POBLACION**	**EN LA MUESTRA**
Tamaño	**N**	**n**
Media	μ **(mu)**	$\bar{X}$
Mediana	$\bar{\mu}$	**Me**
Moda	$\dot{\mu}$	**Mo**
Varianza	σ^2	S^2
Desviación estándar	σ **(tao)**	**S**

Existen aún tres símbolos de amplio uso para representar la media, la varianza y la desviación estándar de distribuciones de muestreo de medias:

$\mu_{\bar{x}}$ = Media de la distribución muestral de medias (dmm)

$\sigma^2_{\bar{x}}$ = Varianza de la dmm

$\sigma_{\bar{x}}$ = Desviación estándar de la dmm

reflexiona Tras realizar una muestra estadística, los epidemiólogos Richard Doll y Bradford Hill demostraron que fumar era el principal factor de riesgo para desarrollar cáncer pulmonar.

Distribuciones de muestreo de la media

Cuando de una población de tamaño N sacamos muestras de tamaño n, el número de muestras posibles será diferente si el muestreo se hace con *reemplazo o sin reemplazo.*

Muestreo con reemplazo:

Significa que cuando una unidad esencial ha sido seleccionada aleatoriamente, esa unidad se registra y se reintegra, por lo que puede ser seleccionada nuevamente. El número de muestras posibles cuando el muestreo es con reemplazo se obtiene fácilmente mediante la expresión:

$$N^n$$

Donde, como ya sabemos, "N" es el tamaño del universo y "n" el de la muestra. Así, el muestreo **es con reemplazo**, **de un universo de tamaño 5 habrá 25 muestras de tamaño 2, ya que**

$$\boxed{N^n = 5^2 = 25}$$

Muestreo sin reemplazo:

Significa que la unidad esencial, una vez seleccionada, no tiene oportunidad de aparecer de nuevo en la misma muestra. Bajo esta condición el total de muestras posibles es menor que en **el muestreo con reemplazo** y se obtiene mediante la expresión:

$$\boxed{\binom{N}{n} = \frac{N!}{n!\,(N-n)!}}$$

Así, en muestreo sin reemplazo, **el número de muestras de tamaño 2**, tomadas de un **universo de tamaño 5** es 10, ya que:

$$\binom{5}{2} = \frac{5!}{2!\,(5-2)\,!} = \frac{5.4.3.2.1}{2!\ \ (3!\ )} = \frac{120}{2(6)} = \frac{120}{12} = 10$$

Actividad de Aprendizaje No 37

Los datos siguientes representan la antigüedad en el empleo de un universo de trabajadores:

.75, 1, 1.5, 2, 2

a) Bajo la condición **de muestreo sin reemplazo,** determinar el total de muestras posibles de **tamaño 2** y anotar las muestras

2.- Los datos siguientes representan el número de hijos por familia en un universo formado por cuatro familias: No. De hijos por familia **3, 3, 4, 6,**

Bajo la condición de muestreo CON REMPLAZO, determinar el total de muestras posibles **de tamaño 3.**

3.- De un universo **de tamaño 10**, determinar el total de muestras de **tamaño 2, 3, 4, 5, 6.**

a) Con reemplazo

b) Sin reemplazo

4.- Para el universo dado a continuación y bajo la condición de muestreo sin reemplazo: Determinar el total de muestras posibles de **tamaño 2 y anotar las Muestras**

a) Edad de niños de primaria, en años 7, 7, 8, 10, 12

b) Paso de personas, en kilogramos: 60, 62, 68, 70, 75,

Posteriormente se retomarán los resultados de esta actividad, para explicar lo que es el ERROR DE MUESTREO Y LA DISTRIBUCION DE MUESTREO DE LA MEDIA (dmm).

PENSAMIENTO MATEMÁTICO
PROGRESIÓN No. 12

Valora las ventajas y limitaciones de los estudios observacionales y los compara con el diseño de experimentos, a traves de la revisión de algunos ejemplos tomados de diversas fuentes

METAS	CATEGORÍAS	SUBCATEGORÍAS
M1 Observa y obtiene información de una situación o fenómeno para establecer estrategias o formas de visualización que ayuden a entenderlo	C4 Interaccion y lenguaje matemático	S3 Pensamioento Formal

Para iniciar reflexiona

¿Qué es un estudio observacional?

En un estudio observacional (también conocido como estudio epidemiológico), los investigadores observan a un grupo de personas para ver qué les sucede con el tiempo. Aunque los participantes del estudio pueden responder preguntas y completar cuestionarios, los investigadores no realizan ningún experimento y no tienen control sobre los participantes.

Un estudio observacional es básicamente un ejercicio de estadística. Los investigadores intentan encontrar correlaciones entre ciertos comportamientos y ciertos resultados. Por ejemplo, las personas que comen más verduras, ¿tienen un riesgo mayor o menor de desarrollar cierta enfermedad?

Aunque las estadísticas de los estudios observacionales pueden mostrar asociaciones entre ciertos comportamientos y el desarrollo de una enfermedad o afección, estas asociaciones pueden ser, o no, relaciones de causa y efecto.

En la mayoría de los casos un estudio observacional no es suficiente para poder determinar esta relación. A menudo, un estudio observacional puede proporcionar una **evidencia muy débil**.

Se necesita un tipo diferente de estudio, generalmente un estudio experimental, para probar que algo es causa de algo más, por ejemplo, que tomar café puede hacer que las personas pierdan peso.

Hay buenas razones que respaldan esa famosa frase que dice que “hay tres tipos de mentiras: mentiras, mentiras malditas y estadísticas”.

En la mayoría de los casos, los estudios observacionales (es decir, las estadísticas) pueden proporcionar pistas débiles, pero rara vez pueden probar algo. Por supuesto, el estudio no es una mentira, son sólo datos. Pero la forma en que estos estudios suelen utilizarse en los medios de comunicación para “probar” una cosa un día y exactamente lo contrario al día siguiente, puede traernos la frase a la memoria.

¿Qué es un estudio experimental?

En un estudio experimental relacionado con la nutrición (también conocido como estudio clínico o estudio de intervención), los investigadores proporcionan a los participantes una dieta, educación sobre nutrición u otro tipo de intervención, y evalúan sus efectos.

La evidencia experimental se considera más sólida que la evidencia observacional. Los estudios controlados aleatorios (ECA) a menudo se conocen como el “estándar de oro” para la evidencia. Están diseñados para probar una intervención contra una intervención diferente (es decir, una dieta baja en carbohidratos frente a una dieta baja en grasa), o contra un grupo de control que no cambia sus comportamientos (es decir, una dieta baja en carbohidratos frente a la dieta estadounidense estándar), todo en condiciones muy controladas.

Asignar participantes al azar al grupo experimental o al grupo de control ayuda a garantizar que ambos grupos sean similares en aspectos que no se están probando (como ingresos, educación, nivel de ejercicio, etc.). Esto hace que estos estudios (en el mejor de los casos) tengan una comparación justa, y hace que la evidencia que proporcionan sea mucho más sólida: frecuentemente una **evidencia moderadamente sólida**.

Los mejores ECA utilizan el desarrollo real de la enfermedad en el estudio o la muerte del participante como el resultado a tomar en cuenta. Debido a que las condiciones médicas pueden tardar muchos años en desarrollarse, los ECA que duran décadas son muy costosos, lo que los hace poco prácticos en la mayoría de los casos. Por lo tanto, muchos ECA son mucho más cortos y, en lugar de medir los resultados de salud, miden los cambios en los marcadores de salud que reflejan el riesgo de enfermedad, tales como los cambios en el nivel de azúcar en sangre, la insulina o los niveles de inflamación.

Sin embargo, esto asume que los cambios en un marcador subordinado significan un impacto positivo o negativo en la salud. Como hemos visto en muchos estudios, esto no siempre es así.

Una posible solución definitiva para el futuro en relación con preguntas muy importantes relacionadas con la dieta y la salud humana, puede ser la financiación de muchos menos estudios experimentales, pero de mayor amplitud. Estudios que sean lo suficientemente grandes y amplios para medir consecuencias reales en la salud. Por ejemplo, estudios que comparen una dieta estricta baja en carbohidratos con una dieta convencional en cientos o miles de personas con diabetes tipo 2, durante 5 o 10 años, y midan resultados tales como enfermedad y muerte. Un estudio de alta calidad como este podría fácilmente costar cientos de millones de dólares, pero los resultados podrían valer mucho más para la humanidad

Pros y contras de los estudios observacionales

Pros:

- Son mucho menos costosos que los estudios clínicos
- Pueden durar varios años o incluso décadas
- Pueden incluir decenas de miles de participantes en el estudio
- Pueden considerar el desarrollo de una enfermedad o la muerte como un resultado

Contras:

- Se basan en datos auto informados que no se pueden confirmar
- No tienen grupo de control para la comparación
- No se pueden tener en cuenta todos los factores que pueden influir en los resultados
- No son relaciones causa y efecto

Pros y contras de los estudios experimentales

Pros:

- Están estrechamente controlados y monitoreados
- Comparan los resultados entre los que reciben una intervención y los que no
- Pueden usar aleatorización para hacer frente a factores desconocidos que podrían influir en los resultados

Contras:

- Son caros y consumen mucho tiempo
- Usan marcadores de salud en lugar de desarrollo de enfermedad o muerte como parámetros
- Son típicamente más pequeños que la mayoría de los estudios observacionales
- Suelen ser más cortos que la mayoría de los estudios observacionales

Sabías que….

Hacemos estudios para recolectar información y sacar conclusiones. El tipo de conclusión que sacamos depende del método de estudio utilizado:

- **En un *estudio observacional*,** medimos o sondeamos a los miembros de una muestra sin tratar de afectarlos.
- **En un *experimento controlado,*** asignamos aleatoriamente personas o cosas a grupos y uno de los grupos sigue cierto tratamiento mientras que el otro no recibe el tratamiento.

Actividad de Aprendizaje No 38

Escoge una respuesta

Problema 1: beber té antes de dormir

Se hizo un estudio en el que se tomó una muestra aleatoria de adultos y se les preguntó sobre sus hábitos para dormir. Los datos mostraron que la gente que tomaba una taza de té antes de ir a dormir tendía a irse a dormir más temprano que aquellas personas que no tomaban té

¿Qué tipo de método de estudio es este?

A) Estudio observacional
B) Experimento

En otro estudio se tomó un grupo de adultos y se dividió aleatoriamente en dos grupos. A un grupo se le pidió beber té cada noche por una semana, mientras que al otro grupo se le pidió no beber té esa semana. Luego los investigadores compararon el momento en el que los individuos de cada grupo se quedaban dormidos.

¿Qué tipo de método de estudio es este?

A) Estudio observacional
B) Experimento

Problema 2: redes sociales y felicidad

En un estudio se asignaron aleatoriamente voluntarios a uno de los dos grupos:

- A un grupo se le indicó usar las redes sociales como de costumbre.
- Al otro grupo se le bloqueó el acceso a las redes sociales.

Los investigadores observaron cuál grupo tendía a ser más feliz.

¿Qué tipo de método de estudio es este?

A) Estudio observacional
B) Experimento

En otro estudio se tomó una muestra aleatoria de gente y se examinaron sus hábitos en redes sociales. Cada persona se clasificó como usuario esporádico, moderado o frecuente de las redes sociales. Los investigadores observaron cuál grupo tendía a ser más feliz.

¿Qué tipo de método de estudio es este?

A) Estudio observacional
B) Experimento

https://www.youtube.com/watch?v=w7UY8gi29fk&t=41s

Problema 3: Píldoras Placebo y Píldoras reales

Un grupo de doctores se intereso en comparar la efectividad de las píldoras placebo (hechas con azúcar) y píldoras reales para tratar migraña. Tomaron un grupo de 300 pacientes que sufren de migraña y se dividieron al azar en dos grupos. A un grupo le asignaron la píldora real y al otro el placebo.

Antes y después de tomar las píldoras, a los pacientes se les dio un cuestionario con preguntas acerca de su condición. Luego los doctores analizaron y compararon los cuestionarios de cada grupo.

¿Qué tipo de método de estudio es este?

A) Estudio observacional

B) Experimento

Problema 4: Vacas y Producción de Leche

Investigadores Británicos hicieron un estudio sobre la relación entre el trato de los granjeros con sus vacas y la producción de leche de las vacas. Prepararon una encuesta con respecto a la percepción que tienen los granjeros sobre la capacidad mental de las vacas, el trato hacia las vacas, y la producción de leche de las vacas, La encuesta se contestó en todas las granjas de Gran Bretaña.

Después de analizar los resultados, encontraron que las granjas donde a las vacas se les llamaba por su nombre, la producción era, en promedio, 258 litros mayor que en las granjas en donde no se les llamaba por su nombre.

¿Qué tipo de método de estudio es este?

A) Estudio observacional
B) Experimento

Problema 5: Tabaco

Supongamos que queremos estudiar el efecto del tabaquismo sobre la capacidad pulmonar de varones mayores de 60 años

Por tanto, buscamos 100 voluntarios en ese rango de edad, de los cuales 50 han fumado un paquete al día durante 10 años, mientras que los otros 50 han estado libres de humo durante ese lapso de tiempo

Luego medimos la capacidad pulmonar de cada uno de los 100 hombres.

Finalmente, analizamos, interpretamos y derivamos conclusiones a partir de los resultados

¿Qué tipo de método de estudio es este?

A) Estudio observacional
B) Experimento

Problema 6: Tabaco

Siguiendo con el ejemplo del tabaquismo…

Supongamos que encontramos 100 hombres de 60 años que son fumadores habituales

Asignamos aleatoriamente a 50 de los 100 hombres a un tratamiento contra el tabaquismo y a los otros 50 se les administra un placebo

Luego de un período de X tiempo que dura el experimento, contabilizamos cuántos de ellos han logrado superar el hábito del tabaquismo y cuántos no

Analizamos, interpretamos y sacamos conclusiones acerca de la efectividad del tratamiento

¿Qué tipo de método de estudio es este?

A) Estudio observacional
B) Experimento

PENSAMIENTO MATEMÁTICO
PROGRESIÓN No. 13

Describe un fenómeno, problemática o situación de interés para el estudantado utilizando las medidas de tendencia central(media, mediana, y moda) y de dispersión (desviación estándar, varianza, rango intercuartil, etc.) adecuadas al contexto y valora que tipo de conclusiones puede extraer a partir de dicha información .

METAS	CATEGORÍAS	SUBCATEGORÍAS
M4 Argumenta a favor o encontra de afirmaciones acerca de situaciones, fenómenos o problemas propios de la matemática, de las ciencias o de su contexto.	C2 Procesos de intuición y razonamiento	S1 Capacidad par observar y conjeturar S2 Pensamiento Intuitivo S3 Pensamioento Formal
M3 Aplica procedimientos, técnicas y lenguaje matemático para la solución de problemas propios del Pensamiento Matemático, de Áreas de Conocimiento, Recursos Socioemocionales y de su entonrno	C3 Solución de problemas y modelación	S1 Capacidad par observar y conjeturar S3 Pensamioento Formal

Aprende Mas

MEDIDAS DE TENDENCIA CENTRAL

Las medidas de tendencia central son categorías o puntos dentro del recorrido de una variable; se les llama de tendencia central porque en torno a ellas parecen agruparse los datos.

Sirven para resumir todo un conjunto de valores; por ello bien se les puede considerar como sintetizadores. En general, cualquier medida, de tendencia cantidad es un valor medio, ya que este, por definición es todo valor que se halla entre dos extremos. Por esta razón a cada indicador de tendencia central se le nombra y define de manera diferente.

Notación sigma

En la inmensa mayoría de los cálculos estadísticos siempre hay un paso que consiste en sumar conjuntos de cantidades. Por esta razón se ha convenido en utilizar un símbolo propio, la letra griega sigma ($\sum$), para indicar que hay que sumar lo que aparece a su derecha. Este símbolo se conoce escuetamente como sumatoria y se debe leer: “efectuar la sumatoria de” o, simplemente, “la sumatoria de”.

También por conversión se usan en mayúsculas las últimas letras de nuestro alfabeto para denotar variables (X, Y, Z, W, etc.) y las primeras para constantes (A, B, C, etc.)

Por ejemplo, X puede representar el salario de un grupo de trabajadores, el número de hijos por familia, el número de accidentes de carretera por entidad federativa, etc.

Punto que algunos autores estadística suelen asociar subíndices tanto a los símbolos que denotan variables como a la sumatoria, explicaciones con este ejemplo:

Supongamos que se indagan dos variables de un universo de familias: el número de hijos de sexo masculino y el de sexo femenino en cada una, como se muestra a continuación.

Número de Hijos

	Mujeres	Hombres
Familia 1	1	2
Familia 2	3	1
Familia 3	5	6
Familia 4	2	3
Familia 5	4	4

Sea X la variable “H de mujeres” y la variable “H de hombres” entonces podemos escribir los datos o categorías de estas variables así:

$X_1 = 1$ $X_2 = 3$ $X_3 = 5$ $X_4 = 2$ $X_5 = 4$

$Y_1 = 2$ $Y_2 = 1$ $Y_3 = 6$ $Y_4 = 3$ $Y_5 = 4$

Tenemos que X_2 representa el dato de la variable X correspondiente a la familia 2 o sea $X_2 = 3$ Y_3, el de la variable Y correspondiente a la familia 3, o sea $Y_3 = 6$, si, para generalizar, escribimos Xi, estamos denotando el valor de la variable X en la i – ésima familia, o sea, en cualquiera. Se sabrá a que familia nos referimos cuando asignemos a i un número entre 1 y N, es decir, entre 1 y el total de datos.

En resumen: un símbolo de variable con subíndice numérico representa cierto valor de esa variable. Si el subíndice es literal *(Es habitual el uso de las minúsculas i, j),* el símbolo no se refiere a ningún valor definido sino a cualquiera, como X o Yi, veamos cómo se interpretaría.

$\sum_{i=1}^{N} Xi$ *Ordena sumar las categorías de la variable X, empezando por la primera y terminando por la enésima.*

$\sum_{i=1}^{5} Xi$ *Indica que se debe sumar sus categorías desde la primera hasta la quinta, es decir todas.*

$$\sum_{i=1}^{5} Xi = X_1 + x_2 + X_3 + X_4 + X_5 = 1 + 3 + 5 + 2 + 4$$

$$\sum_{i=1}^{5} Xi = 15$$

Si la expresión fuera como estas

$$\sum_{i=1}^{4} X_i \qquad \textit{Tiene que sumar, desde el primero hasta el cuarto.}$$

$$\sum_{i=3}^{5} X_i \qquad \textit{Sumar desde el tercero hasta el último}$$

MEDIA ARITMETICA

Es la más conocida de las medidas de tendencia central, aunque no con ese nombre. A ella nos referíamos desde que pasamos por las aulas de la enseñanza elemental, cuando nos preguntábamos, al término de un ciclo escolar, cuál era el *"promedio"* que habíamos obtenido dándonos la respuesta con solo sumar las calificaciones alcanzadas en las diferentes asignaturas y dividir el resultado entre el número de ellas. Se le conoce también con los nombres de *valor medio, promedio aritmético* o simplemente media, como nosotros la llamaremos. Se le simboliza con cualquiera de las letras convenidas para representar variables, coronada con una barrita. *Se le define como la suma de un conjunto de cantidades dividida entre el número de ellas. En símbolos,*

$$\overline{X} = \frac{\sum X}{N}$$

Donde X simboliza los datos de una variable y N, el número de ellos, ejemplo.

Si X = 6, 7, 8, 9, 10, su promedio aritmético es

$$\overline{X} = \frac{\sum X}{N} = \frac{6 + 7 + 8 + 9 + 10}{5} = \frac{40}{5}$$

$$\overline{X} = 8$$

Resulta evidente que la suma de los datos de una variable es igual al producto de su valor medio por el número de datos.

$$\sum X = N\overline{X}$$

sustituyendo

$$\sum X = (5)\ (8) = 40$$

MEDIANA

Llamada también valor mediano, *es el punto dentro del recorrido de una variable que supera a no más de la mitad de los datos y es superado por no más de la otra mitad*. Dicho de otra manera: es un punto dentro de una distribución de datos que tiene la característica <u>de dividirla en dos partes iguales</u>. La identificamos con el símbolo Me.

Tratándose de series de datos sin frecuencia asociada, no se necesita ninguna fórmula para hallarla, pero es <u>preciso ordenarlos</u> de menor a mayor o viceversa, ejemplos:

Determina el valor mediano de las distribuciones siguientes.

1. 10, 6, 7, 9, 8

   ordenando de menor a mayor

   6, 7, $\boxed{8}$, 9, 10

   Me = 8 ya que divide en dos partes iguales.

2. 5, 6, 7, 8, 9, 10

   5, 6, $\boxed{7,\ 8}$ 9, 10

   $$Me = \frac{7+8}{2} = \frac{15}{2} = 7.5$$

Si el número de datos de la variable es par, la mediana es la semisuma de los dos valores intermedios que satisfacen su definición.

Número de Orden de la Mediana

Una forma rápida de obtención de la mediana, es una serie de ordenada de datos, es determinar el número de orden que le corresponde. Para hallarlo basta con sumar una unidad al total de datos (N) y dividir entre dos.

$$Me = \frac{N+1}{2}$$

Determinar el valor mediano de las distribuciones siguientes, hallando primero el número de orden que les corresponde:

X = 12, 13, 17, 21, 23, 25, 30,

Y = 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65

SOLUCION:

Para X

$$\text{No.} = \frac{N+1}{2} \qquad \text{No.} = \frac{7+1}{2} = \frac{8}{2} = 4$$

** El valor mediano tiene número de orden 4, es decir, es el cuarto del conjunto de datos ordenados de menor a mayor, por lo tanto

$$Me = 21$$

Para Y

$$\text{No.} \ \frac{N+1}{2} = \frac{12+1}{2} = 6.5$$

Como no puede haber número ordinal, fraccionario, 6.5 indica que el valor mediano está comprendido entre los datos que ocupan el sexto y el séptimo lugar, lo cual implica hallar la semisuma de 35 y 40

$$Me = \frac{35+40}{2} = \underline{37.5}$$

Actividad de Aprendizaje 39

1,- Siete jóvenes compitieron en la carrera de 100 metros planos. Los siguientes datos representan los tiempos, en segundos, que necesitaron para realizar el recorrido,

12 14 15 13 11 12 16

Encontrar: a) La Media b) La Mediana

2.- Las edades de las diez personas que acuden a solicitar empleo a una Institución Bancaria son las siguientes: 18 30 25 23 25 24 18 25 20 18

Encontrar: a) La Media b) La Mediana

MODA (MO)

También llamada *moda o valor modal, es el dato de variable que aparece más veces en una distribución.* En otras palabras: *es el dato de variable que tiene mayor frecuencia.* Se trata, de hecho, del caso más notorio o típico de una distribución de datos.

En la distribución

2, 3, 2, 4, 4, 2, 5

La moda es 2, pues es el dato que tiene mayor frecuencia: aparece 3 veces.

En la distribución

3, 4, 3, 4, 3, 4

No hay moda, porque ningún dato parece más veces que otro.

Una distribución puede tener más de una moda si 2 o más datos, o clases de datos, tienen la misma frecuencia y esta es la más alta de la distribución. Así, según el número de moda que tengan, las distribuciones reciben nombres específicos: con solo una moda, unimodal; con dos modas, bimodal; con tres modas, trimodal y con más de tres modas, multimodal.

Actividad de Aprendizaje 40

Las edades de las diez personas que acuden a solicitar empleo a una Institución Bancaria son las siguientes: 18 30 25 23 25 24 18 25 20 18 obtener la moda de estos valores.

https://edu.gcfglobal.org/es/estadistica-basica/media-mediana-y-moda/1/#

MEDIA GEOMETRICA (Mg)

En algunas ocasiones, cuando la variable depende del tiempo, como sucede en casos del comercio y la economía, las medidas de posición anteriores no son las más adecuadas, y se recurre a otra denominada media geométrica (mg).

Esta se define como la raíz N-ésima del producto de las N observaciones y se expresa simbólicamente así:

$$Mg = \sqrt[N]{X1 \cdot X2.\ X3 \cdots XN}$$

La expresión anterior nos lleva a concluir que ningún valor de la variable debe ser nulo o negativo y este se cumple en los fenómenos económicos. Lo anterior se puede ilustrar por medio de una tabla, calculando la media geométrica de los precios del producto Z en los años 1970 – 1975.

Años	Precio del producto Z
1970	12.50
1971	14.20
1972	16.70
1974	20.00
1975	25.10

$$mg = \sqrt[.5]{(12.50)\ (14.20)\ (16.70)\ (20.00)\ (25.10)}$$

$$mg = \sqrt[.5]{1{,}488{,}053.50}$$

mg = $ 17.16

Actividad de Aprendizaje 41

Determina la media, la mediana y la Moda de las distribuciones siguientes.

W = 5, 5, 4, 7, 9, 10, 3, 6, 5, 8
X = 8, 9, 6, 8, 4, 5, 7, 3, 10, 8
Y = 10, 11, 15, 14, 12, 12, 13, 12, 16
Z = 11, 15, 8, 9, 14, 10, 10, 13, 13, 12

2.- Resolver el siguiente problema apoyándote en las diferentes medidas de tendencia central.

El problema de la campaña publicitaria. Los siguientes datos corresponden al número de kilómetros recorridos por litro de gasolina en cinco pruebas para tres diferentes marcas de autos compactos:

Marca de auto	Km recorridos por litro de gasolina				
A	12.0	10.8	13.5	12.0	13.0
B	10.5	12.8	12.9	12.8	14.0
C	11.5	14.0	12.5	10.0	12.5

a) Se te ha contratado para impulsar una campaña publicitaria para promocionar que el auto marca A es el que ofrece el mayor rendimiento de kilometraje por litro de gasolina, ¿en qué medida de centralización te apoyarías? ___________________

¿Por qué?
____
____
____
____

Escribe de forma breve el texto de la campaña:

____
____
____

b) Considera ahora que eres el responsable de publicitar a la auto marca C como el de mayor rendimiento de kilómetros por litro de gasolina, ¿En qué medida estadística basarás tu estrategia? ____

¿Por qué?

____
____
____
____

c) Para el caso del auto compacto marca B, ¿Cuál es tu estrategia de publicidad?

____
____
____
____

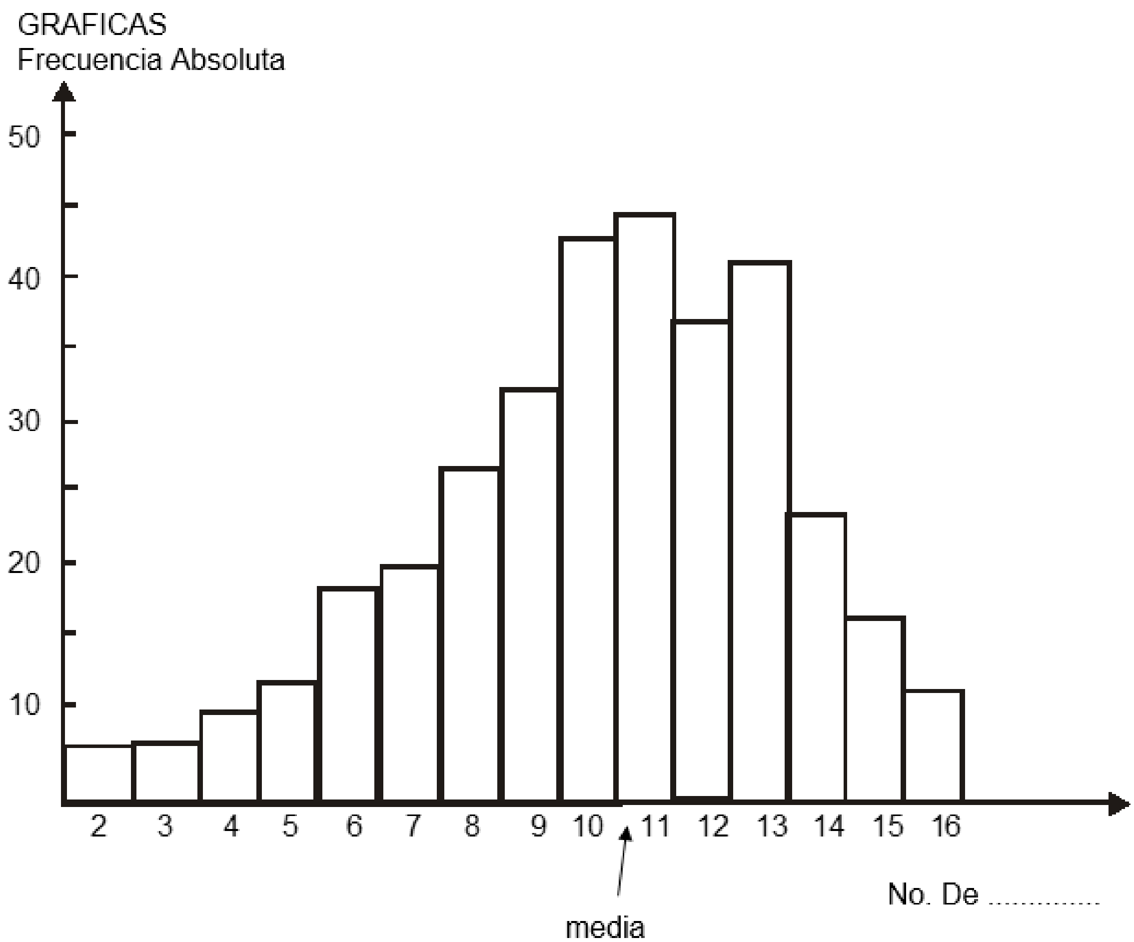
d) Después de las consideraciones anteriores, si tienes que elegir sólo una de las marcas de auto como el de mayor rendimiento ¿Cuál eliges?

____

¿Por qué?

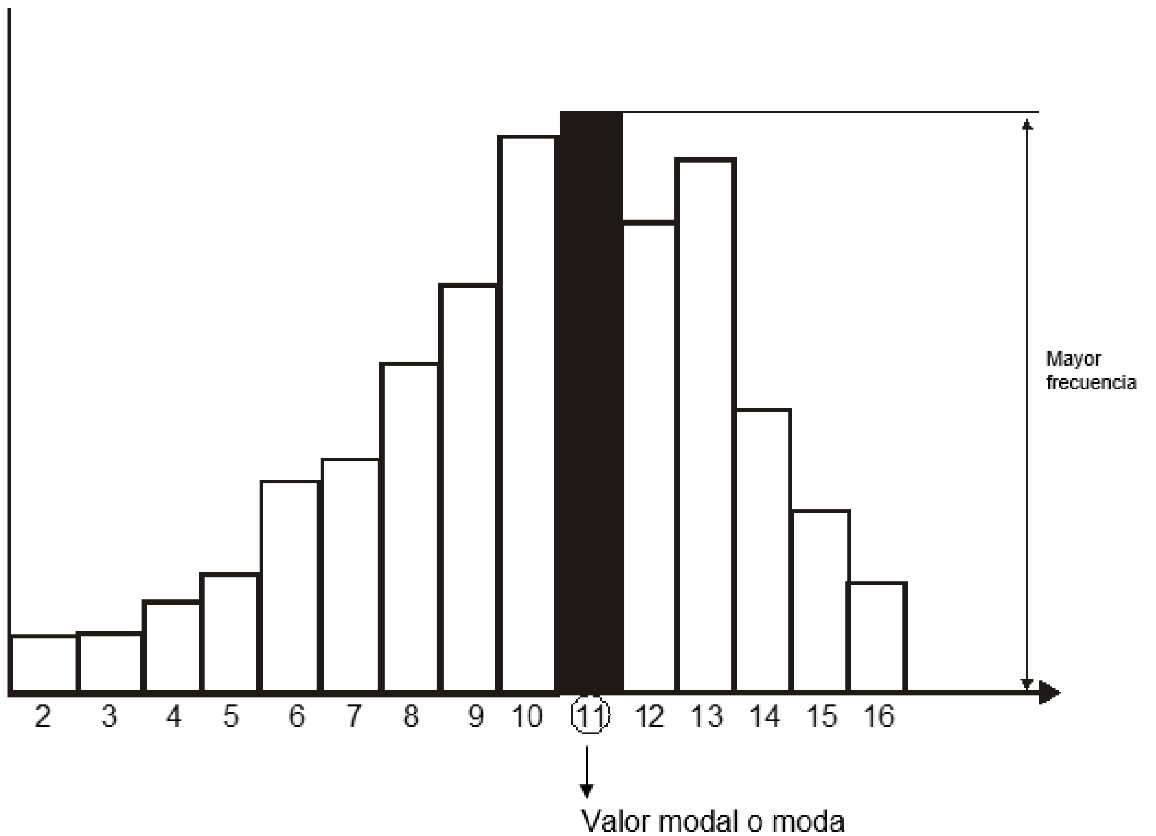
____
____
____
____

Comparte tus respuestas con los demás integrantes de tu comunidad, reflexionen, intercambien opiniones y concluyan.

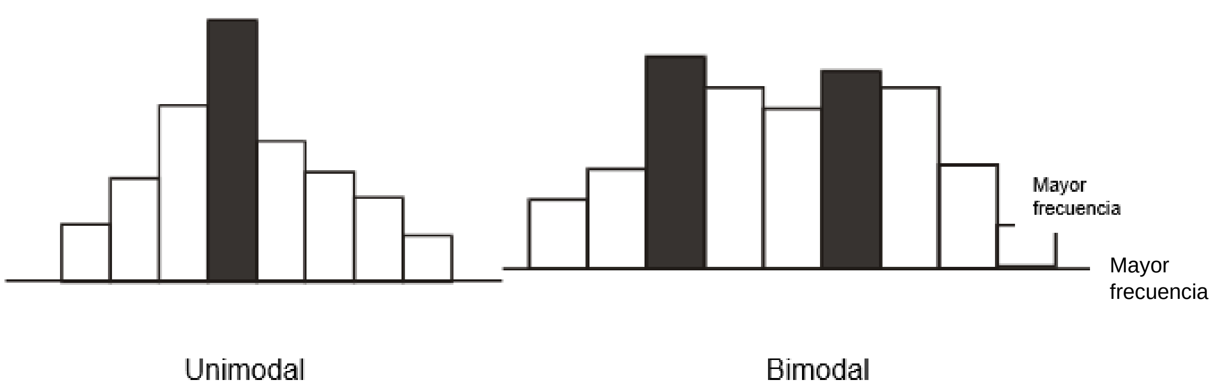


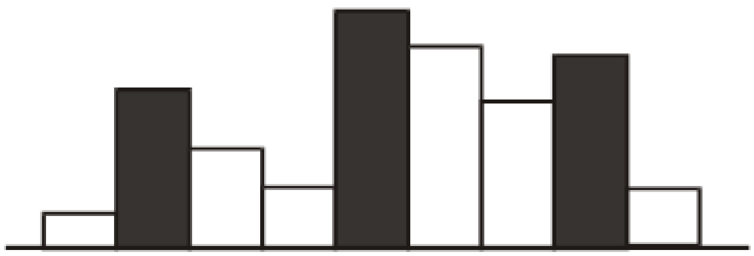
No. De ....

Media es un valor centrado entre los límites del rango

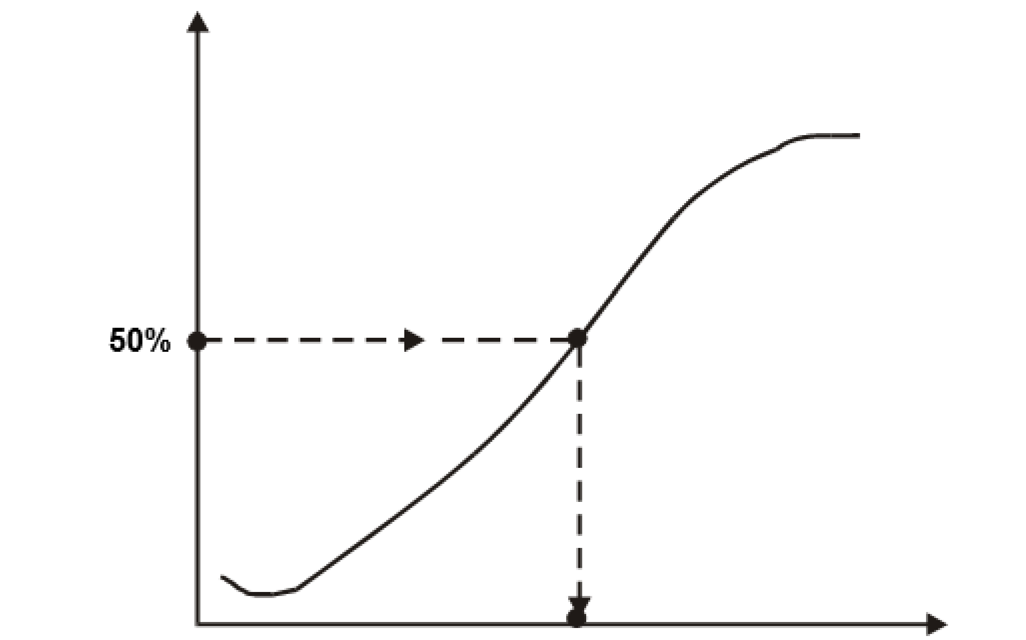


Para fines jurídicos, se considera a la moda como la marca de clase del intervalo que presenta mayor altura. (más frecuencia) que los intervalos vecinos.





Multimodal



CALCULO DE LA MEDIA, LA MEDIANA Y LA MODA
(Datos No Agrupados)

La siguiente tabla presenta una distribución de datos no agrupados, conocida también como distribución simple de frecuencias.

Jornada de Trabajo
De los obreros de la Fábrica A

Horas	No. De obreros
	72
4	5
6	22
8	18
10	15
12	12

Si preguntamos ¿Cuántos datos tiene esta distribución? Algunos podrían responder: *cinco*. Esta respuesta errónea tendría como origen el haber olvidado que se trata de una estructura que muestra las distintas categorías de la variable acompañadas de la frecuencia que les corresponde.

Obsérvese que la *4 aparece 5 veces; la 6, 22 veces, etc*. Podemos afirmar que existen cinco categorías repetidas una y otras de tal manera que sumadas dan 72. La distribución consta de 72.

Las diversas formas en que se pueden distribuir los datos de una variable numérica determinan los valores de sus tendencias centrales. Por esta razón, para hallar, por ejemplo, la mediana, no podríamos limitarnos a sumar las categorías (4, 6, 8, 10, 12) y dividir entre 5; *COMETERIAMOS UN ERROR* de procedimiento inadmisible, por no tener en cuenta sus frecuencias. Incurriríamos en otro error del mismo tipo si dijéramos que la mediana es 8 horas por el solo hecho de ser la categoría que está en medio de todas. Puede ser 8 ciertamente, pero también 6 u otro valor. Esto depende del número de casos que corresponden a cada categoría.

Despleguemos ahora las observaciones tabuladas como una serie de datos sin frecuencia asociada, de la tabla tenemos que (º lugar)

1º al 5º	6º al 27º	28° al 45º	46º al 60º	61º al 72º
4	6	8	10	12

Esto es a lo que llamaremos dfa (+) (distribución frecuencia acumulada).

A los datos que se presentan como en este caso las horas le designamos una letra en este ejemplo X, para encontrar fx se multiplican las horas por la frecuencia, y tendríamos la siguiente información de la tabla anterior.

X = No. De horas f = No de Obreros

X	f	dfa(+)	fx
4	5	5	20
6	22	27	132
8	18	45	144
10	15	60	150
12	12	72	144
	$\sum f = 72$		$\sum FX = 590$

En distribución de datos no agrupados, la **media** se obtiene mediante la fórmula.

$$\bar{X} = \frac{\sum fx}{\sum f}$$

donde X representa cada dato de la variable; f la frecuencia que le corresponde y $\sum f$ el total de datos (N). por lo tanto:

$$\bar{X} = \frac{\sum fx}{\sum f} = \frac{590}{72} = 8.19$$

$$\bar{X} = 8.19$$

Interpretación: La Jornada de trabajo **" en promedio"(**no debe de faltar esta leyenda) de los obreros de la Fabrica A es de 8.19 hrs.

Para hallar la mediana se siguen dos pasos:

Se determina el número de orden que le corresponde sumando una unidad al total de datos N y dividiendo entre dos.

$$\text{No.} = \frac{N+1}{2} = \text{(posición que ocupa)}$$

Mediana (Me)

No. $= \dfrac{N+1}{2} = \dfrac{72+1}{2} = \dfrac{73}{2}$ 36.5 del dfa = 45 se localiza en tercer lugar

Equivalente en valor de "X" A 8 horas por lo tanto la **Me = 8 Hrs.**

Se ordenan los datos, la mediana es el dato de la variable cuya frecuencia acumulada contiene o señala su número ordinal (posición que ocupa).

La Moda es el dato de mayor frecuencia.

Actividad de Aprendizaje 42

CALCULAR E <u>INTERPRETAR</u> SUS MEDIDAS DE TENDENCIA CENTRAL

*** Entiéndase por "vida útil" a la variable "tiempo de duración"

1. Un estudio ausentismo de los obreros de la fábrica A en cierto mes del año, condujo a la tabla siguiente:

Número De días	Total 72
0	17
1	25
2	15
3	7
4	4
5	2
6	2

2.-Una Fábrica lanza al mercado un nuevo tipo de lámpara incandescente. Para determinar la vida útil promedio de esas lámparas, se escogieron al azar 120 y se les sometió a prueba de duración, obtendríamos los datos siguientes:

Miles De Horas	Total 120
6.0	10
6.5	20
7.0	40
7.2	28
7.6	14
8.2	8

CALCULO DE LA MEDIA, MEDIANA Y LA MODA (Datos Agrupados)

Para realizar este cálculo, es necesario hacer las siguientes consideraciones.

- De las tablas anteriores a este tema, se seguirán calculando de igual forma, con las siguientes observaciones

- La columna de Datos ahora son los intervalos (X)
- La columna de frecuencia se mantiene igual
- La columna dfa de igual forma.
- Agregamos una nueva columna PUNTOS MEDIOS (X) los cuales se obtienen precisamente del punto medio de cada intervalo.
- (fx) se obtiene en cada renglón de multiplicar de la frecuencia (f) por puntos medios (x).

Para aclararlo más ampliamente contestaremos el siguiente ejemplo:

Calcular e interpretar la media, la mediana y la moda de la tabla siguiente:

Alumnado según tiempo dedicado
Al estudio fuera de clases
1984

Horas Semanarias	Total 188
1 – 3	50
4 – 6	38
7 – 9	26
10 – 12	36
13 – 15	19
16 – 18	7
19 – 21	7
22 – 28	5

Solución: complementando la tabla.

X = Horas Semanarias

X	f	dfa	PUNTOS MEDIOS (X)	fx
1 – 3	50	50	2	100
4 – 6	38	88	5	190
7 – 9	26	114	8	208
10 – 12	36	150	11	396
13 – 15	19	169	14	266
16 – 18	7	176	17	119
19 – 21	7	183	20	140
22 – 28	5	188	25	125
	$\sum f = 188$			$\sum fX = 1544$

$$\text{MEDIA} = \bar{X} = \frac{\sum fX}{\sum f}$$

Sustituyendo valores

$$\bar{X} = \frac{1544}{188} = 8.2 \text{ horas semanarias}$$

INTERPRETACIÓN: En 1984, el alumno sometido a investigación dedicada al estudio fuera de clases 8.2 Horas Semanarias en promedio.

MEDIANA: Primero identificamos en que intervalo se encuentra

$$\text{No.} = \frac{N+1}{2} = \frac{188+1}{2} = \frac{189}{2} = 94.5, \text{ la frecuencia acumulada}$$

es 114, señalando el intervalo tercero. Esto significa que *<u>LA MEDIANA</u> será el valor comprendido entre los limites reales del intervalo encontrado y el anterior*. Que en este caso los limites reales son 6.5 y 9.5 encontrado estos valores aplicamos la siguiente fórmula.

$$\boxed{Me = L + \left(\frac{N/2 - \sum fd}{f\,j} \right) j}$$

Identificando tenemos

L = es el límite real inferior (*6.5)*
N = es el total de los datos de la distribución (*188)*
$\sum$fd = la frecuencia acumulada inmediatamente menor a la de intervalo *(en este caso, segunda) (88).*
fj= frecuencia del intervalo encontrado (*tercero) (26)*
j = la anchura real del intervalo (*3)*
() = Lo que se encuentra entre paréntesis es alusivo a este ejemplo por lo tanto tendríamos sustituyendo los valores.

$$Me = 6.5 + \left(\frac{188/2 - 88}{26} \right) 3 =$$

$$= 6.5 + \left(\frac{94 - 88}{26} \right) 3 =$$

$$= 6.5 + \left(\frac{6}{26} \right) 3 =$$

$$= 6.5 + \left(0.23 \right) 3 =$$

$$= 6.5 + 0.69 =$$

$$= 7.19$$

Me = 7.2 horas

INTERPRETACIÓN: En 1984, la mitad *(en datos agrupados utilizamos esta expresión, del contrario sería incorrecto)* del alumnado dedicaba al estudio fuera de clases entre 1 y 7.2 horas a la semana (o 7.2 horas como máximo)

MODA: En cuanto a la moda, no hay nada que computar; es simplemente la marca de clase *del primer intervalo*, puesto que es la de más alta frecuencia.

Mo. = 2 horas semanarias

INTERPRETACIÓN: Entre el alumnado en 1984, el caso más notorio era el de los que dedicaban 2 horas semanarias al estudio fuera de clases; el 26.6% de los alumnos.

Nota: En todos los ejemplos vistos a este punto son variables cardinales.

Actividad de Aprendizaje 43

Calcular e interpretar la media, la mediana y la moda de la tabla siguiente:

Profesorado según horas Semanales frente a grupo	
Horas	Total 23
4 – 6	5
7 – 9	2
10 – 12	3
13 – 18	8
19 – 21	3
22 – 24	2

La Media, La Mediana Y La Moda En Variable Nominal Y Ordinal

Se han comentado los procedimientos de obtención de la media, la mediana y la moda en variable cardinal, tanto en distribuciones simples como en datos agrupados.

Surge ahora la pregunta de si esos indicadores son calculables en los otros dos tipos de variables, para responder con fundamento, basta con conocer las

características de las nominales y las ordinales y las propiedades de los símbolos asignados a sus categorías, así como lo inherente a las operaciones que conducen a los valores de la media, la mediana y la moda.

Nominales: Son mutuamente excluyentes, no requieren orden.

Ordinal: Guardan relaciones “mayor que” necesitan orden apropiado en estas dos si se asignan números a sus categorías ellos sirven para distinguir una de otra; *pero no son susceptibles de ninguna operación aritmética, pues carecen de propiedades numéricas,* para presentarlo de manera sencilla a manera de tabla las variables y sus promedios que pueden ser medibles, tenemos:

Variable	Promedios Medibles
Nominal	Moda
Ordinal	Moda, Mediana
Cardinal	Moda, Mediana, Media

Ejemplos: Determinar e interpretar las medidas de tendencia central de las siguientes tablas

Estaciones radiodifusoras según
Bandas de frecuencia en que operan
México, 1991.

Banda	Total 965
Amplitud Modulada	712
Frecuencia Modulada	242
Onda Corta	11

Fuente: INEGI, Agenda Estadística, 1992. P. 30

SOLUCIÓN: Variable Nominal "Bandas de Frecuencia" no tiene más medida de tendencia central que su valor modal.

Mo. = amplitud modulada

INTERPRETACIÓN: Su significado es claro: el caso más notorio es el de las radiodifusoras que operan con banda de frecuencia de amplitud modulada. El 73.8% de las radiodifusoras del país tienen esa característica.

Empleados Solteros por
Grado de Contribución al Gasto Familiar

Grado de Contribución	total 79
No contribuye	23
Mínimo	16
Con la mitad	21
Con más de la mitad	9
Con todo	10

Complementando la tabla

X	f	dfa (+)
No contribuye	23	23
Mínimo	16	39
Con la mitad	21	60
Con más de la mitad	9	69
Con todo	10	79

Solución: Variable Ordinal: “grado de contribución al gasto familiar”

Tiene, por lo tanto, valores modal y mediano

Moda: **Mo = No contribuye**

INTERPRETACIÓN: El caso más notorio es el de los empleados solteros que no contribuyen con su salario al gasto familiar.

MEDIANA: $\text{No.} = \frac{N+1}{2} = \frac{79+1}{2} = \frac{80}{2} = 40 \quad \text{dfa }(+) = 60$

por lo tanto, **Me** = con la mitad

INTERPRETACIÓN: Por lo menos el 50% de los empleados solteros no contribuyen, o contribuyen máximo con la mitad de sus ingresos al gasto familiar.

Actividad de Aprendizaje 44

Determinan las medidas de tendencia central de la siguiente tabla.

Estudiantes por tipo de actividad
Artística o cultural que practican

Actividad	Total 77
Ninguna	40
Instrumentos musicales	17
Pintura	4
Ajedrez	2
Teatro	2
Otras	12

MEDIDAS DE DISPERSION

A menudo escuchamos que en los países latinoamericanos existe mucha *diferencia* entre los ingresos que perciben las diversas capas sociales de la población. En cambio, en países más desarrollados, esa diferencia es menor.

Esas diferencias tienen sus raíces en distintos fenómenos sociales, políticos y económicos; sin embargo, un economista diría que “el ingreso per cápita en los países latinoamericanos esta más disperso que el ingreso per cápita de los países más desarrollados”

El concepto de *dispersión* resulta importante para los estudios económicos, ya que puede darse el caso de poblaciones con igual valor central, pero una puede estar mas *dispersa* que la otra.

Dos o más distribuciones pueden tener iguales valores de tendencia central y no obstante, mostrar grados de dispersión diferentes, por ejemplo:

Tenemos dos conjuntos de datos de variable cardinal.

Distribución X

1, 2, 3, 4, 4, 4, 5, 6, 7,

Distribución M

3, 4, 4, 4, 5

Calculemos la media, la mediana y la moda para cada conjunto.

Distribución X

$$\bar{X} = \frac{\Sigma X}{N} = \frac{1 + 2 \ldots 7}{9} = \frac{36}{9}$$

$$\bar{X} = 4$$

$$Me = No = \frac{N+1}{2} = \frac{9+1}{2} = \frac{10}{2} = 5$$

Me = 4 (ocupa 5º lugar)

MODA = 4

Distribución M

$$\bar{M} = \frac{\Sigma X}{N} = \frac{3 + 4 \ldots 5}{5} = \frac{20}{5}$$

$$\bar{M} = 4$$

$$Me = No = \frac{N+1}{2} = \frac{5+1}{2} = 3$$

Me = 4 (ocupa 4º lugar)

MODA = 4

Las dos distribuciones tienen los mismos promedios y, no obstante, muestran una diferencia notoria: la X tiene más dispersos sus datos en torno a la tendencia central que la M. Si los datos representaran salarios en miles de pesos, por ejemplo, sería evidente que la distribución M reflejaría una mejor distribución del ingreso.

MEDIDAS DE DISPERSION

RANGO:

Se trata de la más simple de las medidas de dispersión. *Representa la distancia entre el menor y el mayor de los datos de una distribución*, por lo cual puede ser interpretado como la dispersión total de todos ellos. *Como es "distancia", se le obtiene restando el dato menor del mayor*; consecuentemente, es calculable únicamente en variable cardinal, ejemplo

DISTRIBUCIÓN P

2, 3, 4, 4, 5, 5, 5, 6, 7

DISTRIBUCION Q

2, 3, 4, 5, 6, 6, 7

En ambas el rango es el mismo DATO MAYOR – DATO MENOR (7 – 2) Pero observándoles y comparándolas descubrimos que la distribución P muestra una mayor concentración de datos en torno a sus tendencias centrales que la Q.

Por esta razón no sirve, por si solo, para dar cuenta objetivamente de la desviación en su conjunto; más que nada se le debe de usar como complemento de otras medidas de dispersión que estudiaremos más adelante.

DESVIACION MEDIA (D. M.)

:
Hasta fines del siglo pasado, la *desviación media* fue la medida de dispersión de más uso.

Su desplazamiento del arsenal estadístico se debió a la aparición del concepto de desviación estándar, otra medida de dispersión con mejores propiedades algebraicas y que genera valores numéricos muy parecidos a los que se obtienen con la desviación media. Aunque haya caído en desuso, conviene estudiarla debido a que su significado - fácil de comprender - facilita la comprensión de la desviación estándar concepto de gran utilidad en el quehacer estadístico y con el cual esta muy aparentada.

La desviación media se define como la desviación promedio de los valores absolutos de las desviaciones de los datos de una variable con respecto a su media y se expresa en las mismas unidades de la variable (años, horas, pesos, etc.)

$$\boxed{D.M. = \frac{\sum |X - \bar{X}|}{N}}$$

Para encontrar la desviación media de una serie de datos sin frecuencia asociada, basta con dar tres pasos sencillos:

A) Se calcula la media

B) Se resta la media de cada dato de la variable, lo cual produce la separación de cada dato respecto a la media y...

C) Se divide la sumatoria de los valores absolutos de esas separaciones entre el total de datos.

Ejemplo:

Hallar e interpretar la desviación media de los datos de la variable X dados a continuación.

$$X = 1, 2, 3, 4, 4, 4, 5, 6, 7.$$

Solución: *(Paso A)*

Primero encontramos la MEDIA

$$Me = \quad \bar{X} = \frac{1+2+ .....7}{9} = \frac{36}{9} = \underline{4}$$

(Paso B) Construimos una estructura considerando que

$$N = 9 \quad \sum X = 36$$

X	$/X - \bar{X}/$
1	$/-3/$
2	$/-2/$
3	$/-1/$
4	0
4	0
4	0
5	$/+1/$
6	$/+2/$
7	$/+3/$
	$\sum /X - \bar{X}/ = 12$

Para obtener del primero al último valor de X

$/X - X/ = /1 - 4/ = /-3/$ al $/7 - 4/ = /+3/$

Entonces

$$D.\ M. = \frac{\sum /X - \bar{X}/}{N} = \frac{12}{9} = 1.33$$

INTERPRETACION:

Es decir, lo datos de la *variable X, se desvían 1.33 unidades en promedio con respecto a su media.*

Nota: Los números negativos en la tabla nos indican que están separados o desviados (según el número –3, -2, -11 unidades del valor medio del conjunto, de igual forma los valores positivos indican que están separados (según el número +1, +2, +3) unidades para arriba del valor medio.

Actividad de Aprendizaje 45

Hallar e interpretar el rango y la desviación media de los siguientes datos.

a) **4,8,12,16**
b) **65,60,55,70**

DESVIACION ESTANDAR Y VARIANZA
(serie de datos simples, sin frecuencia asociada)

La *desviación estándar* es la medida de dispersión mas adecuada por sus propiedades algebraicas; se le conoce también como *desviación típica*. Su símbolo es S y se le define así:

$$S = \sqrt{\frac{\sum(X - \bar{X})^2}{N}}$$

Como la desviación promedio de los datos de una distribución respecto a su media.

Una vez hallada en un caso concreto, debe ser expresada en las mismas unidades de la variable estudiada (años, pesos, etc.)

PROCEDIMIENTOS PARA

ENCONTRAR "S"

a) Encontramos primero la media

b) Calculamos el conjunto de desviaciones $(X - \bar{X})$ se eleva al cuadrado para *obtener las desviaciones cuadráticas* $(X - X)^2$ y eliminar los valores negativos.

c) La suma de los valores $(X - X)$ es siempre 0

$$\sum (X - \bar{X}) = 0$$

d) Se efectúa la suma de las desviaciones cuadráticas respecto a la media $\sum(X - \bar{X})^2$. Este *valor se conoce brevemente como la suma de los cuadrados.*

e) Se divide la suma de los cuadrados entre el número de datos de la distribución, el cociente <u>representa la media de las desviaciones cuadráticas</u> *se le conoce como VARIANZA y su símbolo es* S^2

$$\boxed{S^2 = \frac{\sum (X - \bar{X})^2}{N}}$$

f) Finalmente, para hallar la desviación estándar se extrae raíz cuadrada a la varianza.

$$\boxed{S = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}}$$

Ejemplo:

Hallar e interpretar la varianza y la desviación estándar de la variable X, cuyos datos son:

6, 7, 7, 7, 8, 9, 10, 11, 12, 15

Solución:

a) $Me = \bar{X} = \frac{6+7+7 \ldots 15}{10} = \frac{92}{10} = 9.2$

b) $(X - \bar{X})$ iniciamos con 6 y terminamos con 15

$(6 - 9.2) = -3.2$ al $(15 - 9.2) = 5.8$

$(X - \bar{X})^2$ $(-3.2)^2 = 10.24$ al $(5.8)^2 = 33.64$

Construimos la tabla correspondiente

X	$(X-\overline{X})$	$(X-\overline{X})^2$
6	- 3.2	10.24
7	- 2.2	4.84
7	- 2.2	4.84
7	- 2.2	4.84
8	- 1.2	1.44
9	- 0.2	0.04
10	0.8	0.64
11	1.8	3.24
12	2.8	7.84
15	5.8	33.64
$\sum X = 92$	$\sum(X-\overline{X}) = 0$	$\sum(X-\overline{X})^2 = 71.6$

$N = 10$ $\quad \overline{X} = 9.2$

Aplicando la fórmula de la <u>VARIANZA</u>

$$S^2 = \frac{\sum (x-\overline{x})^2}{N}$$

$$S^2 = \underline{71.6}$$

LA DESVIACION TIPICA

$S^2 = 7.16$

$S = 7.16$

$S = \sqrt{2.67}$

$S = 2.7$

INTERPRETACIÓN: Los datos de la variable X se separan 2.7 unidades en promedio respecto a la Media.

Podremos presentar ahora una forma abreviada para S^2 y S de donde tenemos sus fórmulas abreviadas.

$$S^2 = \frac{\sum X^2}{N} - \bar{X}^2$$

$$S = \sqrt{\frac{\sum X^2}{N} - \bar{x}^2}$$

Del ejemplo anterior. Hallar la varianza y la desviación estándar de la distribución de datos X, por el método abreviado.

X = 6, 7, 7, 7, 8, 9, 10, 11, 12, 15

$$\bar{X} = \frac{6 + 7 + 7 \ldots 15}{10} = \frac{92}{10} = 9.2 \qquad \underline{\bar{X} = 9.2}$$

Solamente tomamos para la tabla los siguientes valores.

X	X^2
6	36
7	49
7	49
7	49
8	64
9	81
8	100
9	121
12	144
15	225
$\sum X = 92$	$\sum X^2 = 918$

Donde N = 10

De las fórmulas

$$S^2 = \frac{\Sigma X^2}{N} - \bar{X}^2$$

$$S^2 = \frac{918}{10} - 84.64$$

$$S^2 = 91.8 - 84.64$$

$$S^2 = 7.16 \text{ VARIANZA}$$

$$S = \sqrt{7.16}$$

Actividad de Aprendizaje 46

Hallar e interpretar la varianza y la desviación estándar de los datos siguientes.

a) 7,8,10,6,7,8,11,12
b) 65,69,66,68,66,68,67,67,67

Varianza Y Desviación Estándar En Distribuciones De Datos No Agrupados Y Agrupados.

1. Calcular e interpretar la desviación estándar de la distribución siguiente

Alumnos según número de personas
Con quienes comparten su habitación

Personas	Total 187
0	40
1	76
2	44
3	12
4	9
5	3
6	2
7	1

Solución: la variable es "número de personas" $\sum f = 187$

Te recuerdo que son DATOS NO AGRUPADOS por lo que.

X	X^2	f	fx	fx^2
0	0	40	0	0
1	1	76	76	76
2	4	44	88	176
3	9	12	36	108
4	16	9	36	144
5	25	3	15	75
6	36	2	12	72
7	49	1	7	49
		$\sum f = 187$	$\sum FX = 270$	$\sum fx^2 = 700$

MEDIA

$$\bar{X} = \frac{\sum fX}{\sum F}$$

$$\bar{X} = \frac{270}{187} = 1.44 \text{ PERSONAS} \qquad \bar{X} = 1.44$$

Aplicando las fórmulas tenemos

** No confundir con el tema anterior $S^2 = \dfrac{\sum FX^2}{N} - \bar{X}^2$

$$S^2 = \frac{700}{187} - (1.44)^2$$

$$S^2 = 3.743 - 2.0736$$

$$S^2 = 1.6694 \text{ VARIANZA}$$

$$S = +\sqrt{1.6694}$$

$$S = 1.29$$

$S = 1.3$ personas DESVIACIÓN TIPICA

INTERPRETACIÓN: La desviación promedio de los datos de la variable X (número de personas) es de 1.3 unidades con el valor medio

Actividad de Aprendizaje 47

Hallar e interpretar la desviación estándar de la información dada de la tabla siguiente.

Niños de Primaria rural
Por edad

Años	Total 69
8	1
9	2
10	8
11	19
12	17
13	11
14	8
15	3

1. **Calcular e interpretar la desviación típica de los datos de la tabla siguiente.**

Alumnado según tiempo dedicado
Al estudio fuera de clases
1984

Horas Semanarias	Total 188
1 – 3	50
4 – 6	38
7 – 9	26
10 – 12	36
13 – 15	19
16 – 18	7
19 – 21	7
22 – 28	5

Construimos la nueva tabla para encontrar los valores que necesitamos.

Solución:

Datos X	f	Puntos Medios (X)	fx	fx^2
1 – 3	50	2	100	200
4 – 6	38	5	190	950
7 – 9	26	8	208	1664
10 – 12	36	11	396	4356
13 – 15	19	14	266	3724
16 – 18	7	17	119	2023
19 – 21	7	20	140	2800
22 – 28	5	25	125	3125
	$\sum f = 188$	$N = 188$	$\sum fx = 1544$	$\sum fx^2 = 18{,}842$

MEDIA:

$$\bar{X} = \frac{\sum FX}{\sum f} = \frac{1544}{188} = 8.2$$

$\bar{X} = 8.2$ horas semanarias

$$S^2 = \frac{\sum fx^2}{N} - \bar{X}^2$$

$$S^2 = \frac{18,842}{188} - (8.2)^2$$

$$S^2 = 100.22 - 67.24$$

$S^2 = 32.98$ VARIANZA

$$S = \sqrt{32.98}$$

$S = 5.7$ horas semanarias

INTERPRETACIÓN: Los datos de la variable "tiempo dedicado al estudio, se alejan 5.7 horas en promedio de su media.

Actividad de Aprendizaje 48

Del ejercicio anterior (niños de primaria rural) encontrar la desviación típica.

Coeficiente de Variabilidad

Tomemos las dos distribuciones propuestas a continuación y calculemos sus desviaciones típicas respectivas.

X = 10, 15, 20 Y = 50, 55, 60

Para X

X	X^2	
10	100	
15	225	
20	400	
$\sum X=45$	$\sum X^2=725$	N=3

Para Y

Y	Y^2	
50	2500	
55	3025	
60	3600	
$\sum Y=165$	$\sum Y^2=9125$	N=3

MEDIA

$$\overline{X} = \frac{10+15+20}{3} = \frac{45}{3}$$

$$\overline{x} = 15$$

de la Fórmula

$$S = \sqrt{\frac{\sum X^2}{N} - \overline{X}^2}$$

$$S= \sqrt{\frac{725}{3} - (15)^2}$$

$$S= \sqrt{241.7 - 225}$$

$$S= \sqrt{16.7}$$

$$S= 4.08$$

MEDIA

$$\overline{Y}= \frac{50+55+60}{3} = \frac{165}{3} =$$

$$\overline{Y}= 55$$

de la Fórmula

$$S= \sqrt{\frac{\sum Y^2}{N} - \overline{X}^2}$$

$$S= \sqrt{\frac{9125}{3} - (55)^2}$$

$$S= \sqrt{3041.7 - 3025}$$

$$S= \sqrt{16.7}$$

$$S= 4.08$$

La desviación típica es insuficiente para dar cuenta objetivamente de la dispersión de dos o más conjuntos de datos que se comparan entre sí. En consecuencia, es necesario disponer de un indicador que tome en cuenta la tendencia central de la distribución. Este indicador ha sido definido como la *razón de la desviación estándar a la media de una distribución dada.*

Se le conoce como *coeficiente de variabilidad coeficiente de variación o desviación estándar relativa.*

Cuyo símbolo es CV y fórmula:

$$CV = \frac{S}{\bar{X}}$$

Del ejemplo anterior tenemos

Para X $\quad CVx = \frac{S}{\bar{X}} = \frac{4.08}{15} = 0.272 = 27.2\%$

Para Y $\quad CVy = \frac{S}{\bar{X}} = \frac{4.08}{15} = 0.74 \quad = 7.4\%$

Del resultado que nos dé al aplicar esta fórmula el MENOR de ellos es la DISTRIBUCIÓN MAS HOMOGENEA. (es decir de 10 al 20 su diferencia es 100% para x, y de 50 a 60 es 20% su diferencia) por lo tanto.

SOLUCIÓN: La distribución Y es más homogénea que la X

- Por “más homogénea” se entiende que los datos bajos, intermedios y altos de la distribución son menos dispares o disímbolos entre sí.

Actividad de Aprendizaje 49

Dadas las distribuciones de datos de las variable W y Z, es decir con fundamento cuál de las dos es más homogénea.

W = 8, 9, 11, 15, 20 $\qquad\qquad$ Z = 4, 5, 7, 11, 16

La Media y la Desviación Estándar de las Distribuciones de Muestreo

El estudio de las distribuciones teóricas de muestreo permite demostrar que si se mantiene la condición del muestreo aleatorio simple, es decir, si todos los elementos de una población tienen la misma oportunidad de pertenecer a la muestra, la distribución de muestreo de la media (dmm) *tiene una media igual a la media de la población* μ *y una desviación estándar igual a:*

$$\boxed{\sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}}$$

Es decir:

$$\mu_{\bar{x}} = \mu$$

$$\sigma = \sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}$$

La expresión $\frac{N - 2}{N - 1}$ *se conoce como factor de corrección de población finita;*

se puede prescindir de ella cuando el muestreo se hace con reemplazo, ya que, bajo esta condición la población que se muestra se considera teóricamente infinitamente, o cuando la muestra recoge solo una pequeña parte del universo, pues entonces dicha expresión se acerca notoriamente a la unidad.

Esto sucede más o menos desde $n/N \leq .05$, es decir, cuando la fracción de muestreo es menor o igual a 5% ($f \leq .05$)

Ejemplos de la progresión No. 11

1. Los datos siguientes representan la antigüedad en el empleo de un universo de trabajadores:

.75, 1, 1.5, 2, 2

a) Bajo la condición de muestreo sin reemplazo, determinar el total de muestras posibles de tamaño 2 y anotar las muestras

Solución:

Puesto que el muestreo es sin reemplazos, el número de muestras de tamaño 2 queda determinado por.

$$\left(\frac{N}{N}\right) = \frac{N!}{n!\ (N-n)!} = \left(\frac{5}{2}\right) = \frac{5!}{2!\ (5-2)!} = \frac{5.43!}{2!3!} = \boxed{\text{10 muestras posibles}}$$

La distribución de muestreo es la siguiente:

.75, 1	**.75, 1.5**	**.75, 2**	**.75,2**	**1, 1.5**
1, 2	**1, 2**	**1.5, 2**	**1.5, 2**	**2,2**

Cada pareja de datos forma una muestra posible de tamaño 2 y que cualquiera de ellas se habría obtenido seleccionándola aleatoriamente.

b) Si a cada una de las muestras anteriores le calculamos *su media* y anotamos todas las medias resultantes, obtenemos *la distribución de muestreo de la media.*

$$\bar{X} = \frac{.75 + 1}{2} = \frac{1.75}{2} = .875$$

$$\bar{X} = \frac{.75 + 1.5}{2} = \frac{2.25}{2} = 1.125$$

$$\bar{X} = \frac{.75 + 2}{2} = \frac{2.75}{2} = 1.375$$

$$\bar{X} = \frac{.75 + 2}{2} = \frac{2.75}{2} = 1.375$$

$$\bar{X} = \frac{1 + 1.5}{2} = \frac{2.5}{2} = 1.25$$

$$\bar{X} = \frac{1 + 2}{2} = \frac{3}{2} = 1.5$$

$$\bar{X} = \frac{1 + 2}{2} = \frac{3}{2} = 1.5$$

$$\bar{X} = \frac{1.5 + 2}{2} = \frac{3.5}{2} = 1.75$$

$$\bar{X} = \frac{2 + 2}{2} = \frac{4}{2} = 2$$

$$\bar{X} = \frac{1.5 + 2}{2} = \frac{3.5}{2} = 1.75$$

c) La media de la dmm ($\mu_{\bar{x}}$) es:

$$\mu x = \frac{.875 + 1.125 + \ldots .2}{10} = \frac{14.5}{10} = \boxed{1.45 \text{ años}}$$

Para hallar la desviación estándar ($\mu_{\bar{x}}$) , basta con aplicar el método de cálculo abreviado.

$\bar{X}$	$\bar{X}^2$
.875	.766
1.125	1.266
1.375	1.891
1.375	1.891
1.25	1.562
1.5	2.25
1.5	2.25
1.75	3.062
1.75	3.062
2	4

$\sum \bar{X} = 14.5$ $\sum \bar{X}^2 = 22$

N = 10

De la fórmula:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum \bar{X}^2}{N} - \bar{X}^2}$$

Sustituyendo valores

$$\sigma_{\bar{x}} = \sqrt{\frac{22}{10} - (1.45)^2} = \sqrt{2.2 - 2.1025} = \sqrt{.0975}$$

$$\sigma_{\bar{x}} = \boxed{.312 \text{ años}}$$

En resumen, la dmm tiene una MEDIA IGUAL a 1.45 años y una DISPERSION PROMEDIO de sus datos con respeto a su media de .31 años d)Calcular la media y la desviación estándar en la población *En cuanto a la población, la media (μ) es:*

$$\mu = \frac{.75 + 1 + \dots\dots\dots 2}{5} = \frac{7.25}{5} = \boxed{1.45 \text{ años}}$$

y su desviación estándar, de:

X	X^2
.75	.562
1	1
1.5	2.25
2	4
2	4
$\sum x = 7.25$	$\sum x^2 = 11.81$ n=5

$$\sigma = \sqrt{\frac{\sum X^2}{N} - \mu^2}$$

$$\sigma = \sqrt{\frac{11.81}{5} - (1.45)^2} = \sqrt{2.362 - 2.1025}$$

$$\sigma = \sqrt{0.26}$$

σ = .51 AÑOS

e) Utilizar la desviación estándar en el universo para calcular la dmm.

Se ha calculado los indicadores de tendencia central y de dispersión tanto para la dmm como para el universo, podemos utilizar la dispersión en el universo para calcular la correspondiente en la dmm, por lo que tenemos:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma 2}{n}\left(\frac{N-n}{N-1}\right)} = \sqrt{\frac{.26}{2}\left(\frac{5-2}{5-1}\right)} = \sqrt{.13\,(3/4)} = \sqrt{.975}$$

$$\sigma_{\bar{x}} = \boxed{.312 \text{ años}}$$

CONCLUSION:

Del ejercicio anterior permite hacer hallazgos que son fundamentales para la estadística inferencial por las aplicaciones a que da lugar.

PRIMERO:

Se confirma que la MEDIA de una distribución de muestreo de medias es siempre igual a la MEDIA DE LA POBLACIÓN.

$$\boxed{(\mu_{\bar{x}} = \mu)}$$

SEGUNDO:

Que la división estándar en la dmm, que se conoce como *ERROR DE MUESTREO,* es determinable a partir de la desviación estándar en la población. En otras palabras, si se conoce esta dispersión, se puede conocer también el error de muestreo sin necesidad de tratar de construir una distribución de muestreo de medias, lo cual, dicho sea de paso, sería imposible en universos grandes.

TERCERO:

Que la mayoría de las muestras de todas las muestras extraíbles de un universo determinado arrojan promedios muy próximos al parámetro correspondiente.

Todo lo expuesto anteriormente conduce a un teorema fundamental de la estadística inferencial conocido como *TEOREMA DEL LIMITE CENTRAL*, que puede ser enunciado como sigue:

Si se mantiene la condición básica del muestreo aleatorio, y si la muestra de tamaño "η" es suficientemente grande, entonces la distribución de muestreo de la media es aproximadamente normal con una media igual a la media de la población (μ) y una desviación estándar igual a

$$\boxed{\sqrt{\frac{\sigma^2}{N}} \qquad \sqrt{(N - n)(N - 1)}}$$

Para casi cualquier distribución poblacional que surge de la práctica de investigación por muestreo, una muestra de tamaño n = 30 se considera suficientemente grande para aplicar los resultados del teorema del límite central con bastante seguridad.

2. Los datos siguientes representan el número de hijos por familia en un universo formado por cuatro familias:

No. De hijos por familia

3, 3, 4, 6,

a) Bajo la condición de muestreo CON REMPLAZO, determinar el total de muestras posibles de tamaño 3.

Como es CON REMPLAZO tenemos de la fórmula Nn

Por lo tanto

$$\mathbf{N^n = 4^3 = 64 \text{ muestras}}$$

b) Calcular la media y la desviación estándar de la población *para la media,*

$$\mu = \frac{3+3+4+6}{4} = \frac{16}{4} = 4 \text{ hijos por familia}$$

Para la desviación estándar

X	X^2
3	9
3	9
4	16
6	36
$\sum X = 16$	$\sum X^2 = 70$
$N = 4$	$\mu = 4$

$$\sigma = \sqrt{\frac{\sum X^2}{N} - \mu^2}$$

$$\sigma = \sqrt{\frac{70}{4} - 4^2} = \sqrt{17.5 - 16}$$

$$\sigma = \sqrt{1.5}$$

$$\sigma = \boxed{1.2 \text{ hijos por familia}}$$

N= 4 **μ = 4** **σ = 1.2 hijos por familia**

Es decir, los datos de la variable número de hijos por familia se dispersan 1.2 unidades en promedio con respecto a la media, que es 4 hijos por familia.

b) Calcular la media y la desviación estándar de la distribución de muestreo de la media.

Por el teorema del límite central, sabemos que la media en el universo es igual a la media de la dmm, por lo tanto,

$$\sigma_{\overline{x}} = \mu = 4 \text{ hijos por familia}$$

En cuanto a la desviación estándar de la dmm, el teorema dice que es igual a

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} = \left(\frac{N-n}{N-1}\right)}$$

Pero como ya vimos, se puede prescindir del factor de corrección en esta expresión cuando <u>el muestreo es con reemplazo,</u> entonces

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{1.5}{3}}$$

$$\sigma_{\bar{x}} = \boxed{0.71 \text{ hijos por familia}}$$

En la distribución muestral de medias, estos se desvían 0.71 unidades en promedio con respecto a su media, que es 4 hijos por familia.

En otras palabras, EL ERROR DE MUESTREO de la dmm es 0.71

a) Bajo la condición de muestro SIN REMPLAZO, resolver los incisos a y c

COMO ES SIN REMPLAZO, el total de muestras posibles de tamaño 3 extraídas del universo de tamaño 4 es:

$$\binom{N}{n} = \frac{n!}{n!\,(N-n)}$$

$$\binom{4}{3} \frac{4!}{3!\;(4-3)!} = \frac{4.3.2!}{3!\;1!} = \frac{24}{6} = \boxed{4 \text{ muestras}}$$

En cuanto a la media de la dmm, sabemos ya que es <u>la misma que la del universo</u>

μ_x = 4 hijos por familia

***** EL ERROR DE MUESTREO, en cambio se calcula la expresión conocida, pero sin prescindir del factor de corrección, ya que el muestreo es son reemplazo, así,*

$$\sigma_{\overline{x}} = \sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}$$

$$\sigma_{\overline{x}} = \sqrt{\frac{(1.2)^2}{3}\;\frac{4-3}{4-1}} = \sqrt{\frac{1.5}{3}\;(.33)}$$

$$\sigma_{\overline{x}} = \sqrt{0.165}$$

$$\sigma_{\overline{x}} = \boxed{0.41 \text{ hijos por familia}}$$

Métodos de Inferencia Estadística

Independientemente del parámetro que se quiera conocer, existen dos tipos de estimación:

Estimación de Punto:

Es la que adopta la forma de un valor único

Estimación de Intervalo:

Es aquella que adopta la forma de un intervalo de valores.

En ambos casos, tratan de métodos de inferencia estadística, de los cuales veremos lo mas fundamental para medias y proporciones únicamente.

Ejemplos:

Para la estimación de punto de un parámetro.

Cuando lo que se generaliza al universo es una descripción numérica de un solo valor calculado en la muestra, ejemplo:

Si una porción de trabajadores extraída de un universo A, se calcula que el ingreso promedio diario es de 32 pesos, y si se dice: "se estima que en el universo A el ingreso promedio diario es de 32 pesos"

NOTA: La estimación de punto, es decir de un solo, valor, del parámetro MEDIA DE INGRESO en el universo en este ejemplo, o bien también se podría establecer de esta otra forma

"Se estima que, en A el porcentaje de trabajadores con compromisos conyugales es de 68%"
en este ejemplo el parámetro correspondiente: proporción de trabajadores con....

INCONVENIENTES:

- No aportan información respecto a su precisión
- Rara vez son exactas
- Los valores posibles de la estimada se distribuyen alrededor del verdadero valor del parámetro de acuerdo con la distribución de muestreo correspondiente.

Una medida directa de la precisión de la media de una muestra aleatoria de tamaño "n" está dada por su error de muestreo

$$\boxed{\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}}$$

Sin embargo, en la práctica, lo común es que se desconozca el valor de la desviación estándar en la población (σ); esto impide calcular el error de muestreo de la media

Sin embargo, se suele estimar usando la dispersión de la muestra mediante la fórmula.

$$\boxed{\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}}$$

donde s y n representan la desviación estándar y el tamaño de la muestra, respectivamente.

EL ERROR DE MUESTREO ESTIMADO, proporciona información respecto a la precisión de la media de una muestra aleatoria; CUANTO MAS PEQUEÑO es dicho error, más precisa es la media de la muestra como estimada de la media del universo.

Para mostrar cuan precisa es una estimación de punto, se acostumbra reportarla bajo la forma

Estimación de punto ± desviación estándar

Ejemplo:

La media de ingresos mensuales por propinas reportadas por 69 niños empaquetadores al servicio de los supermercados de una ciudad en 1983, fue en unidades monetarias actuales, de 185 pesos. Si generalizamos este valor al universo (formado por 220 elementos) diciendo que los niños empaquetadores de los supermercados de esa ciudad recibían un ingreso mensual promedio, por propinas de 185 pesos estamos haciendo una inferencia estadística, una estimación de punto, del parámetro media de ingresos. Pero ¿cuán precisa es la estimación? Dos factores determinan su precisión:

El primero es que la estimada está basada en una muestra de 69 datos, pues es mucho más precisa que una basada, digamos, en 10 datos, y no tan precisa que otra basada en 100.

El segundo es que la desviación estándar de la muestra haya sido de 56 pesos, ya que, si este valor fuese menor, la precisión sería mayor y viceversa.

Estimamos ahora con esta desviación el error de muestreo.

$$\sigma_{\overline{x}} = \sqrt{\frac{S^2}{n}\left(\frac{N-n}{N-1}\right)} = \sqrt{\frac{(56)^2}{69}\left(\frac{220-69}{220-1}\right)} = \sqrt{\frac{3136}{69}\left(\frac{151}{219}\right)}$$

$$\sigma_{\overline{x}} = \sqrt{31.34}$$

$$\sigma_{\overline{x}} = 5.6 \text{ pesos}$$

Tenemos una estimada de 185 pesos con un error de muestreo estimado de 6 pesos.

Este último dato ayuda a confiar bastante, en este caso, en la precisión de la estimada, ya que la variabilidad seria pequeña en términos relativos.

$$185 \pm 6$$

En el caso de proporciones que implican atributos dicotómicos (dos categorías mutuamente excluyentes y exhaustivas), la desviación estándar de la muestra se calcula así:

$$S^2 = p\,(1 - p\,) = pq \qquad p = \%$$

$$S = pq$$

Y la desviación estándar de la distribución muestral de proporciones, se estima así, cuando no se conoce la desviación estándar de la población:

$$\sigma_p = \sqrt{\frac{pq}{n}\,\frac{N-n}{N-1}}$$

En una muestra aleatoria de 200 elementos extraída de una población de tamaño 2,800 resuelta que el 72% tienen estudios por debajo del grado de licenciatura y el 28% de licenciatura en adelante.

Estimar la desviación estándar de la distribución muestral de proporciones, o sea, el error de muestreo y hacer un reporte más completo de la estimación de punto. Para la fracción de muestreo $\frac{n}{N} = \frac{200}{2800} = .07$

Usaremos en el cálculo el factor de corrección:

$$\sigma_P = \sqrt{\frac{pq}{n}\;\frac{N-n}{N-1}}$$

$$\sigma_p = \sqrt{\frac{.72\,(.28)}{200}\left(\frac{2800-200}{2800-1}\right)} = \sqrt{\frac{.72\,(.28)}{200}\left(\frac{2000}{2799}\right)} = \sqrt{.001008\,(.923)} =$$

$$\sigma_p = \sqrt{.00094} = .03$$

$$\sigma_p = 3\%$$

Así diríamos que en la población investigada el porcentaje de los que tienen estudios de Licenciatura es de 72% con un error de muestreo estimado de 3%

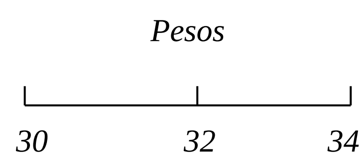
$$72\% \pm 3\%$$

ESTIMACIONES DE INTERVALOS

Partimos para ellos nuevamente de los datos que sirvieron para el ejemplo de las estimaciones de punto.

Si se dice: *"Se estima con un 95% de probabilidad que en el universo A el ingreso promedio diario es un valor comprometido entre $30 y $34, estamos efectuando una estimación de intervalo del parámetro media de ingresos en A.*

Si observamos que la estimación de intervalo es siempre un conjunto de valores entre dos límites.



Cuyo punto medio es el estadígrafo que sirve de base para una estimación de punto.

Se diferencia de esta última en que añade el grado de certeza (probabilidad) de que el verdadero valor del parámetro que se estima caiga dentro de los dos limites dados, en síntesis.

Una estimación de intervalo es un conjunto de valores determinado con los datos de una muestra y acompañado de un enunciado que señala la probabilidad de que el intervalo incluya el verdadero valor del parámetro que se estima.

Al intervalo (30 – 34) se le conoce como INTERVALO DE CONFIANZA; a sus extremos (30 y 34) LIMITES DE CONFIANZA, y al valor de probabilidad, NIVEL DE CONFIANZA (95% en el ejemplo). En las investigaciones por muestreo los niveles de confianza más usados, son 95% y 99%

Actividad de Aprendizaje 50

Para el universo dado a continuación y bajo la condición de muestreo sin reemplazo:

- Escribir la distribución de muestreo de la media (dmm)
- Determinar el total de muestras posibles de tamaño 2 y anotar las muestras.
- Calcular la media y la desviación estándar (error de muestreo) en la distribución muestral de medias utilizando el teorema del límite central

a) Edad de niños de primaria, en años 7, 7, 8, 10, 12

b) Paso de personas, en kilogramos: 60, 62, 68, 70, 75,

PENSAMIENTO MATEMÁTICO
PROGRESIÓN No. 14

Explica un evento aleatorio cuyo comportamiento puede describirse a traves del estudio de la distribución normal y cacula la probabilidad de dicho evento suceda.

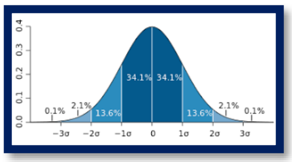
METAS	CATEGORÍAS	SUBCATEGORÍAS
M4 Argumenta a favor o encontra de afirmaciones acerca de situaciones, fenómenos o problemas propios de la matemática, de las ciencias o de su contexto.	C2 Procesos de intuición y razonamiento	S1 Capacidad par observar y conjeturar S2 Pensamiento intuitivo S3 Pensamioento Formal
M3 Aplica procedimientos, técnicas y lenguaje matemático para la solución de problemas propios del Pensamiento Matemático, de Áreas de Conocimiento, Recursos Socioemocionales y de su entonrno	C3 Solución de problemas y modelación	S1 Capacidad par observar y conjeturar S3 Pensamioento Formal

Mate mayeko

Aprende Mas

Distribución de Probabilidad Normal

La normal estándar, es de las distribuciones de probabilidad, la más importante, ayuda a administrar muchos fenómenos, pues varias poblaciones tienen distribución normal o pueden ajustarse muy bien a ella. Esta clase de distribución es comúnmente utilizada en el campo de la industria cuando se tiene una muestra grande.



Se identifica a través de una curva simétrica. Esta gráfica también recibe el nombre distribución o campana de Gauss, pues al representar su función de probabilidad, tiene forma de campana.

El profesor Julio Vargas (2010) afirma que esta distribución nos da la probabilidad de que, al elegir un valor, éste tenga una medida contenida en unos intervalos definidos. Esto permitirá predecir de forma aproximada, el comportamiento futuro de un proceso, mediante los datos del presente.

La distribución normal es utilizada comúnmente cuando es relevante considerar:

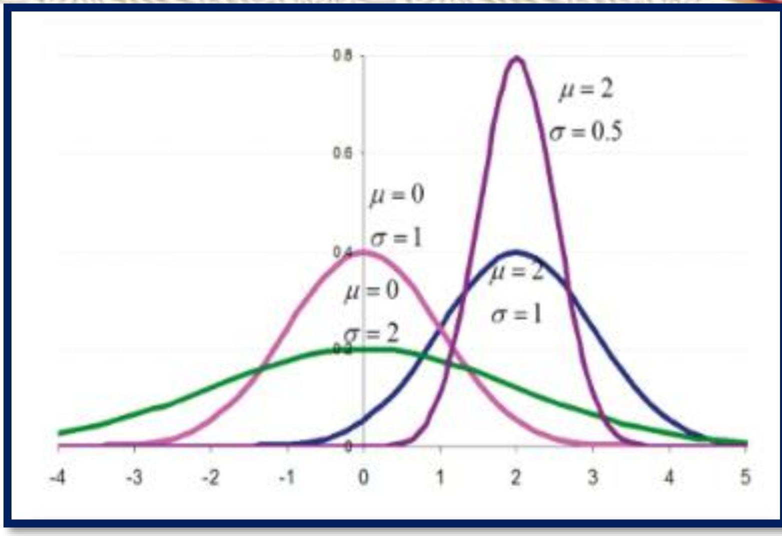
•Características de forma, segura o configuración de personas, animales, planta, tales como las tallas, pesos, perímetros, etc.

•Particularidades Fisiológicas, como pueden ser, los efectos de una misma dosis de un fármaco o de una misma cantidad de abono para las plantas.

•Corregir errores que pueden presentarse al realizar medidas en determinadas magnitudes.

•Elementos escolares como promedios, calificaciones, desempeño, etc.

Todos los tipos de distribución normal se representan en curvas simétricas, cada una de ellas con su respectiva media y desviación estándar.



Para poder estudiar este tipo de distribuciones fue creada la distribución normal estandarizada, llamada así por que utiliza las puntuaciones estándar Z, que en el gráfico aparece en color magenta.

La representación gráfica se caracteriza por ser:

•Una distribución simétrica.

•Es asintótica, es decir, sus extremos nunca tocan el eje horizontal, cuyos valores tienden a infinito.

•En el centro de la curva se encuentran la media, la mediana y la moda.

•El área total bajo la curva representa el 100% de los casos.

•Los elementos centrales del modelo son la media y la varianza.

•Debido a que una distribución normal es simétrica, debido a que el eje que pasa por x = μ, deja un área igual a 0.5 de lado izquierda y otro igual a 0.5 ala derecha.

Una variable aleatoria continua, x debe seguir una distribución normal de media μ con valor cero y desviación estañar σ con valor de 1 es decir N (0,1) o N(μ, σ).Dicho en otras palabras, necesitamos conocer la media, y las desviación estándar o la varianza **para tener definida la distribución normal**

Sabías que...

La distribución normal es también un caso particular de probabilidad de variable aleatoria continua, fue reconocida por primera vez por el francés Abraham de Moivre (1667-1754). Posteriormente, Carl Friedrich Gauss (1777-1855) elaboró desarrollos más profundos y formuló la ecuación de la curva; de ahí que también se le conozca, más comúnmente, como la “campana de Gauss”.

Para poder utilizar la gráfica tenemos que transformar la variable x que sigue una distribución N (μ, σ) en otra variable Z que siga una distribución N (0, 1).Es decir, utilizaremos la fórmula

$$Z = \frac{X - \mu}{\sigma}$$

<table>
<tr><th colspan="3">Distribución binomial</th></tr>
<tr><th>Ejemplo 1</th><th>Valores</th><th>Fórmula</th></tr>
<tr><td>1. Una población normal tiene una media de 80 y una desviación estándar de 14.0

Calcula la probabilidad de un valor localizado entre 75.0 y 90.0 , es decir, p(75 ≤ x ≤ 90)</td><td>μ = 80

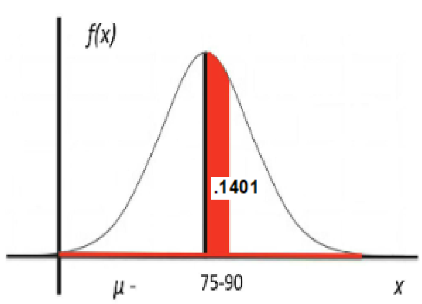
σ = 14.0</td><td>$Z = \frac{X - \mu}{\sigma}$</td></tr>
</table>

Distribución binomial

$Z = \frac{90-80}{14} = \frac{10}{14} = 0.71 =$ **0.2580**

$Z = \frac{75-80}{14} = \frac{-5}{14} = -0.36 =$ **0.1179**

P(75=X=90) = 0.2580 – 0.1179=.1401=14.01 %



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389

Vemos los valores **en la tabla** 0.71 = **0.2580**

Y – 0.36 = **0.1179**

P(75≤X≤90) = 0.2580 – 0.1179=.1401=14.01 %

Ejemplo No. 2

La selección de aspirantes para ocupar cargos directivos en un importante corporativo, es a través de un examen de conocimientos que en promedio los aspirantes obtienen una puntuación de 75 con una desviación estándar de 5.

Si la empresa selecciona un aspirante al azar, cual sería la probabilidad de que resuelva el examen y obtenga:

a) Una puntuación entre 75 y 80
b) Una puntuación de 80 o más
c) Una puntuación entre 70 y 80
d) Una puntuación entre 80 o menos

Solución

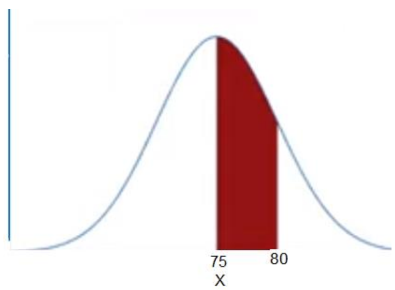
a) Una puntuación entre 75 y 80

Datos

μ = 75

σ = 5

P(75≤X≤80)

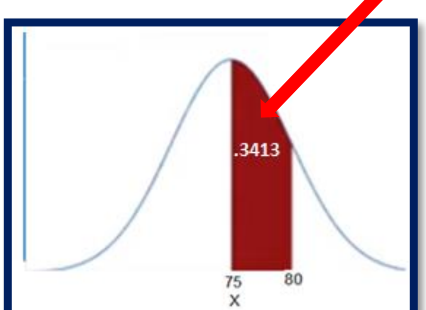


$$Z = \frac{X-\mu}{\sigma} = \frac{80-75}{5} = 1$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015

Vemos la tabla de distribución normal.

Por lo tanto su probabilidad es .3413 **su Área es de .3413**



P(75≤X≤80) = 0.3416= 34.13%

b) Una puntuación de 80 o más

Sabemos que entre la media y 80 la probabilidad es de .3413 también sabemos que la mitad es de .5 o el 50% por lo que a esta mitad le restamos el área encontrada.

Datos

μ = 75

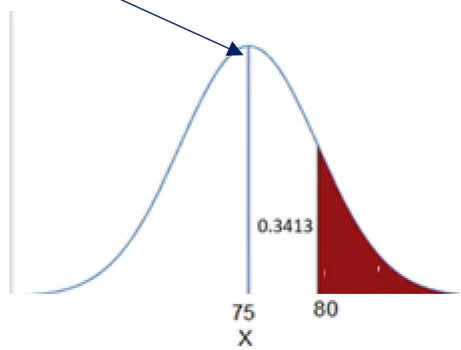
σ = 5

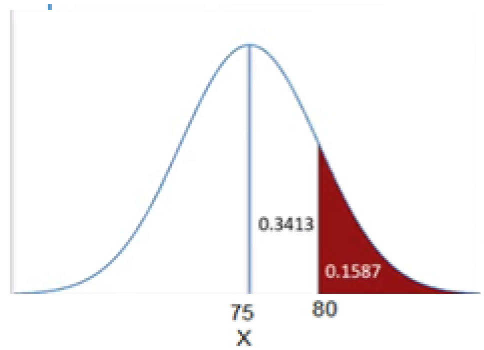
$$Z = \frac{X - \mu}{\sigma} =$$

0.5 - .3413 = **0.1587**

Por lo que tendríamos

P(X≥80) = 0.1587 = **15.87%**





c) Una puntuación entre 70 y 80

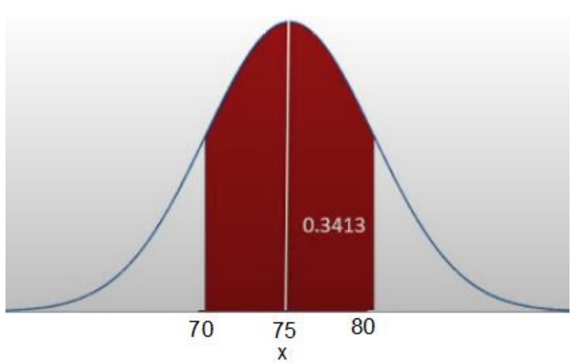
Datos

μ = 75

σ = 5

$$Z = \frac{X - \mu}{\sigma} =$$

P(70≤X≤80)

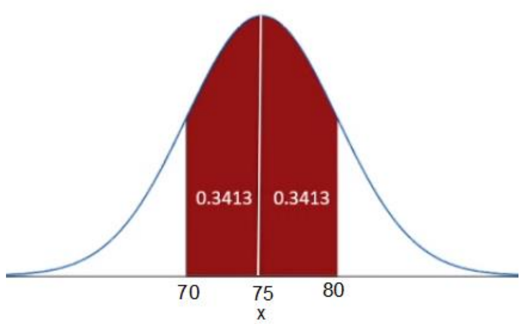


$$Z = \frac{X - \mu}{\sigma} = \frac{70 - 75}{5} = -1$$

se toma el valor absoluto de -1 = 1 consultamos la tabla

como se ve es la misma , que el inciso **a) .3413**

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015



Teniendo ambas posibilidades se procede a sumarlas .3403 + .3413 = **0.6826**

P(70≤X≤80) = **0.6826 = 68.26%**

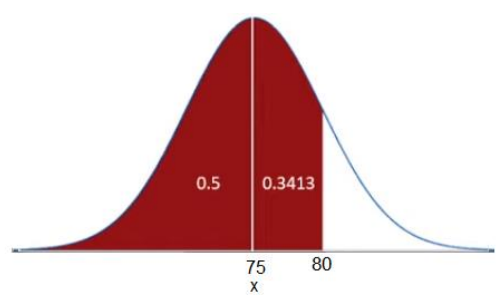
d) Una puntuación entre 80 o menos
μ = 75
σ = 5

$$Z = \frac{X - \mu}{\sigma} =$$

P(X≤80)

Sabemos que la mitad es igual 0.5 y entre 75 y 80 es 0.3413

P(X≤80)= 0.5 + 0.3413 = **0.8413= 84.13%**



Actividad de Aprendizaje 51

Resuelve los siguientes problemas y comparen sus resultados

1.- Una panadería tiene una producción diaria que se distribuye normalmente con una media de 158 panes y una desviación estándar de cuatro panes. Encentre la probabilidad de que el numero de panes producidos por día :

a) Sea menor que 163 ´panes

b) Sea mayor que 164 panes

c) Este entre 150 y 165 panes

d) Esté entre 160 y 168 panes

Actividad de Aprendizaje 52

La Media de los pasos de 500 habitantes de un conjunto residencial es de 70Kg., Y la desviación estándar es de 4Kg., suponiendo que los valores se distribuyen normalmente, hallar cuantos habitantes pesan:

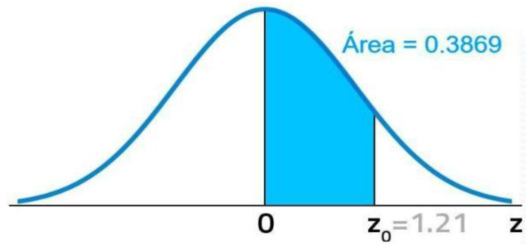
a) Menos de 75 Kg.
b) Menos de 62 Kg.
c) Mas de 79 Kg.

Tabla de Distribución Normal Estandarizada

$$z = \frac{x - \mu}{\sigma}$$

$$\mu = media$$

$$\sigma = desviación\ estándar$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
4.0	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA
2023 Francisco VILLA
DGETI Guanajuato CBTis 75

PENSAMIENTO MATEMÁTICO
PROGRESIÓN No. 15

Valora la posibilidad de hacer inferencias a partir de la revisión de algunas propiedades de distribuciones y del sentido de la estadisrtica inferencial con la finalidad de modelar y entender algunos fenomenos.

METAS	CATEGORÍAS	SUBCATEGORÍAS
M3 Comprueba los procedimientos usados en la realizacion de problemas utilizando diversos métodos ,empleandos recursos tecnológicos o la interaccion con sus pares.	C1 Procedimental	S4 Manejo de datos e incertidumbre
M4 Argumenta a favor o en contra de afirmaciones a serca de situaciones, fenomenos o problemas propios de la matemática, ciencias o de su contexto.	C2 Procesos de intuicion y razonamiento	S1 Capacidad par observar y conjeturar S2 Pensamiento intuitivo S3 Pensamioento Formal
M4 Connstruye y plantea posible solucines a ´roblemas de Áreas de Conocimiento, recursos Sociocognitivos , Recursos Sociemocionales y de su entorno , empleando técnicas y Lenguaje matemático.	C3 Solucion y Problemas de Modelación	S2 Construcción de Modelos S3 Estrategias Heurísticas y ejecución de Procedimientos no rutinarios

Mate mayeko

Para iniciar reflexiona

Es frecuente, especialmente en el campo de la salud, que un profesional en ejercicio de su actividad se detenga en un fenómeno que lo hace pensar que el grupo al que está observando tiene un comportamiento especial respecto a una determinada variable.

Así, por ejemplo, un kinesiólogo puede pensar que los pacientes sometidos a una secuencia especial de ejercicios demoran menos en recuperar la función muscular que aquellos tratados con el método tradicional. A un médico psiquiatra radicado en Punta Arenas le puede parecer que los suicidios adolescentes son más frecuentes en su región. Un profesional de la nutrición puede creer que los pacientes con problemas de absorción intestinal responden mejor a una alimentación con verduras que con carnes. El director de salud de una municipalidad puede pensar que su consultorio tiene mejor resolución de problemas complejos que el consultorio del municipio vecino. Detrás de todas estas situaciones se esconde una hipótesis que espera para ser verificada.

Definición de hipótesis

Una hipótesis se define como una afirmación transitoria que debe ser sometida a prueba. La inferencia estadística propone un procedimiento para llevar a cabo la prueba de las hipótesis. Propone, primero, enunciarlas formalmente y luego contrastarlas con la evidencia de los datos. Son los datos, entonces, con su coro de características, los que dirán si una hipótesis es falsa o verdadera.

Este procedimiento se realiza considerando a los parámetros, que ya sabemos corresponden al universo, como los objetos para los cuales se enuncian las hipótesis. Dicho de otro modo, una hipótesis se enuncia para una característica del universo o población y se origina en la observación del comportamiento de la misma característica en un grupo restringido o muestra.

Una hipótesis por ejemplo, al decir: “estos enfermos demoran en promedio 25 días en recuperarse” está afirmando que, en el universo, el promedio de los pacientes tardan 25 días en mejorar. Será tarea del investigador probar la veracidad o falsedad de dicha afirmación contrastando el valor propuesto para el parámetro del universo (25 días), con los datos reales provenientes de una muestra cualquiera. Si luego de esta comparación resulta que el promedio obtenido en la muestra es de 22 días, se le encarga a la estadística que resuelva el dilema de si la diferencia entre el promedio muestral (22 días) y el poblacional (25 días) permite aceptar como verdadera la hipótesis planteada. Será el método estadístico el que permita en definitiva resolver este dilema, evaluando la significación de la diferencia entre 22 y 25.

¿Azar o no?

El método de las pruebas de hipótesis consiste fundamentalmente en establecer la probabilidad de que sea consecuencia del azar la diferencia existente entre dos cantidades. Se pueden distinguir dos situaciones:

a) Diferencia entre un valor muestral y un valor poblacional, o valor teórico.

b) Diferencia entre dos o más valores muestrales.

En el caso a se tratará de evaluar la diferencia entre un valor obtenido en la muestra (estadístico) y un valor correspondiente en el universo (parámetro), y en el caso b se evaluará la diferencia entre dos valores provenientes de dos muestras (estadísticos). Los valores que se comparen, ya sean de la muestra o del universo, pueden ser promedios, porcentajes u otros. Nosotros nos ocuparemos sólo de promedios y porcentajes.

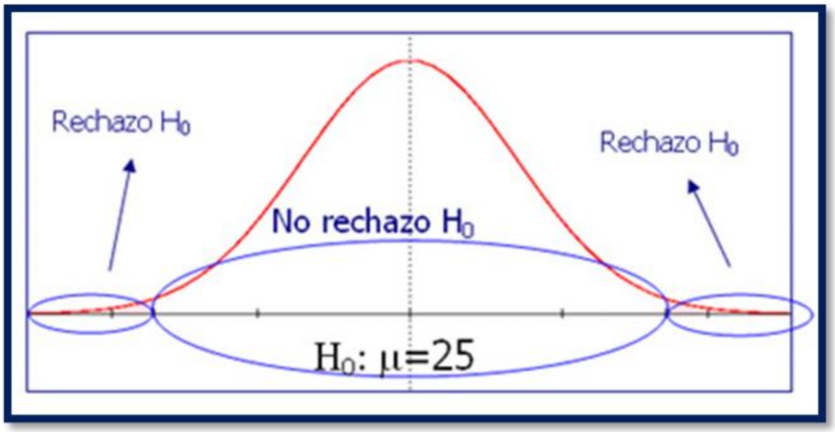
En general, lo que hace una prueba estadística es evaluar la diferencia entre dos o más valores (dos promedios, dos porcentajes). Respecto de esta diferencia se elabora una hipótesis previa y se plantea formalmente en términos estadísticos.

Luego, usando la distribución de probabilidad adecuada, se calcula la probabilidad de la diferencia entre los valores comparados. Si la probabilidad de obtener tal diferencia es pequeña, diremos que dicha diferencia es significativa.

Una diferencia es significativa cuando la probabilidad de que se produzca por azar es pequeña.

Estadístico de prueba

Para realizar tan delicada operación debemos utilizar el instrumento apropiado: le llamaremos estadístico de prueba, el que podremos calcular con los datos de nuestra muestra. Luego buscaremos la probabilidad de ocurrencia del valor calculado en la tabla correspondiente (Normal, t de student u otra) y a la luz de la probabilidad obtenida tomaremos una decisión respecto de nuestra hipótesis.



El esquema a seguir:

1. Plantear la hipótesis en términos estadísticos

Esta etapa consiste en representar el problema de investigación bajo la forma de dos hipótesis excluyentes: la Hipótesis Nula y la Hipótesis Alternativa.

- **Hipótesis Nula**. Esta hipótesis plantea que los valores comparados son iguales. Dependiendo del problema podrá presentarse como:

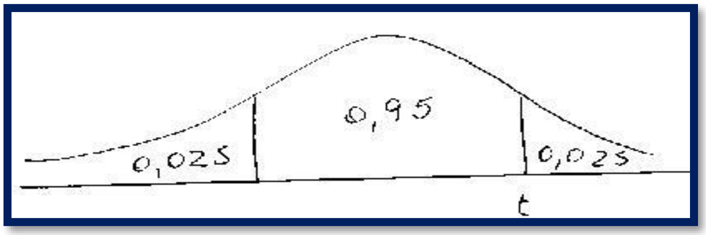
$$H_0 : \mu = \mu_0 \text{ o } H_0 : P = P_0$$
$$H_0 : \mu_1 = \mu_2 \text{ o } H_0 : P_1 = P_2$$

- **Hipótesis Alternativa**. Esta hipótesis plantea que los valores comparados son distintos y por lo tanto pertenecen a universos distintos. Dependiendo del problema podrá presentarse como:

$$H_1 : \mu \neq \mu_0 \text{ o } H_1 : P \neq P_0$$
$$H_1 : \mu_1 \neq \mu_2 \text{ o } H_1 : P_1 \neq P_2$$

2. Elegir un nivel de significación

El nivel de significación es la probabilidad de que la diferencia observada se deba al azar. Interesa que esta probabilidad sea pequeña, por eso, en la práctica se utilizan valores iguales o inferiores a 0,05. El valor más usado es 0,05 pero también puede ser 0,04; 0,02; 0,01; etc. Al nivel de significación se le identifica con la letra griega alfa (α). Al elegir un valor de alfa concreto, estamos dejando la mitad de alfa en cada extremo de la distribución de probabilidades (α/2).



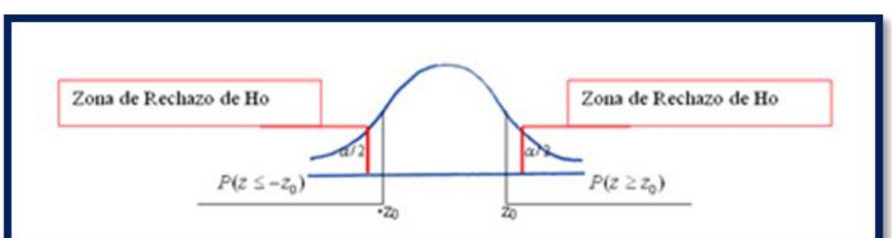
3. Calcular el estadístico de prueba a base de los datos muestrales

El estadístico que se utilice para la prueba de la hipótesis dependerá de los elementos que participan en él. Así, cuando se trate de comparar 2 promedios usaremos el estadístico t de *student*, cuando necesitemos comparar dos porcentajes muestrales usaremos Z, etc. Pero ya iremos viendo a cada uno de estos estadísticos actuar en terreno.

4. Buscar en la tabla correspondiente

Será necesario buscar a continuación:

- La probabilidad de obtener un valor igual o mayor al estadístico calculado, cuando éste sea positivo, o
- La probabilidad de obtener un valor menor o igual, cuando el estadístico sea negativo.
- En resumen: $P(z \geq z_0)$ cuando z_0 sea positivo o $P(z \leq z_0)$ cuando z_0 sea negativo.



5. Comparar la probabilidad obtenida en la tabla con el nivel de significación elegido en el punto 2 y tomar una decisión respecto de las hipótesis planteadas

Parece evidente que para tomar una buena decisión es conveniente disponer de criterios. Debemos decidir si la hipótesis nula es verdadera o falsa. Entonces, de acuerdo a la evidencia aportada por los datos de la muestra aceptaremos o rechazaremos la hipótesis nula según el siguiente criterio:

Se rechazará la *hipótesis nula* si la probabilidad encontrada en la tabla es inferior a la mitad del nivel de significación(α/2).

6. Elaborar una conclusión derivada de la decisión

Una vez tomada la decisión sobre las hipótesis debemos exponer lo que esto significa en el contexto de nuestro problema particular.

7. Apoyar todo el proceso de análisis con un gráfico del problema

A la hora de tomar la decisión es muy útil y orientador un buen gráfico donde se consigne el nivel de significación, el valor del estadístico y la probabilidad asociada a él.

Ejemplo No. 1

Una empresa fabrica y arma escritorios y otros muebles para oficina. La producción semanal del escritorio modelo A 325 tiene una distribución Normal, con una media de 200 y una desviación estándar de 16. Hace poco con motivo de expansión del mercado, se introdujeron nuevos métodos de producción y se contrató a más empleados. El vicepresidente de producción pretende investigar si hubo algún cambio en la producción semanal del escritorio. En otras palabras ¿La cantidad media de escritorios que se produjeron es diferente de 200 escritorios semanales?, utilice un nivel de significancia de 0.01.En una muestra de 50 semanas la cantidad media de escritorios que se produjeron fue de 203.5 .

Solución:

En base al problema este presenta una distribución Normal, por esa razón se ocupará el estadístico de prueba "**Z".**

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Datos

$\mu = 200$

$\sigma = 16$

$\alpha = 0.01$ (nivel de significancia, alfa)

$\bar{X} = 203.5$

$n = 50$

Paso No.1

Se establece la Hipótesis nula (H_o) : y La Hipótesis Alternativa (H_a)

$H_{o:}\ \mu = 200$

H_a: $\mu \neq 200$ es lo diferente o contrario de H_o

La hipótesis Nula se formula con el fin de rechazarse o no rechazarse es decir al final vamos a probar si la producción semanal sigue siendo de 200 o es diferente de 200, en este último caso si fuera diferente de 200 entonces estaríamos rechazando la Hipótesis nula y aceptando la Hipótesis Alternativa, pero si no es diferente de 200 y es igual a 200 entonces estaríamos afirmando la Hipótesis nula,

Paso No. 2

Seleccionar nivel de significancia y el valor de alfa , en este ejercicio ya nos menciona que debemos de utilizar un $\alpha = 0.01$, esta es la probabilidad de rechazar la Hipótesis nula cuando es verdadera, en una situación real nosotros somos los que debemos de seleccionar el nivel de significancia, los niveles de significancia mas utilizados son 0.05, 0.01, 0.10 que seria el 5%,1% y 10%.

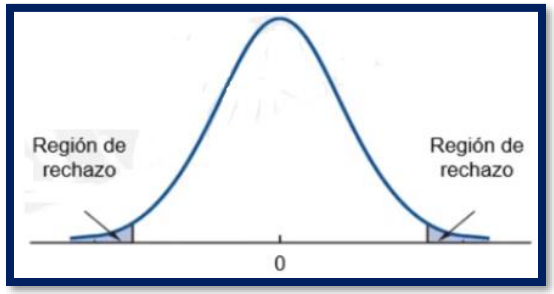
Paso No. 3

Seleccionar el estadístico de prueba; cuando se comporta normal será **“Z”**

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Paso No.

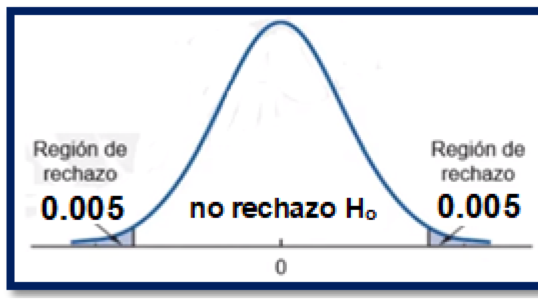
Formular la regla de decisión, dibujamos la curva de grafica normal.



Si la Hipótesis de rechazo fuera menor de 200 apuntaría hacia la izquierda

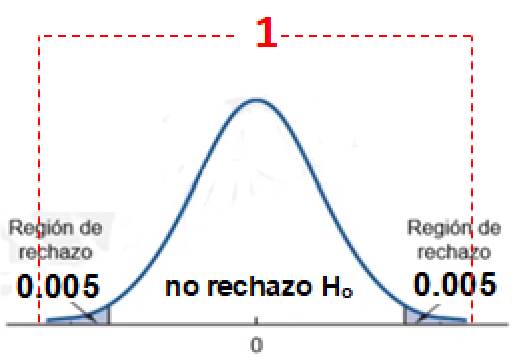
Si la Hipótesis alternativa fuera mayor a 200 entonces estaría apuntado hacia la derecha

Pero como en este ejemplo solo menciona que es diferente a 200 tenemos una gráfica de dos colas, dos regiones de rechazo , una ubicada en el lado izquierdo y otra en el derecho. El valor de la región de rechazo es alfa 0.01 pero como en este caso al existir dos regiones de rechazo, por lo que este valor se divide en forma equitativa por 2 es decir 0.01/2 = **0.005** lo colocamos en nuestra gráfica.



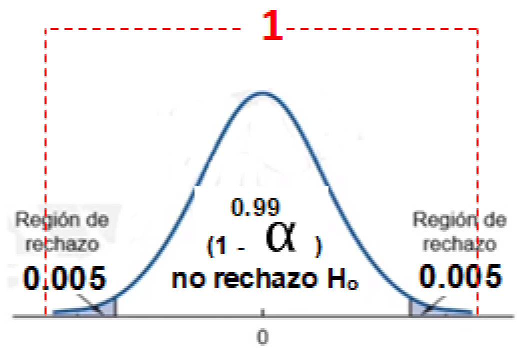
En la parte del centro de la curva será la región de **no rechazo H_o**,

El valor del área debajo de la curva incluyendo el área de rechazo y no rechazo tiene el valor de 1,

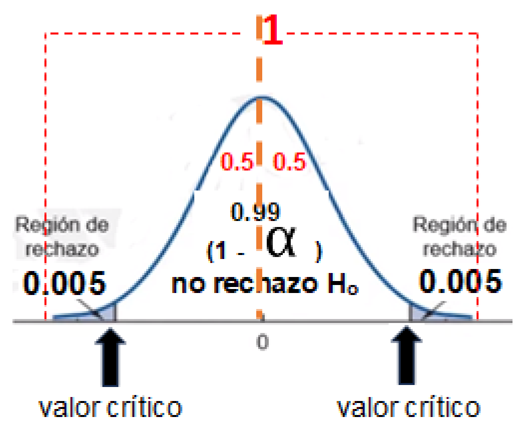


Si a 1 le quito la región de rechazo que es alfa tendríamos 1 – 0.01= 0.99 encontramos el área de no rechazo.

Si sumamos 0.005 + 0.005 + 0.99= 1
Que sería el valor de toda el área debajo de la curva.
En este caso al existir dos regiones de Rechazo, existen dos puntos de división Entre la región de rechazo y la de no rechazo, a esas divisiones se les conoce como valor crítico.



Para encontrar el área bajo la curva utilizaremos la tabla de valores de la curva, por lo que se partirá la curva a la mitad, como vale 1 la curva la mitad valdría **0.5**



Como la región de rechazo del lado derecho vale 0.005 y la mitad vale 0.5 basta con restar a 0.5 – 0.005 para encontrar el área de no rechazo de la mitad. **0.5 – 0.005 = 0.495**

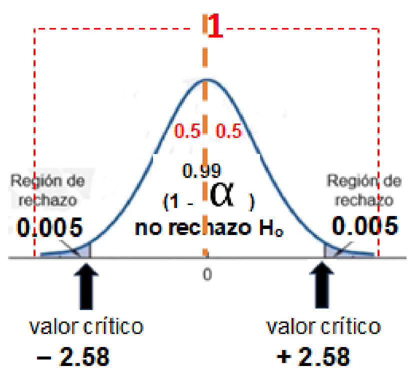
Checamos nuestra tabla de distribución normal, pero como no hay un valor igual a 0.495 el mas cercano para el valor de “Z” sería

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952

el

valor critico sería 2.58

Hacia la derecha son + 2.58 hacia la izquierda – 2.58



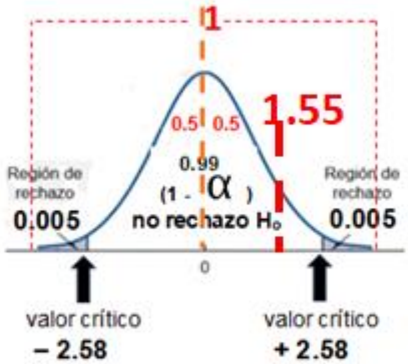
Paso No. 5

Se toma una decisión y se interpretan los resultados por lo que nos regresamos al paso No. 3

$$\boxed{z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}}$$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{203.5 - 200}{16 / \text{v}50} = 1.55$$

Este valor Z = 1.55 se ubica entre el 0 y el valor critico 2.58



Si el valor de **“Z”** no es mayor de 2.58 su valor crítico, se considera que esta dentro de la zona de NO rechazo.

Si estuviera por arriba de 2.58 estaría entrando a la región nula o de recházo.

En conclusión: Es no rechazar la Hipótesis nula, que menciona que la media poblacional es igual a 200, informaremos al vicepresidente de fabricación que la evidencia de la muestra no indica que la tasa de producción haya cambiado de 200 escritorios semanales, la diferencia de promedio muestral de 200 a 203.5 se debe al error de muestreo,

Para responder la pregunta planteada al inicio:

¿La cantidad media de escritorios que se produjeron es diferente de 200 escritorios semanales?

Según los datos estadísticos mostrados NO es diferente, en otras palabras no rechazamos la Hipótesis nula de media = 200

Actividad de Aprendizaje 53

La agencia de protección al consumidor, realiza periódicamente estudios estadísticos con objeto de comprobar las afirmaciones de los fabricantes acerca de sus productos. Por ejemplo, en la etiqueta de una lata de chocolate dice que la lata contiene 3 libras. La agencia sabe que el proceso de producción no permite llenar las latas con 3 libras exactas de chocolate, incluso si la media poblacional de peso de llenado de todas las latas es de 3 libras por lata ($\mu = 3$) con una σ 0.18. Sin embargo, mientras la media poblacional sea por lo menos 3 libras por lata, los derechos del consumidor estarán protegidos. Por tanto, la agencia interpreta que la información de la etiqueta de una lata de chocolate tiene una media poblacional del peso de llenado de por lo menos de 3 libras por lata. Suponga que en la muestra de las 36 latas de chocolate, la media obtenida es de 2.92 libras con un $\alpha = 0.01$