UNIVERSIDAD DE BUENOS AIRES

ESTADÍSTICA - CÁTEDRA I

CÓDIGO 60

PROF. ADJUNTA A CARGO: DRA. MARÍA SILVIA GALIBERT

TRABAJO PRÁCTICO Nro. 2

AUTORES

Profesores y Docentes de la Cátedra I

"Cuando las estadísticas nos dicen que la familia mexicana tiene un promedio de cuatro hijos y medio, nos explicamos por qué siempre hay uno chaparrito."

Marco Antonio Almazán

PRÁCTICA 2: Resúmenes estadísticos

1.- En séptimo grado de cierta escuela los alumnos decidieron el color del buzo de egresados por simple mayoría en una votación. Hay dos divisiones: una en el turno mañana y otra en el turno tarde. Cada división eligió su propio diseño. Los resultados se organizaron en la siguiente tabla:

Color	Cantidad de alumnos	
	Turno Mañana	Turno Tarde
amarillo	-	3
azul	16	6
negro	6	5
rojo	3	9
verde	4	2
violeta	1	-

a) Mencione y clasifique las variables que intervienen en este problema. Observe que no tienen el mismo rol: ¿cuál es la variable dependiente y cuál la clasificatoria (o variable independiente)? Liste los valores que toman cada una de estas variables.

Conceptualice: Distinga el concepto de "variable" del de "valores de la variable"; evite el error de referirse a "los valores de la variable" como "las variables".

b) ¿Qué color se elegirá para el buzo de los alumnos del turno mañana y cuál para el turno tarde?

c) ¿Qué parte (proporción o razón) del grupo logró su preferencia en cada turno?

d) ¿Qué parte (proporción o razón) del total de alumnos de ambos turnos logró su preferencia?

e) ¿En cuál de los dos grupos las preferencias de los alumnos quedan mejor representadas por el color elegido? ¿Por qué?

f) ¿Cuál es el resumen estadístico que expresa la respuesta dada en b)?

g) ¿Cuál grupo presenta mayor variabilidad en sus preferencias? ¿Qué resumen estadístico expresa dicha variabilidad?

h) **Conceptualice:** ¿Podría hablar de "preferencia promedio" de los alumnos y hallar la media aritmética del color preferido? ¿Podría considerar el color mediano (la mediana de la distribución)? Justifique estas respuestas.

2.- Los mismos alumnos aludidos en **1.-** completaron una encuesta al regresar de su viaje de egresados donde expresaban su nivel de satisfacción con los servicios que les prestaron las respectivas empresas de turismo que contrataron y también el nivel de agrado con sus respectivos animadores de las actividades recreativas. Se les pidió que expresaran con 1, 2 y 3 el nivel de satisfacción según fuera bajo, medio o alto. Para el nivel de agrado, se los interrogó en una escala de 1 a 5. Los resultados fueron:

Nivel de satisfacción con los servicios de la empresa	Cantidad de alumnos	
	Turno Mañana	Turno Tarde
1.- Bajo	6	3
2.- Medio	12	7
3.- Alto	12	15

Nivel de agrado con el animador	Cantidad de alumnos	
	Turno Mañana	Turno Tarde
1.- Muy desagradable	1	3
2.- Bastante desagradable	3	7
3.- Indiferente	6	5
4.- Bastante agradable	12	2
5.- Muy agradable	8	8

a) Mencione y clasifique las variables dependientes.
b) ¿Qué resúmenes estadísticos son adecuados, en rigor, para dar una idea global de las opiniones de los grupos? Discuta en cada caso la utilidad de la moda y de la mediana.
c) **Conceptualice:** Muchos textos de Estadística se refieren a la mediana como el valor que deja la mitad de los datos por debajo y la otra mitad por encima de ella. Observe que no es así en los ejemplos anteriores. ¿Cómo podría precisar la definición de mediana?
d) Tomando en cuenta las estadísticas, elabore una conclusión en términos cualitativos acerca de las opiniones de los alumnos.
e) **Conceptualice:**
   i. ¿Pueden promediarse los niveles de satisfacción o agrado en términos de sus valores cualitativos (bajo, medio, alto, etc.)? ¿Por qué?
   ii. Si se consideran los niveles codificados numéricamente la media aritmética puede ser calculada e interpretada como si las variables fueran aproximadamente cuantitativas. ¿En cuál de las dos variables esa aproximación es mejor y por qué?
   iii. ¿Qué supuesto es necesario para fundar dicha aproximación?
   iv. ¿Qué limitación tiene la unidad de medida en este tipo de variables para que puedan ser consideradas de nivel intervalar?

3.- Considere los datos del ejercicio 7 de la práctica 1.
a) Para el tiempo de reacción antes del entrenamiento
   i. calcule las medidas de tendencia central y la desviación estándar en cada grupo. Observe estas medidas en relación con la forma de cada distribución. Indique el tipo de asimetría en cada caso.
   ii. calcule los cuartiles y construya los diagramas de caja y bigotes. Relacione la forma de los mismos con la de los diagramas de tallo-hoja.
   iii. calcule para cada grupo el rango percentilar y el puntaje Z correspondiente al tiempo de reacción 7,5. Interprete.
b) En cuanto a los tiempos luego del entrenamiento, no calcule pero diga, según puede observarse en los histogramas,
   i. cuáles son aproximadamente la media y la mediana.
   ii. en cuál de los dos casos la desviación estándar es mayor.

4.- a) Calcule todas las medidas de tendencia central y de variabilidad adecuadas para el ej. 9 de TP 1.
b) Teniendo en cuenta estos resultados y los obtenidos en la Práctica 1, si tuviera que sintetizar a modo de conclusión lo que globalmente aconteció con las recaídas de los episodios depresivos en este grupo, ¿qué diría?

5.- Conceptualice:
a) ¿Cuál es el rol esencial de las medidas de tendencia central?
b) Si las variables son cuantitativas
   i. ¿de qué dan idea las medidas de tendencia central?
   ii. ¿cuál de ellas suele preferirse para sintetizar todos los datos en un solo valor que los represente y por qué?

iii. ¿en qué casos se prefiere la mediana sobre la media? (Estamos hablando de variables cuantitativas).

c) ¿Qué expresa la mediana si la variable es cuantitativa continua y cuál es la diferencia si es discreta o cualitativa ordinal?

d) Resumir todos los valores en uno solo que pretende representar al grupo es una simplificación muy grande ¿Por qué podría ser necesario aceptar esa simplificación?

6.- Un grupo de personas fueron evaluadas en su motricidad fina y gruesa a través de dos tareas, y se midió el tiempo, en segundos, que les llevó la realización de cada una de ellas. La tarea 1 consistió en enhebrar una aguja y la tarea 2 en dirigirse saltando en una sola pierna a recoger 5 manzanas esparcidas en el suelo y ponerlas en una canasta. Las medias y desviaciones estándar para cada una de estas tareas fueron:

	Tarea 1	Tarea 2
Media	4	50
Desviación estándar	1,2	10

¿En cuál de las dos tareas las personas tuvieron un desempeño más parejo en cuanto al tiempo empleado en ellas? Fundamente sobre la base del cálculo de un resumen estadístico adecuado.

7.- A continuación se muestran tres conjuntos de datos A, B y C, sólo A está completo. Complete los datos faltantes en B y en C para que todos tengan la misma media y para que C tenga mayor desviación estándar que A.

A: 10 – 15 – 20 – 25 – 30
B: 12 – 13 – 15 – 18 - ……
C: …… - 15 – 20 – 25 - ……

8.- Para ciertos experimentos de laboratorio se necesitan ratones que difieran entre sí lo menos posible en su peso, entre otras características de homogeneidad. Para los experimentos de tres investigaciones sobre distintos temas se utilizaron ratones cuyos pesos se detallan a continuación. Sobre la base de un resumen estadístico adecuado, decida en cuál de los tres casos se cumplió mejor el requerimiento de homogeneidad de los pesos.

Investigación 1: 13 – 13 – 14 – 15 – 16 – 16
Investigación 2: 20 – 20 – 21 – 22 – 23 – 23
Investigación 3: 20 – 20 – 22 – 24 – 26 – 26

9.- Llamando X, Y y Z a los datos (ficticios) de las investigaciones 1, 2 y 3 del Ejercicio 8.- observe que $Y = X + 7$ y $Z = 2X - 6$. Relacione la media y la desviación estándar de X con las de Y y las de Z. Generalice formulando una propiedad: $Y = a\,X + b$ (donde a y b son dos números fijos –constantes, con $a > 0$) entonces media de Y es …… y la desviación estándar de Y es ……

10.- Realice una lectura comprensiva de la siguiente nota y complete la columna de rango percentilar de la tabla.

NOTA: RANGO PERCENTILAR DE UN PUNTAJE

El rango percentilar de un puntaje es el porcentaje de puntuaciones por debajo del mismo (lo consideraremos representando a un continuo, por lo que el puntaje en sí tiene teóricamente frecuencia nula). Supóngase que un sujeto tiene un puntaje 9 en una prueba de memoria proveniente de la cantidad de palabras recordadas de una lista de 12 palabras, y su rango percentilar es 72, eso significa que el 72% de

las personas tiene un puntaje en memoria inferior a 9 y el 28% restante superior a 9. La frecuencia que podría corresponderle a 9 en el conjunto de datos, se considera repartida, mitad entre los valores comprendidos entre 8,5 y 9 (exclusive) y la otra mitad entre 9 (exclusive) y 9,5.

Ejemplo:

Cantidad de palabras recordadas	Frecuencia Porcentual	Frecuencia porcentual acumulada	Rango percentilar
2	5	5	
4	8	13	
5	10	23	
7	15	38	
8	26	64	
9	16	80	72
10	10	90	
11	8	98	
12	2	100	

Según los datos tal cual se observaron, el 64% de las personas tuvieron un puntaje estrictamente inferior a 9 (hasta 8 inclusive), el 20% tuvo un puntaje estrictamente superior (a partir de 10) y el 16% tuvo exactamente 9 puntos. Pero al remitirnos a un continuo teórico en memoria, repartimos ese 16% en todo un intervalo de números reales entre 8,5 y 9,5. En dicho continuo el porcentaje de los valores inferiores a 9 es el correspondiente hasta 8 (64%) más la mitad de la frecuencia de 9 (16/2); por tanto el rango percentilar de 9 es 64% + 8% = 72%.

Maneras equivalentes de expresar lo mismo:

"El rango percentilar de 9 es 72" es equivalente a decir "9 es el percentil 72 de la distribución", todo lo cual significa que "el 72% de los sujetos tiene un puntaje inferior a 9".

11.- En En la página 289 de Carrasco Ortiz, Holgado Tello y del Barrio Gandara (2005), se muestran las medias y desviaciones típicas para cada dimensión del Big Five Questionnaire (Barbaranelli, C., Caprara, G. y Rabasca, A., 1998) según el baremo que obtuvieron en la población de niños españoles. Los datos son:

	Conciencia	Apertura	Extraversión	Cordialidad	Neuroticismo
Media	67,28	27,35	40,85	37,17	26,99
Desviación típica	11,93	5,98	5,24	5,81	7,09

a) Andrés obtuvo los puntajes directos 60, 37, 37, 45 y 27 en las respectivas dimensiones.

a1) Ubique el artículo en el material de la cátedra y, al inicio de la página 278, lea una breve descripción de estas dimensiones.

a2) i) ¿Puede afirmarse que Andrés es igualmente abierto que extrovertido porque sus puntajes son iguales en estas dimensiones de la personalidad?

ii) ¿Puede, de manera similar, afirmarse que es mucho más consciente que Neurótico, comparando 60 con 27?

iii) ¿Tiene sentido la comparación de niveles de variables de diferente naturaleza?

a3) Las variables Apertura y Extraversión son diferentes, como lo son, por ejemplo, la estatura y el peso; por lo que, en principio, no tendría sentido comparalas directamente. ¿En qué sentido, entonces puede establecerse la comparación de los puntajes mencionados en a2)?

a4) Halle las puntuaciones de Andrés en las escalas Z, T y CI.
a5) ¿Qué ventaja proporciona expresar los puntajes en escala T?

b) Matilde obtuvo un puntaje T = 51 en Cordialidad y T = 27 en Apertura.
   b1) ¿En cuál de las dos características se halla más cerca de la media del grupo? Justifique.
   b2) Obtenga las puntuaciones Z correspondientes.

c) El puntaje de Mario, expresado en escala CI, fue 110 en Conciencia y 90 en Extraversión. ¿Qué puntuaciones directas tuvo?
d) Piense un ejemplo de puntuación directa en Apertura y en Cordialidad que le permita interpretar que la persona es más cordial que abierta en relación con la media de la población considerada. Justifique.
e) **Conceptualice:**
   ¿Qué se entiende por baremos de un test?
   e1) ¿Por qué es importante tenerlos en cuenta?
   e2) ¿Qué información proporciona el puntaje Z?
   e3) ¿Qué significa que es adimensional?
   e4) ¿Cuáles son la media y la desviación estándar de las escalas que suelen usarse en psicometría: Z, T y CI?
   e5) ¿Qué significa estandarizar las puntuaciones?

12.- Un profesor de Inglés tiene almacenados exámenes que utiliza en diversas ocasiones. Él lleva el registro histórico de la distribución de frecuencias porcentuales (redondeadas) de las calificaciones que sus alumnos obtuvieron con esos exámenes. La nota de aprobación es 6 y para promoción directa se necesita 7 en cada parcial. A continuación se muestran dichas distribuciones para dos exámenes parciales y un examen final.

	1er Parcial	2do Parcial	Final
Notas	Frecuencia porcentual		
1	4	1	2
2	7	4	2
3	15	6	4
4	24	15	6
5	25	24	11
6	10	25	13
7	8	17	29
8	4	4	19
9	2	3	9
10	1	1	5

a) Halle para cada distribución el diagrama de caja y bigotes. A partir de los mismos describa el tipo de asimetría de estas distribuciones y relaciónelas con el nivel de dificultad de los exámenes. ¿En cuál de ellos quedan mejor discriminados los bajos rendimientos y en cuál los altos rendimientos?
b) i) ¿Qué porcentaje de alumnos desaprobó cada examen y qué porcentaje obtuvo notas que habilitan la promoción en cada parcial?
   ii) ¿Qué nota de aprobación habría que establecer para asegurar que el primer parcial lo desaprobase a lo sumo el 30% de los alumnos?
c) Cano de Becerra (1971) sugiere elaborar exámenes que produzcan distribuciones asimétricas positivas en las instancias parciales y asimétricas negativas en las instancias de final. ¿Qué significa eso en términos de la dificultad de los exámenes y por qué le parece que hará esa sugerencia?

13.- Cierta empresa de desarrollos informáticos aplica un índice de eficacia (de 0 a 100) a los integrantes de sus diversos equipos de trabajo. Los datos (ficticios) tal cual se tomaron son:

92 – 90 – 30 – 91 – 91 – 94 – 93 – 90 – 91 – 90 – 92 – 91 – 93 – 90 – 93 – 90 – 93 – 92

a) Obtenga el diagrama de caja y bigotes con una aplicación y reconozca allí el dato atípico. (Por razones simplicidad didácticas el conjunto de datos es pequeño y el dato atípico es tan evidente que no necesita del uso de diagramas para su detección pero en los casos reales, donde se trabaja con muchos datos, es útil).

b) ¿Con qué resumen estadístico representaría el índice de eficacia típico de las personas del equipo? Justifique.

c) Suponga que la empresa debe elegir a un equipo entre varios para un desarrollo especial, con el requerimiento de que al menos la mitad de sus integrantes tenga un índice de eficacia de 80 o más.

i. ¿Qué información le resultaría más útil: la media o la mediana de cada equipo? ¿Por qué?
ii. Como una manera de afianzar con un ejemplo lo que respondió en i., invente dos conjuntos de 10 datos cada uno, correspondientes a sendos equipos de la empresa, donde uno solo de los dos equipos satisfaga el requerimiento a pesar de tener una media menor que el otro. Piense: ¿qué pasaría si los directivos de la empresa hubieran elegido entre los equipos que Ud. inventó basándose sobre la media en vez de la mediana?

d) **Conceptualice:** Discuta la representatividad de la media o de la mediana para datos cuantitativos según la situación.

Remitirse a la Clase Teórica 3.

14.- Conceptualice: A partir de lo trabajado en esta práctica explique de qué dan idea y para qué puede ser útil conocer la media y la desviación estándar de una distribución.

En los siguientes ítems elija la única opción correcta

15.- Las medidas de tendencia central que tienen significado en una variable cualitativa de nivel ordinal son:
a. solamente la moda.
b. solamente la moda y la mediana.
c. la moda, la mediana y la media.

16.- Si una variable es cuantitativa, tienen significado todas las medidas de tendencia central pero suele preferirse como representante la media porque:
a. es la única que utiliza toda la información.
b. no es sensible a los valores extremos.
c. es la más próxima a la mayoría de los valores.

17.- Si el puntaje Z del individuo A es mayor que el puntaje Z del individuo B en un mismo test para un mismo grupo normativo entonces el rango percentilar de A es mayor que el de B
a. siempre que la distribución de los puntajes sea asimétrica negativa.
b. siempre que la distribución de los puntajes sea asimétrica positiva.
c. siempre, independientemente de la distribución de los puntajes.

18.- Al realizar un diagrama de caja y bigotes se obtuvo que el primer bigote se extiende desde el valor 12 al 15, el segundo bigote desde el valor 18 a 25 y que la mediana es 13. Ello indica que la distribución es

a. asimétrica negativa.
b. asimétrica positiva.
c. aproximadamente simétrica.

19.- La desviación estándar resume la variabilidad de los datos teniendo en cuenta las distancias entre
a. los valores y la mediana.
b. los valores y la media.
c. los cuartiles 1 y 3.

20.- El coeficiente de variación permanece igual si a todos los valores de la variable
a. se los transforma con la raíz cuadrada.
b. se les suma una constante.
c. se los multiplica por una constante.

Referencias

Barbaranelli, C., Caprara, G. y Rabasca, A. (1998). *Manuale del BFQC. Big Five Questionnaire Children. O.S.* Organizzaaioni Speciali-Firenze. Barbaranelli, C., Carpara, G., Rabasca, A., Pastorelli,

Cano de Becerra, F. (1971). Elementos de estadística al servicio de la evaluación del rendimiento. *Revista de Psicología, 16*(1), 61 - 77.

Carrasco Ortiz, M. Á., Holgado Tello, F. P., & del Barrio Gandara, M. V. (2005). Dimensionalidad del cuestionario de los cinco grandes (BFQ-N) en población infantil española. *Psicothema*, *17*(2), 286-291.

RESOLUCIÓN

1.- En séptimo grado de cierta escuela los alumnos decidieron el color del buzo de egresados por simple mayoría en una votación. Hay dos divisiones: una en el turno mañana y otra en el turno tarde. Cada división eligió su propio diseño. Los resultados se organizaron en la siguiente tabla:

Color	Cantidad de alumnos	
	Turno Mañana	Turno Tarde
amarillo	-	3
azul	16	6
negro	6	5
rojo	3	9
verde	4	2
violeta	1	-

a) Mencione y clasifique las variables que intervienen en este problema. Observe que no tienen el mismo rol: ¿cuál es la variable dependiente y cuál la clasificatoria (o variable independiente)? Liste los valores que toman cada una de estas variables.

Variable 1: Color de buzo de egresados elegido por alumnos de séptimo grado de cierta escuela. **Valores:** amarillo, azul, negro, rojo, verde y violeta. **Clasificación:** variable cualitativa de nivel nominal.

Variable 2: Turno Escolar de alumnos de séptimo grado de cierta escuela. **Valores:** Turno Mañana y Turno Tarde. **Clasificación:** variable cualitativa de nivel nominal.

Variable dependiente: color de buzo de egresados elegido.

Variable clasificatoria: Turno Escolar.

Conceptualice: Distinga el concepto de "variable" del de "valores de la variable"; evite el error de referirse a "los valores de la variable" como "las variables".

Variable: característica de las unidades de análisis que puede asumir diferentes valores en cada una de ellas y resulta de interés estudiar.

Categorías/valores de una variable: las distintas modalidades que la variable puede asumir.

Por ejemplo, la variable es "Color de buzo" y sus valores, los distintos colores. Es incorrecto referirse a amarillo, azul, etc. como "las variables".

b) ¿Qué color se elegirá para el buzo de los alumnos del turno mañana y cuál para el turno tarde?

Por simple mayoría en una votación, se elegirá el color azul para el Turno Mañana (16 alumnos) y el color rojo para el Turno Tarde (9 alumnos).

c) ¿Qué parte (proporción o razón) del grupo logró su preferencia en cada turno?

Turno Mañana: El 53,33% ((16/30)x100) del grupo logró su preferencia (o p= 0,53)

Turno Tarde: El 36% ((9/25)x100) del grupo logró su preferencia (o p= 0,36).

d) ¿Qué parte (proporción o razón) del total de alumnos de ambos turnos logró su preferencia?

Total de Alumnos (Turno Mañana + Turno Tarde): El 45,45% ([(16+9)/55]x100) de los alumnos (o p= 0,45) logró su preferencia.

e) ¿En cuál de los dos grupos las preferencias de los alumnos quedan mejor representadas por el color elegido? ¿Por qué?

Las preferencias de los alumnos quedan mejor representadas en el Turno Mañana ya que el color seleccionado fue elegido por más de la mitad de los alumnos (53,33%) mientras para el Turno Tarde el color seleccionado fue elegido sólo por el 36%.

f) ¿Cuál es el resumen estadístico que expresa la respuesta dada en b)?

El resumen estadístico es la moda (valor/categoría de la variable que se repite con más frecuencia).

g) ¿Cuál grupo presenta mayor variabilidad en sus preferencias? ¿Qué resumen estadístico expresa dicha variabilidad?

El Turno Tarde presenta mayor variabilidad en sus preferencias. El resumen estadístico es la Entropía (medida de dispersión adecuada para variables cualitativas).

h) **Conceptualice:** ¿Podría hablar de "preferencia promedio" de los alumnos y hallar la media aritmética del color preferido? ¿Podría considerar el color mediano (la mediana de la distribución)? Justifique estas respuestas.

Considerando que la variable mencionada tiene un nivel de medición nominal:

No sería adecuado calcular la media ya que esta medida de tendencia central sólo se puede utilizar para variables métricas (con nivel de medición intervalar o proporcional).

Tampoco sería adecuado calcular la mediana porque esta medida de tendencia central se puede utilizar para variables con un nivel de medición desde ordinal en adelante.

2.- Los mismos alumnos aludidos en **1.-** completaron una encuesta al regresar de su viaje de egresados donde expresaban su nivel de satisfacción con los servicios que les prestaron las respectivas empresas de turismo que contrataron y también el nivel de agrado con sus respectivos animadores de las actividades recreativas. Se les pidió que expresaran con 1, 2 y 3 el nivel de satisfacción según fuera bajo, medio o alto. Para el nivel de agrado, se los interrogó en una escala de 1 a 5. Los resultados fueron:

Nivel de satisfacción con los servicios de la empresa	Cantidad de alumnos	
	Turno Mañana	Turno Tarde
1.- Bajo	6	3
2.- Medio	12	7
3.- Alto	12	15

Nivel de agrado con el animador	Cantidad de alumnos	
	Turno Mañana	Turno Tarde
1.- Muy desagradable	1	3
2.- Bastante desagradable	3	7
3.- Indiferente	6	5
4.- Bastante agradable	12	2
5.- Muy agradable	8	8

a) Mencione y clasifique las variables dependientes.

Variable 1: "Nivel de Satisfacción con los servicios de la empresa de turismo contratada de alumnos de séptimo grado de cierta escuela". **Clasificación:** variable cualitativa de nivel ordinal.

Variable 2: "Nivel de agrado con el animador de las actividades recreativas del viaje de egresados que presentan los alumnos de séptimo grado de cierta escuela al regresar del mismo" **Clasificación:** variable cualitativa de nivel ordinal.

b) ¿Qué resúmenes estadísticos son adecuados, en rigor, para dar una idea global de las opiniones de los grupos? Discuta en cada caso la utilidad de la moda y de la mediana.

Considerando el nivel de medición de las variables mencionadas en el punto a) (nivel de medición ordinal), resulta adecuado obtener moda y mediana.

La **moda** es útil en la medida en que concentra una porción substancial de la frecuencia total que la diferencia del resto de las categorías; de lo contrario, no tiene mucho valor de representación. En el caso de Satisfacción con la empresa, esto no se da en el turno mañana, donde hay dos categorías con la mayor frecuencia (12) y sí se da en el turno tarde. Para el Nivel de agrado con el animador, por el contrario, la moda se diferencia bien en el turno mañana y no en el turno tarde. En los casos como éstos, de variables medidas en nivel ordinal (o superior), puede utilizarse la mediana, que representa la posición central, en el sentido de que no hay más de la mitad de los valores a cada lado de la misma. Entonces, por ejemplo, en los casos mencionados donde la moda no resultaba útil, se puede resumir la opinión general con la mediana. En el Nivel de satisfacción con la empresa del turno mañana, la mediana es "Medio" y en cuanto al Nivel de agrado con el animador en el turno tarde es "Indiferente". En este último caso queda exactamente la misma cantidad de observaciones a cada lado de la mediana (10) aunque no es siempre así, ya que, para el caso del Nivel de satisfacción, los valores a ambos lados de la mediana fueron 6 y 12 (pero ninguno de ellos supera la mitad: 15).

c) **Conceptualice:** Muchos textos de Estadística se refieren a la mediana como el valor que deja la mitad de los datos por debajo y la otra mitad por encima de ella. Observe que no es así en los ejemplos anteriores. ¿Cómo podría precisar la definición de mediana?

La mediana es el valor que "supera a" y "es superado por" no más de la mitad de las observaciones.

Otra manera equivalente de decir lo mismo es:

A lo sumo la mitad de las observaciones es inferior a la mediana y a lo sumo la mitad de las observaciones es superior a la mediana.

d) Tomando en cuenta las estadísticas, elabore una conclusión en términos cualitativos acerca de las opiniones de los alumnos.

Con respecto a la variable "*Nivel de Satisfacción con los servicios de la empresa de turismo contratada de alumnos de séptimo grado de cierta escuela*", la mayoría de los alumnos del turno mañana manifestó un nivel de satisfacción entre medio y alto (ambas categorías presentan igual frecuencia y entre ambas suman el 80% de las observaciones de ese grupo) mientras que en el Turno Tarde, un 60% de los individuos se ubicaron en la categoría Alto. Que el turno tarde tuvo mayores niveles de satisfacción que el turno mañana, queda también en evidencia comparando las respectivas medianas. Los niveles de satisfacción presentaron mayor variabilidad (estuvieron más repartidos) en el turno mañana que en el turno tarde.

Respecto de la variable "*Nivel de Agrado con el animador de las actividades recreativas de la empresa de turismo contratada de alumnos de séptimo grado de cierta escuela*", gran parte (67%) de los alumnos del turno mañana manifestaron respuestas positivas, ubicándose en las categorías Bastante agradable y Muy agradable mientras que las opiniones de los alumnos del turno tarde están polarizadas: las dos categorías con frecuencias más altas han sido Muy agradable (32%) seguido de Bastante desagradable (28%). Tomando en cuenta las respectivas medianas (Bastante agradable para el turno mañana e Indiferente para el turno tarde), puede decirse que el Nivel de Agrado con el animador, en términos generales, fue mayor en el turno mañana que en el turno tarde. Además los alumnos de la mañana coincidieron más en su nivel de agrado que los de la tarde, donde se repartieron más entre los distintos niveles. En otras palabras, la variable Nivel de Agrado con el animador presenta mayor variabilidad a la tarde que a la mañana.

e) **Conceptualice:**

i. ¿Pueden promediarse los niveles de satisfacción o agrado en términos de sus valores cualitativos (bajo, medio, alto, etc.)? ¿Por qué?

No sería adecuado calcular la media ya que esta medida de tendencia central se puede utilizar para variables métricas (con nivel de medición intervalar o proporcional). Las variables antes enunciadas tienen nivel de medición ordinal.

ii. Si se consideran los niveles codificados numéricamente la media aritmética puede ser calculada e interpretada como si las variables fueran aproximadamente cuantitativas. ¿En cuál de las dos variables esa aproximación es mejor y por qué?

Para que esta aproximación no sea tan grosera se debería trabajar con una cantidad de categorías no menor a 4. Por eso que la aproximación sería mejor para la variable Nivel de Agrado con el animador.

iii. ¿Qué supuesto es necesario para fundar dicha aproximación?

Para poder hacer la aproximación antes mencionada, estaríamos suponiendo que la distancia entre los valores es aproximadamente igual, aunque no estén objetivadas.

iv. ¿Qué limitación tiene la unidad de medida en este tipo de variables para que puedan ser consideradas de nivel intervalar?

Que no existe una unidad de medición objetiva que permita definir las distancias; por tanto las distancias no están precisadas de manera objetiva, la equidistancia entre las categorías es "un supuesto", una aproximación.

3.- Considere los datos del ejercicio 7 de la práctica 1.

a) Para el tiempo de reacción antes del entrenamiento

i. calcule las medidas de tendencia central y la desviación estándar en cada grupo. Observe estas medidas en relación con la forma de cada distribución. Indique el tipo de asimetría en cada caso.

Variables	**Media**	**DS**	**Mediana**	**Moda**
Tiempo de Reacción de los Adultos	8,0250	1,0387	8,2000	Interv. [8,1-8,9]
Tiempo de Reacción de los Adolescentes	7,3750	1,0387	7,2000	Interv. [6,1-6,9]

En el grupo de los adolescentes la media es levemente mayor que la mediana y mayor que el máximo del intervalo modal, lo que indicaría que nos encontramos ante una distribución asimétrica positiva (hacia la derecha): la mayor cantidad de observaciones se ubican en los tiempos de reacción menores.

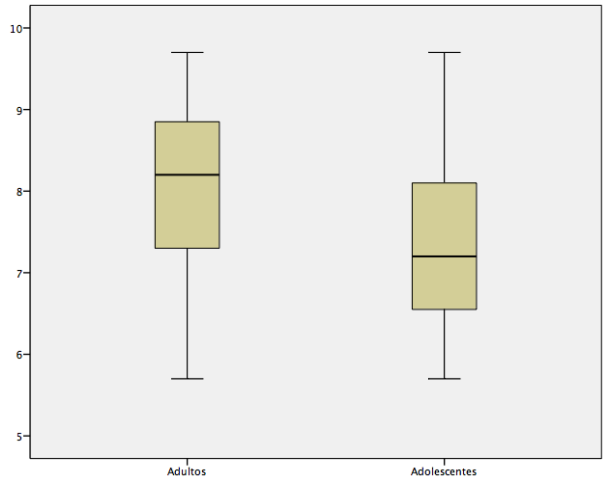
En el grupo de los adultos la media es levemente menor que la mediana y menor que el mínimo del intervalo modal, lo que indicaría que se trata de una distribución asimétrica negativa (hacia la izquierda): las observaciones de concentran en los tiempos mayores.

ii. calcule los cuartiles y construya los diagramas de caja y bigotes. Relacione la forma de los mismos con la de los diagramas de tallo-hoja.

Lo haremos mediante el programa Statistix (si los cuartiles se calculan a mano, pueden dar ligeramente diferentes).

Descriptive Statistics

Variable	N	1st Quarti	Median	3rd Quarti
TR_adultos	20	7.2500	8.2000	8.8750
TR_adolescentes	20	6.5250	7.2000	8.1500



Adultos

Adultos Stem-and-Leaf Plot

Frequency	Stem &	Leaf
1,00	5 .	7
2,00	6 .	48
6,00	7 .	224677
8,00	8 .	13556899
3,00	9 .	237

Stem width: 1,0
Each leaf: 1 case(s)

Adolescentes

Adolescentes Stem-and-Leaf Plot

Frequency	Stem &	Leaf
1,00	5 .	7
8,00	6 .	12556899
5,00	7 .	13778
4,00	8 .	0226
2,00	9 .	07

Stem width: 1,0
Each leaf: 1 case(s)

Se puede observar en los diagramas de caja y bigotes, la caja delimitada por el 1º y 3º cuartil y la mediana en la línea horizontal media (interior de la caja). A partir de la posición de dicha línea interna a la caja y de la longitud de los bigotes, tenemos una idea de la asimetría de cada distribución, que corrobora gráficamente lo dicho en i): en la caja del gráfico de los adultos, la horizontal está más cerca de los valores mayores y el bigote más largo hacia los valores menores; lo contrario en el caso de los adolescentes. Esto también se puede ver en el diagrama de tallo-hojas. Si lo viéramos "*de costado*" representaría un histograma (donde el largo de cada renglón representaría la altura de cada barra). El descenso gradual desde la moda hacia el extremo refleja "la cola" de la asimetría que se expresa en la longitud del bigote de mayor longitud.

iii. Calcule para cada grupo el rango percentilar y el puntaje Z correspondiente al tiempo de reacción 7,5. Interprete.

Acumulando la frecuencia hasta 7,5 se obtienen, para el grupo de adultos 6 y para el de adolescentes

11. El rango percentilar es esa frecuencia expresada porcentualmente; esto es 30% y 55% respectivamente.

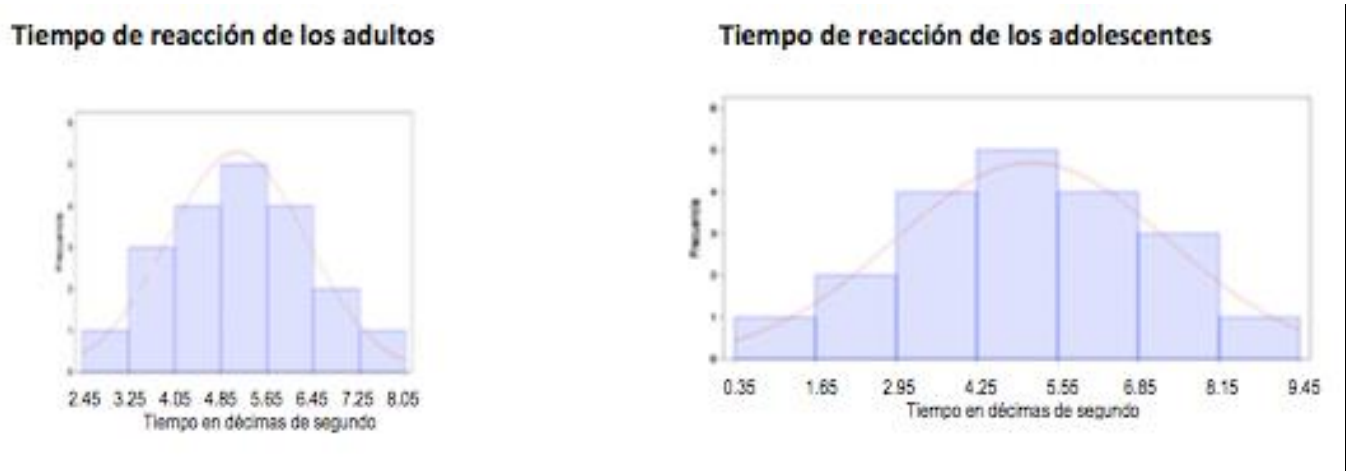
El puntaje $Z_{adultos}$ = (7,5 – 8,025)/1,0387 = -0,51

El puntaje $Z_{adolescentes}$ = (7,5 – 7,375)/1,0387 = 0,12

Un mismo tiempo de reacción, de 7,5 décimas de segundo, a un adulto lo posiciona a 0,51 desviaciones estándar por debajo de la media (más rápido que la media de su grupo) y a un adolescentes a 0,12 desviaciones estándar por encima de ella (más lento que la media de su grupo). Esto se corresponde con los respectivos rangos percentilares, por los cuales en el caso de los adultos este puntaje este puntaje de 7,5 deja menor porcentaje de valores debajo de sí.

b) En cuanto a los tiempos luego del entrenamiento, no calcule pero diga, según puede observarse en los histogramas:

iii. cuáles son aproximadamente la media y la mediana.



Siendo dos distribuciones casi simétricas, la media y la mediana están cercanas al valor central de la escala, que es 5,25 y 4,9 respectivamente.

iv. en cuál de los dos casos la desviación estándar es mayor.

La desviación estándar es mayor en el caso de los adolescentes. En efecto, en el histograma que representa la distribución de estos sujetos, se aprecia que el recorrido de la variable es allí mayor (Xmín=0,35 y Xmáx=9,45) y la curva más achatada que la de los adultos, donde el rango de variación es menor (Xmín=2,45 y Xmáx=8,05). Es decir que, en el caso de los adolescentes, las observaciones se encuentran más dispersas con respecto a la media.

4.- a) Calcule todas las medidas de tendencia central y de variabilidad adecuadas para el ej. 9 de TP 1.

Medidas de tendencia central

Media: 2,9 Mediana: 3 Moda: 4

La moda como resumen representativo del total de los datos no es de interés en este caso porque hay mucha dispersión; no concentra una frecuencia que se diferencie sustancialmente de las de los otros valores. La media y la mediana resultaron muy similares.

Medidas de variabilidad

Amplitud = 6 – 0 = 6 Amplitud intercuartílica = $Q_3 – Q_1$ = 4 - 1,5 = 2,5

Desviación estándar = 1,889 CV = 65,139

b) Teniendo en cuenta estos resultados y los obtenidos en la Práctica 1, si tuviera que sintetizar a modo de conclusión lo que globalmente aconteció con las recaídas de los episodios depresivos en este grupo, ¿qué diría?

Un 20% de los individuos no presenta recaídas. El 80% restante recae de 1 a 6 veces. Algo más de la mitad (55%) recae no más de 3 veces (incluyendo la no recaída) y el 45% más de 3 veces. El porcentaje de los que recaen 2 o más veces, entre los que tienen recaídas, es muy alto: 93,75%.

La variabilidad es amplia, como indica el CV, ya que la desviación estándar representa una parte importante de la media (65,139%). Esto se debe a que una parte sustantiva de la frecuencia total (20%) se concentra en 0, las no recaídas, y el resto se distribuye entre los demás valores alejados de cero.

5.- Conceptualice:

a) ¿Cuál es el rol esencial de las medidas de tendencia central?

b) Si las variables son cuantitativas

i. ¿de qué dan idea las medidas de tendencia central?

ii. ¿cuál de ellas suele preferirse para sintetizar todos los datos en un solo valor que los represente y por qué?

iii. ¿en qué casos se prefiere la mediana sobre la media? (Estamos hablando de variables cuantitativas)

c) ¿Qué expresa la mediana si la variable es cuantitativa continua y cuál es la diferencia si es discreta o cualitativa ordinal?

d) Resumir todos los valores en uno solo que pretende representar al grupo es una simplificación muy grande ¿Por qué podría ser necesario aceptar esa simplificación?

Para responder a estas preguntas consulte la Clase Teórica 3.

6.- Un grupo de personas fueron evaluadas en su motricidad fina y gruesa a través de dos tareas, y se midió el tiempo, en segundos, que les llevó la realización de cada una de ellas. La tarea 1 consistió en enhebrar una aguja y la tarea 2 en dirigirse saltando en una sola pierna a recoger 5 manzanas esparcidas en el suelo y ponerlas en una canasta. Las medias y desviaciones estándar para cada una de estas tareas fueron:

	Tarea 1	Tarea 2
Media	4	50
Desviación estándar	1,2	10

¿En cuál de las dos tareas las personas tuvieron un desempeño más parejo en cuanto al tiempo empleado en ellas? Fundamente sobre la base del cálculo de un resumen estadístico adecuado.

Teniendo en cuenta que se trata de comparar dos variables distintas, cuyas medias también lo son, y considerando, a su vez, que ambas variables están medidas en un nivel proporcional, usaremos el CV para concluir acerca de la homogeneidad o heterogeneidad de las distribuciones.

El CV de los tiempos de las personas que realizaron la tarea 1 es 30 [(1,2/4)x100], mientras que el CV de los tiempos de quienes realizaron la tarea 2 es 20 [(10/50)x100]. Por eso, en la tarea 2 las personas tuvieron un desempeño más parejo.

7.- A continuación se muestran tres conjuntos de datos A, B y C, sólo A está completo. Complete los datos faltantes en B y en C para que todos tengan la misma media y para que C tenga mayor desviación estándar que A.

A: 10 – 15 – 20 – 25 – 30

B: 12 – 13 – 15 – 18 - ……

C: ---- - 15 – 20 – 25 - …….

Si los tres conjuntos tienen que tener la misma media, y dado que tienen la misma cantidad de datos cada uno, entonces la sumatoria de sus valores debe ser igual; en este caso, tomando como referencia el conjunto que está completo, la sumatoria de los valores de cada uno de los conjuntos debe ser igual a 100. Luego, en el conjunto B, el dato faltante es 42.

En el conjunto C, a lo ya mencionado, se agrega otra condición, y es que el conjunto debe tener un mayor desvío, es decir, los valores deben estar más distanciados de la media. Así, los datos faltantes podrían ser los valores 3 y 37 (atender a que no son los únicos valores posibles que cumplen ambas condiciones).

8.- Para ciertos experimentos de laboratorio se necesitan ratones que difieran entre sí lo menos posible en su peso, entre otras características de homogeneidad. Para los experimentos de tres investigaciones sobre distintos temas se utilizaron ratones cuyos pesos se detallan a continuación. Sobre la base de un resumen estadístico adecuado, decida en cuál de los tres casos se cumplió mejor el requerimiento de homogeneidad de los pesos.
Investigación 1: 13 – 13 – 14 – 15 – 16 – 16
Investigación 2: 20 – 20 – 21 – 22 – 23 – 23
Investigación 3: 20 – 20 – 22 – 24 – 26 – 26

Debe calcularse y comparar el CV. Lo haremos utilizando Statistix.

Descriptive Statistics

Variable	N	Mean	SD	C.V.
Inves_1	6	14.500	1.3784	9.5062
Inves_2	6	21.500	1.3784	6.4112
Inves_3	6	23.000	2.7568	11.986

El grupo más homogéneo respecto del peso corresponde a la Investigación 2, ya que presenta un CV menor.

9.- Llamando X, Y y Z a los datos (ficticios) de las investigaciones 1, 2 y 3 del Ejercicio 8.- observe que $Y = X + 7$ y $Z = 2X - 6$. Relacione la media y la desviación estándar de X con las de Y y las de Z. Generalice formulando una propiedad: $Y = a\,X + b$ (donde a y b son dos números fijos –constantes, con $a > 0$) entonces media de Y es …… y la desviación estándar de Y es ……

$\bar{x} = 14{,}5 \quad \bar{y} = 21{,}5 \quad$ por tanto $\quad \bar{y} = \bar{x} + 7$
$\bar{x} = 14{,}5 \quad \bar{z} = 23 \quad$ por tanto $\quad \bar{z} = 2\bar{x} - 6$
$s_x = 1{,}3784 \quad s_y = 1{,}3784 \quad$ por tanto $\quad s_y = 1.s_x$
$s_x = 1{,}3784 \quad s_z = 2{,}7568 \quad$ por tanto $\quad s_z = 2.s_x$

Generalizando:

$$\bar{Y} = a * \bar{X} + b$$
$$s_y = a * s_x$$

10.- Realice una lectura comprensiva de la siguiente nota y complete la columna de rango percentilar de la tabla.

NOTA: RANGO PERCENTILAR DE UN PUNTAJE

El rango percentilar de un puntaje es el porcentaje de puntuaciones por debajo del mismo (lo consideraremos representando a un continuo, por lo que el puntaje en sí tiene teóricamente frecuencia nula). Supóngase que un sujeto tiene un puntaje 9 en una prueba de memoria proveniente de la cantidad de palabras recordadas de una lista de 12 palabras, y su rango percentilar es 72, eso significa que el 72% de las personas tiene un puntaje en memoria inferior a 9 y el 28% restante superior a 9. La frecuencia que podría corresponderle a 9 en el conjunto de datos, se considera repartida, mitad entre los valores comprendidos entre 8,5 y 9 (exclusive) y la otra mitad entre 9 (exclusive) y 9,5.

Ejemplo:

Cantidad de palabras recordadas	Frecuencia Porcentual	Frecuencia porcentual acumulada	Rango percentilar
2	5	5	2,5 = 5/2
4	8	13	9 = 5 + 8/2
5	10	23	18 = 13 + 10/2
7	15	38	30,5 = 23 + 15/2
8	26	64	51 = 38 + 26/2
9	16	80	72 = 64 + 16/2
10	10	90	85 = 80 + 10/2
11	8	98	94 = 90 + 8/2
12	2	100	99 = 98 + 2/2

11.- En En la página 289 de Carrasco Ortiz, Holgado Tello y del Barrio Gandara (2005), se muestran las medias y desviaciones típicas para cada dimensión del Big Five Questionnaire (Barbaranelli, C., Caprara, G. y Rabasca, A., 1998) según el baremo que obtuvieron en la población de niños españoles. Los datos son:

	Conciencia	Apertura	Extraversión	Cordialidad	Neuroticismo
Media	67,28	27,35	40,85	37,17	26,99
Desviación típica	11,93	5,98	5,24	5,81	7,09

a) Andrés obtuvo los puntajes directos 60, 37, 37, 45 y 27 en las respectivas dimensiones.

a1) Ubique el artículo en el material de la cátedra y, al inicio de la página 278, lea una breve descripción de estas dimensiones.

a2) i) ¿Puede afirmarse que Andrés es igualmente abierto que extrovertido porque sus puntajes son iguales en estas dimensiones de la personalidad?

No, los puntajes directos tienen un valor prácticamente nulo para realizar estar comparaciones, ya que cada uno de estos valores corresponden a escalas distintas.

ii) ¿Puede, de manera similar, afirmarse que es mucho más consciente que Neurótico, comparando 60 con 27?

No, por la misma razón dada en i).

iii) ¿Tiene sentido la comparación de niveles de variables de diferente naturaleza?

No, no es posible establecer la comparación directa, precisamente porque son de diferente naturaleza.

a3) Las variables Apertura y Extraversión son diferentes, como lo son, por ejemplo, la estatura y el peso; por lo que, en principio, no tendría sentido compararlas directamente. ¿En qué sentido, entonces puede establecerse la comparación de los puntajes mencionados en a2)?

Puede compararse la relación que tiene cada puntaje con su grupo de referencia, por ejemplo, cuánto se desvía de la media en unidades de desviación estándar.

a4) Halle las puntuaciones de Andrés en las escalas Z, T y CI.

	X	Z	T	CI
Conciencia	60	-0,6102	43,90	90,85
Apertura	37	1,6137	66,14	124,21
Extroversión	37	-0,7347	42,65	88,98
Cordialidad	45	1,3477	63,48	120,22
Neuroticismo	27	0,0014	50,01	100,02

a5) ¿Qué ventaja proporciona expresar los puntajes en escala T?

Los puntajes quedan expresados en valores positivos y, muchas veces, en números enteros lo que facilita su interpretación.

b) Matilde obtuvo un puntaje T = 51 en Cordialidad y T = 27 en Apertura.

b1) ¿En cuál de las dos características se halla más cerca de la media del grupo? Justifique.

En cordialidad, ya que la media de los Puntajes T es 50.

b2) Obtenga las puntuaciones Z correspondientes.

$Z_{cordialidad} = (51 – 50)/10 = 0,1$ $Z_{apertura} = (27 – 50)/10 = -2,3$

c) El puntaje de Mario, expresado en escala CI, fue 110 en Conciencia y 90 en Extraversión. ¿Qué puntuaciones directas tuvo?

$Z_{Conciencia} = (110\text{-}100)/15 = 2/3$ $Z_{Extraversión} = (90 – 100)/15 = - 2/3$

$X_{Conciencia} = 67,28 + 2/3 \times 11,93 = 75,23$ $X_{Extraversión} = 40,85 – 2/3 \times 5,24 = 37,36$

d) Piense un ejemplo de puntuación directa en Apertura y en Cordialidad que le permita interpretar que la persona es más cordial que abierta en relación con la media de la población considerada. Justifique.

La pregunta tiene múltiples respuestas, una posibilidad es, por ejemplo, asignar un valor por encima de la media en Cordialidad y uno por debajo de la media en Apertura, como 40 y 25 respectivamente. Otra posibilidad es elegir dos valores por encima de la media pero a mayor distancia en términos de unidades de desviación estándar; por ejemplo, 45,89 (Z = 1,5) y 33,33 (Z = 1) respectivamente.

e) **Conceptualice:**
¿Qué se entiende por baremos de un test?

e1) ¿Por qué es importante tenerlos en cuenta?

e2) ¿Qué información proporciona el puntaje Z?

e3) ¿Qué significa que es adimensional?

e4) ¿Cuáles son la media y la desviación estándar de las escalas que suelen usarse en psicometría: Z, T y CI?

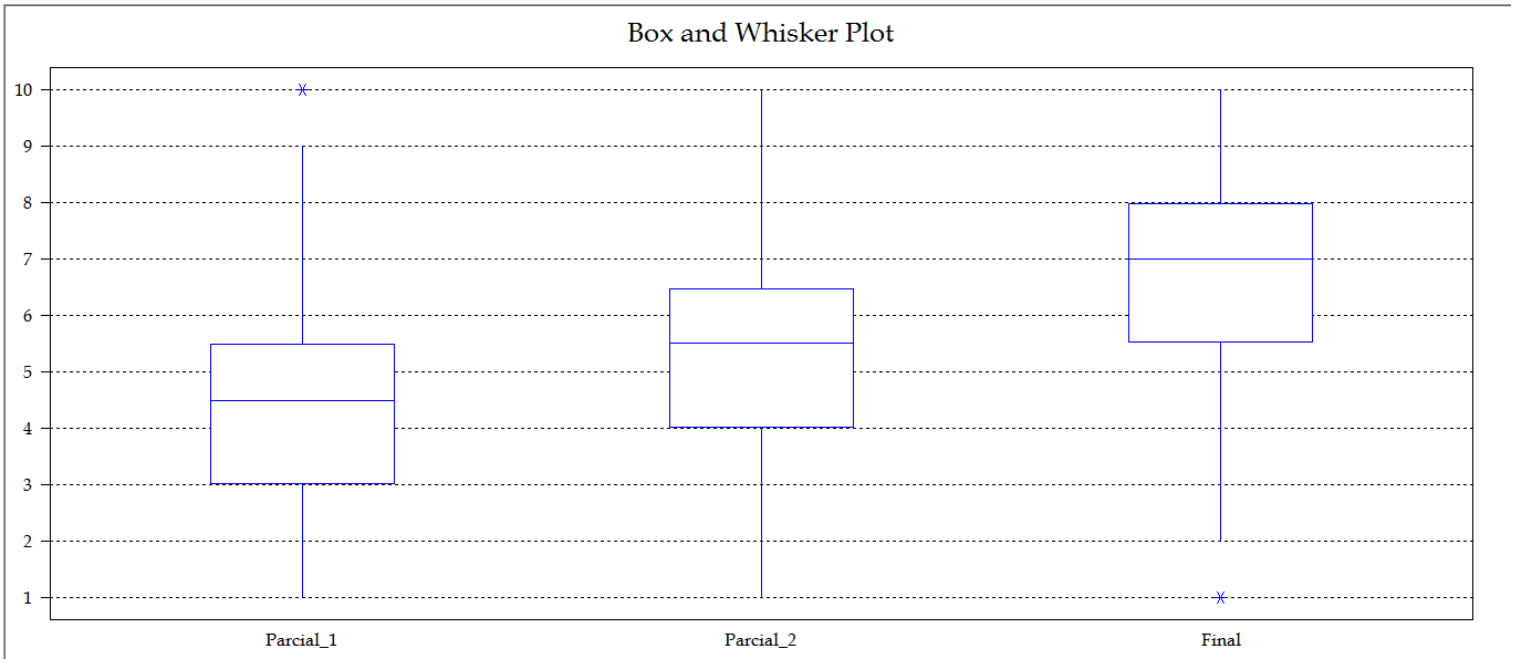
e5) ¿Qué significa estandarizar las puntuaciones?

Algunas respuestas a estas preguntas se hallan en el Cap. 3 de Bologna y otras en la Clase Teórica 3.

12.- Un profesor de Inglés tiene almacenados exámenes que utiliza en diversas ocasiones. Él lleva el registro histórico de la distribución de frecuencias porcentuales (redondeadas) de las calificaciones que sus alumnos obtuvieron con esos exámenes. La nota de aprobación es 6 y para promoción directa se necesita 7 en cada parcial. A continuación se muestran dichas distribuciones para dos exámenes parciales y un examen final.

	1er Parcial	2do Parcial	Final
Notas	Frecuencia porcentual		
1	4	1	2
2	7	4	2
3	15	6	4
4	24	15	6
5	25	24	11
6	10	25	13
7	8	17	29
8	4	4	19
9	2	3	9
10	1	1	5

a) Halle para cada distribución el diagrama de caja y bigotes. A partir de los mismos describa el tipo de asimetría de estas distribuciones y relaciónelas con el nivel de dificultad de los exámenes. ¿En cuál de ellos quedan mejor discriminados los bajos rendimientos y en cuál los altos rendimientos?



En el primer parcial se observa una asimetría hacia la derecha: las observaciones tienden a concentrarse en las calificaciones más bajas. Esto hace pensar en que se trató de un examen más difícil. Aquí quedan mejor discriminados los estudiantes con alto rendimiento (obsérvese, por ejemplo, el asterisco en

el extremo superior del recorrido de la variable que representa a un sujeto con una nota excepcional respecto de lo que ha sido el rendimiento grupal).

En el segundo parcial la distribución pareciera ser simétrica: la mayoría de las observaciones se concentran en las calificaciones medias, lo que hace pensar en un examen de dificultad media, y quedan discriminados por igual los estudiantes con alto y bajo rendimiento.

En el examen final se observa una asimetría hacia la izquierda: las observaciones tienden a concentrarse en las calificaciones altas. Esto hace pensar en un examen de dificultad baja; quedan mejor discriminados los estudiantes con bajo rendimiento (obsérvese, por ejemplo, el asterisco en el extremo inferior del recorrido de la variable que representa a un sujeto con una nota excepcional con respecto a lo que ha sido el rendimiento grupal).

b) i) ¿Qué porcentaje de alumnos desaprobó cada examen y qué porcentaje obtuvo notas que habilitan la promoción en cada parcial?
Se resuelve acumulando las frecuencias *hasta* 6 y *desde* 7.

	1º parcial	2º parcial	Final
Desaprobaron	75%	50%	25%
Promocionaron	15%	25%	

ii) ¿Qué nota de aprobación habría que establecer para asegurar que el primer parcial lo desaprobase a lo sumo el 30% de los alumnos?

Se necesita acumular la frecuencia y ver dónde se acumula hasta el 30%.

Notas x_i	$f_i\%$	$F_i\%$
1	4	4
2	7	11
3	15	26
4	24	50
5	25	
6	10	
7	8	
8	4	
9	2	
10	1	

Se observa que hasta 3 inclusive se acumula 26% y hasta 4, ya se pasa al 50%. Luego, habría que poner como nota de aprobación mínima a 4, en cuyo caso el porcentaje de desaprobados sería 26%.

c) Cano de Becerra (1971) sugiere elaborar exámenes que produzcan distribuciones asimétricas positivas (hacia la derecha) en las instancias parciales y asimétricas negativas (hacia la izquierda) en las instancias de final. ¿Qué significa eso en términos de la dificultad de los exámenes y por qué le parece que hará esa sugerencia?

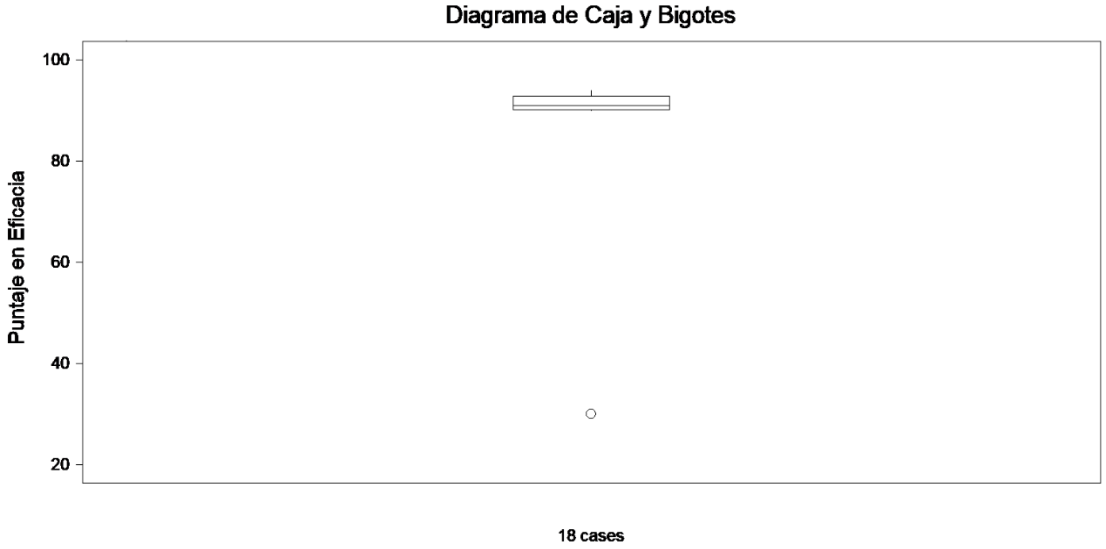
Significa hacer parciales con cierto grado de dificultad, que no sean muy fáciles, a fin de que las notas más altas sean menos frecuentes; es decir, que no muchos alumnos obtengan esas notas. Por el contrario,

en los finales se buscaría que fueran pocos los de notas bajas. El fundamento de esto es exigir a los alumnos durante la cursada para que se esfuercen por tener los resultados altos, que las notas muy altas no sean "regaladas". Una vez que llegan a la instancia de final, se desea que aprueben, a menos que haya suficiente evidencia de que no saben lo mínimo necesario; es decir, se desea discriminar mejor a los de bajo rendimiento. En otras palabras, que si se desaprueba un alumno en la instancia de final, no sea porque el examen es difícil sino porque el alumno no sabe lo suficiente.

13.- Cierta empresa de desarrollos informáticos aplica un índice de eficacia (de 0 a 100) a los integrantes de sus diversos equipos de trabajo. Los datos (ficticios) tal cual se tomaron son:

92 – 90 – 30 – 91 – 91 – 94 – 93 – 90 – 91 – 90 – 92 – 91 – 93 – 90 – 93 – 90 – 93 – 92

Obtenga el diagrama de caja y bigotes con una aplicación y reconozca allí el dato atípico. (Por razones simplicidad didácticas el conjunto de datos es pequeño y el dato atípico es tan evidente que no necesita del uso de diagramas para su detección; pero en los casos reales donde se trabaja con muchos datos, es útil).



b) ¿Con qué resumen estadístico representaría el índice de eficacia típico de las personas del equipo? Justifique y calcúlelo.

Con la mediana, dado que se trata de una variable cuantitativa que presenta un valor extremo sin compensar. La mediana es 91.

Obsérvese que, salvo 30, todos los datos son 90 o más. Mientras que la media es 88,111; por debajo de todos los valores, excepto el 30. Por eso en este caso la mediana es más parecida a la mayoría de los valores que la media.

c) Suponga que la empresa debe elegir a un equipo entre varios para un desarrollo especial, con el requerimiento de que al menos la mitad de sus integrantes tenga un índice de eficacia de 80 o más.

i. ¿Qué información le resultaría más útil: la media o la mediana de cada equipo? ¿Por qué?

Con este requerimiento, es más útil la mediana porque va a dar cuenta de la localización de a lo sumo la mitad de las observaciones a la derecha de la misma (que es lo que se desea) mientras que la media no asegura que haya una cantidad suficiente (la mitad) a la derecha de la misma, según el requerimiento de la empresa.

ii. Como una manera de afianzar con un ejemplo lo que respondió en i., invente dos conjuntos de 10 datos cada uno, correspondientes a sendos equipos de la empresa, donde uno solo de los dos equipos satisfaga el requerimiento a pesar de tener una media menor que el otro. Piense: ¿qué pasaría si los

directivos de la empresa hubieran elegido entre los equipos que Ud. inventó basándose sobre la media en vez de la mediana?

Equipo 1: 30 – 40 – 50 – 60 – 80 – 81 – 82 – 83 – 85 – 90

Equipo 2: 73 – 74 – 75 – 76 – 77 – 77 – 78 – 78 – 95 – 100

En el primer caso la media es 68,1 y la mediana 80,5.

En el segundo caso la media es 80,3 y la mediana 77.

Si los directivos de la empresa eligieran al equipo tomando como referencia la media, elegirían el equipo 2, pero este equipo no contiene, como se requiere, la mitad de los integrantes con un índice de eficacia de 80 o más. Sólo dos personas reúnen este requisito y, sus puntajes altos compensan los del resto. Pero no hay 5 personas con 80 o más. Mientras que, si la elección es tomando en cuenta la mediana, se elegiría el equipo 1, donde puede asegurarse que la mitad de los datos es mayor que 80,5; luego se cumple con el requerimiento.

d) **Conceptualice:** Discuta la representatividad de la media o de la mediana para datos cuantitativos según la situación. Remitirse a la Clase Teórica 3.

15.- Las medidas de tendencia central que tienen significado en una variable cualitativa de nivel ordinal son:
a. solamente la moda.
b. solamente la moda y la mediana.
c. la moda, la mediana y la media.

16.- Si una variable es cuantitativa, tienen significado todas las medidas de tendencia central pero suele preferirse como representante la media porque:
a. es la única que utiliza toda la información.
b. no es sensible a los valores extremos.
c. es la más próxima a la mayoría de los valores.

17.- Si el puntaje Z del individuo A es mayor que el puntaje Z del individuo B en un mismo test para un mismo grupo normativo entonces el rango percentilar de A es mayor que el de B
a. siempre que la distribución de los puntajes sea asimétrica negativa.
b. siempre que la distribución de los puntajes sea asimétrica positiva.
c. siempre, independientemente de la distribución de los puntajes.

18.- Al realizar un diagrama de caja y bigotes se obtuvo que el primer bigote se extiende desde el valor 12 al 15, el segundo bigote desde el valor 18 a 25 y que la mediana es 13. Ello indica que la distribución es
a. asimétrica negativa.
b. asimétrica positiva.
c. aproximadamente simétrica.

19.- La desviación estándar resume la variabilidad de los datos teniendo en cuenta las distancias entre
a. los valores y la mediana.
b. los valores y la media.
c. los cuartiles 1 y 3.

20.- El coeficiente de variación permanece igual si a todos los valores de la variable
a. se los transforma con la raíz cuadrada.
b. se les suma una constante.
c. se los multiplica por una constante.