ESTADÍSTICA DESCRIPTIVA

Apuntes de la Profesora:
Est. Ana María Rossi de Salvagni

Universidad Nacional de La Patagonia S. J. B.
Comodoro Rivadavia

El presente no es un “Apunte tradicional” sino lo que hemos llamado “facilitador”.

Se ha ido dando forma a través del tiempo a **“facilitadores”**, que consisten en desarrollos alternativamente teóricos y / o prácticos, con complementos anexos, con citación de bibliografía específica, que cumplen la función de permitir que el alumno, solo o en grupo y de manera “guiada”, encuentre, analice, discuta, compare, relacione, etc, los contenidos de algunas unidades, partiendo de la teoría para luego aplicarla, o bien, por el contrario, deduciendo contenidos a partir del “hacer” y “resolver” situaciones problemáticas.

Estos **“facilitadores”** se ponen a disposición de los alumnos, para que sean empleados si ellos lo desean, en forma voluntaria, entendiendo que el trabajo a realizar con ellos significa reconocer que el esfuerzo recae en el estudiante. Es así que el estudiante se transforma en hacedor de su propio saber, al construir poco a poco y respetando ”su propio tiempo” aquellos contenidos que le interesen, al redescubrir ideas que posiblemente sin saberlo maneja diariamente en sus actividades, al encontrar otras formas de razonar diferentes. Lo que el estudiante debe tener en claro es que trabajar con estos facilitadores lleva tiempo; tal vez más del que tenga disponible para la asignatura y que también es él mismo, como estudiante universitario, el que debe seleccionar la forma en que se apropia de los contenidos y el tipo de material con el que le es más útil lograr las competencias que trata de conseguir.

ESTADÍSTICA

(1) a) Interprete las siguientes expresiones, obtenidas de diferentes autores y enumere términos desconocidos y/o poco empleados para luego tenerlos en cuenta e ir definiéndolos

a.a) Un sentido moderno de la palabra Estadística está relacionado con "el desarrollo y aplicación de métodos y técnicas para recoger datos para que, mediante su análisis, se pueda llegar a la interpretación de un fenómeno ya sea biológico, social como también un proceso de producción, todos ellos a través de conceptos probabilísticos".

a.b) "Como ciencia, la estadística utiliza los resultados y la metodología del cálculo de Probabilidades y tiene por objeto la elaboración de métodos y modelos que permitan la descripción, el análisis, la interpretación y la predicción de fenómenos reales cuando estos son de carácter aleatorio".

a.c) La estadística es una ciencia intelectualmente y prácticamente de la mayor utilidad y aplicación en campos tan diversos como las ciencias naturales y sociales, la tecnología, la política, etc. Precisamente hoy resulta impensable una estadística en cuya elaboración no se haya tomado en cuenta los conocimientos y técnicas de la ESTADÍSTICA.

a.d) Se define a la Estadística como "un campo de ensayo en el que se recogen y analizan los datos con el propósito de sacar conclusiones". Se dice que "proporciona instrumentos para la toma de decisiones cuando prevalecen condiciones de incertidumbre".

.....**b)** Relacione lo que ha interpretado con lo discutido en clase y elabore su propia definición de Estadística.

ESTADÍSTICA DESCRIPTIVA E INFERENCIAL

La **Estadística descriptiva** incluye la obtención, presentación y descripción de datos numéricos y la **Estadística Inferencial** se ocupa de las técnicas para tomar decisiones con base en el análisis de los datos obtenidos.

Así es que, por ejemplo:
Un fabricante de medicinas afirma que una nueva vacuna contra el catarro desarrollada por su compañía tiene una efectividad del 90%; esto es, "en promedio", 90 de cada 100 personas que emplean la vacuna pasarán el invierno sin contagiarse el catarro.

Como resulta imposible al fabricante probar la vacuna en todas las personas, selecciona correctamente 30 personas y les aplica la vacuna. Observa luego que 25 de las 30 personas no se contagiaron el catarro. Esto es, el 83% de las 30 personas no se contagiaron. Ahora bien, toda la información que sobre las 30 personas estudiadas ha obtenido, le permiten "describir" lo sucedido, sin obtener conclusiones más amplias.

Puede escribir la información, ordenarla y graficarla (Estadística Descriptiva)

Por otra parte, el fabricante quisiera saber si el número observado de 25 personas que no se contagiaron (debieron haber sido 27 de las 30),es suficiente para hacer que se rechace su afirmación inicial. Tendrá que encontrar métodos basados en el concepto de probabilidad que le permitan analizar su hipótesis y luego arribar a alguna conclusión general acerca del efecto de su vacuna (Estadística Inferencial).

(2) a) A partir de las siguientes situaciones se han redactado afirmaciones. Identifique las que se realizan en base a métodos descriptivos y las que se derivan de métodos inferenciales.(Tenga cuidado con las expresiones que se emplean. Las usamos en "nuestra vida diaria", pero...)

a.a) Para optar por un préstamo hipotecario para la vivienda se presentaron el día lunes tres familias de la zona oeste y tres de la zona norte de la ciudad. Los ingresos familiares declarados por las flias. fueron 800,450 y 820$ y 730,920 y 450$ respectivamente para ambas zonas.

1) el ingreso familiar promedio de la zona oeste es de $690.
2) el ingreso familiar promedio de las tres familias de la zona oeste es menor que el de las tres familias de la zona norte.
3) el mayor ingreso familiar observado fue de $920 y corresponde a una familia de la zona norte.
4) es posible que los ingresos familiares sean más altos en la zona norte que en la oeste.

a.b) Se tienen datos acerca de la contaminación del aire que fueron proporcionados por 42 de las grandes ciudades del país

1) el número promedio de partículas de materia en suspensión (microgramos por metro cúbico)en las muestras de aire tomadas de las 42 ciudades es 32.
2) resulta evidente que la mitad de las 42 ciudades presentan un promedio mayor a lo que se determinó en un estudio realizado el año anterior.
3) la contaminación subió en promedio 2 puntos en los últimos 3 años.

4) probablemente la contaminación en las ciudades del este sea mayor que en las de la costa del país.
5) si la contaminación subió mas de 3 puntos, se recomendará que salga del Congreso una nueva ley.

.....**b)** De los siguientes enunciados, identifique los problemas que exigen del empleo de la Estadística Descriptiva y los que exigen el método Inferencial.

b.a) (1) un profesor de inglés emplea diferentes métodos con cada uno de sus dos cursos. Compara las calificaciones obtenidas por algunos de sus alumnos con el fin de establecer cuál es el método más efectivo y decidirse por uno de ellos.
(2)el mismo profesor compara las calificaciones obtenidas por sus dos grupos de alumnos a fin de establecer cuál de los dos métodos es el más efectivo.

b.b) Se quiere saber cuáles son los niveles educativos por los que han atravesado las 530 personas de una empresa.

POBLACIÓN Y MUESTRA

En las situaciones presentadas anteriormente hemos utilizado algunos conceptos que es necesario definir.

Población: es un conjunto cuyos elementos tienen una o más características comunes. Este conjunto puede estar formado por elementos animados o no, o por las mediciones que sobre esos elementos se realicen.

Puede tratarse de una población pequeña y finita, grande pero finita o infinita; también puede ser tan grande que se la suele considerar como infinita. Sobre ella se realizan los censos. Sobre ella se realizan los censos.

Por otra parte:

Muestra: es una parte o subconjunto de la población previamente especificada. Sus componentes deben tener la o las mismas características comunes que los de la población de donde ella proviene.

Para que veamos si se han comprendido los conceptos anteriores le proponemos el siguiente ejercicio.

(3) a) A partir de las definiciones anteriores resuelva las siguientes consignas.

a.a) nombrar una población que no esté definida por límites geográficos y que no esté integrada por seres vivos.
a.b) nombrar una población de mediciones.

a.c) ¿puede definir censo?

a.d) ¿qué características comunes le está exigiendo a los elementos o mediciones de los puntos a.a) y a.b)?

a.e) en una de las poblaciones elegidas, agregue una nueva característica a las exigidas anteriormente. ¿la población se agranda o se achica? ¿es siempre así?

.....**b)** suponga que se está interesado en la edad promedio de los estudiantes de segundo año de la Facultad de Ciencias Naturales e Ingeniería y considere las edades de los alumnos en la clase de Estadística como muestra.

b.a)describa cuidadosamente la población bajo consideración y enumere características comunes de los elementos

b.b) describa la muestra (tal como está definida) y enumere las características comunes de los elementos.

.....**c)** Suponga que la clase de Álgebra hay 3 estudiantes extranjeros, todos casados.

¿Es adecuado afirmar que todos los compañeros extranjeros de la carrera están casados? Comente la respuesta.

.....**d)** Analice los siguientes casos:

d.a) en 1983 una revista predijo la elección presidencial por muestreo hecho al azar a partir de guías telefónicas y de una lista de suscriptores a la revista. La predicción fue totalmente equivocada. Explique por qué puede haber sucedido esto.

d.b) durante cierto período de tiempo, un investigador realizó 200 análisis de laboratorio obteniendo la cantidad de glucosa en gramos/litros de "muestras" de sangre de pacientes diabéticos. Explique las circunstancias bajo las cuales estos 200 análisis pueden considerarse:

- ♦ una muestra
- ♦ una población.

d.c) Dado el caso de (3.b). Comente si la muestra es representativa.

PARÁMETRO Y ESTIMADOR

Si le informamos que:
- El "promedio" de materias aprobadas de todos los alumnos de 2do. Año de las carreras de la Facultad de Ciencias Económicas es 3 (tres)

a este valor 3 (tres) se lo considera un **parámetro**.

Por otro lado, si ahora le decimos que:
- El "promedio" de materias aprobadas de un grupo de 25 (veinticinco) alumnos seleccionados entre los alumnos de 2do.Año de las carreras de la Facultad de Ciencias Económicas es 3 (tres)

a este número 3(tres) se lo considera un **estimador**, **estadígrafo** o **estadístico.**

(4) a) ¿Podría decir a qué conjunto identifica o caracteriza un parámetro? ¿y un estimador?

.....**b)** Teniendo en cuenta lo anterior, identifique parámetro (P), estimador (E) y dato (D) en los siguientes enunciados.

b.a) índice de latidos cardíacos medio de 20 personas del club evaluados luego de un ejercicio moderado.

b.b) índice de latidos cardíacos medio de las personas del club evaluados luego de un ejercicio moderado.

b.c) número de personas que vio el programa especial de la televisión emitido el viernes último.

b.d) proporción de personas, de las encuestadas, que vio el programa especial de televisión emitido el viernes último.

b.e) los mejores tres estudiantes del colegio obtuvieron en la última evaluación las siguientes calificaciones: 93,98 y 100 puntos.

b.f) el promedio de los tres mejores estudiantes en la última evaluación fue de 97 puntos.

b.g) la variabilidad de los TRC (tubos de rayos catódicos)de la marca Piripipí es de 2250 horas.

(5) a) Diga cuál es el parámetro y el estimador en el siguiente caso. Analice y discuta

Se sabe que el medicamento estándar usado para tratar a cierta enfermedad ha resultado ser efectivo en un lapso promedio de 3 días. Para evaluar la efectividad de un nuevo medicamento para tratar esa misma enfermedad se ha dispuesto seleccionar 150 personas a las que se dará ese nuevo medicamento.

.....**b)** Dado el siguiente caso responda:

Supóngase que el 60% de todos los socios del club "El Perejil" son varones y el resto mujeres. A partir de una muestra de 50 socios, se encuentra que 25 son varones.

b.a) ¿cuál es la proporción de socios, de la muestra, que son varones?

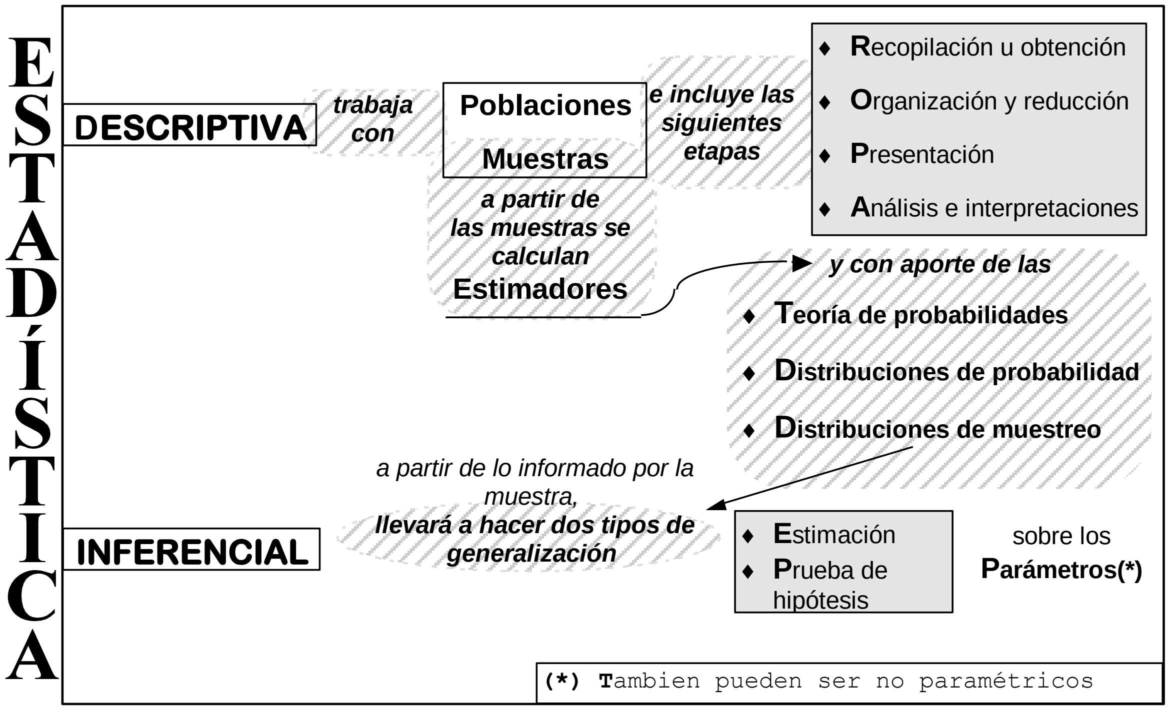
b.b) ¿cuál es la proporción de socios de la población, que son varones?

b.c) ¿cuál es la población? ¿es finita o infinita?

b.d) ¿cuál es el parámetro?

b.e) ¿cuál es el estimador?

.....**c)** Analice el siguiente cuadro.
(Complete el sentido de posibles flechas, si corresponde)



Nos ocuparemos ahora de estudiar mas detalladamente algunos de los conceptos definidos con anterioridad.

MUESTREO

(6) a) En algunas oportunidades,
* cuando la población es **infinita**,
* cuando la población es **homogénea**,
* cuando el **proceso** es **destructivo**,

se utiliza por necesidad el muestreo.

Explique las situaciones planteadas y ejemplifique cada caso.

.....**b)** Tenga en cuenta que el muestreo es:

Aleatorio: si garantiza que todos los elementos de la población tienen la misma posibilidad de integrar la muestra;
Opinático: si el investigador se deja dominar por la propia opinión o intención;
Sin norma: cuando se hace sin regla fija, sin usar un método determinado.

Distinga cada uno de los tipos de muestreo definido anteriormente como correcto o incorrecto. Elabore un ejemplo para cada caso.

LEER: Recurra a la bibliografía señalada para informarse un poco más acerca de qué se entiende por muestreo aleatorio y para reconocer cuáles son los diferentes tipos de muestreo, además de distinguir cuándo se usa cada uno y cuáles son sus características.

.....**c**) Luego de haberse enterado a través de la bibliografía de cuáles son los diferentes tipos de muestreo y de las características de cada uno de ellos (aleatorio simple, estratificado, por conglomerados, sistemático, en dos o más etapas) elabore un ejemplo sencillo para cada uno de ellos

ANALICE LO QUE ENTIENDE A PARTIR DE LAS DISCUSIONES EN CLASE. . .

¿Qué es, entonces, lo que estudia la teoría de probabilidades, que da lugar a esta diversidad de aplicaciones?

Para contestar esta pregunta, comenzaremos por definir una propiedad que tienen en común fenómenos como el de números de individuos que poseen cierta característica genética; el número de llamadas telefónicas que se hace en una ciudad a horas determinadas del día; el nivel de calidad de los artículos manufacturados por cierto proceso; el número de accidentes automovilísticos que ocurren en un día cualquiera en una carretera dada; y muchos otros que son tan diferentes entre sí.

Es frecuente que se pueda considerar cada uno de los anteriores como un **fenómeno aleatorio** en el sentido de la definición siguiente.

> Un fenómeno aleatorio (fortuito o al azar) es un fenómeno empírico que se caracteriza por la propiedad de que, al observarlo bajo determinado conjunto de condiciones, **no siempre se obtiene el mismo resultado** (de manera que **no existe regularidad determinista)** sino que **los diferentes resultados ocurren con regularidad estadística.** Esto quiere decir que existen números entre 0 y 1 que representan la frecuencia relativa con la que se observan los diferentes resultados en una serie de repeticiones independientes del fenómeno.

Dos conceptos estrechamente ligados al de fenómeno aleatorio son los de evento aleatorio y de la probabilidad de un evento aleatorio.

Un evento aleatorio tiene la propiedad de que la frecuencia relativa con la que aparece en una sucesión muy larga de observaciones realizadas al azar, se acerca a un valor límite estable a medida que el número de observaciones tiende a infinito; **el valor límite de la frecuencia relativa se llama probabilidad del evento aleatorio.**

Para arrojar más luz sobre el significado de un evento aleatorio, consideremos uno que sea característico; por ejemplo, un accidente de automóvil. Es evidente que el lugar, el momento y la forma de este accidente depende de un número enorme de factores, y que un ligero cambio en cualquiera de ellos puede alterar considerablemente la naturaleza del accidente, y aun evitarlo por completo. Por ejemplo, si tomamos un choque entre dos autos, vemos que si uno de los dos conductores hubiera iniciado su recorrido diez segundos antes o después, si se hubiera detenido un instante a hacer, por ejemplo, alguna compra si en el camino hubiera aminorado su velocidad para no atropellar a algún animal, o se hubiera alterado su movimiento por cualquier otra razón, no habría ocurrido dicho accidente; quizá una forma ligeramente diferente de maniobrar al momento de la colisión también podría haber evitado el choque o alterar, para bien o para mal, sus consecuencias.

Es imposible predecir que un automovilista en particular que inicia un viaje por carretera se vea o no envuelto en algún accidente. Sin embargo, si observamos a todos los automovilistas (o a un gran número de ellos) que viajan por la misma carretera en un día, es posible determinar la proporción de ellos que sufrirán accidentes. Si esta proporción permanece constante de un día a otro, podemos considerar que lo que acontece a un automovilista en particular en esa carretera es un fenómeno aleatorio y que el accidente es un evento aleatorio.

(7) Le proponemos el siguiente juego mental:
.....**a)** corte 10 cartoncitos del mismo tamaño y enumérelos de 0 a 9.(imagínelo)
.....**b)** colóquelos en una caja ó bolsita, y mézclelos bien. (mentalmente, por supuesto)
.....**c)** elija a un compañero para que extraiga uno de ellos en forma totalmente aleatoria (esto es, <u>sin</u> <u>observar</u> el número del cartoncito que extrae hasta que lo tiene fuera de la caja)
(¿ Qué número extrajo? ¿Lo imaginó?)
.....**d)** Anote el número "observado" y vuelva a colocar el cartoncito en la caja.
.....**e)** repita el procedimiento un gran número de veces
(sea paciente).

Ha obtenido así un "conjunto o listado" de "<u>dígitos aleatorios</u>" de tal forma que, si se repite el procedimiento "muchísimas" veces (¿cuántas serán "muchísimas"?) tendrá los dígitos repetidos aproximadamente el mismo número de veces (en realidad, tal vez usted prefiera usar una calculadora que genere números aleatorios). Este juego lo ha capacitado para construir una tabla de dígitos aleatorios o de números al azar como las que encontrará en las últimas secciones de cualquier libro de Estadística.(Se adjunta copia de una hoja extraída del texto de Guenther, citado en la bibliografía)

TABLA6. Números aleatorios*

07018	31172	12572	23968	55216	85366	56223	09300	94564	18172
52444	65625	97918	46794	62370	59344	20149	17596	51669	47429
72161	57299	87521	44351	99981	55008	93371	60620	66662	27036
17918	75071	91057	46829	47992	26797	64423	42379	91676	75127
13623	76165	43195	50205	75736	77473	07268	31330	07337	55901
27426	97534	89707	97453	90836	78967	00704	85734	21776	85764
96039	21338	88169	69530	53300	29895	71507	28517	77761	17244
68282	98888	25545	69406	29470	46476	54562	79373	72993	98998
54262	21477	33097	48125	92982	98382	11265	25366	06636	25349
66290	27544	72780	91384	47296	54892	59168	83951	91075	04724
53348	39044	04072	62210	01209	43999	54952	68699	31912	09317
34482	42758	40128	48436	30254	50029	19016	56837	05206	33851
99268	98715	07545	27317	52459	75366	43688	27460	65145	65429
95342	97178	10401	31615	95784	77026	33087	65961	10056	72834
38556	60373	77935	64608	28949	94764	45312	71171	15400	72182
39159	04795	51163	84475	60722	35268	05044	56420	39214	89822
41786	18169	96649	92406	42773	23672	37333	85734	99886	81200
95627	30768	30607	89023	60730	31519	53462	90489	81693	17849
98738	15548	42263	79489	85118	97073	01574	57310	59375	54417
75214	61575	27805	21930	94726	39454	19616	72239	93791	22610
73904	89123	19271	15792	72675	62175	48746	56084	54029	22296
33329	08896	94662	05781	59187	53284	28024	45421	37956	14252
66364	94799	62211	37539	80172	43269	91133	05562	82385	91760
68349	16984	86532	96186	53891	48268	82821	19526	63257	14288
19193	99621	66899	12351	72438	99839	24228	32079	53517	18558
49017	23489	19172	80439	76263	98918	59330	20121	89779	58862
76941	77008	27646	82072	28048	41589	70883	72035	81800	50296
55430	25875	26446	25738	32962	24266	26814	01194	48587	93319
33023	26895	65304	34978	43053	28951	22676	05303	39725	60054
87337	74487	83196	61939	05045	20405	69324	80823	20905	68727
81773	36773	21247	54735	68996	16937	18134	51873	10973	77090
74279	85087	94186	67793	18178	82224	17069	87880	54945	73489
34968	76028	54285	90845	35464	68076	15868	70063	26794	81386
99696	78454	21700	12301	88832	96796	59341	16136	01803	17537
55282	61051	97260	89829	69121	86547	62195	72492	33536	60137
31337	83886	72886	42598	05464	88071	92209	50728	67442	47529
94128	97990	58609	20002	76530	81981	30999	50147	93941	80754
06511	48241	49521	64568	69459	95079	42588	98590	12829	64366
69981	03469	56128	80405	97485	88251	76708	09558	86759	15065
23701	56612	86307	02364	88677	17192	23082	00728	78660	74196
09237	24607	12817	98120	30937	70666	76059	44446	94188	14060
11007	45461	24725	02877	74667	18427	45658	40044	59484	59966
60622	78444	39582	91930	97948	13221	99234	99629	22430	49247
79973	43668	19599	30021	68572	31816	63033	14597	28953	21162
71080	71367	23485	82364	30321	42982	74427	25625	74309	15855
09923	26729	14573	16583	37689	06703	21846	78329	98578	25447
63094	72826	65558	22616	33472	67515	75585	90005	19747	08865
19806	42212	41268	84923	21002	30588	40676	94961	31154	83133
17295	74244	43088	27056	86338	47331	9737	83735	84058	12382
59338	27190	99302	84020	15425	14748	42380	99376	30496	84523

*The Rand Corporation, A *Mlillion Random Digits with 100,000 Deviates,* The Free Press, 1955.

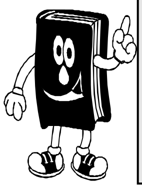
(8) a) Elabore un ejemplo para ilustrar cada uno de los siguientes casos.

a.a) una población finita.
a.b) una población infinita.
a.c) una población de mediciones.
a.d) un parámetro de cierta población.
a.e) un estimador de cierta muestra.
a.f) una muestra representativa de una población.
a.g) una muestra no representativa de una población.

.....b) Explique por qué los siguientes procedimientos podrían no dar por resultado muestras aleatorias de las poblaciones correspondientes.

b.a) para determinar los gastos promedio de una persona durante las vacaciones, un investigador entrevista a los diputados y senadores.

b.b) para determinar la opinión de los alumnos respecto a la implantación de las correlatividades, un investigador pregunta a los entrevistados: ¿creen que esta medida injusta debería ser anulada?



NOTA : en el "Bloque Anexo" tiene un problema interesante para aplicar los distintos tipos de muestreo. No es obligación resolverlo y no es demasiado simple...pero, tal vez le interese el desafío...

EXPERIMENTO

En este momento es necesario definir:

Experimento: es una acción mediante la cuál se obtiene un resultado y que implica la observación de éste.
Experimento aleatorio: es aquel en que el resultado no se puede predecir con exactitud.
Experimento determinista: es aquel en el que el resultado puede predecirse con total exactitud.

(9) Dé ejemplos de experimentos aleatorios y deterministas.

VARIABLES - ESCALAS

Recuerde que los resultados o datos que se obtienen de un experimento pueden ser caracteres cualitativos o cuantitativos.

- En el caso de los **caracteres cualitativos,** los elementos de que consta el material en estudio se distinguen por alguna cualidad o atributo.
- En los **caracteres cuantitativos,** los elementos se distinguen por una medida o por el conteo. En ellos la variación puede ser continua o discreta.
- Una cantidad que varía de elemento a elemento estudiado se llama **variable.**
- En una **variación** continua la variable puede tomar cualquier valor comprendido en la amplitud de variación, al menos teóricamente.
- En una **variación discreta o discontinua** la variable sólo puede tomar ciertos valores en su amplitud de variación, generalmente enteros.
- Una **variable aleatoria** es la cualidad, medida o conteo, resultado de un experimento aleatorio.
- Una **observación** es un valor o dato observado de la variable aleatoria; $x_1, x_2, \ldots, x_n$ son los valores ó datos de la variable aleatoria, x_n es el valor ó dato de la n-ésima observación.
- La variable aleatoria "X" puede tomar valores en un conjunto determinado llamado **dominio** de la variable.
- **Variable aleatoria cualitativa** es aquella que puede tomar valores que son modalidades del atributo o carácter.
- **Variable aleatoria continua** es la que puede tomar cualquier valor del intervalo de la definición, lo puede corresponder cualquier número real.
- **Variable aleatoria discontinua o discreta** es la que sólo puede tomar algunos valores del intervalo en el que está definida.

Teniendo en cuenta que en todo experimento se obtienen resultados, se ha elaborado el siguiente cuadro:

RESULTADOS

	Clasificación		Escalas
Constantes			
Variables	No numéricas, de atributos o cualitativas		Nominales Ordinales
	Numéricas o cuantitativas	Discretas	Cardinales
		Contínuas	Intervalos (card.) Cocientes o razones (card.)

(10) Identifique si los siguientes enunciados corresponden a una constante (**C**) o una variable (**V**).

- ☐ - número de camas de la sala A del Hospital Zonal.
- ☐ - número de camas ocupadas diariamente durante el mes de Agosto en la sala A del Hospital Zonal.
- ☐ - color de los ojos de los alumnos de una escuela elemental.
- ☐ - estado civil de los 50 empleados ocupados en la Secretaría de Relaciones.
- ☐ - distancia entre el Club "El Perejil" y la Municipalidad
- ☐ - edad mínima exigida para comenzar la escuela primaria
- ☐ - cantidad de glóbulos rojos en un mm^3 de sangre.
- ☐ - milímetros de precipitación pluvial mensual en una ciudad durante un año.
- ☐ - velocidad de un móvil en millas por hora.
- ☐ - curso del secundario al que asisten los hijos de los empleados de la empresa "Manito".
- ☐ - temperatura registrada cada media hora en cierto observatorio.

(11) a) Recuerde la diferencia entre número y numeral, y su relación.

El "10" puede ser la edad de un niño, la ubicación de mi casa en la manzana, el dibujo que aparece en la camiseta de fútbol, el canal de televisión, etc.

.....**b)** ¿Puede agregar otros "usos" del 10?

.....**c)** Identifique como número o numeral.

c.a) el número de la casilla de correo.
c.b) el precio de un remedio determinado.
c.c) el número de serie del motor del auto.
c.d) el número del teléfono.

.....**d)** Retome la actividad (10) y a las variables indicadas clasifíquelas en numéricas y numerales.

.....**e)** A las variables numéricas del ítem d, clasifíquelas en discretas y continuas.

.....**f)** Los datos que se obtienen al realizar un experimento ¿son discretos o continuos? ¿por qué? Ejemplifique.

.....**g)** Busque un ejemplo para demostrar que " no toda variable discreta toma siempre valores numéricos enteros" (Nosotros conocemos sólo un ejemplo y para encontrarlo hay que empezar por los pies) (hasta 1992 ya conocimos dos , y por eso le preguntamos si recuerda sus "faltas")

(12) ¿Recuerda para qué se usan los numerales y cómo se llaman en cada caso? Elabore un ejemplo para cada caso.

(13) a) ¿qué escalas existen y qué relaciones se establecen en cada una? Enumérelas y descríbalas.

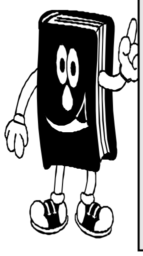
.....**b)** ¿Qué entiende por Cero arbitrario y Cero real? ¿en qué escalas se emplean?. Recuerde algún ejemplo.

.....**c)** ¿Cuándo una escala se denomina discreta o discontinua? ¿Cuál es su característica básica? Dé un ejemplo. Compare con una escala continua.

.....**d)** Enumere ejemplos de datos tomados en escala:

1) **N**ominal 2) **O**rdinal 3) **C**ardinal

.....**e)** Dé ejemplos de datos que corresponda ubicar en escalas continuas y otros en escalas discontinuas.
.....**f)** Retome la actividad 10 y complete indicando cuál es la escala correspondiente en cada caso.



NOTA: si le interesa conocer, (y criticar, si no está de acuerdo) algunos ejemplos de ***variables discretas especiales***,..., lea las propuestas que se presentaron en un curso de estadística, en el tercer ciclo de E.G.B. actual, en una Institución de Comodoro Rivadavia, en 1998.

PASOS EN LA TAREA DESCRIPTIVA

El siguiente enumerado de ítems puede ser una buena guía a tener en cuenta con la finalidad de cumplir en forma adecuada con los grandes cuatro pasos que incluye la tarea descriptiva

1. Recopilación u obtención.
   - 1.1 plan de recopilación de datos
   - 1.2 recogida material de los datos
   - 1.3 recuento de datos
     - enumeración
     - clasificación
     - operaciones materiales de recuento

2. Sistematización de datos u organización.
   - cuadros estadísticos simples y complejos
   - tablas (distribuciones de frecuencia)

3. Representación de los datos.
   - representación gráfica
   - utilidad e inconvenientes
   - errores en las representaciones

4. Descripción de fenómenos estadísticos.
   - medidas de posición y dispersión (concepto, propiedades, usos)
   - proporciones, índices, etc.

Se ve entonces que en primer lugar se debe contar con un plan de recopilación de datos. Esto es, decidir si trabajaremos con población o muestra; si se trabaja con una muestra, determinar su tamaño y el método o tipo de muestreo a usar; analizar cuáles son los medios reales físicos, económicos con que contamos, etc.

Ya determinado el plan, se recoge la información. Una vez que se ha obtenido la muestra (o bien se han tenido en cuenta los elementos de la población, si se decidió hacer un censo) y se han realizado las observaciones (por conteo o medición) de todas las unidades consideradas, sólo se tiene un conjunto de datos "en bruto" que, generalmente, no da idea concreta del "total" analizado. Esto es, se tienen datos aún sin organizar ni tabular y ello no permite obtener y/o dar alguna información sobre el tema en estudio, no es posible aún identificar algún o algunos rasgos de modo tal de obtener conclusiones válidas.

Se debe entonces organizar y resumir los datos en alguna forma.

ORGANIZACIÓN Y RESUMEN DE LA INFORMACIÓN

Se puede hacer una "combinación ordenada" o "arreglo", colocando las observaciones en orden a su magnitud, ascendente o descendente. Es interesante observar una combinación ordenada sobre la recta numérica real. Suelen detectarse situaciones atípicas.

Es posible algún "resumen", que es la condensación de varias observaciones en forma gráfica o numérica.

Se puede asimismo organizar los datos en un arreglo "tronco-hojas" o "tallos-hojas"

Veamos estos temas a partir de un problema.

ACTIVIDAD PARA IR RESOLVIENDO (semi – resuelta)

En un puesto caminero instalado en la entrada de un pueblo hay una curva; allí se mide la velocidad de 50 automóviles. Las velocidades obtenidas están volcadas en el siguiente cuadro:

45	61	50	65	70	60	95	65	60	65
48	64	54	65	67	64	56	60	61	62
58	62	57	75	53	59	56	54	67	68
61	60	63	56	53	62	69	70	44	47
55	65	56	57	58	51	43	74	72	48

En este listado los valores observados son:

$x_1 = 45$, $x_2 = 48$, $x_3 = 58$, ... , $x_{13} = 57$, ..., = 48

Como primer paso, útil para continuar con mayor facilidad, organizamos los datos en **combinación ordenada**.

Para ello buscamos el menor valor observado, que corresponde en este caso a la trigésima quinta observación (x_{35}=43) y el mayor valor observado que corresponde a la trigésima primer observación (x_{31}= 95). Estas pasarán a ser las observaciones $x_{(1)}$ y $x_{(50)}$ respectivamente. Ordenaremos entre ellas las observaciones restantes, obteniendo el siguiente cuadro:

43	48	54	56	58	60	62	65	67	70
44	50	54	56	59	61	62	65	67	72
45	51	55	57	60	61	63	65	68	74
47	53	56	57	60	61	64	65	69	75
48	53	56	58	60	62	64	65	70	95

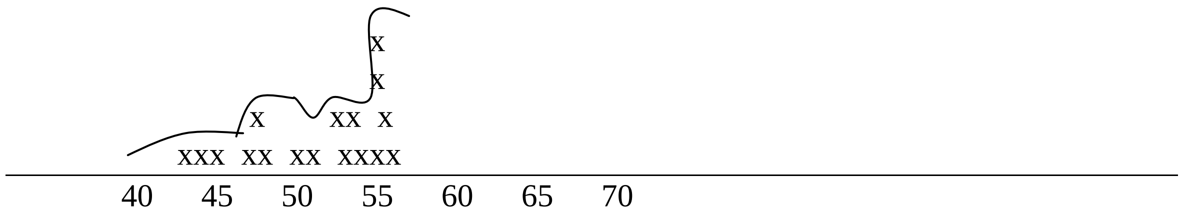
En este ordenamiento, los valores aparecen ordenados de menor a mayor y los identificamos como

$x_{(1)}$= 43, ,$x_{(2)}$ = 44, ...,
$x_{(27)}$= 61, ... , $x_{(50)}$= 95.

Para visualizar la información que nos da el cuadro, se puede representar los valores observados sobre la recta real, ubicando los datos como pequeñas cruces sobre los valores correspondientes.

En nuestro ejemplo tenemos una observación mínima apenas mayor a 40 y una observación máxima menor a 100. Tomemos entonces una escala de 40 a 100, cada 5 unidades.(Es más rápido y fácil)

Dibujemos una cruz por cada valor, "acomodándola" en forma rápida, tal como se ha comenzado a hacer. Continuemos hasta ubicar la cruz correspondiente al mayor de los valores de la variable. Luego completemos el trazo por encima de la combinación ordenada sobre la recta real.



(14) a) ¿Se atreve a analizar la información? Le ayudamos con los siguientes interrogantes:

-¿entre qué valores se encuentra la "mayor parte" de las observaciones?

-¿cuál o cuáles valores se repiten más veces?

-¿observa algo raro?

.-¿cree necesario revisar la forma de obtener la información?

-¿se producen aberturas o huecos en el conjunto de datos?

-¿es simétrica la distribución?

También se pueden organizar los datos en un **arreglo tronco-hojas** (o tallo-hojas). La ventaja de éste es que no se pierde la información original y a su vez podemos obtener información sobre la distribución de los datos en el momento de la recolección de los mismos. Para realizar el arreglo procedemos así: trazamos una línea vertical y a su izquierda ubicamos los troncos (o tallos) y a su derecha las hojas. En nuestro ejemplo tomamos como troncos las decenas y como hojas las unidades. Como cada tallo se coloca una sola vez, se dice que se tienen tallos simples y así obtenemos:

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

9	5

.....**b)** Compare este diagrama con la combinación ordenada sobre la recta real. ¿qué sucede con la concentración de los datos? ¿Es mayor o menor que la que se observaba sobre la recta real?

Si se quisiera tener los mismos datos pero con una mayor dispersión, se subdivide cada tallo en dos y entonces se realiza un diagrama de tallos y hojas con tallos dobles, es decir que cada tallo se coloca dos veces. Para distinguir un tallo de otro le colocamos un símbolo que lo identifique, por ejemplo:

* para las hojas 0, 1, 2, 3 y 4
. para las hojas 5, 6, 7, 8 y 9

.....**c)** Complete el diagrama con tallos dobles para nuestro ejemplo.

4.	3 4
4*	5 8 7 8
5.	
5*	
6.	
6*	
7.	
7*	
8.	
8*	
9.	
9*	

También suele presentarse este diagrama ordenando las hojas correspondientes a cada tallo, por ejemplo:

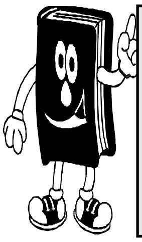
.....**d)** A partir del diagrama del ítem anterior, complete el diagrama de tallo-hojas ordenado que se agrega a la derecha

.....**e)** ¿qué diagrama se parece más a la combinación ordenada sobre la recta real? ¿por qué es así?

4*	3 4
4.	5 7 8 8
5*	
5.	
6*	
6.	
7*	
7.	
8*	
8.	
9*	
9.	

e.a) Compare el primer diagrama de tallo-hojas, con el que usted completó en el ítem c). ¿Cuál da mejor información sobre la distribución?

e.b) Ahora que conoce algunas técnicas para "ordenar" la información numérica, ¿cree usted que si cambia el procedimiento con que hemos resuelto este ejemplo puede agilizar algún paso de la tarea? ¿Qué pasos seguiría?



NOTA: este problema referido a la velocidad de loss vehículos, será retomado más adelante, en el "Bloquee Anexo", para analizar salidas de computadora con , distribuciones de frecuencia, medidas de posición yy dispersión, diagramas de frecuencia y box-plot.

DIAGRAMA DE PUNTOS

Proporciona una forma rápida de visualizar la información, siempre que no se tenga un gran número de datos. Es prácticamente idéntico al mencionado como “representación sobre el eje real”, aquella vez utilizando cruces.

Es muy buena idea graficar los datos de la forma más simple posible y detectar características anormales, si las hay.

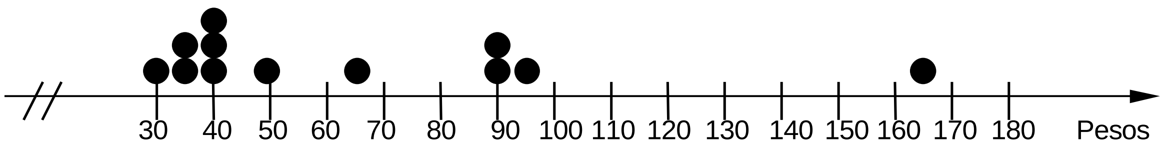
Consiste en representar cada observación con un punto “λ” (lleno, vacío o con otro símbolo) sobre un eje en el que se ha tomado el rango de la variable. Se trata de una representación aproximada que nos muestra la situación general de las observaciones y la dispersión de las mismas.

EJEMPLO PARA ANALIZAR (resuelto)

(15) El grupo Promo de jóvenes de 4^to^ año ha comenzado a realizar actividades a fin de recaudar dinero para su próximo viaje de egreso. En las últimas 12 semanas lograron (en pesos):

165	90	35	40
95	50	40	35
90	40	65	30

El diagrama de puntos es



y en él se advierten algunos datos “apiñados” y unos pocos “alejados”. Estos últimos se deben posiblemente a los montos recaudados a principio de mes, luego del cobro de sueldos. se ve también un valor “raro”, muy diferente a los demás.

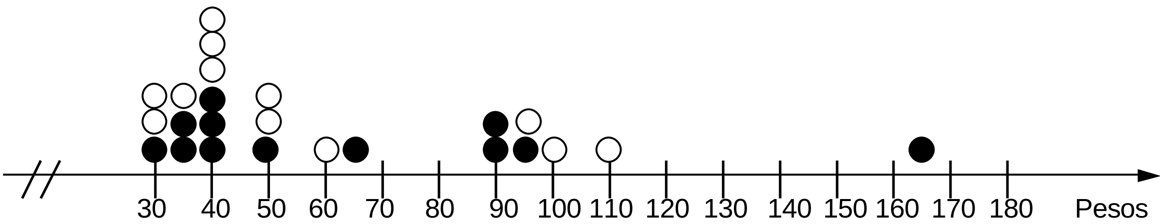
A valores atípicos, diferentes al resto de la información, se los suele llamar “outliers”.

Es necesario recordar que los “alejamientos” deben ser analizados, prestándoles una mayor atención, con la finalidad de detectar errores.

Continuando con el ejemplo, supongamos que "Los Duendes", otro grupo de adolescentes que también está trabajando por su viaje de egreso, logró en las 12 semanas los siguientes montos:

100	60	35	30
110	50	50	40
95	30	40	40

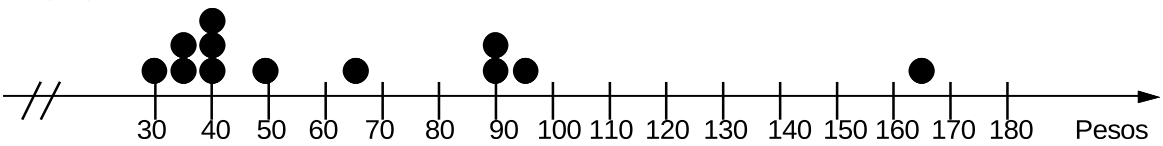
Si superponemos la información de los Promo y Los Duendes, la representación de puntos se verá así:



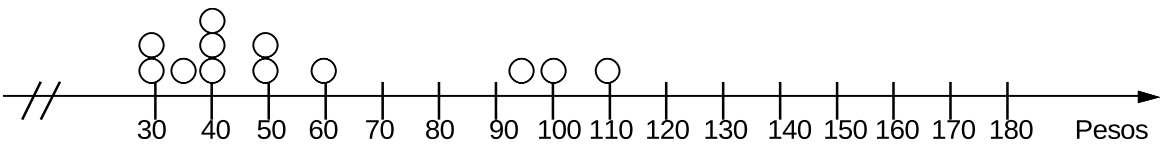
donde "λ"es un monto recaudado por los Promo y "μ"es un monto recaudado por Los Duendes.

También podrían presentarse los datos así:

Promo



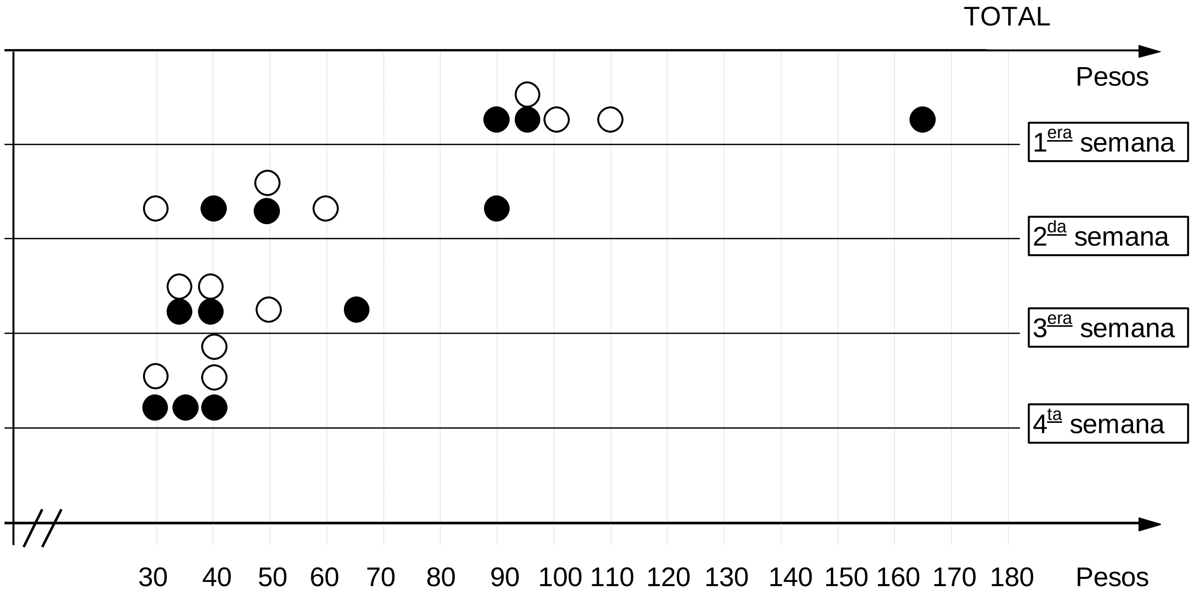
Los Duendes



Parecería, observando las distribuciones de puntos, que los Promo han tenido más éxito, o al menos en una de las semanas han recaudado un monto significativamente mayor y que el resto de los montos están más apiñados o concentrados. ¿Es así? Analice las distribuciones.

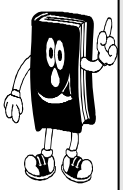
Se nos ocurre otra representación a través de puntos, considerando también por separado a las cuatro semanas de cada uno de los tres meses en que los chicos trabajaron.

Observemos e interpretemos.



Todas estas técnicas rápidas que permiten visualizar la distribución de la información integran el llamado **Análisis Exploratorio de datos.**

Se incluyen el diagrama sobre el eje real, el diagrama de tallo-hojas, el diagrama de puntos y otros que veremos más adelante como el diagrama de caja y el box-plot.



NOTA: Recuerde que tendrá más elementos para incrementar su dominio del tema. Contará con ejemplos para resolver, con salidas de computadora, con las respuestas a sus consultas, con la bibliografía,...

-TABLAS DE FRECUENCIA

Una forma más precisa de ordenar, sistematizar los datos es disponerlos en **tablas o cuadros** de tal forma que sea práctico y sencillo usarlos para obtener conclusiones.

Existen cuadros estadísticos generales o de referencia y resumen o de texto.

En general, un cuadro es un arreglo de filas y columnas.

En este momento sólo presentaremos un tipo muy especial de tablas estadísticas, dejando las otras representaciones tabulares y gráficas para la última parte del curso.

Existen las denominadas **tablas de frecuencia**, que son aquellas que se usan justamente para "tener idea de la distribución de las frecuencias con que aparecen (o se dan) los diferentes valores (o grupos de valores) de la variable".

Para construir o interpretar tablas de frecuencia se deben tener claros ciertos conceptos.

Estos son:

- título
- clasificación
- variable, definición, clases, valores de la variable
- escalas, intervalos de clase, tamaño o amplitud, extremo o frontera y límite, centro o marca
- frecuencia absoluta
- frecuencia relativa
- frecuencia acumulada
- frecuencia relativa acumulada

A los conceptos anteriores puede intentar comprenderlos leyendo el tema en algún libro de los mencionados en la bibliografía o cualquier otro texto. Entonces estará en mejores condiciones para participar en el desarrollo de la siguiente actividad.

ACTIVIDAD PARA IR RESOLVIENDO (semi – resuelta)

(16) ***Se tienen los siguientes datos de 15 alumnos de 5to. año de una Escuela Técnica, obtenidos en los dos primeros meses de clase, por un grupo de docentes que quisieron tener un diagnóstico actualizado, a fin de ajustar objetivos y planificaciones.***

Alumno	**E**valuac.gral cuerpo doc.	**N°** de faltas	**E**dad (en meses)	**C**alificación. promedio de las asignaturas
1	**A**	1	216	75
2	**D**	0	219	62
3	**I**	4½	219	79
4	**A**	0	214	68
5	**A**	0	219	47
6	**I**	7	223	58
7	**M**	0	217	81
8	**D**	5	214	73
9	**I**	3½	217	84
10	**A**	3½	217	53
11	**M**	1	223	57
12	**M**	5	219	38
13	**D**	3½	216	65
14	**D**	1	223	58
15	**D**	0	219	53

La evaluación general del grupo de docentes surgió de una reunión en la cual subjetivamente se opinó acerca de la dedicación puesta de manifiesto por cada alumno.

Se usan los siguientes códigos:

M: molesto (se niega a participar, molesto en clase)

D: desinteresado (no se interesa, pero no se niega)

I: inconstante (a veces participa y se interesa)

A: aplicado (se interesa y muestra preocupación)

Vemos que el equipo docente ha tenido en cuenta cuatro variables:

.....**a)** ¿Puede usted identificar cada variable, clasificarla y determinar su **rango** de variación?

Llámelas X_i, Y_i, W_i, Z_i, en forma ordenada de izquierda a derecha. Así podremos continuar resolviendo el ejemplo juntos.

Una vez que haya definido las variables, comenzaremos por su análisis individual, pues seguramente usted admitió que las hay de distinto tipo.

Si **consideramos X_i,** vemos que ésta se presenta en cuatro **clases** o **categorías** que podemos identificar desde C_1 hasta C_k...(k= 4)

.....**b)** ¿Cree usted que esas clases deben ordenarse de alguna manera especial? ¿Por qué?

Elijamos, sólo para ponernos de acuerdo, la forma de ordenar que nos lleva de la situación peor a la mejor; y entonces tendremos que C_1 corresponde a la clasificación A (aplicado).

.....**c)** Continúe usted con los tres restantes.
- ¿Cuánto vale k y qué significa?, ¿A qué clase corresponde k=3?
- Observe en el cuadro cuál es la clase que más abunda.
- ¿Cuántos alumnos han recibido esa clasificación?
- ¿Cómo son las clases restantes? ¿cuál es la proporción de A en el total?

Si al resultado que obtuvo lo multiplica por 100, seguramente verá que se trata de un 26,6%.

Toda la información que usted está obteniendo con sólo estudiar el listado de calificaciones se puede conseguir de una tabla en la cual puede encontrar los datos de acuerdo con las clases o categorías que hemos tenido en cuenta. Para ello construimos una tabla de frecuencias.

.....**d)** ¿Podría completarla? Nosotros le hemos ayudado a tabular hasta el quinto alumno inclusive (al que le corresponde x_5 = A, usando palotes. Usualmente cuando se llegan a cinco observaciones de una misma clase se indica así: ~~////~~ , o bien como en el truco:

El objeto de agrupar de a cinco es similar al que tenemos al jugar.

Además de recordar los razonamientos que hemos hecho previamente, tenga en cuenta las siguientes notaciones:

f_i = frecuencia absoluta o frecuencia de clase
h_i = frecuencia relativa $\quad h_i = f_i/n$

En nuestro ejemplo

$$\sum_{i=1}^{k=4} f_i = 15 \qquad \text{y} \qquad \sum_{i=1}^{k=4} h_i = 1$$

Calificación (categoría)	**T**abulación	**N°** de alumnos **f_i**	**N°** de alumnos relativos al total	
A	//	4	4/15 = 0,266	(26,6 %)
I	/	...	...	(...)
D	/	...	...	(...)
M		...	...	(...)
		15		(...)

Ahora le proponemos ...

.....**e)** responder las siguientes preguntas:

1. ¿Cuántos alumnos han recibido la calificación D? ¿Es esta frecuencia absoluta o relativa?

Los docentes aseguraron que casi el 75% de los alumnos debe ser "atraído o recuperado".

2. ¿Considera usted posible esta conclusión? ¿Cómo se la obtuvo?

3. ¿Es correcto asegurar que la calificación más frecuente es la de "desinteresado"? ¿Por qué?

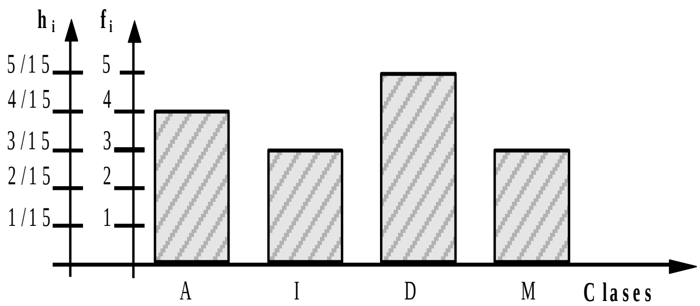
4. ¿Cuál es el menor valor que podría tomar una frecuencia absoluta? ¿Y el mayor? ¿Y la frecuencia relativa?

5. ¿Puede definir ahora en forma general, clase o categoría, frecuencia absoluta y frecuencia relativa, y además concretamente refiriéndose al problema que nos ocupa?

Hágalo de las dos formas.

Resulta también de importancia representar gráficamente la información para extraer conclusiones de los datos analizados.

En el siguiente gráfico, se ha volcado la información de las frecuencias absolutas, pero también, puede ser empleado como gráfico de frecuencias relativas



Comentarios:

Observe que en el gráfico...

* los rectángulos tienen el mismo ancho y se separan por un espacio comprendido entre la mitad y el total de ese ancho.

** se puede cambiar el orden de las calificaciones, por ejemplo utilizando el orden alfabético.

*** ¿cuál es la dimensión del rectángulo que representa a la frecuencia?

.....**f)** Por último le pedimos, para completar el análisis de la variable Xi: característica o calificación otorgada por el equipo docente a cada uno de los quince alumnos tomados para la muestra, que rehaga el gráfico de frecuencias absolutas de manera tal que quien no conoce de Estadística pueda conocer e interpretar la información.

Le sugerimos que elimine iniciales y símbolos.

Consideremos ahora la **variable** $\mathbf{Y_i}$, que usted definió y clasificó en el ítem a).

Construyamos una tabla de frecuencias, como la que aparece a continuación y en la que se han agregado otras dos columnas que no tendremos en cuenta por ahora.

y'_i	.....	.....	.....	F_i	H_i
0			... ( )	5	5/15 ( )
1			... ( )	8	... ( )
3½			... ( )	...	... ( )
4½			... ( )	...	... ( )
5			... ( )	...	... ( )
7			... ( )	15	... (100%)
		15	... ( )		

Con el símbolo y'_i representamos a los valores "distintos" de la variable sin considerar los repetidos, pues al contar cuántas veces se repite cada valor obtenemos la frecuencia absoluta que le corresponde.

Con y_i representamos a cada uno de los quince valores observados de la variable, y con y'_i a cada uno de los seis valores diferentes observados de la variable.

Observe que podríamos definir una variable que depende funcionalmente de las que ya hemos definido y que sería:

$\mathbf{M_i}$ = número de asistencias que el alumno tiene hasta la fecha de iniciado el estudio.

Luego de estas aclaraciones le pedimos que:

.....**g)** Complete los encabezados de las columnas y tabule los datos.

Obtenga y defina las frecuencias absolutas y relativas. Analice la información.

Definimos ahora **frecuencia acumulada** que, como su nombre lo indica es la frecuencia acumulada o sumada **hasta el i-ésimo valor de la variable**, por ejemplo

$$F_2 = F_{(y\ '\leq 1)} = 8 = f_1 + f_2 = .\ \ .\ \ . = \sum_{i=1}^{2} f_i$$

La información que obtenemos es el número de alumnos que tienen hasta una falta.

En general:

$$F_i = f_1 + f_2 + \ldots + f_i$$
$$F_k = n$$

Por otra parte **frecuencia relativa acumulada** es la frecuencia acumulada hasta cierto valor de la variable, pero relativa al total de elementos observados.

Por ejemplo:

$$H_2 = F_2 / n = 8/15 = 0{,}533\ (53{,}3\%)$$

El número encontrado nos indica que más de la mitad de los alumnos no han faltado o lo han hecho una vez.

En general:

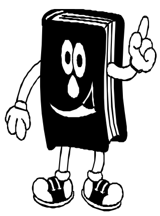
$$H_i = F_i / n$$

Está usted ahora en condiciones de completar las columnas de la última tabla que incorporamos.

.....**h)** Hágalo.

¿Interpreta usted el motivo por el cual no se pueden sumar estas nuevas columnas?

Para representar gráficamente, en este caso, se usan diagramas de barras (o le parece que debería llamarse "de bastones") en los que interesa (al igual que en la tabla) el orden en el que han sido ubicados los valores de la variable.



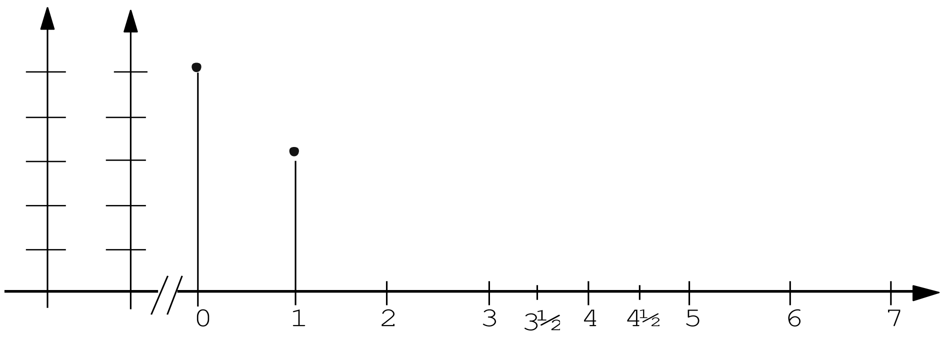
BIBLIOGRAFÍA – NOTA –

Recurra a la bibliografía, a los ejemplos que se agregan en el "Bloque Anexo". Distinga los distintos tipos de representaciones gráficas que se van desarrollando, hasta tener en claro cómo se construyen los gráficos, cómo se leen, cuáles son los errores más comunes que se pueden cometer, etc..

Observe el diagrama que sigue y que apenas está iniciado.
Analice lo que aparece dibujado y . . .

.....**i)** Realice las siguientes actividades:

1. Complete el siguiente diagrama para frecuencias absolutas y relativas.



Observe que el orden de las barras sólo podría invertirse (de mayor a menor números de faltas) pero no intercambiarse.

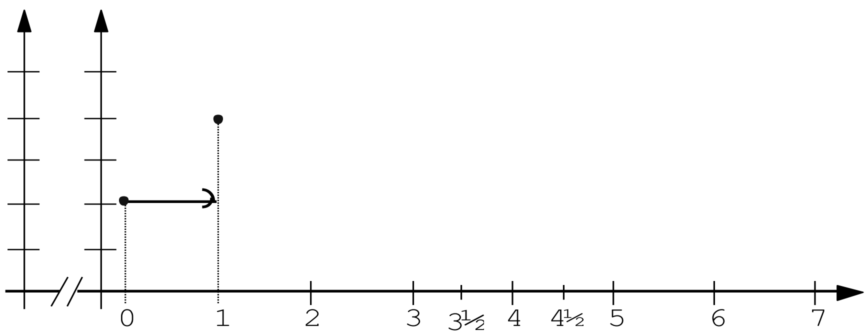
2. ¿Por qué es así?

3. ¿Qué representa la altura de la barra ubicada sobre el valor de la variable y'_2?

4. ¿Podría unir los extremos superiores de las barras para hacer un trazo continuo? ¿Por qué?

5. ¿Por qué cree que hemos cortado el eje vertical a la derecha del origen y no hemos hecho coincidir el origen con el 0 de dicha escala? ¿Siempre será así?

Recuerde ahora cómo se definió a la frecuencia acumulada y observe el diagrama que sigue y que corresponde al "diagrama de frecuencia acumulada y frecuencia relativa acumulada".



.....**j)** Complételo, teniendo en cuenta que hemos dibujado para cada punto una línea punteada sólo para que los puntos (cuya distancia al eje vertical indica la frecuencia acumulada hasta ese valor de la variable) no se pierdan y además para que quede claro que esa frecuencia acumulada se mantiene hasta el próximo valor de la variable que fue observada.

Consideremos ahora la siguiente **variable W_i** ,que usted clasificó como variable . . . . . . ; que toma como valor mínimo a

w_m = . . . y como valor máximo a w_M = . . . , y entre ellos los valores . . . , ... , . . . ; esto es, que toma k=. . .valores diferentes.

Está usted en condiciones de realizar la siguiente actividad:

.....**k)** Construya la tabla de frecuencias similar a la del caso anterior.

Considere la segunda fila y explique qué significa cada número, es decir: interprételа.

.....**l)** ¿Es posible encontrar algún valor de la variable entre 214 meses y 216 meses? ¿Y entre 216 meses y 217 meses?

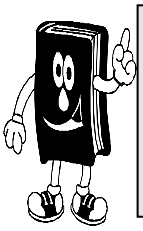
(recuerde de qué tipo de variable se trata)

(piense detenidamente en el hecho de que usted fijó una norma para "discretizar" a la variable, pero...)(¿es meses ó edad?)

Si ahora se quiere graficar la información, ¿cree usted que podrá haber diferencias con el procedimiento que empleó en el caso anterior? ¿por qué ?(le pedimos nuevamente que recuerde qué tipo de variable es W_i)

¿Cómo cree que deberá interpretar la frecuencia? ¿Será ésta una altura o un área? Si no está seguro del procedimiento, déjelo pendiente hasta que analice el próximo caso y luego vuelva hasta acá.

Ahora le pedimos que recuerde lo que se ha estado analizando en el ítem l) anterior y deje pendiente la graficación de la información correspondiente al caso de una variable continua que ha sido tratada considerando los "valores diferentes" que puede tomar.



NOTA: Tenga en cuenta que se volverá a esta variable en el ítem "o", de la página 35, y además cuenta con material para analizar, en la bibliografía citada, y en el elaborado por la cátedra, en las páginas 40, 47, 48.

Ahora, **consideremos la variable Z ,** siendo
z_i = calificación promedio obtenida por el i-ésimo alumno y construyamos la tabla de frecuencias.

Se observa que en este caso sólo el valor 58 se repite en dos oportunidades, es decir tiene frecuencia 2, y entonces no tendremos ventaja alguna si sólo ordenamos de menor a mayor o de mayor a menor, pues deberíamos tener en cuenta catorce diferentes valores de la variable. En este caso resulta conveniente "agrupar" la información y para ello formaremos <u>intervalos de clase</u> y luego deberemos ubicar el palote correspondiente a cada observación en el intervalo adecuado.

Nuestro desafío consiste ahora en decidir cuántos intervalos determinaremos y de qué amplitud.

En este sentido y volviendo al ejemplo, si observamos que las calificaciones promedio toman valores reales desde 38 a 84, podríamos entonces, pensando rápidamente, considerar valores entre 30 y 90, a fin de incluir los valores extremos y de esta forma tomar los intervalos de 30 a 40 puntos, de 40 a 50 puntos, etc. De esta manera quedan determinados seis intervalos. Aclaramos que ésta es sólo una posibilidad de trabajo.

Existe la llamada fórmula de Sturges que nos permite determinar el número de intervalos y consiste en buscar la potencia de 2 más adecuada, a partir de la expresión:

$$2^{k-1} < n \leq 2^k ,$$

donde

k: cantidad de intervalos **n**: total de observaciones

En nuestro caso esta fórmula con n=15 nos daría k=4 intervalos.

Existen otras fórmulas para calcular el número de intervalos que se pueden considerar, pero...todo depende del autor que estemos considerando...

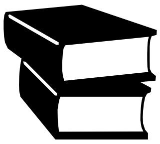
Por ejemplo, otra forma es

$$K = \text{parte entera de } \left[1 + (\log n / \log 2) \right]$$

En realidad, la cantidad de intervalos que es conveniente utilizar estará dada por la experiencia, el conocimiento del problema en estudio, la necesidad o conveniencia de agrupar la información, la cantidad de información que se tiene, el rango de la variable, etc. Como todos estos factores influyen en la determinación del número de intervalos, aceptamos que en realidad no existe una fórmula mágica que lo determine.

Volviendo a nuestro ejemplo, consideremos k=6 intervalos y como nuevo rango al conjunto de valores de la variable entre 30 y 90.(Luego criticaremos esta forma de trabajar)

Para la determinación de los intervalos consideraremos, sin embargo, como valor mínimo..... y como valor máximo.....
¿ Por qué? Si no está seguro, consulte la bibliografía.



LEER: si usted quiere profundizar sobre lo que diferencia a un extremo real o verdadero de uno aparente, busque en la bibliografía sugerida, especialmente en Haber-Runyon.

Le presentamos la siguiente tabla que se ha construido teniendo en cuenta los intervalos posibles y en la que se agregó una columna en la que se indicará con la notación z_i^* al valor medio del intervalo o marca de clase.

Intervalos	**M**arca z_i	**T**abulación f_i	h_i	F_i	H_i
.... -					
.... -					
.... -					
.... -					
.... -					
.... -					

Nótese que el primer intervalo se construye con una amplitud C_1 =..., su límite real inferior es l_1 = ..., su límite real superior es L_1 = ...; su marca de clase o centro es z_1^*=....
En general:

$$z_i^* = \frac{l_i - L_i}{2}$$

Los límites aparentes son... y ...,que son el mínimo y el máximo valor que puede tomar la variable en ese intervalo debido a la precisión con que hemos trabajado.

Recordemos que los límites verdaderos de una variable continua (nro) es igual a ese número(su valor aparente) más o menos una mitad de la unidad de medida.

Observamos que la amplitud de intervalos (C_i) debe coincidir con un múltiplo de la precisión utilizada en la medición de la variable. En el ejemplo que nos ocupa debe ser múltiplo de 1, puesto que tomamos calificaciones promedio que difieren en la unidad.

Por último, al completar la tabulación vemos que se ha perdido la información original, ya que no "tenemos" los valores observados, sino el número de valores que ha correspondido a cada intervalo.

Ahora sí le solicitamos que...

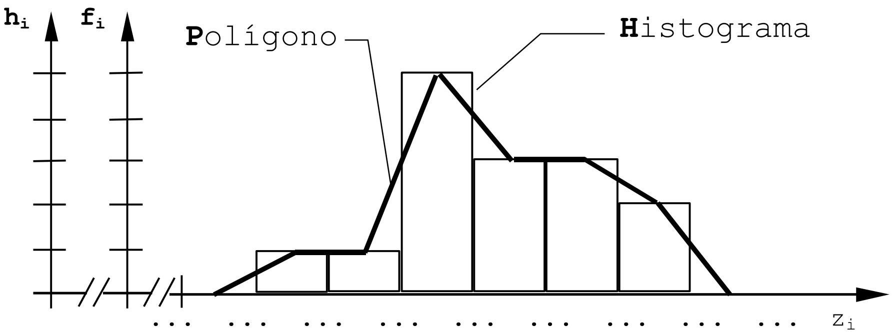
.....**m)** realice lo siguiente:

1. Complete la tabla anterior.

2. ¿Cuál es el valor de la variable que a partir de este momento representa a los valores de la variable que están en el tercer intervalo? ¿Por qué?

¿Cómo pensamos que están distribuidos los valores a lo largo del intervalo?

En el siguiente gráfico se ha construido un histograma de frecuencias absolutas y relativas, como así también un polígono de frecuencias absolutas y relativas.



Observando y analizando le pedimos que:

......**n)** ahora...

1. Complete la representación gráfica y la explique en términos del problema.

2. La frecuencia absoluta del intervalo
¿está representada por la altura o el área del rectángulo? ¿Por qué?

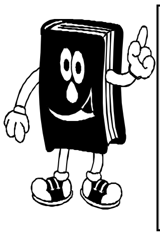
3. ¿Por qué para la construcción del polígono se tienen en cuenta los puntos medios de cada intervalo?

4. ¿Cuánto vale el área total? ¿Por qué?

5. ¿Por qué se han agregado intervalos vacíos a la izquierda y a la derecha?

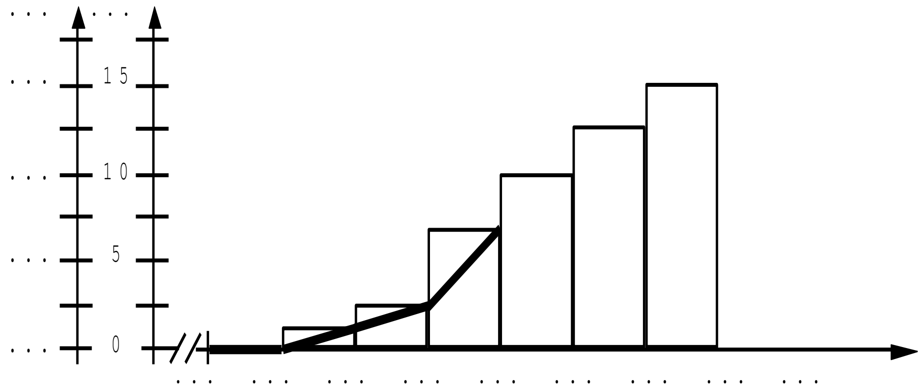
6. Suponga que se nos ocurra considerar a todos los alumnos con calificaciones entre 70,5 y 90,5 puntos. Son cinco,¿ no es cierto ?

¿Será correcto reemplazar los "dos rectángulos" dibujados por uno solo de altura cinco y de base igual a: 90,5 - 70,5? ¿ Por qué? ¿Qué cuidados le sugiere el razonamiento que debe tener en cuenta a fin de "leer" correctamente lo que expresa el diagrama?



NOTA :A partir de la página 48 encontrará sugerencias y ejemplos que le ayudarán a responder éstas y otras preguntas y le aclararán algunas dudas que seguramente tiene sobre la forma de trabajar con intervalos de diferente ancho.

La siguiente representación gráfica corresponde a un histograma y polígono de frecuencias absolutas acumuladas y relativas acumuladas.



.....ñ) Complete la representación gráfica y luego...

1. Analice y compare este último gráfico con el que usó para representar la frecuencia absoluta y relativa.

2. ¿Por qué en esta oportunidad se consideran los extremos derechos de los intervalos y no la marca de clase para la construcción del polígono de frecuencia?

Entonces...¿está correctamente dibujado el polígono ?

3. ¿Qué representa la altura del rectángulo en el histograma de frecuencias acumuladas? ¿O es el área?

Retome el tema de la construcción de los intervalos y de sus límites. Discuta sobre posibilidades diferentes y criterios de selección de alguna de ellas.

Ahora sí es el momento de retomar el caso de la variable edad en meses, ya que antes no la habíamos graficado pero sabemos ahora cómo representar gráficamente una variable continua.

¿Leyó lo explicado a partir de la página 39?

Le pedimos entonces que...

.....**o)** explique (no realice), cómo construiría un gráfico de barras para la frecuencia absoluta y la frecuencia acumulada en ese caso.(Diga qué cambiaría)

Analice y compare con los que ha hecho hasta ahora.

Teniendo en cuenta el estudio que hemos realizado para las variables Y_i, W_i y Z_i , le pedimos que...

.....**p)** responda:

1. ¿Podría tomar intervalos para analizar todas las variables? Fundamente su respuesta.

2. ¿Podría cambiar las representaciones gráficas que hizo para Y_i y hacerlas para Z_i? ¿y para W_i ?

Ya hemos terminado de resolver un ejemplo en el cual trabajamos con variable cualitativa, cuantitativa discreta y continua y definimos, graficamos e interpretamos.

Estamos ahora en condiciones de resolver algunos problemas, lo que ayudará a aclarar o reforzar los conceptos, comparar métodos y/o formas de trabajar, decidir la mejor manera de solucionar una cuestión concreta, ...

PROBLEMAS – PREGUNTAS – EJERCICIOS (a resolver)

(17) A veinte (20) obreros de una empresa se les hizo elegir entre 4 tipos de carne para la comida. Sus elecciones fueron las siguientes:

pollo, pescado, pescado, hígado, pollo, pollo, torta de carne, pollo, pescado, torta de carne, hígado, pescado, torta de carne, pollo, pollo, pollo, hígado, pescado, torta de carne, pollo.

¿Cuál es la variable de interés? ¿de qué tipo es?

Construya una tabla de frecuencia y analice la información.

(18) En un muestreo realizado en una fábrica se determinó el número de elementos defectuosos de cada caja de artículos producidos seleccionados. Se muestrearon 40 cajas y los resultados fueron:

0	6	1	4	1	5	2	5	2	4
3	4	3	3	3	3	3	3	4	1
4	2	5	2	5	1	2	3	6	0
3	3	3	3	2	2	3	3	3	4

1. ¿En qué consiste el experimento?

2. ¿Cuántos experimentos simples forman el experimento completo?

3. ¿Cuál es el objetivo del experimento?

4. ¿Cuál es la variable de interés, de qué tipo es y qué valores puede tomar?

5. ¿Existe alguna otra variable, además de la definida en el ítem anterior? Si la hay diga cuál es y cuál es la relación entre ellas.

6. Construya una tabla de frecuencia (f_i, h_i, F_i, H_i), y comente qué significa cada frecuencia calculada.

7. Elija una fila de la tabla de frecuencias y explique el significado de cada uno de los elementos de la fila, en términos del problema.

8.Represente gráficamente la información e interprete.

9.Explique mediante un ejemplo el significado de una ordenada
.-una diferencia de ordenadas
.-un salto en el gráfico acumulativo

(19) El número de accidentes automovilísticos que ocurrieron en los sesenta cruces más transitados de cierta ciudad en el fin de semana largo de diciembre fueron:

0	2	5	0	1	4	1	0	2	1	5	0	1	3	0	0	2	1	3	1
1	4	0	2	4	1	4	4	0	4	3	5	0	1	3	6	4	2	0	2
0	2	3	0	4	2	5	1	1	2	2	1	6	5	0	3	3	0	0	4

- Responda a los mismos interrogantes que en el ejemplo anterior.

(20) Suponga que tiene 18 observaciones las que varían entre 14,3 y 34,5.

1. determine el número de clases.
2. determine los límites y marcas de clase seleccionadas.

(21) Suponga que tiene 125 observaciones, que toman valores entre 20,4 y 85,7. Determine el número de clases, los límites y las marcas de las mismas.

(22) Las marcas de clase de una distribución de lecturas de temperaturas (dadas al grado celcius más cercano) son 16, 25, 34, 43, 52 y 61.
Calcular:
1. los límites de clase (superior e inferior).
2. formar los intervalos.

(23) Los límites superiores de las cuatro primeras clases de una distribución de frecuencias son:
- 12,2 - 14,9 - 17,6 - 20,3 -
¿Cuáles son los límites inferiores y las marcas de clase?

(24) Los límites inferiores de las 3 primeras clases de una distribución de frecuencias son 31,6 ; 36,1 y 40,6. ¿cuáles son los límites superiores y las marcas de esas clases?

(25) Analice el ejemplo siguiente:
Los siguientes datos corresponden a millas por galón obtenidas en veinte recorridos, en un perímetro citadino con un automóvil mediano.

19,7	21,5	22,5	22,2	22,6

21,9	20,5	19,3	19,9	21,7
22,8	23,2	21,4	20,8	19,4
22,0	23,0	21,1	20,9	21,3

Construya la distribución de frecuencias y opine sobre ella.

Le proponemos ahora una serie de interrogantes que le permitirán autoevaluarse en ciertas cuestiones.

(26)....a) ¿cuáles son las ventajas y desventajas de emplear un arreglo como una descripción de los datos en bruto obtenidos?

....b) si la variable es cualitativa, ¿cuál o cuáles ordenamientos puede emplear? ¿tiene sólo una posibilidad? ¿por qué?

....c) si está en el caso de la pregunta anterior ¿qué tipo de información puede obtener? ¿puede agregar algo más?

....d) ¿puede usar intervalos si la variable es cuantitativa discreta? Compare con el caso de tener variable cuantitativa continua.

....e) ¿puede tener intervalos diferentes? ¿le conviene tenerlos? ¿qué facilidades o dificultades se le presentan?

....f) ¿pensó en qué tipos de problemas se le presentarían si el primero o el último (o ambos) de los intervalos fuera abierto? Piénselo. Ejemplifique.

....g) ¿cuántos datos debe tener como mínimo para construir intervalos, cuantos intervalos y de qué amplitud puede construir? ¿conoce alguna regla fija, algunas consideraciones generales? Recuérdelas.

....h) ¿qué ventajas y desventajas enfrenta si construye intervalos?

(27) La siguiente tabla muestra la distribución de partículas de materias en suspensión (microgramos por metro cúbico) en muestras de aire tomadas en 57 grandes ciudades.

X: microgramos/metro^3 de partículas de materia en suspensión

Inter	% Ciudades
menos de 16	1
16 - 17	6
18 - 19	8
20 - 21	7
22 - 25	12

Inter	% Ciudades
26 - 29	14
30 - 39	19
40 - 49	25
50 o más	8

Esta información fue utilizada por un miembro del Comité de Preservación del Medio Ambiente de la ciudad de Córdoba para asegurar: "es un error creer que las ciudades en las que se presenta un bajo número de partículas en suspensión representan un número relativamente alto dentro de las 57 ciudades estudiadas". El miembro del Comité hizo notar que en las ciudades en las que se presentan 40 y 49 partículas representan un porcentaje considerablemente mayor que en los otros grupos. Por otra parte dijo: "las cifras más frecuentes se encuentran en las ciudades que tienen entre 40 y 49".

¿Está de acuerdo con esta afirmación que realizó el miembro del Comité? Justifique su respuesta.

(28) Se sabe que 6 hornos rotatorios de fabricación de cemento han sido cargados y reglados para producir la misma calidad en todos ellos.

El control de las 6 fabricaciones consiste en la extracción de muestras de probetas de cada horno, para medir la carga de ruptura por tracción, en kg/cm², de estas probetas.
Sabiendo que los resultados fueron

13,1	12,8	12,7	13,6	13,5
11	12,9	13,7	14,7	14,5
12,8	11,5	14,1	14	15
14	12,5	14	14	14,7
10	13,7	14,7	14,1	14,7

analice los datos; opine sobre la distribución.

(29) En una prueba para seleccionar mecánicos para un gran taller se obtuvo la siguiente información:

Obrero	Tiempo (min)
A	35
B	22,5
C	45
D	20
E	45
F	47,5
G	20
H	40
I	20
J	35
K	45
L	23,5
M	25
N	40
Ñ	30
O	47,5

¿Qué puede opinar sobre la distribución de los tiempos que tardan los mecánicos en resolver la prueba?

PASOS PARA ARMAR LOS INTERVALOS DE UNA TABLA DE FRECUENCIAS

El siguiente es un enumerado de pasos y conceptos que seguramente le servirán como una ayuda para construir intervalos, tal como lo hicimos en el desarrollo del ejemplo anterior.

1) Localice los valores mínimo y máximo de la variable y calcule el rango

$$\boxed{R = x_M - x_m}$$

2) Tenga en cuenta que, cualquiera sea el número de intervalos que usted construirá, deberá **ampliar al menos mínimamente ese rango**, para garantizar que los valores mínimo y máximo "pertenezcan" a un intervalo (Si no lo hace, uno de los dos valores se "escapará" de los intervalos).

3) Decida cuántos intervalos va a construir (tenga en cuenta que no existe una receta mágica sino que debe usar su conocimiento del tema y su experiencia)

K = número de intervalos.

4) Comience a "probar" cuál será la amplitud de los intervalos, realizando el cociente entre el "rango mínimamente corregido" y el número de intervalos, hasta encontrar que ese **cociente de un número entero** (si es su objetivo) y se den algunas otras condiciones que a usted le interesen.

Las opciones son $\dfrac{x_M - (x_m - 1)}{K}$; $\dfrac{(x_M + 1) - x_m}{K}$;

o también $\dfrac{(x_M + 1) - (x_m - 1)}{K}$

Esto es, "corrija" uno ó los dos extremos, **ampliando** el rango y logre

$$\boxed{\frac{x_M^* - x_m^*}{K} = a}$$

Donde:

x_M^*: valor máximo corregido ($x_M^* \geq x_M$)
x_m^*: valor m'nimo corregido ($x_m^* \geq x_m$)
K : número de intervalos
a : amplitud de cada intervalo

Notas:

a. Observe que en el ejemplo se suma y/o resta "1". Esto se hace si los valores con los que está trabajando son enteros.
Puede también trabajar con décimos, centésimos, etc.

b. Algunos prefieren que el número "a" sea un número entero e **impar**, para garantizar que el número o valor que está en la mitad del intervalo, que se llama "centro o marca de clase", sea un número entero.

c. Algunos tratan, además, de que la marca de clase sea un número terminado en cero o cinco, para facilitar las operaciones aritméticas que luego se harán con ellos.

5) Tome el valor mínimo x^*_m (ya sea este un valor original o corregido) y construya el primer intervalo

$$\left.\begin{array}{r} x^*_m + a = l_s \\ x^*_m = l_i \end{array}\right\} \qquad l_i \quad ; \quad l_s$$

6) Considere al límite superior del primer intervalo como límite inferior del segundo y repita el procedimiento hasta lograr el último intervalo, cuyo límite superior debe coincidir con

$$x^*_M$$

Al terminar de construir los intervalos de acuerdo con los pasos descriptos, seguramente ha advertido que el valor numérico del límite superior de cada intervalo (excluyendo el último) coincide con el valor numérico del límite inferior del intervalo siguiente.

¿Cómo procederá entonces si entre los valores observados de la variable, alguna coincide con esos valores numéricos?. ¿Dónde ubica ese valor de la variable?.

Puede resolver el problema de diferentes formas.

Si usted sólo modificó el valor mínimo pero no el máximo, esto es si trabaja con x^*_m y $x^*_M = x_M$, entonces deberá considerar cada intervalo semiabierto por la izquierda y hacer $\longrightarrow$ **(** l_i , l_s **]** para cada intervalo.

Ej.: **(** 28 , 33 **]** lo que implica que se incluya en ese intervalo el valor 33 pero no el 28.

Si por el contrario trabaja con x_m y x^*_M , entonces considerará $\longrightarrow$ **[** l_i , l_s **)** para cada intervalo.

Ej.: **[** 28 , 33 **)** lo que implica que se incluye el valor 28 pero no el 33.

Otra posibilidad es tomar el valor numérico de los límites y sumarles (o restarles) una mitad de la unidad de medida.

Ej.: **(**28,5 - 33,5**)** lo que implica que no se incluye el 28 pero si el 33, o bien (27,5 - 32,5) lo que implica que se incluye el 28 y no el 33.

Cualquiera sea la convención seguida para determinar límites de los intervalos, esta debe ser respetada a la larga de todo el análisis descriptivo de la información.

Los ítems mencionados significan (y en algunos casos seremos intencionalmente reiterativos) que

* tanto el x_m como el x_M **deben** pertenecer al primero y último intervalo respectivamente.

* todos los datos **deben** pertenecer a algún intervalo.

* los intervalos **no deben** tener valores comunes (deben ser mutuamente excluyentes).

* **de ser posible los intervalos se construirán de igual longitud**.

* es conveniente que la “marca de clase” o valor representativo del intervalo tenga la mayor cantidad de ceros posibles o esté terminado en cinco.

* el número de intervalos a determinar no será recibido como un “mensaje divino” “ni será” “adivinado graciosamente”; el número dependerá del problema del que se trate y del conocimiento que el “dueño de la información” tenga del tema. (Recordemos que en Estadística **no hay recetas mágicas**).

* es conveniente tratar de que no queden intervalos vacíos, si bien esto no siempre se puede lograr.

EJEMPLO PARA ANALIZAR (resuelto)

En una reunión de profesores jubilados de Estadística, Ana y sus amigos se encontraban comparando alegremente sus experiencias en los exámenes finales de Estadística de sus alumnos universitarios. Cada uno insistía en que sus alumnos eran los que "duraban menos" en los exámenes.

Antes de que la discusión se fuera a las manos, Ana propuso llevar a cabo una estadística macabra: para eso necesitaban tabular los resultados en forma adecuada (resultados obtenidos también en forma adecuada). He aquí los datos del "tiempo necesario para desaprobar a un alumno", eufemismo utilizado por la malvada Ana por "tiempo de duración de los alumnos" (expresión que "no suena" bien, ¿no es cierto?), en minutos:

27	24	25	17	3
43	9	34	48	33
28	28	0	40	27
19	27	15	38	32
45	19	35	32	0
33	21	35	16	13
16	39	13	19	35
43	44	11	11	31

x_m = 0 [min] (éstos duraron "lo que un respiro")
x_M^* = 48 [min] (fue el que opuso tenaz resistencia, aunque ya tenía su destino escrito)

$$\text{Rango} = R = x_M - x_m = 48 - 0 = 48 min$$

Corrijamos este rango agregando una (1) unidad, esto es, un minuto más, al valor de x_M :

$$\text{Rango corregido} = R' = (x_M + 1) - x_m = (48 + 1) - 0 = 49 min$$

Probemos con 10 intervalos: $k = 10$.

Entonces

$$a = \frac{R'}{K} = \frac{49}{K} = \frac{49}{10} = 4{,}9 min$$

No nos conviene demasiado para hacer cálculos →entonces probemos con $K = 7$

$$a = \frac{R'}{K} = \frac{49}{K} = \frac{49}{7} = 7 min$$

Esto significa que podremos trabajar con 7 intervalos de amplitud 7[min.].

Hallemos los límites inferiores y superiores de esos intervalos:

1º intervalo: $l_i = 0\ [min.]$

$l_i + 7\ [min.]$

2º intervalo: $l_i = 7\ [min.]$

$l_s = 7 + 7[min.] = 14\ [min.]$

3º intervalo: $l_i = 14\ [min.]$

$l_s = 21\ [min.]$

4º intervalo: $l_i = 21\ [min.]$

$l_s = 28\ [min.]$

5º intervalo: $l_i = 28\ [min.]$

$l_s = 35\ [min.]$

6º intervalo: $l_i = 35\ [min.]$

$l_s = 42\ [min.]$

7º intervalo: $l_i = 42\ [min.]$

$l_s = 49\ [min.]$

Decidamos ahora qué sucederá con los valores que coincidan con los límites de los intervalos. Como el l_s del 7° intervalo es 49 [min.], está claro que no será un valor observando de la variable, porque x_M = 48 [min.]. Luego decidimos que los intervalos serán de la forma

$$[l_i\ ;\ l_s)$$

Esto significa que el valor "conflictivo" se colocará en el intervalo en que aparezca como límite inferior.

Encolumnemos nuestros intervalos y hallemos su marca de clase o punto medio:

0 – 7	3,5
7 – 14	10,5
14 – 21	17,5
21 – 28	24,5
28 – 35	31,5
35 – 42	38,5
42 – 49	45,5

Esto podría significar un problema, si nos resulta difícil trabajar con decimales (por ejemplo, si no tenemos una calculadora a mano). Si es así, podemos cambiar la amplitud o el número de intervalos, o ambos.

Si ésta es su situación, le sugerimos que corrija nuevamente el rango ya corregido, sumando una nueva unidad; obtendrá así un R′ = 50 min. , y podrá armar 5 intervalos de amplitud 10 min. o bien 10 intervalos de amplitud 5.

¿Con cuál de las opciones solucionará su problema?

Sigamos nosotros con el desarrollo anterior.

Contemos ahora las observaciones que caen en cada intervalo; usemos para ello palotes (o cruces, puntos, asteriscos, ...)

Intervalos	**M**arca de clase	**N**úmero de observaciones
0 – 7	3,5	III
7 – 14	10,5	~~IIII~~ I
14 – 21	17,5	~~IIII~~ III
21 – 28	24,5	~~IIII~~ II
28 – 35	31,5	~~IIII~~ III
35 – 42	38,5	~~IIII~~ II
42 – 49	45,5	~~IIII~~ I

Finalmente, podemos pasar en limpio los resultados obtenidos, con un poco más de formalismo

Límites de Intervalos (min.)	**M**arca de clase (min.)	**F**recuencia
0 – 7	3,5	3
7 – 14	10,5	5
14 – 21	17,5	7
21 – 28	24,5	6
28 – 35	31,5	8
35 – 42	38,5	6
42 – 49	45,5	5
	Total n =	40

Si hubiésemos decidido tomar 5 intervalos de amplitud 10 min., habríamos obtenido:

Límites de Intervalos (min.)	**M**arca de clase (min.)	**F**recuencia
0 a menos de 10	5	4
10 a menos de 20	15	11
20 a menos de 30	25	8
30 a menos de 40	35	11
40 a menos de 50	45	6
	Total n =	40

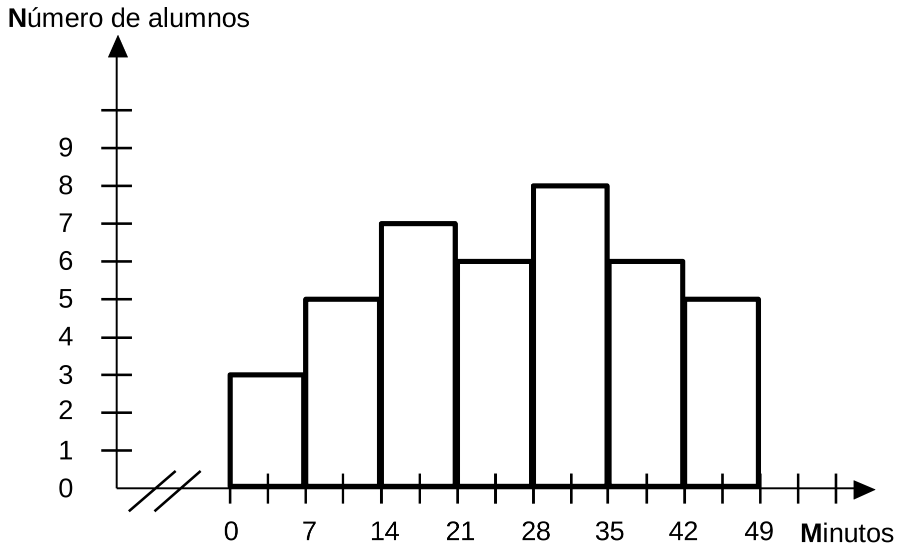
Observemos que al decir "0 a menos 10," 10 a menos 20", etc., estamos usando una expresión equivalente de [0, 10) ; [10, 20) ; etc., y mucho más adecuada para la representación ante personas que no manejen el idioma matemático. Además, nos estamos evitando el trabajo de aclarar con una nota al pie de la tabla la convención usada en el caso de las observaciones "problemáticas": en la primer tabla construida, deberíamos haber escrito al pie algo así como "los valores coincidentes con los límites de un intervalo se tomarán en el intervalo siguiente", o una expresión equivalente.

Observemos también que en las unidades agregadas para corregir el rango inicial sólo podían tomarse al final, y no al principio del conjunto de datos, puesto que no tendría sentido hablar de "alumnos que duraron - 1 min.". En otros problemas podrían tomarse al principio (x_m - 1 unidad , o x_m - 2 unidades), o uno al principio y otro al final (x_m -1 unidad y x_M + 1 unidad).

Finalmente, no debemos olvidar que estamos trabajando con datos reales, que tienen unidades: éstas deben colocarse siempre, para que las interpretaciones puedan hacerse correctamente sin necesidad de buscar más información que la dada por la tabla construida.

A pesar de que los intervalos que construimos tienen amplitud a = 7 min, pensemos por un instante que corresponden a 1 unidad (esto significa que: 7 min equivale a 1 unidad).

Si a cada intervalo (ahora "unitario") lo hacemos corresponder a la base de un rectángulo, y a la frecuencia la hacemos corresponder a la altura del rectángulo, vemos entonces que el área de ese rectángulo considerado tiene el mismo valor numérico que la frecuencia. Con esta transformación de frecuencias en áreas de intervalos de amplitud unitaria, podemos construir el histograma para esos datos:



Las dos líneas paralelas sobre el eje horizontal se colocan para indicar un "corte" en la escala que permita dibujar el histograma adecuadamente (sino habría quedado "pegado" al eje vertical). Si calculáramos el área del histograma, tendríamos:
que es precisamente el número de observaciones con las que contaban Ana y sus colegas al comienzo de este estudio estadístico.

$$\text{Area} = \sum_{i=1}^{7} (\text{base.altura})_i = (1.3) + (1.5) + (1.7) + (1.6) + (1.8) + (1.6) + (1.5) = 40$$

ALGUNAS INDICACIONES ACERCA DE CÓMO ELABORAR UN HISTOGRAMA DE FRECUENCIAS

Como ya se ha visto, el histograma no es más que una representación gráfica de una distribución de frecuencias de una variable cuantitativa continua, consistente en un conjunto de "rectángulos pegados".

El gráfico debe construirse teniendo en claro que **el área de cada rectángulo (y no su altura) representa a la frecuencia absoluta o a la proporción de elementos que pertenecen al intervalo**

El primer paso consiste en elaborar la escala vertical ($\vec{x}$) correspondiente a la variable en estudio.

Si observamos la tabla de frecuencias, los intervalos que hemos construido y por supuesto, tenemos en cuenta los valores mínimo (x_m) y máximo (x_M) de la variable, podemos armarla muy rápidamente, siempre teniendo en cuenta que la escala es numérica y que en consecuencia las distancias iguales deben representar siempre conjuntos iguales de valores.

Estamos suponiendo ahora que según "los pasos para armar los intervalos de una tabla de frecuencias" estos ya están listos y supongamos por el momento que son iguales (se siguieron los pasos del punto anterior).

Debemos ahora construir los rectángulos. La primera intención será, seguramente, igualar en forma directa la frecuencia absoluta o proporción (o porcentaje) correspondiente con la altura del rectángulo.

Es posible que así lo consideremos y que el hecho no sea un error, pero sólo será así si la base de cada uno de los rectángulos, **aún cuando coincida con un conjunto determinado de valores de la variable, sea considerada la unidad**; en este caso, la frecuencia no es realmente la altura, aún cuando la leamos en la escala vertical, sino que la frecuencia es el área del rectángulo.

Recordemos que en el histograma representamos las frecuencias absolutas o porcentajes como si estos se distribuyeran en forma homogénea a lo largo del intervalo considerado. Recordemos también que ésta es la razón por la cual el valor central, que llamamos "centro" o "marca de clase", es el valor de la variable que representa a todos los valores del intervalo.

¿CÓMO PROCEDEMOS SI LAS AMPLITUDES DE LOS INTERVALOS SON DIFERENTES?

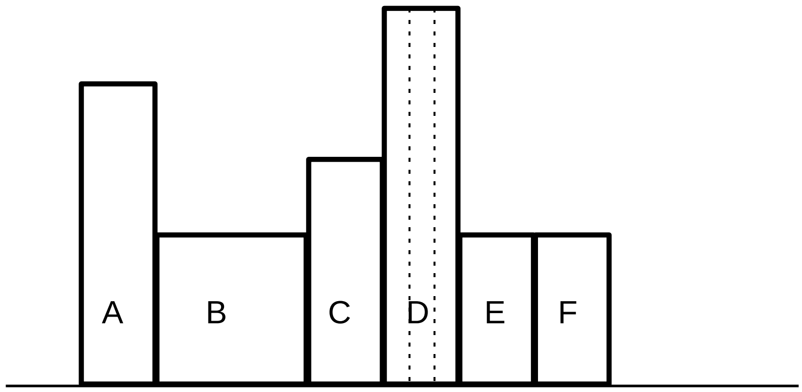
Para construir los intervalos en este caso y luego los rectángulos del histograma, es necesario reiterar que hay que tener en cuenta que la frecuencia del intervalo es el área del rectángulo correspondiente, y no la altura.

Se debe trabajar de modo que el área del rectángulo sea proporcional a la frecuencia del intervalo de clase.

RECUERDE . . . OBSERVE . . . RAZONE . . .
Importa el ÁREA y no la ALTURA . . .

Recordemos también que la diferencia entre un diagrama de barras (separadas) y un histograma (pegadas) es que en el primero, (que se emplea para variable cualitativa o cuantitativa pero en los casos en que los valores de la variable son considerados como “carteles” o rótulos que se diferencian no por su valor sino por su “tipo”) la frecuencia es directamente la altura de las barras. En un histograma, en cambio, por tratarse de variable cuantitativa continua, los rectángulos deben estar pegados y se lee la altura pero de acuerdo con la base del rectángulo considerado.

Si se tiene, por ejemplo



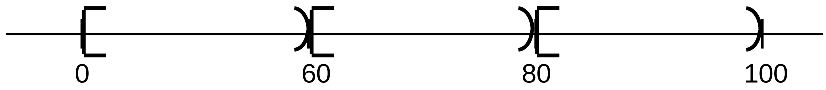
Se puede ver que la cantidad de elementos de A es prácticamente igual a la cantidad de elementos de B, ya que el primer intervalo es el doble de alto pero el segundo tiene una base doble a la anterior. Las áreas son similares por lo que podemos admitir que hay aproximadamente la misma cantidad de elementos en ambos rectángulos.

En el caso D, se observa que cada una de las tres partes tiene 1/3 de los elementos de D, ya que la base de cada rectangulito es la tercera parte del total.

Para tener en claro cómo proceder para construir un histograma cuando los intervalos a considerar son distintos, veamos un ejemplo.

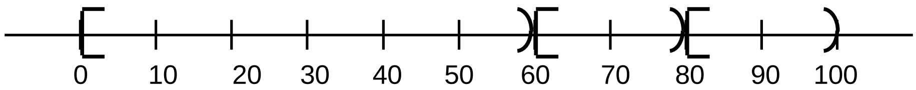
Supongamos que en la clase de Bioestadística, cursada 1997, 30 de los 60 alumnos inscriptos reprobaron, es decir que obtuvieron calificaciones menores a 60 puntos; 28 de los alumnos obtuvieron concepto, con calificación desde 60 a menos de 80 y el resto, los otros 12 alumnos, fueron promovidos sin necesidad de rendir el examen final. Además sabemos que ningún alumno obtuvo la calificación 100.

Si comenzamos a trabajar según lo visto anteriormente y “no pensamos mucho”, dibujaríamos una escala de puntajes como sigue

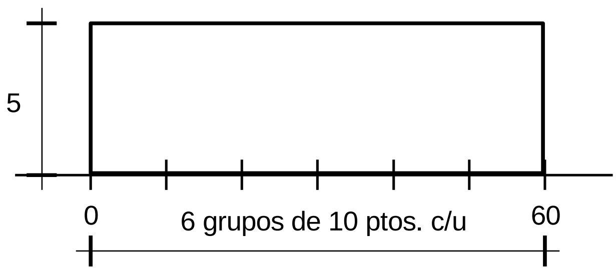


lo que supuesto no es correcto porque los segmentos dibujados son de igual longitud pero los puntajes incluidos no son iguales.

La escala correcta es:

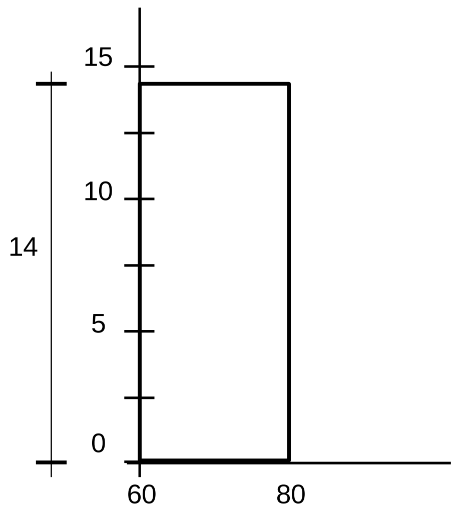


Razonemos ahora que 30 de los 60 alumnos obtuvieron calificaciones entre 0 y 60, pertenecientes a[0, 60), por lo que, si pensamos un poco y recordamos que los elementos se distribuyen en forma homogénea a la largo de cada intervalo, tendremos que si hay 30 elementos sobre 6 intervalos, tendremos 5 en cada uno de los intervalos, cada uno de ellos con 10 puntos de calificaciones como base, esto es

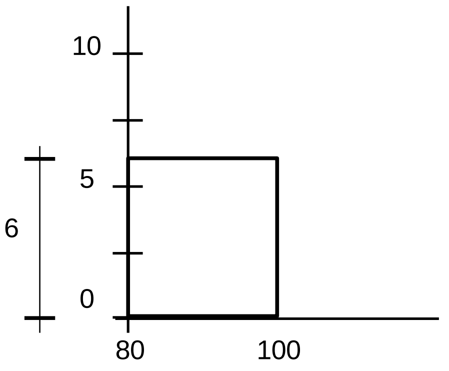


Nótese que se está considerando un conjunto de “10 puntos” de calificación como una unidad.

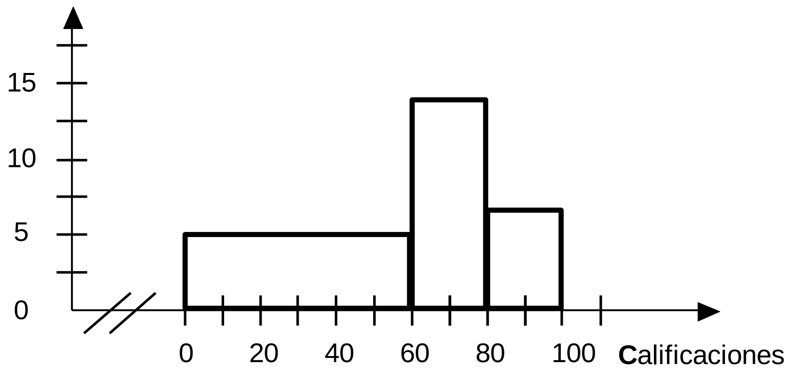
Se tiene además 28 alumnos con calificaciones entre 60 y 80 puntos, por lo que, con el mismo razonamiento, será:



y por último será, para los 12 alumnos con calificaciones de 80 a 100 puntos:



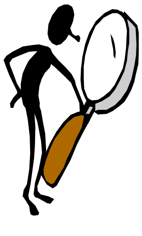
En definitiva, el histograma completo será:



Cuando se interpreta un histograma de estas características es necesario tener mucho cuidado en "leer" la escala vertical. Esta escala puede presentar la frecuencia absoluta o bien el % de casos analizados.

Algunos autores opinan que:
Cuando se trabaja con una serie de frecuencias de variable continua marcadamente asimétrica, con unos pocos valores extremos, conviene construir intervalos cada vez de mayor amplitud cuando se van encontrando valores de la variable con escasa frecuencia, con la finalidad de evitar que aparezcan intervalos vacíos.

¿Usted que opina acerca de esta propuesta?



ATENCIÓN...
Lea los conceptos que sobre estos temas se emplean en LEVIN Y RUBIN - (citado en la bibliografía)
CRITIQUE - Razone y justifique sus críticas.

PASOS PARA ARMAR LOS INTERVALOS DE UNA TABLA DE FRECUENCIAS

El siguiente es un enumerado de pasos y conceptos que seguramente le servirán como una ayuda para construir intervalos, tal como lo hicimos en el desarrollo del ejemplo anterior.

1) Localice los valores mínimo y máximo de la variable y calcule el rango

$$\boxed{R = x_M - x_m}$$

2) Tenga en cuenta que, cualquiera sea el número de intervalos que usted construirá, deberá **ampliar al menos mínimamente ese rango**, para garantizar que los valores mínimo y máximo "pertenezcan" a un intervalo (Si no lo hace, uno de los dos valores se "escapará" de los intervalos).

3) Decida cuántos intervalos va a construir (tenga en cuenta que no existe una receta mágica sino que debe usar su conocimiento del tema y su experiencia)

K = número de intervalos.

4) Comience a "probar" cuál será la amplitud de los intervalos, realizando el cociente entre el "rango mínimamente corregido" y el número de intervalos, hasta encontrar que ese **cociente de un número entero** (si es su objetivo) y se den algunas otras condiciones que a usted le interesen.

Las opciones son $\dfrac{x_M - (x_m - 1)}{K}$; $\dfrac{(x_M + 1) - x_m}{K}$;

o también $\dfrac{(x_M + 1) - (x_m - 1)}{K}$

Esto es, "corrija" uno ó los dos extremos, **ampliando** el rango y logre

$$\boxed{\frac{x_M^* - x_m^*}{K} = a}$$

Donde:

x_M^*: valor máximo corregido ($x_M^* \geq x_M$)
x_m^*: valor m'nimo corregido ($x_m^* \geq x_m$)
K : número de intervalos
a : amplitud de cada intervalo

Notas:

a. Observe que en el ejemplo se suma y/o resta "1". Esto se hace si los valores con los que está trabajando son enteros.
Puede también trabajar con décimos, centésimos, etc.

b. Algunos prefieren que el número "a" sea un número entero e **impar**, para garantizar que el número o valor que está en la mitad del intervalo, que se llama "centro o marca de clase", sea un número entero.

c. Algunos tratan, además, de que la marca de clase sea un número terminado en cero o cinco, para facilitar las operaciones aritméticas que luego se harán con ellos.

5) Tome el valor mínimo x^*_m (ya sea este un valor original o corregido) y construya el primer intervalo

$$\left.\begin{array}{r} x^*_m + a = l_s \\ \\ x^*_m = l_i \end{array}\right\} \qquad l_i \quad ; \quad l_s$$

6) Considere al límite superior del primer intervalo como límite inferior del segundo y repita el procedimiento hasta lograr el último intervalo, cuyo límite superior debe coincidir con

$$x^*_M$$

Al terminar de construir los intervalos de acuerdo con los pasos descriptos, seguramente ha advertido que el valor numérico del límite superior de cada intervalo (excluyendo el último) coincide con el valor numérico del límite inferior del intervalo siguiente.

¿Cómo procederá entonces si entre los valores observados de la variable, alguna coincide con esos valores numéricos?. ¿Dónde ubica ese valor de la variable?.

Puede resolver el problema de diferentes formas.

Si usted sólo modificó el valor mínimo pero no el máximo, esto es si trabaja con x^*_m y $x^*_M = x_M$, entonces deberá considerar cada intervalo semiabierto por la izquierda y hacer
$\longrightarrow$ **(** l_i , l_s **]** para cada intervalo.

Ej.: **(** 28 , 33 **]** lo que implica que se incluya en ese intervalo el valor 33 pero no el 28.

Si por el contrario trabaja con x_m y x^*_M , entonces considerará $\longrightarrow$ **[** l_i , l_s **)** para cada intervalo.

Ej.: **[** 28 , 33 **)** lo que implica que se incluye el valor 28 pero no el 33.

Otra posibilidad es tomar el valor numérico de los límites y sumarles (o restarles) una mitad de la unidad de medida.

Ej.: **(**28,5 - 33,5**)** lo que implica que no se incluye el 28 pero si el 33, o bien (27,5 - 32,5) lo que implica que se incluye el 28 y no el 33.

Cualquiera sea la convención seguida para determinar límites de los intervalos, esta debe ser respetada a la larga de todo el análisis descriptivo de la información.

Los ítems mencionados significan (y en algunos casos seremos intencionalmente reiterativos) que

* tanto el x_m como el x_M **deben** pertenecer al primero y último intervalo respectivamente.

* todos los datos **deben** pertenecer a algún intervalo.

* los intervalos **no deben** tener valores comunes (deben ser mutuamente excluyentes).

* **de ser posible los intervalos se construirán de igual longitud**.

* es conveniente que la "marca de clase" o valor representativo del intervalo tenga la mayor cantidad de ceros posibles o esté terminado en cinco.

* el número de intervalos a determinar no será recibido como un "mensaje divino" "ni será" "adivinado graciosamente"; el número dependerá del problema del que se trate y del conocimiento que el "dueño de la información" tenga del tema. (Recordemos que en Estadística **no hay recetas mágicas**).

* es conveniente tratar de que no queden intervalos vacíos, si bien esto no siempre se puede lograr.

EJEMPLO PARA ANALIZAR (resuelto)

En una reunión de profesores jubilados de Estadística, Ana y sus amigos se encontraban comparando alegremente sus experiencias en los exámenes finales de Estadística de sus alumnos universitarios. Cada uno insistía en que sus alumnos eran los que "duraban menos" en los exámenes.

Antes de que la discusión se fuera a las manos, Ana propuso llevar a cabo una estadística macabra: para eso necesitaban tabular los resultados en forma adecuada (resultados obtenidos también en forma adecuada). He aquí los datos del "tiempo necesario para desaprobar a un alumno", eufemismo utilizado por la malvada Ana por "tiempo de duración de los alumnos" (expresión que "no suena" bien, ¿no es cierto?), en minutos:

27	24	25	17	3
43	9	34	48	33
28	28	0	40	27
19	27	15	38	32
45	19	35	32	0
33	21	35	16	13
16	39	13	19	35
43	44	11	11	31

x_m = 0 [min] (éstos duraron "lo que un respiro")
x_M^* = 48 [min] (fue el que opuso tenaz resistencia, aunque ya tenía su destino escrito)

$$\text{Rango} = R = x_M - x_m = 48 - 0 = 48 min$$

Corrijamos este rango agregando una (1) unidad, esto es, un minuto más, al valor de x_M :

$$\text{Rango corregido} = R' = (x_M + 1) - x_m = (48 + 1) - 0 = 49 min$$

Probemos con 10 intervalos: $k = 10$.

Entonces

$$a = \frac{R'}{K} = \frac{49}{K} = \frac{49}{10} = 4{,}9 min$$

No nos conviene demasiado para hacer cálculos →entonces probemos con $K = 7$

$$a = \frac{R'}{K} = \frac{49}{K} = \frac{49}{7} = 7 min$$

Esto significa que podremos trabajar con 7 intervalos de amplitud 7[min.].

Hallemos los límites inferiores y superiores de esos intervalos:

1º intervalo: $l_i = 0\ [min.]$

$l_i + 7\ [min.]$

2º intervalo: $l_i = 7\ [min.]$

$l_s = 7 + 7[min.] = 14\ [min.]$

3º intervalo: $l_i = 14\ [min.]$

$l_s = 21\ [min.]$

4º intervalo: $l_i = 21\ [min.]$

$l_s = 28\ [min.]$

5º intervalo: $l_i = 28\ [min.]$

$l_s = 35\ [min.]$

6º intervalo: $l_i = 35\ [min.]$

$l_s = 42\ [min.]$

7º intervalo: $l_i = 42\ [min.]$

$l_s = 49\ [min.]$

Decidamos ahora qué sucederá con los valores que coincidan con los límites de los intervalos. Como el l_s del 7° intervalo es 49 [min.], está claro que no será un valor observando de la variable, porque x_M = 48 [min.]. Luego decidimos que los intervalos serán de la forma

$$[l_i \ ; \ l_s)$$

Esto significa que el valor "conflictivo" se colocará en el intervalo en que aparezca como límite inferior.

Encolumnemos nuestros intervalos y hallemos su marca de clase o punto medio:

0 – 7	3,5
7 – 14	10,5
14 – 21	17,5
21 – 28	24,5
28 – 35	31,5
35 – 42	38,5
42 – 49	45,5

Esto podría significar un problema, si nos resulta difícil trabajar con decimales (por ejemplo, si no tenemos una calculadora a mano). Si es así, podemos cambiar la amplitud o el número de intervalos, o ambos.

Si ésta es su situación, le sugerimos que corrija nuevamente el rango ya corregido, sumando una nueva unidad; obtendrá así un $R' = 50$ min. , y podrá armar 5 intervalos de amplitud 10 min. o bien 10 intervalos de amplitud 5.
¿Con cuál de las opciones solucionará su problema?

Sigamos nosotros con el desarrollo anterior.

Contemos ahora las observaciones que caen en cada intervalo; usemos para ello palotes (o cruces, puntos, asteriscos, ...)

Intervalos	Marca de clase	Número de observaciones
0 – 7	3,5	III
7 – 14	10,5	~~IIII~~ I
14 – 21	17,5	~~IIII~~ III
21 – 28	24,5	~~IIII~~ II
28 – 35	31,5	~~IIII~~ III
35 – 42	38,5	~~IIII~~ II
42 – 49	45,5	~~IIII~~ I

Finalmente, podemos pasar en limpio los resultados obtenidos, con un poco más de formalismo

Límites de Intervalos (min.)	Marca de clase (min.)	Frecuencia
0 – 7	3,5	3
7 – 14	10,5	5
14 – 21	17,5	7
21 – 28	24,5	6
28 – 35	31,5	8
35 – 42	38,5	6
42 – 49	45,5	5
	Total n =	40

Si hubiésemos decidido tomar 5 intervalos de amplitud 10 min., habríamos obtenido:

Límites de Intervalos (min.)	Marca de clase (min.)	Frecuencia
0 a menos de 10	5	4
10 a menos de 20	15	11
20 a menos de 30	25	8
30 a menos de 40	35	11
40 a menos de 50	45	6
	Total n =	40

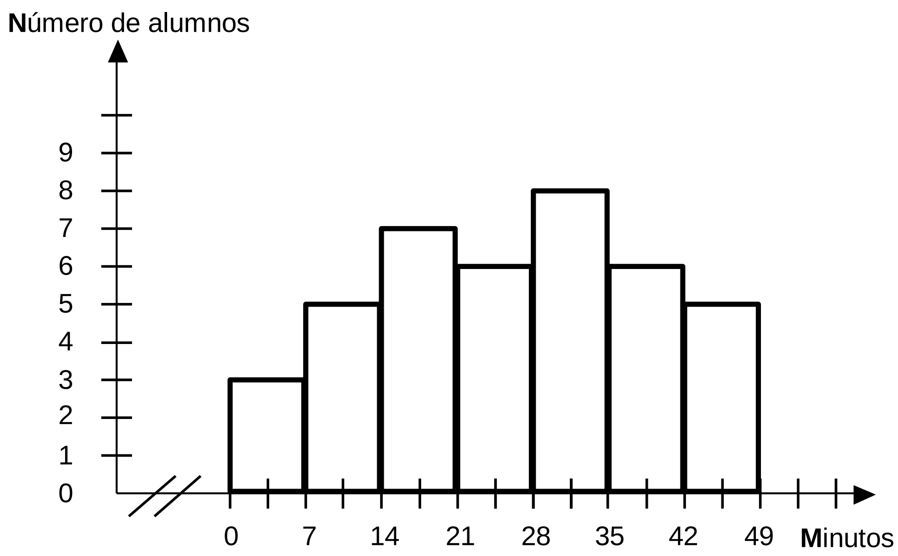
Observemos que al decir "0 a menos 10," 10 a menos 20", etc., estamos usando una expresión equivalente de [0, 10) ; [10, 20) ; etc., y mucho más adecuada para la representación ante personas que no manejen el idioma matemático. Además, nos estamos evitando el trabajo de aclarar con una nota al pie de la tabla la convención usada en el caso de las observaciones "problemáticas": en la primer tabla construida, deberíamos haber escrito al pie algo así como "los valores coincidentes con los límites de un intervalo se tomarán en el intervalo siguiente", o una expresión equivalente.

Observemos también que en las unidades agregadas para corregir el rango inicial sólo podían tomarse al final, y no al principio del conjunto de datos, puesto que no tendría sentido hablar de "alumnos que duraron - 1 min.". En otros problemas podrían tomarse al principio (x_m - 1 unidad , o x_m - 2 unidades), o uno al principio y otro al final (x_m -1 unidad y x_M + 1 unidad).

Finalmente, no debemos olvidar que estamos trabajando con datos reales, que tienen unidades: éstas deben colocarse siempre, para que las interpretaciones puedan hacerse correctamente sin necesidad de buscar más información que la dada por la tabla construida.

A pesar de que los intervalos que construimos tienen amplitud a = 7 min, pensemos por un instante que corresponden a 1 unidad (esto significa que: 7 min equivale a 1 unidad).

Si a cada intervalo (ahora "unitario") lo hacemos corresponder a la base de un rectángulo, y a la frecuencia la hacemos corresponder a la altura del rectángulo, vemos entonces que el área de ese rectángulo considerado tiene el mismo valor numérico que la frecuencia. Con esta transformación de frecuencias en áreas de intervalos de amplitud unitaria, podemos construir el histograma para esos datos:



Las dos líneas paralelas sobre el eje horizontal se colocan para indicar un "corte" en la escala que permita dibujar el histograma adecuadamente (sino habría quedado "pegado" al eje vertical). Si calculáramos el área del histograma, tendríamos:
que es precisamente el número de observaciones con las que contaban Ana y sus colegas al comienzo de este estudio estadístico.

$$\text{Area} = \sum_{i=1}^{7} (\text{base.altura})_i = (1.3) + (1.5) + (1.7) + (1.6) + (1.8) + (1.6) + (1.5) = 40$$

ALGUNAS INDICACIONES ACERCA DE CÓMO ELABORAR UN HISTOGRAMA DE FRECUENCIAS

Como ya se ha visto, el histograma no es más que una representación gráfica de una distribución de frecuencias de una variable cuantitativa continua, consistente en un conjunto de "rectángulos pegados".

El gráfico debe construirse teniendo en claro que **el área de cada rectángulo (y no su altura) representa a la frecuencia absoluta o a la proporción de elementos que pertenecen al intervalo**

El primer paso consiste en elaborar la escala vertical ($\vec{x}$) correspondiente a la variable en estudio.

Si observamos la tabla de frecuencias, los intervalos que hemos construido y por supuesto, tenemos en cuenta los valores mínimo (x_m) y máximo (x_M) de la variable, podemos armarla muy rápidamente, siempre teniendo en cuenta que la escala es numérica y que en consecuencia las distancias iguales deben representar siempre conjuntos iguales de valores.

Estamos suponiendo ahora que según "los pasos para armar los intervalos de una tabla de frecuencias" estos ya están listos y supongamos por el momento que son iguales (se siguieron los pasos del punto anterior).

Debemos ahora construir los rectángulos. La primera intención será, seguramente, igualar en forma directa la frecuencia absoluta o proporción (o porcentaje) correspondiente con la altura del rectángulo.

Es posible que así lo consideremos y que el hecho no sea un error, pero sólo será así si la base de cada uno de los rectángulos, **aún cuando coincida con un conjunto determinado de valores de la variable, sea considerada la unidad**; en este caso, la frecuencia no es realmente la altura, aún cuando la leamos en la escala vertical, sino que la frecuencia es el área del rectángulo.

Recordemos que en el histograma representamos las frecuencias absolutas o porcentajes como si estos se distribuyeran en forma homogénea a lo largo del intervalo considerado. Recordemos también que ésta es la razón por la cual el valor central, que llamamos "centro" o "marca de clase", es el valor de la variable que representa a todos los valores del intervalo.

¿CÓMO PROCEDEMOS SI LAS AMPLITUDES DE LOS INTERVALOS SON DIFERENTES?

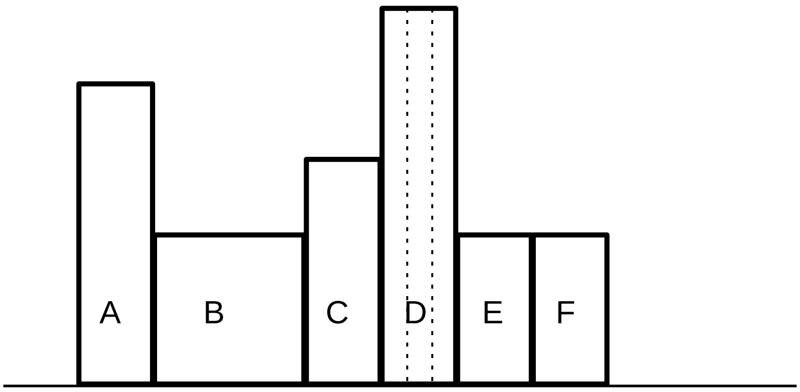
Para construir los intervalos en este caso y luego los rectángulos del histograma, es necesario reiterar que hay que tener en cuenta que la frecuencia del intervalo es el área del rectángulo correspondiente, y no la altura.

Se debe trabajar de modo que el área del rectángulo sea proporcional a la frecuencia del intervalo de clase.

RECUERDE . . . OBSERVE . . . RAZONE . . .
Importa el ÁREA y no la ALTURA . . .

Recordemos también que la diferencia entre un diagrama de barras (separadas) y un histograma (pegadas) es que en el primero, (que se emplea para variable cualitativa o cuantitativa pero en los casos en que los valores de la variable son considerados como "carteles" o rótulos que se diferencian no por su valor sino por su "tipo") la frecuencia es directamente la altura de las barras. En un histograma, en cambio, por tratarse de variable cuantitativa continua, los rectángulos deben estar pegados y se lee la altura pero de acuerdo con la base del rectángulo considerado.

Si se tiene, por ejemplo



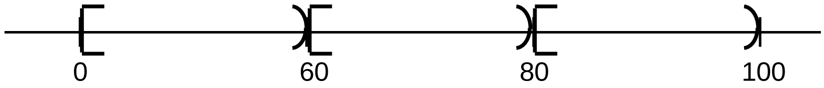
Se puede ver que la cantidad de elementos de A es prácticamente igual a la cantidad de elementos de B, ya que el primer intervalo es el doble de alto pero el segundo tiene una base doble a la anterior. Las áreas son similares por lo que podemos admitir que hay aproximadamente la misma cantidad de elementos en ambos rectángulos.

En el caso D, se observa que cada una de las tres partes tiene 1/3 de los elementos de D, ya que la base de cada rectangulito es la tercera parte del total.

Para tener en claro cómo proceder para construir un histograma cuando los intervalos a considerar son distintos, veamos un ejemplo.

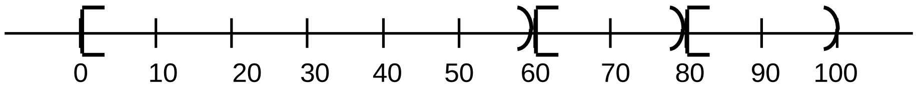
Supongamos que en la clase de Bioestadística, cursada 1997, 30 de los 60 alumnos inscriptos reprobaron, es decir que obtuvieron calificaciones menores a 60 puntos; 28 de los alumnos obtuvieron concepto, con calificación desde 60 a menos de 80 y el resto, los otros 12 alumnos, fueron promovidos sin necesidad de rendir el examen final. Además sabemos que ningún alumno obtuvo la calificación 100.

Si comenzamos a trabajar según lo visto anteriormente y "no pensamos mucho", dibujaríamos una escala de puntajes como sigue

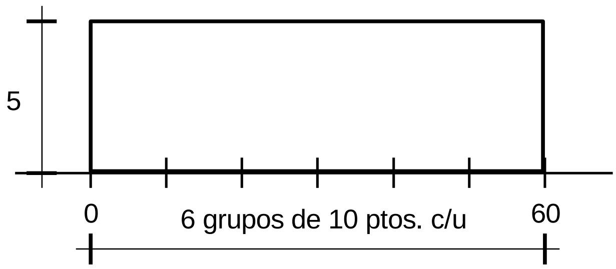


lo que supuesto no es correcto porque los segmentos dibujados son de igual longitud pero los puntajes incluidos no son iguales.

La escala correcta es:

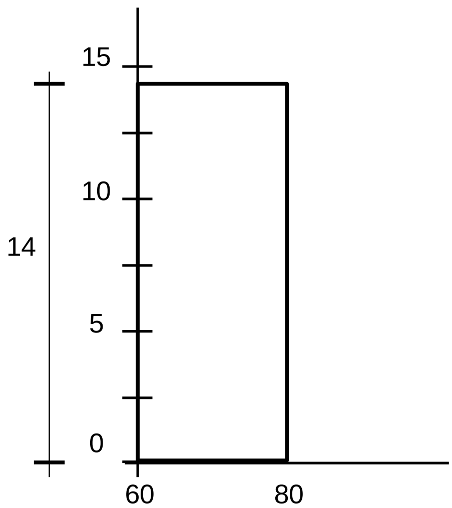


Razonemos ahora que 30 de los 60 alumnos obtuvieron calificaciones entre 0 y 60, pertenecientes a[0, 60), por lo que, si pensamos un poco y recordamos que los elementos se distribuyen en forma homogénea a la largo de cada intervalo, tendremos que si hay 30 elementos sobre 6 intervalos, tendremos 5 en cada uno de los intervalos, cada uno de ellos con 10 puntos de calificaciones como base, esto es

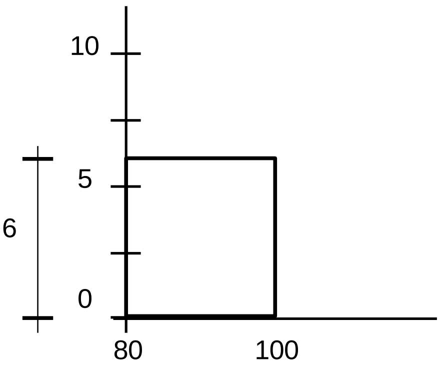


Nótese que se está considerando un conjunto de “10 puntos” de calificación como una unidad.

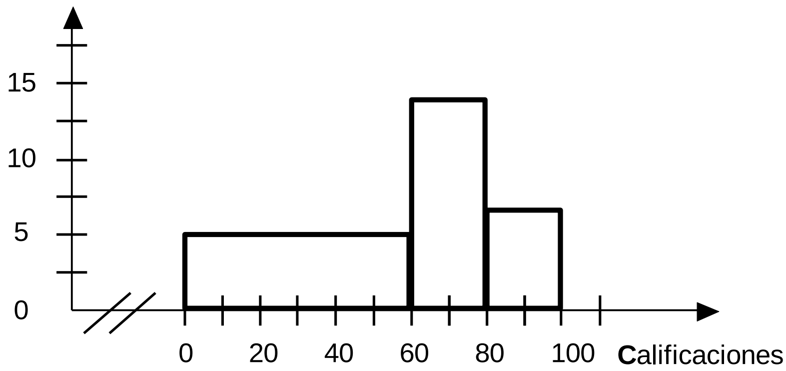
Se tiene además 28 alumnos con calificaciones entre 60 y 80 puntos, por lo que, con el mismo razonamiento, será:



y por último será, para los 12 alumnos con calificaciones de 80 a 100 puntos:



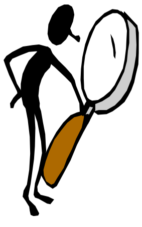
En definitiva, el histograma completo será:



Cuando se interpreta un histograma de estas características es necesario tener mucho cuidado en "leer" la escala vertical. Esta escala puede presentar la frecuencia absoluta o bien el % de casos analizados.

Algunos autores opinan que:
Cuando se trabaja con una serie de frecuencias de variable continua marcadamente asimétrica, con unos pocos valores extremos, conviene construir intervalos cada vez de mayor amplitud cuando se van encontrando valores de la variable con escasa frecuencia, con la finalidad de evitar que aparezcan intervalos vacíos.

¿Usted que opina acerca de esta propuesta?



ATENCIÓN...
Lea los conceptos que sobre estos temas se emplean en LEVIN Y RUBIN - (citado en la bibliografía)
CRITIQUE - Razone y justifique sus críticas.