



UNIVERSIDAD NACIONAL EXPERIMENTAL DEL TÁCHIRA
VICE-RECTORADO ACADÉMICO
DECANATO DE DOCENCIA
DEPARTAMENTO DE MATEMÁTICA Y FÍSICA
UNIDAD CURRICULAR ESTADÍSTICA APLICADA A LA PSICOLOGÍA I

TEMA 2. CONCEPTOS BÁSICOS

La Estadística se ha convertido en un aliado en las investigaciones científicas con enfoque cuantitativo, por esta razón es relevante que el estudiante de Psicología aprenda y comprenda algunos términos que se utilizan durante el desarrollo de una investigación, teniendo en cuenta que la Estadística nos permite reunir datos sobre un tema, organizarlos para entenderlos mejor y, con ello, tomar mejor nuestras decisiones. A continuación, se describen algunos términos básicos que son utilizados en el ámbito estadístico y por supuesto, forman parte del contexto de la Psicología.

2.1. Variable

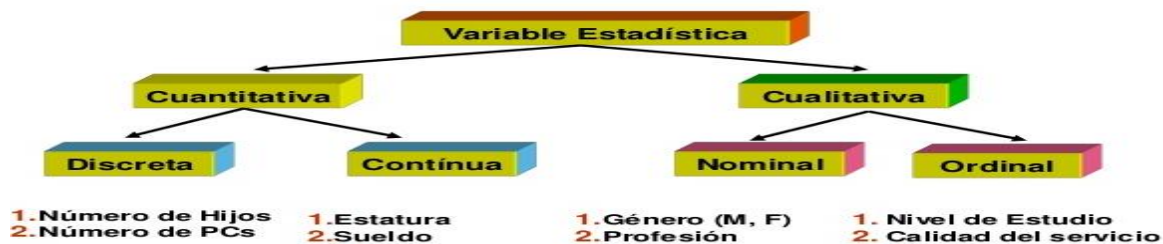
Es una característica de la población o de la muestra cuya medida puede cambiar de valor. Se representa simbólicamente mediante las letras del alfabeto. Según su naturaleza puede ser cualitativa y cuantitativa.

1. Variable cualitativa. Es aquella que representa cualidades, atributos o características no numéricas y estas pueden ser nominales y ordinales.

2. Variable cuantitativa. Es aquella característica de la población o de la muestra que es posible representar numéricamente. Éstas pueden ser continua y discreta.

3. Variable continua. Es una variable cuantitativa. Es la característica de la población, cuyos valores están representados mediante el conjunto de los números reales. Puede tomar cualquier valor real dentro de un intervalo. Por ejemplo, la velocidad de un vehículo puede ser 80,3 km/h, 94,57 km/h.

4. Variable discreta. Es una variable cuantitativa. Es la característica de la población, cuyos valores están representados mediante el conjunto de los números naturales. Por ejemplo, el número de alumnos de un aula.



5. Variable aleatoria. Conocida también como variable estocástica o probabilística. Es la característica considerada en un experimento aleatorio cuyo valor de ocurrencia sólo puede saberse con exactitud una vez observado.

6. Variable bidimensional. Es aquella que proporciona información sobre dos características de la población (por ejemplo: edad y altura de los alumnos de una clase).

7. Variable determinística. Es aquella cuyo valor puede ser predicho con exactitud.

2.2. Dato

Conocido también como información, es el valor de la variable asociada a un elemento de una población o una muestra.

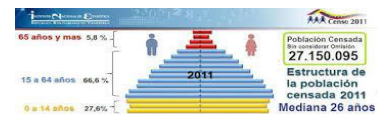
1. Dato cualitativo. Es aquel que representa alguna característica de los elementos de una muestra o una población que presentan, atributos, actitudes o son opiniones. Son datos no numéricos.

2. Dato cuantitativo. Es aquel dato numérico que representa aspectos de una muestra o una población que es medible o que se puede contar.

3. Dato estadístico. Es cada uno de los valores que se ha obtenido al realizar un estudio estadístico. Por ejemplo, el sexo de una persona (masculino y femenino), la edad de la persona.

2.3. Censo

Es una investigación estadística que consiste en el recuento de la totalidad de los elementos que componen la población por investigar. Es necesario que se especifique el espacio y el tiempo al que se refiere el recuento.



2.4. Población o universo (N)

Es cualquier conjunto de unidades o elementos claramente definido, en el espacio y el tiempo, donde los elementos pueden ser personas, granjas, hogares, manzanas, condados, escuelas, hospitales, empresas, y cualquier otro. Las poblaciones pueden ser finitas e infinitas.



1. Población finita. Es aquella en la que es posible enumerar (contar) físicamente los elementos que pertenecen a la población. Por ejemplo, la cantidad de empleados de una organización, la cantidad de estudiantes de una institución educativa.

2. Población infinita. Es aquella en la que no es posible enumerar (contar) físicamente los elementos que pertenecen a la población. Dicho de otra manera, cuando los elementos de la población son ilimitados. Por ejemplo, la cantidad de personas que visitan un centro comercial.

Nota. Se considera que una población finita a toda población formada por menos de 100.000 unidades, e infinita a aquella que tiene 100.000 o más.



3. Población objetivo. Es el conjunto de elementos de los que nos interesa obtener información o tomar decisiones.

2.5. Muestra

Es un subconjunto representativo de la población a partir del cual se pretende realizar inferencias respecto a la población de donde procede. Los elementos seleccionados con cierta técnica reúnen ciertas características que la hacen ser representativa, significativa y confiable y que en base a ella se pueden hacer inferencias respecto a la población. La muestra puede ser probabilística y no probabilística.



1. Muestra probabilística. Es aquella muestra obtenida por un mecanismo de probabilidades, en el cual cada elemento de la población total o universo tiene una probabilidad conocida de selección. (Ver muestreo aleatorio).



2. Muestra no probabilística. Es aquella que se obtiene mediante juicio de la persona que selecciona los elementos de la muestra que usualmente es un experto en la materia. Este método está basado en los puntos de vista subjetivos de una persona y la teoría de la probabilidad no puede ser empleada para medir el error de muestreo. Las principales ventajas de una muestra de juicio son la facilidad de obtenerla y que el costo usualmente es bajo.



¿Cómo se determina el tamaño de la muestra de la población objeto de estudio?

Uno de los procedimientos a utilizar para este propósito es haciendo uso de fórmulas estadísticas, dependiendo si población es infinita o finita, y que se va a estimar, si es media poblacional o proporción poblacional. También existen programas estadísticos para este propósito. Sea la manera de calcular el tamaño de la muestra, es importante conocer la condición de la variable, si es cualitativa binomial o cuantitativa. A continuación, se mencionan las respectivas fórmulas:

Población	Para estimar la media (variables cuantitativas)	Para estimar la proporción (variables categóricas)
Infinita	$n = \left(\frac{Z_{1-\alpha/2} * \sigma}{e} \right)^2$	$n = \frac{Z_{1-\alpha/2}^2 * P * Q}{e^2}$
Finita	$n = \frac{N * Z_{1-\alpha/2}^2 * \sigma^2}{e^2(N-1) + Z_{1-\alpha/2}^2 * \sigma^2}$	$n = \frac{N * Z_{1-\alpha/2}^2 * P * Q}{e^2(N-1) + Z_{1-\alpha/2}^2 * P * Q}$

Siendo **n** tamaño de la muestra, **N** tamaño de la población, **Z** valor utilizado para el grado o nivel de confianza especificado (**1- α**) (generalmente se utilizan 90%, 95% y 99%) cuyo valor se busca en la tabla de distribución normal (siendo éstos los siguientes: 1.65, 1.95 y 2.58, respectivamente), **s** es la desviación estándar de la población, **P** es la proporción del hecho que se está estudiando (en ocasiones cuando se desconoce, se asume 0,5), **Q** = 1 - P y, **e** es el factor de error más o menos permitido asumido por el investigador. En las investigaciones sociales se puede utilizar un valor máximo admisible de 5%, 10% o 15%.

Cabe destacar que, además del nivel de confianza, está el nivel de significación que se simboliza con **α**, por tanto, es dato que se debe calcular, tal como se demuestra con el siguiente ejemplo: Supongamos que tenemos un nivel de confianza del 95%, entonces el nivel de significación es:

$$1 - \alpha = 0,95 \rightarrow \alpha = 1 - 0,95 = 0,05$$

Una vez obtenido **α** tenemos que $Z_{1-0,05/2} = Z_{0,975}$. Entonces, haciendo uso de la tabla de distribución normal (ver página siguiente) procedemos a buscar el valor 0,975 y luego se busca de manera horizontal y vertical los valores de Z; en este caso, es 1,96 como se muestra en la figura.

z	.00	.01	.02	.03	.04	.05	.06	.07
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756

También podemos hacer uso del Excel mediante la siguiente función:

$$=INV.NORM.ESTAND(probabilidad)$$

$$=DISTR.NORM.ESTAND(0,975) = 1,96$$

Nota. En ocasiones no se obtiene un valor percentil directamente en la tabla, por lo que se debe realizar una interpolación lineal para obtener el valor verdadero. Este método trata de aproximar el valor de la probabilidad buscada partiendo de dos puntos conocidos. Ambos puntos definen una función lineal y con dicha función aproximamos el valor.

Por ejemplo: Se tiene un valor de P igual a 0,95. Esta cifra la buscamos en la tabla de distribución normal y observamos que no aparece, en su lugar, está entre dos valores de Z, como se muestra en la figura.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633

$z_1 = 1,64$	$p_1 = 0,9495$
$z = ?$	$p = 0,95$
$z_2 = 1,65$	$p_2 = 0,9505$

En general, sean z_1 y z_2 y sus probabilidades correspondientes, p_1 y p_2 , y sea z el valor cuya probabilidad p nos piden, entonces se cumple:

$$\frac{z - z_1}{p - p_1} = \frac{z_2 - z_1}{p_2 - p_1}$$

De aquí obtenemos la probabilidad $P(Z \leq z) = p$ con $z \in (z_1, z_2)$

$$p = (z - z_1) \cdot \frac{p_2 - p_1}{z_2 - z_1} + p_1$$

$$z = (p - p_1) \cdot \frac{z_2 - z_1}{p_2 - p_1} + z_1$$

Sustituyendo los valores en la formula nos queda:

$$0,95 = (z - 1,64) \cdot \frac{0,9505 - 0,9495}{1,65 - 1,64} + 0,9495$$

$$0,95 = (z - 1,64) \cdot 0,1 + 0,9495$$

$$0,95 - 0,9495 = 0,1z - 0,164$$

$$0,1z = 0,0005 + 0,164$$

$$z = 1,645$$

Así que el valor de z para una probabilidad de 0,95 es 1,645.

Cabe destacar que, el tamaño de la muestra debe ser:

a. Representativa: Hace referencia a que todos los miembros de un grupo de personas tengan las mismas oportunidades de participar en la investigación.

b. Adecuada: Se refiere a que el tamaño de la muestra debe de ser obtenido mediante un análisis que permite resultados como disminuir el margen de error.

A normal distribution curve is shown with a horizontal axis. The mean is marked as 0 with a vertical dashed line. A point z is marked on the axis to the right of 0. The area under the curve to the left of z is shaded and labeled "Area".

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.					

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

2.6. Muestreo

Es un conjunto de métodos y procedimientos estadísticos destinados a la selección de una o más muestras es la técnica seguida para elegir muestras. El objetivo principal de un diseño de muestreo es proporcionar procedimientos para la selección de muestras que sean representativas de la población en estudio. Existen dos métodos de muestreo: probabilístico y no probabilístico.

2.6.1. Muestreos Probabilísticos

Muestreo aleatorio simple . También llamado irrestrictamente aleatorio. Es un método de muestreo donde una muestra aleatoria simple es seleccionada de tal manera que cada muestra posible del mismo tamaño tiene igual probabilidad de ser seleccionada de la población. Una muestra aleatoria es también llamada una **muestra probabilística** es aquella cuyos elementos se seleccionan individualmente de la población en forma aleatoria, y es preferida por los estadísticos porque la selección de las muestras es objetiva y el error muestral puede ser medido en términos de probabilidad bajo la curva normal. Por conveniencia, este método puede ser reemplazado por una tabla de números aleatorios cuando una población es infinita.



Muestreo estratificado aleatorio. Es un método de muestreo que se aplica cuando se divide la población en grupos, llamados estratos, donde los datos son más homogéneos pero un estrato frente al otro muy distintos. Para extraer la muestra aleatoria se aplica el muestreo aleatorio simple a cada estrato y el tamaño es la suma de los tamaños de todos los estratos. Para determinar los tamaños de los estratos se puede utilizar la asignación proporcional, óptima y óptima económica. Si no se conoce la variabilidad de los datos se aplica la asignación proporcional.



Afijación igual. Todos los estratos tienen el mismo número de elementos en la muestra. Se elige el mismo número de individuos de cada estrato, basta entonces con dividir el número de elementos que queremos en la muestra entre el número de estratos y obtendremos el número de elementos a elegir en cada estrato.

$$n_i = \frac{n}{k}$$

Siendo n_i la cantidad de individuos en el estrato i ($i = 1, 2, 3, \dots, k$), n es el tamaño de la muestra, k es la cantidad estrato.

Afijación proporcional. Cada estrato tiene un número de elementos en la muestra proporcional a su tamaño. Para cada estrato hay que hacer una proporción, por tanto, basta con dividir el tamaño de la muestra entre el tamaño de la población y multiplicar por el tamaño de cada estrato. En la escena basta con introducir el tamaño de la población, el tamaño de la muestra y el tamaño de un estrato y obtendremos los elementos de la muestra que debemos elegir en ese estrato.

$$n_i = \frac{n}{N} * N_i$$

Siendo n_i tamaño de cada estrato i , n es el tamaño de la muestra total, N es el tamaño de la

población y N_i es el tamaño de muestra de cada estrato i .

Ejemplo. Supongamos que, en un centro universitario de 1.100 estudiantes, hay 350 estudiantes que cursan ingeniería, 300 en administración, 250 en psicología, y 200 en educación; y queremos estudiar el número de estudiantes que aprueban todas las asignaturas. Pensamos que puede haber diferencias entre las distintas especialidades, por lo que nos interesa que en la muestra estén representados todos los cursos.

1. ¿Cuál será el tamaño de la muestra, para un nivel de confianza del 95% y un error muestral del 5%?

Calculamos el tamaño de la muestra mediante la estimación proporcional, asumiendo que la variable "Aprueba las asignaturas" es categórica y binomial, por lo que valor de P se desconoce; sin embargo, lo asumimos, en este caso sería 0,50 siendo $Q = 1 - P = 0,50$. El valor de Z para un nivel de confianza del 95% sería 1,96. Además, se tiene una población de 1.100 estudiantes.

$$n = \frac{N * Z_{\alpha/2}^2 * P * Q}{e^2(N - 1) + Z_{\alpha/2}^2 * P * Q}$$
$$n = \frac{1100 * 1,96^2 * 0,50 * 0,50}{0,05^2(1100 - 1) + 1,96^2 * 0,50 * 0,50} = \frac{1056,44}{2,7475 + 0,9604} = 284,91 = 285$$

Se seleccionarán aleatoriamente 285 estudiantes del centro universitario, lo que representa aproximadamente el 26% (285/1100).

2. ¿Cuántos estudiantes se seleccionarán de cada estrato?

350 estudiantes de ingeniería
300 estudiantes de administración
250 estudiantes de psicología
200 estudiantes de educación

Aplicando el método de afijación igual, el tamaño de cada estrato sería:

$$n_i = \frac{n}{k}$$

Como hay 4 estratos, entonces $k = 4$

$$n_1 = n_2 = n_3 = n_4 = \frac{285}{4} = 71,25 = 72$$

Se tomará de cada estrato una muestra al azar de 72 estudiantes.

Aplicando el método de afijación proporcional, el tamaño de cada estrato sería:

$$n_i = \frac{n}{N} * N_i$$

$$n_1 = \frac{285}{1100} * 350 = 90,68 = 91 \text{ estudiantes de ingeniería}$$

$$n_2 = \frac{285}{1100} * 300 = 77,72 = 78 \text{ estudiantes de administracion}$$

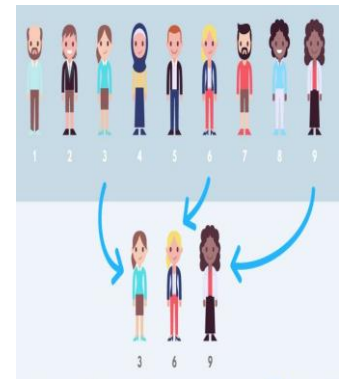
$$n_3 = \frac{285}{1100} * 250 = 64,77 = 65 \text{ estudiantes de psicología}$$

$$n_4 = \frac{285}{1100} * 200 = 51,81 = 52 \text{ estudiantes de educación}$$

Estrato	Ni	n _i
Ingeniería	350	91
Administración	300	78
Psicología	250	65
Educación	200	52
TOTAL	1100	286

En la tabla se indica la cantidad de estudiantes que debemos seleccionar al azar por especialidad del centro universitario para estudiar el número de estudiantes que aprueban todas las asignaturas, para un nivel de confianza del 95% con un error muestral del 5%.

Muestreo sistemático. Es una variante del método aleatorio simple de selección de cada elemento de la muestra. Se aplica cuando la población está listada en algún orden. Consiste en seleccionar un número aleatorio menor que N/n y luego los (n-1) elementos de la muestra se eligen agregando al primer aleatorio: el entero K obtenido por $K=N/n$ y así sucesivamente. El primer elemento de la muestra es seleccionado al azar. Por lo tanto, una muestra sistemática puede dar la misma precisión de estimación acerca de la población que una muestra aleatoria simple cuando los elementos en la población están ordenados al azar.



Ejemplo. Siguiendo con el ejemplo anterior, tenemos que la población de estudiantes (N) es 1.100, y vamos a tomar una muestra (n) de 285 estudiantes, por tanto, al aplicar la fórmula $K=N/n = 1100/285 = 3,85 = 4$, es decir, cada 4 estudiantes es la selección sistemática de la muestra. El procedimiento a seguir es el siguiente: Se numeran los elementos del 1 al 1.100, se toma al azar un número entre este intervalo, por ejemplo 10, a partir de este se comienza sistemáticamente a elegir la cantidad de elementos hasta completar **n**. Lo recomendable es hacerlo sin reemplazamiento, es decir, aquel estudiante seleccionado no se vuelve a introducir a la población.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68
69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	100	101	102	103								11000

Los números resaltados en color amarillo son los estudiantes que se tomaran como muestra de estudio. Si es necesario, debe volverse al principio para completar la muestra porque debe seleccionarse la cantidad de $n = 285$.

Muestreo por conglomerados. Es un método de muestreo en el cual la población está en grupos debido a la organización administrativa u otro (conglomerados). Ejemplo: Colegios, Universidades, manzanas de casas, entre otros. Al interior de los conglomerados se puede garantizar homogeneidad. Cada conglomerado es una unidad donde la muestra se selecciona como en el muestreo aleatorio simple y se aplica la encuesta a todos los elementos del conglomerado.



Una muestra de conglomerados, usualmente produce un mayor error muestral, por tanto, se obtiene menor precisión de las estimaciones acerca de la población, que una muestra aleatoria simple del mismo tamaño. Los elementos individuales dentro de cada "conglomerado" tienden frecuentemente a ser iguales.

Ejemplo. Un grupo de investigadores quiere conocer las características de las estudiantes que ingresan a las universidades públicas en Venezuela. Para realizar la muestra representativa, se siguen los pasos del muestreo por conglomerados:

- Se selecciona a todos los estudiantes universitarios de nuevo ingreso como población objetivo, por ejemplo, de una entidad o región. Cada universidad será un conglomerado.
- Se calcula que de las **N** universidades públicas que hay en el país, se necesitarán **n** conglomerados para elaborar la muestra.
- Se realiza una lista de las universidades públicas.
- Se seleccionan **n** universidades al azar y con un muestreo aleatorio simple se escoge a **X** estudiantes de cada una.

Ventajas y desventajas de los muestreos probabilísticos

Muestreo	Ventajas	Desventajas
Muestreo aleatorio simple	<ul style="list-style-type: none"> – Es un muestreo de comprensión sencilla. – Las medias y varianzas se calculan rápidamente. – Se basan en datos estadísticos para el análisis de resultados 	<ul style="list-style-type: none"> – Requiere un listado previo de toda la población. – Si la muestra es pequeña puede que no sea representativa.
Muestreo estratificado	<ul style="list-style-type: none"> – Este muestreo pretende representar la población en función de las variables de estudio. – Las estimaciones son con un alto nivel de precisión. – Su objetivo es conseguir una muestra lo más semejante a la población en lo que a la o las variables estratificadoras se refiere. 	<ul style="list-style-type: none"> – Se debe conocer la forma en que se encuentra distribuida la población para realizar la estratificación. Muchas veces se requiere la ponderación y complica el análisis.
Muestreo sistemático	<ul style="list-style-type: none"> – Tiene facilidad de aplicación. – No es un requisito fundamental tener la población enlistada. – Si la población se encuentra ordenada vamos a tener una tendencia conocida. 	<ul style="list-style-type: none"> – Se pueden dar estimaciones con nivel sesgado, si el muestreo se asocia al fenómeno de interés.
Muestreo por conglomerados	<ul style="list-style-type: none"> – Es realmente útil cuando la población es dispersa y grande. – Costos reducidos. – No tiene un nivel de precisión de la población en general solo de las muestras. 	<ul style="list-style-type: none"> – Existe mayor error estándar que le muestreo estratificado o simple. – Existe complejidad en el cálculo de error.

2.6.2. Muestreos No Probabilísticos

Muestreo por conveniencia. Es una técnica de muestreo no probabilística donde las muestras de la población se seleccionan solo porque están convenientemente disponibles para el investigador. Estas muestras se seleccionan solo porque son fáciles de reclutar y porque el investigador no consideró seleccionar una muestra que represente a toda la población. Esta técnica de muestreo no probabilística es más común, debido a su velocidad, costo-efectividad y facilidad de disponibilidad de la muestra. Ejemplo de muestreo por conveniencia sería utilizar a estudiantes voluntarios que sean conocidos del investigador. El investigador puede enviar la encuesta a los estudiantes y ellos en este caso actuarían como muestra.

Muestreo intencional u opinático. En este caso los elementos son escogidos con base en criterios o juicios preestablecidos por el investigador. Es decir, el investigador elige solo a aquellos que estos creen que son los adecuados (con respecto a los atributos y la representación de una población) para participar en un estudio de investigación. La desventaja del muestreo intencional es que los resultados pueden estar influenciados por nociones percibidas del investigador. Por lo tanto, hay una gran cantidad de ambigüedad involucrada en esta técnica de investigación. Este tipo de método de muestreo se puede utilizar en estudios piloto. Por ejemplo, para un estudio sobre la atención en el servicio de psicología en un centro de salud, previamente, se establecen como criterios de selección de la muestra los siguientes: ser paciente del servicio, haber acudido mínimo tres consultas, ser mayor de 18 años y menor de 40 años. Por supuesto, la muestra la integrarán sólo aquellos que cumplan con las condiciones anteriores.

Muestreo casual o accidental. Es un procedimiento que permite elegir arbitrariamente los elementos sin un juicio o criterio preestablecido. Por ejemplo, Un encuestador se ubica en un sector y aborda a los transeúntes que pasan por el lugar. Lógicamente, las personas que no circulen por la zona, carecen de toda probabilidad para integrar la muestra.

Muestreo por cuotas. Se basa en la elección de los elementos en función de ciertas características de la población, de modo tal que se conformen grupos o cuotas correspondientes con cada característica, procurando respetar las proporciones en que se encuentran en la población. Por lo tanto, el muestreo por cuotas ayuda a dividir la población en estratos o grupos. En el ejemplo anterior, supongamos que deseamos tomar una muestra de 50 pacientes, por lo que debemos seleccionar un número proporcional de hombres y mujeres, lo que significa que debe haber 25 hombres y 25 mujeres. Como esto es improbable, los grupos o estratos se seleccionan mediante el muestreo por cuota.

Muestreo de bola de nieve. Este tipo de técnica de muestreo ayuda a los investigadores a encontrar muestras cuando son difíciles de localizar. Los investigadores utilizan esta técnica cuando el tamaño de la muestra es pequeño y no está disponible fácilmente. Este sistema de muestreo bola de nieve funciona como el programa de referencia. Una vez que los investigadores encuentran sujetos adecuados, se les pide a estos ayudas para buscar a sujetos similares y así poder formar una muestra de buen tamaño. Por ejemplo, este tipo de muestreo se puede utilizar para realizar investigaciones que involucren una enfermedad particular en pacientes o tal vez una enfermedad rara también. Los investigadores pueden buscar ayuda de las personas enfermas para que estos refieran a otros que sufran de la misma dolencia y con esto formar una muestra subjetiva para llevar a cabo el estudio.

¿Cuándo usar el muestreo no probabilístico?

- Este tipo de muestreo se usa para indicar si existe un rasgo o característica particular en una población.
- Esta técnica de muestreo es ampliamente utilizada cuando los investigadores realizan investigaciones cualitativas, estudios piloto o investigación exploratoria.
- El muestreo no probabilístico se usa cuando los investigadores tienen un tiempo limitado para llevar a cabo la investigación o tienen limitaciones presupuestarias.
- El muestreo no probabilístico se realiza para observar si un tema en particular necesita un análisis en profundidad.

Ventajas del muestreo no probabilístico

- El muestreo no probabilístico es un método práctico para los investigadores que implementan encuestas en el mundo real. Aunque claro, cabe mencionar que los estadistas prefieren el muestreo probabilístico porque arroja datos en forma de números. Pero la realidad es que, si se hace correctamente, el muestreo no probabilístico puede arrojar resultados similares, si no es que de la misma calidad.
- Obtener respuestas utilizando el muestreo no probabilístico es más rápido y más rentable en comparación al muestreo probabilístico ya que el investigador conoce la muestra. Comúnmente los participantes están motivados para responder rápidamente en comparación con las personas que se seleccionan al azar.

Desventajas del muestreo no probabilístico

- En el muestreo no probabilístico, el investigador necesita pensar las posibles razones de los sesgos. Es importante tener una muestra que represente de cerca a la población.
- Al elegir una muestra en un muestreo no probabilístico, los investigadores deben tener cuidado con los reclutas que puedan distorsionar los datos. Al final del día, la investigación se lleva a cabo para obtener información valiosa y datos útiles.

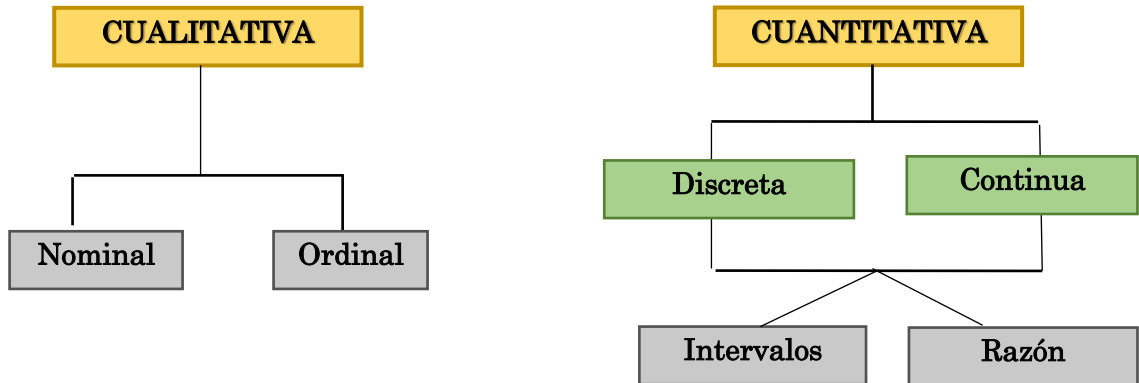
Muestreo con reposición. Es el método para obtener una muestra con reposición. Esta muestra consiste en que, al seleccionar un segundo elemento, el primero debe haber sido devuelto a la población. De este modo un elemento puede repetirse en la muestra. Es decir, con este método una unidad en particular puede quedar incluida más de una vez en la muestra, pudiendo ser hasta "n veces". El universo o población se mantiene permanentemente con un tamaño N.

Muestreo sin reposición. Es el procedimiento para seleccionar cada elemento de la población éste no se repone o considera de nuevo en la población, por lo que no puede ser seleccionado nuevamente. En este caso el tamaño de la población o universo se irá reduciendo en cada selección $N-1$, $N-2$, $N-3$, ..., etc. unidades de muestreo hasta $N - n$ elementos.

2.7. Escala de Medida

Una escala de medición es el conjunto de los posibles valores que una cierta variable puede tomar. Es un continuo de valores ordenados correlativamente, que admite un punto inicial y otro final. El nivel en que una variable puede ser medida determina las propiedades de medición de una variable, el tipo de operaciones matemáticas que puede usarse apropiadamente con dicho nivel, las fórmulas y procedimientos estadísticos que se utilizan para el análisis de datos y la prueba de hipótesis teóricas.

Clasificación de Variables y Escala de Medición



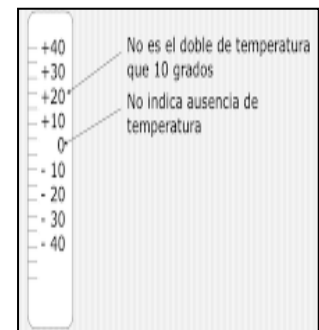
1. Nominal. Describe variables de naturaleza categórica que difieren en cualidad más que en cantidad. Ante las observaciones que se realizan de la realidad, es posible asignar cada una de ellas exclusivamente a una categoría o grupo. Este nivel de medición es exclusivamente cualitativo y sus variables son por lo tanto cualitativas. Este tipo de variables sólo permite establecer relaciones de igualdad/desigualdad entre los elementos de la variable, pero no operaciones aritméticas. La asignación de los valores se realiza en forma aleatoria por lo que NO cuenta con un orden lógico y generan categorías mutuamente excluyentes y exhaustivas. Ejemplos: Género, estado civil, procedencia geográfica, nacionalidad, religión, ideología política, etc.



2. Ordinal. Se establecen categorías con dos o más niveles que implican un orden inherente entre sí. El nivel ordinal incorpora valores numéricos a los caracteres de tal forma que se asignan los números más altos a los elementos que poseen más de la característica que se mide. En este caso las variables no sólo se asignan a grupos, sino que además pueden establecerse relaciones de mayor que, menor que o igual que, entre los elementos. Ejemplos: Rangos militares, categoría de docentes universitarios, orden de llegada en una carrera, la entrega de medalla en una olimpiada, cuestionarios fraseados en escala, etc.



3. Intervalo. Son variables numéricas cuyos valores representan magnitudes y la distancia entre los números de su escala es igual. Con este tipo de variables podemos realizar comparaciones de igualdad/desigualdad, establecer un orden dentro de sus valores y medir la distancia existente entre cada valor de la escala. Las variables en escala de intervalo el valor cero de la escala no es absoluto, sino un cero arbitrario, es decir, no refleja ausencia de la magnitud medida, por lo que las operaciones aritméticas de multiplicación y división no son apropiadas. Ejemplos: Temperatura, escala cronológica, etc.



4. Razón. Las variables de razón poseen las mismas características de las variables de intervalo, con la diferencia que cuentan con un cero absoluto; el valor cero (0) representa la ausencia total de medida, por lo que se puede realizar cualquier operación Aritmética (Suma, Resta, Multiplicación y División) y Lógica (Comparación y ordenamiento). Este tipo de variables permiten el nivel más alto de medición. Ejemplos: altura, salario, peso, ingresos, ventas, unidades producidas, distancia, etc.

RELACIÓN ENTRE LAS ESCALAS DE MEDIDA Y LAS TÉCNICAS ESTADÍSTICAS

TÉCNICA ESTADÍSTICA	ESCALA DE MEDIDA			
	Nominal	Ordinal	Intervalo	Razón
Frecuencia	X	X	X	X
Porcentaje	X	X	X	X
Moda	X	X	X	X
Mediana		X	X	X
Media aritmética			X	X
Media geométrica				X
Media armónica				X
Rango			X	X
Desviación media			X	X
Varianza			X	X
Desviación típica			X	X
Sesgo			X	X
Curtosis			X	X

2.8. Parámetro

Los parámetros son medidas cuantitativas que describen una característica de la población, entre ellas están: media aritmética, desviación estándar, varianza, proporción, coeficiente de variación, entre otras medidas estadísticas.

2.9. Estadístico o Estadígrafo

Los estadígrafos o estadísticas son medidas cuantitativas que describen una característica de la muestra y se consideran estimadores para la población, por ejemplo, la media muestral, desviación estándar muestral.

2.10. Constructo

En la ciencia de la psicología no todo es observable y, por ello, se utiliza el término constructo, que no es sino un término que se aplica a un conjunto de comportamientos que no se pueden observar directamente, por lo que no existe una realidad empírica. Los constructos se utilizan para conocer fenómenos como la inteligencia, la conciencia, la autoestima, las habilidades sociales o la depresión, entre otros muchos.

Actividad de autoevaluación

1. Es un sondeo de opinión acerca de lo que la gente pensaría si se adoptasen medidas restrictivas en el consumo de tabaco. La idea es anticiparse a las reacciones, de forma que interesa conocer la opinión general a partir de las preguntas realizadas en el sondeo. Sin duda, lo más preciso sería preguntarles a todos y cada uno de los españoles por su opinión, pero por razones económicas esto no sería sensato. Decidimos, en consecuencia, seleccionar un grupo de 2.500 españoles de todas las comunidades autónomas y edades, consultándoles acerca de su opinión (a favor/en contra) sobre esta cuestión.

Responder lo siguiente: (a) mencione la variable de estudio, será cuantitativa o cualitativa; (b) ¿cuál es la escala de medida?; (c) ¿qué tipo de muestreo se aplicó?, explique su escogencia.

2. En el siguiente listado de situaciones que se mencionan a continuación, indique si es población (finita o infinita) o muestra (probabilística o no probabilística):

- a) Pacientes que asisten a consulta a un centro médico.
- b) Personas que acuden a una peregrinación.
- c) Selección de pacientes entre 18 y 25 años para un diagnóstico clínico.
- d) Escogencia del 25% de los habitantes de una comunidad.
- e) Proporción de personas con una enfermedad mental.
- f) Rendimiento académico de estudiantes de educación primaria del estado Táchira.
- g) Personas que son tratadas con un fármaco para el control de la tensión arterial en una comunidad.
- h) Personas diagnosticada con COVID-19 en el año 2020.

3. Señale el tipo de variable y escala de medida a la que pertenecen las siguientes variables:

- a) Lugar de la residencia del paciente.
- b) Grado de aceptación (totalmente de acuerdo; de acuerdo; en desacuerdo; totalmente en desacuerdo) de la regulación de la muerte asistida.
- c) Diagnóstico clínico.
- d) Numero de estímulos visuales detectados.
- e) Inteligencia medida con el Test de Inteligencia WAIS III para adultos.
- f) Distancia tolerada a un estímulo fóbico.
- g) Nivel de adecuación (adecuado, neutro, inadecuado) de una conducta de las personas en un estadio de fútbol.
- h) Orientación terapéutica seguida por un psicólogo.
- i) Gravedad del diagnóstico clínico.
- j) Años de ejercicio profesional de un psicólogo clínico.
- k) Sexo del paciente.
- l) Selección de un color (nada adecuado; adecuado; totalmente adecuado) asociado a una emoción.

4. Supóngase que se desea conocer la incidencia de la aplicación de un programa de capacitación a los empleados de una empresa X. Se sabe que la misma dispone de 350 trabajadores, distribuidos en tres departamentos: A = 150, B = 85, C = 115. Ahora bien, el gerente de recursos humanos no desea tomar toda la población de empleados, por lo que decide buscar

un especialista para que realice el cálculo de cuántos empleados debe elegirse aleatoriamente de cada departamento, a un nivel de confianza del 95% y un error del 10%. ¿Aumentará o disminuirá el tamaño de la muestra para un nivel de confianza del 90% y un error del 10%?

5. Indique cuatro ejemplos, donde establezca tipo de variable, codificación y nivel de medición, como se muestra en el ejemplo.

Nombre de la variable	Tipo de variable	Codificación	Nivel de medición

6. Indique el valor de z para una probabilidad de (a) 0,9804 y (b) 0,0065.

7. Indique la probabilidad para un valor de z igual a (a) 0,125 y (b) 1,358.