



UNIVERSIDAD DE BUENOS AIRES

ESTADÍSTICA - CÁTEDRA I

CÓDIGO 60

PROF. ADJUNTA A CARGO: DRA. MARÍA SILVIA GALIBERT

PRÁCTICA DE REPASO – UNIDADES 1, 2 y 3

AUTORES

Profesores y Docentes de la Cátedra I



PRÁCTICA DE REPASO DE LAS UNIDADES 1, 2 y 3

1.- Los siguientes diagramas de tallo-hoja representan la distribución de los puntajes en una encuesta sobre la imagen positiva de dos precandidatos (A) y (B) de un mismo partido político entre los que se quiere decidir para proponer en las elecciones. Ambos diagramas corresponden a la misma muestra de sujetos encuestados, quienes expresaron su opinión en una escala de 0 a 100.

STATISTIX FOR WINDOWS

STEM AND LEAF PLOT OF A	STEM AND LEAF PLOT OF B																																																						
<p>LEAF DIGIT UNIT = 1 4 2 REPRESENTS 42.</p> <table style="margin-left: 40px;"> <tr><td>STEM</td><td>LEAVES</td></tr> <tr><td>2</td><td>4 24</td></tr> <tr><td>4</td><td>4 68</td></tr> <tr><td>7</td><td>5 234</td></tr> <tr><td>12</td><td>5 56789</td></tr> <tr><td>17</td><td>6 01234</td></tr> <tr><td>24</td><td>6 5566789</td></tr> <tr><td>(11)</td><td>7 01223334444</td></tr> <tr><td>25</td><td>7 555556666777889</td></tr> <tr><td>10</td><td>8 012344</td></tr> <tr><td>4</td><td>8 5678</td></tr> </table>	STEM	LEAVES	2	4 24	4	4 68	7	5 234	12	5 56789	17	6 01234	24	6 5566789	(11)	7 01223334444	25	7 555556666777889	10	8 012344	4	8 5678	<p>LEAF DIGIT UNIT = 1 3 2 REPRESENTS 32.</p> <table style="margin-left: 40px;"> <tr><td>STEM</td><td>LEAVES</td></tr> <tr><td>3</td><td>3 234</td></tr> <tr><td>8</td><td>3 56789</td></tr> <tr><td>14</td><td>4 012344</td></tr> <tr><td>20</td><td>4 556789</td></tr> <tr><td>(11)</td><td>5 00112233344</td></tr> <tr><td>29</td><td>5 55556789</td></tr> <tr><td>21</td><td>6 01234</td></tr> <tr><td>16</td><td>6 5679</td></tr> <tr><td>12</td><td>7 0134</td></tr> <tr><td>8</td><td>7 56</td></tr> <tr><td>6</td><td>8 24</td></tr> <tr><td>4</td><td>8 5</td></tr> <tr><td>3</td><td>9 2</td></tr> <tr><td>2</td><td>9 5</td></tr> <tr><td>1</td><td>10 0</td></tr> </table>	STEM	LEAVES	3	3 234	8	3 56789	14	4 012344	20	4 556789	(11)	5 00112233344	29	5 55556789	21	6 01234	16	6 5679	12	7 0134	8	7 56	6	8 24	4	8 5	3	9 2	2	9 5	1	10 0
STEM	LEAVES																																																						
2	4 24																																																						
4	4 68																																																						
7	5 234																																																						
12	5 56789																																																						
17	6 01234																																																						
24	6 5566789																																																						
(11)	7 01223334444																																																						
25	7 555556666777889																																																						
10	8 012344																																																						
4	8 5678																																																						
STEM	LEAVES																																																						
3	3 234																																																						
8	3 56789																																																						
14	4 012344																																																						
20	4 556789																																																						
(11)	5 00112233344																																																						
29	5 55556789																																																						
21	6 01234																																																						
16	6 5679																																																						
12	7 0134																																																						
8	7 56																																																						
6	8 24																																																						
4	8 5																																																						
3	9 2																																																						
2	9 5																																																						
1	10 0																																																						

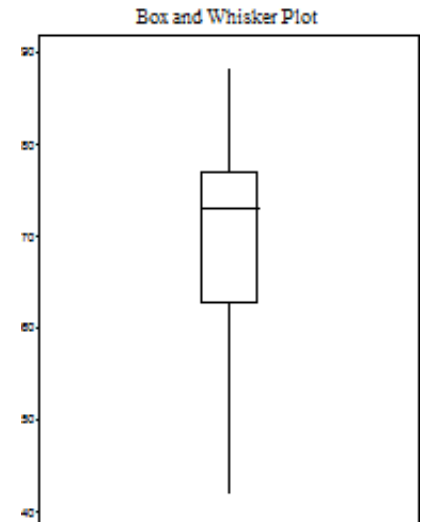
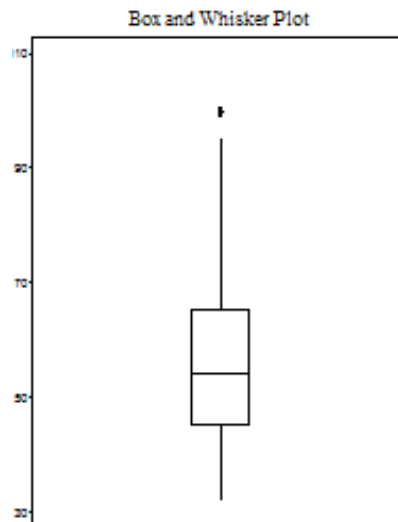
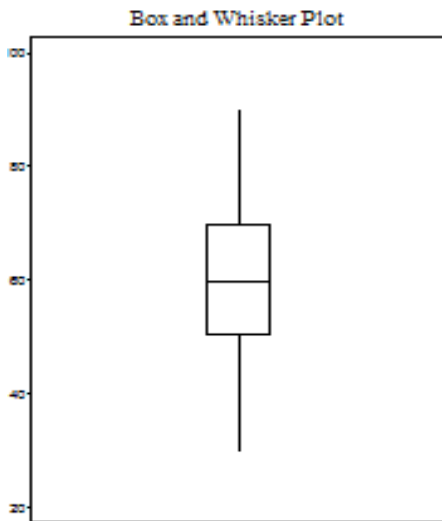
a) ¿Cuál es el tamaño de la muestra de encuestados?

b) La siguiente tabla contiene en cada renglón los resúmenes estadísticos correspondientes a seis conjuntos de datos. **Sin hacer cálculos**, identifique qué renglón corresponde a la distribución A y qué renglón a la distribución B. Hágalo escribiendo A o B al lado del número de la fila elegida: 1), 2), 3), 4), 5) o 6). Sólo un renglón es para A y sólo un renglón es para B. En otras palabras, de las 5 opciones sólo dos son verdaderas. Ayuda: Observe que en cada columna aparecen sólo dos valores diferentes. Se trata, por tanto de encontrar a ojo la única combinación que se ajusta a cada diagrama.

DESCRIPTIVE STATISTICS

VARIABLE	MEAN	SD	1ST QUARTI	MEDIAN	3RD QUARTI
1)	56.583	15.822	45.000	54.000	65.750
2)	69.750	15.822	62.250	73.000	77.000
3)	56.583	11.208	45.000	54.000	65.750
4)	69.750	11.208	62.250	73.000	77.000
5)	69.750	11.208	45.000	54.000	65.750
6)	56.583	15.822	62.250	73.000	77.000

c) Señale con A y con B las cajas que corresponden a la distribución A y B respectivamente.



d) Sobre la base de la información obtenida ¿A cuál propondría como candidato? Fundamente su respuesta.

2.- A continuación se exhiben dos conjuntos de datos. Corresponden a la cantidad de días que tardaron respectivamente los proveedores SALUTEM y VITAE en satisfacer el encargo del mismo producto a cierto hospital. A SALUTEM se le hicieron 10 encargos y a VITAE 20 encargos.

SALUTEM: 4 – 7 – 6 – 3 – 4 – 3 – 5 – 3 – 4 – 5

VITAE: 3 – 5 – 8 – 1 – 4 – 7 – 8 – 6 – 4 – 6 – 7 – 7 – 6 – 5 – 7 – 5 – 8 – 8 – 6 – 7

- Mencione todas las variables que intervienen en el problema y clasifíquelas. Indique también el nivel de medición.
- Organice la información en tablas de frecuencia.
- Exhiba los datos en dos diagramas de tallo – hoja.
- Si quisiera comparar las distribuciones de frecuencia de los tiempos para ambas sucursales en un mismo gráfico ¿qué tipo de gráfico sería apropiado? Puede bosquejarlo a mano o, si está familiarizado con una aplicación, utilizarla para ello.
- Calcule los resúmenes numéricos adecuados para comparar ambas distribuciones en todos los aspectos de las mismas.
- Dé una conclusión breve pero completa de la comparación entre las dos distribuciones.
- Sobre la base del análisis anterior, y suponiendo que ambos venden al mismo precio, ¿recomendaría a alguna de los dos proveedores por sobre el otro?

3.- En un programa familiar de entretenimientos compiten un equipo de 6 niños con un equipo de 4 padres. El entretenimiento consiste en **cada una** de las 10 personas tiene que escribir cuantas palabras pueda en un minuto que comiencen con una letra dada. Gana el equipo que en total haya producido la mayor cantidad de palabras (es decir, se suma la cantidad de palabras producidas por cada miembro del equipo). Se sabe que la cantidad promedio de palabras de las 10 personas fue 8 y que la media del equipo de niños fue 7. ¿Qué equipo ganó la competencia? Justifique con base en los cálculos adecuados.

4.- Leila es psicóloga y trabaja para las obras sociales A y B. Durante el último año llevó registro del tipo de trastorno que presentaba cada uno de los pacientes que era diagnosticado con un trastorno mental o de la conducta para cada obra social y organizó la información en la siguiente tabla.

Tipo de Trastorno	OBRA SOCIAL A	OBRA SOCIAL B	TOTAL
Ataques de pánico	5	3	8
Depresión	8	9	17
Ansiedad	14	4	18
Alteraciones en el sueño	6	1	7
Déficit de atención	7	3	10

- Mencione las variables que intervienen en el problema, clasifíquelas e indique nivel de medición.
- Indique qué diagrama sería adecuado para comparar las distribuciones de frecuencias de los tipos de trastornos en ambas obras sociales. Puede hacer un bosquejo manual del mismo o utilizar una aplicación para obtenerlo. Saque conclusiones de la comparación.
- Si se presenta un paciente de la Obra Social A
 - ¿Qué trastorno mental o de la conducta es más probable que padezca?
 - ¿Y para la Obra Social B?
 - ¿En cuál de los dos casos la incertidumbre es mayor?
- ¿Qué resumen estadístico es adecuado para dar cuenta de la centralidad de los datos del tipo de trastorno?
- Si utilizara un diagrama circular para representar la distribución del tipo de trastorno sin distinguir Obra Social, ¿qué ángulo debería dar a cada sector?
- Calcule un coeficiente adecuado para expresar el grado de asociación muestral entre la Obra Social y el Tipo de Trastorno.

5.- Treinta y cinco personas fueron encuestadas con respecto a la penalización en años de un homicida antes y después de ver un video sobre su vida. Se les preguntó con cuántos años penalizarían al homicida. La distribución de frecuencias se halla en la tabla. Sobre la base de las medidas de tendencia central adecuadas, analice los datos y concluya si el video tuvo efecto en la decisión de los espectadores y cómo incidió.

IMAGEN	ANTES	DESPUÉS
PERPETUA	10	6
LARGA	15	11
MODERADA	7	10
BREVE	3	7
NULA		1

6.- A continuación se muestra el resumen y parte de los resultados de la investigación de Vargas Orozco et al (2014).

En esta investigación presentamos los índices de Depresión y Ansiedad de los estudiantes de la Universidad Autónoma de Tamaulipas campus Reynosa-Aztlán. La muestra estuvo conformada por 500 estudiantes clasificados por género y por programa académico, se aplicaron dos instrumentos estandarizados para medir depresión, se utilizó la Escala de Depresión de Beck y la Escala de Ansiedad de Hamilton, los resultados nos muestran que existe un mayor puntaje en los cuadros depresivos en el sexo femenino así como en los índices de ansiedad.

Sexo	NIVELES DE DEPRESIÓN			
	Normal	Leve	Moderada	Severa
Femenino	176	33	8	2
Masculino	259	15	6	1

Sexo	ANSIEDAD	
	Con ansiedad ¹	Sin ansiedad
Femenino	169	50
Masculino	101	180

- a) Mencione y clasifique todas las variables a las que se alude en el resumen y cuyos resultados de distribuciones de frecuencias se exhiben en las tablas.
- b) b1) Obtenga las tablas de frecuencias porcentuales para las distribuciones condicionales de la depresión según el sexo. ¿Cuál presenta mayor variabilidad? Fundamente.
 - b2) Grafique las distribuciones condicionales de la ansiedad según el sexo en un mismo gráfico de modo que puedan ser comparadas.
 - b3) Sobre la base de b1) y de b2) saque una conclusión de la comparación entre sexos.
 - b4) Obtenga las distribuciones marginales de la Depresión y de la Ansiedad. ¿Qué medidas de tendencia central son adecuadas para resumir estos datos? Calcúlelas e interprete.
 - b5) ¿En qué medida, a nivel muestral, el nivel de depresión, por un lado, y la ansiedad, por el otro, están asociados al sexo?
- c) Si la edad promedio de las mujeres fue de 22,8 años y la de los varones de 23,2 ¿Cuál es fue la edad promedio del total de estudiantes?

7.- A continuación se muestra el resumen e instrumentos de la investigación de Horrillo-Álvarez, C. et al (2019).

Los efectos beneficiosos de la meditación mindfulness (MM) se han observado en numerosos estudios, pudiendo ser la atención y conciencia en el momento presente una variable esencial. Sin embargo, la adherencia al entrenamiento de la práctica meditativa es en ocasiones baja. Se investigaron los efectos del entrenamiento en MM sobre la atención y conciencia y si el formato de registro de la práctica de MM, formato papel (PP) (n = 27) o aplicación móvil (APP) (n = 21), produce diferencias en la adherencia al entrenamiento y en la atención y conciencia. Los resultados mostraron un aumento de la atención y conciencia tras el entrenamiento, independientemente del formato de registro de la práctica. La APP redujo el abandono al entrenamiento y aumentó el tiempo de práctica de MM cada día que se practicó.

Instrumentos

Escala de atención y conciencia en el momento presente (Mindfulness Attention Awareness Scale - MAAS). Compuesta por 15 ítems permite obtener una puntuación entre 15 y 90 puntos.

Aplicación de móvil Meditate. Diseñada ad hoc, registra la fecha y los minutos meditados cada vez que se practica MM durante las 8 semanas de entrenamiento. El tiempo máximo que permite practicar la aplicación son 30 minutos diarios.

¹ Respetamos la clasificación de los autores pero cabe aclarar que siempre hay algún grado de ansiedad; por lo que se entiende que se refieren a ansiedad alta.

Hoja de registro de la práctica de MM. En formato papel y creada ad hoc, recoge la fecha y los minutos de MM cada día que se practica.

- I)
- Mencione las variables bajo estudio, clasifíquelas según sus valores y nivel de medición e indique cuáles son las dependientes y las independientes.
 - ¿Qué resúmenes estadísticos numéricos elegiría y cómo lo utilizaría para:
 - clasificar a los sujetos en dos polos opuestos de menor MM y mayor MM.
 - decidir si el formato de registro de la práctica incide en la atención y conciencia.

II)

Adherencia al entrenamiento		
Tipo de registro	Abandonan	Perseveran
Formato en papel	23	27
Aplicación móvil	4	21

- Represente con un diagrama adecuado la distribución de la adherencia según el tipo de registro de modo que pueda indagar si éste tiene incidencia sobre la misma.
- Obtenga una medida de tendencia central adecuada para comparar los dos grupos. ¿En cuál de ellos se registra mayor variabilidad?
- ¿Proporcionan estos datos indicios de que el tipo de registro esté asociado a la adherencia en algún sentido? Fundamente su respuesta.

III)

Tiempo de práctica durante las 8 semanas de entrenamiento según el tipo de registro

	Formato en papel (n=27)		Aplicación móvil (n=21)	
	Media	Desv. Estándar	Media	Desv. Estándar
Nro. de días de práctica	46,78	8,059	38,10	8,882
Nro. de minutos por día de práctica	8,55	2,299	12,64	6,076
Total hs práctica en los 56 días	7,98	2,145	7,93	4,210

- A partir de la información precedente dé una conclusión sintética pero completa de la comparación de ambos grupos en cuanto al tiempo de práctica.
- Sobre la base del cálculo de un resumen estadístico adecuado, decida cuál de los dos grupos es más homogéneo en cuanto al número de días de práctica.
- Calcule el promedio de días de práctica de la muestra del total de los participantes.

8.- El Inventario Burnout o «síndrome de quemarse por el trabajo» de Maslach-Escala General de Schaufeli et al. (1996) es un constructo psicológico en el que se diferencian tres subescalas. Las medias y desviaciones estándar de los puntajes de dichas subescalas en el grupo normativo se muestran en la siguiente tabla:

	Agotamiento (Fatiga física y emocional)	Cinismo (Indiferencia o actitudes de distanciamiento hacia el trabajo)	Eficacia Profesional (Sentido de autocompetencia al desempeñar el trabajo)
Media	2,15	1,55	4,83
Desv.estándar	1,26	1,16	1

Ordene de modo creciente en una escala común los puntajes obtenidos por un trabajador a quien se le administró este instrumento, según la siguiente información:

Obtuvo un puntaje en *Agotamiento* que está a un desvío por encima de la media.

Obtuvo un puntaje directo de 0,57 en *Cinismo*.

Obtuvo un puntaje en T igual a 40 en *Eficacia Profesional*.

9.- Ubique el artículo de Martín-Ramos, R. et al (2015) en el material de la Cátedra o en la web:

<https://www.medigraphic.com/pdfs/arcneu/ane-2015/ane154a.pdf>. Lea los párrafos de Objetivo (p.234), *Instrumentos de evaluación psicológica* (p. 235), Tabla 2 (p. 235), Tabla 3 (p. 236) y Tabla 4 (p. 236).

- a) En la Tabla 3, encuentre los dos casos de variables más correlacionadas y los dos casos de variables menos correlacionadas e indique cuáles son dichas correlaciones.
- b) La correlación entre los puntajes en el test de Errores Perseverativos y en el Inventario de Depresión de Beck (BDI) fue 0,398.
 - b1) Interprete el signo positivo de esta correlación.
 - b2) ¿Qué porcentaje de la variabilidad de los puntajes en el test de Errores Perseverativos es explicado por el Puntaje en Depresión de Beck? Verifique la correspondencia entre el resultado obtenido en la Tabla 3 con el que se muestra en la Tabla 4.
- c) La correlación entre los puntajes en el test de Fluidez Verbal Semántica (COWA) y en Ansiedad Rasgo (IDARE) fue -0,414.
 - c1) Interprete el signo negativo de esta correlación.
 - c2) ¿Qué porcentaje de la variabilidad de los puntajes en el test de Fluidez verbal semántica es explicado por el Puntaje en Ansiedad rasgo? Verifique la correspondencia entre el resultado obtenido en la Tabla 3 con el que se muestra en la Tabla 4.
- d) ¿De cuántos análisis de regresión dan cuenta los resultados de la Tabla 4? y en cuáles dos de ellos se logró un mejor ajuste?
- e) ¿Cuáles son las dos rectas para las que se obtuvieron los mejores ajustes? Justifique.
- f) Utilice la información sobre las medias de la Tabla 2 y la información sobre las pendientes de la Tabla 4 para deducir las ecuaciones de las dos rectas de regresión mencionadas en e).
- g) Si una persona puntúa 18 en el BDI y obtiene 36 puntos en Diseño con Cubos, ¿cuál es el error de predicción?

En los siguientes ejercicios señale la única opción correcta.

10.- Las tablas muestran salidas de Statistix referidas a las notas de dos grupos de alumnos en un examen de inglés. Tomando como criterio de comparación el porcentaje de aprobados y siendo 7 la nota mínima de aprobación ¿podría decirse que alguno de los grupos tuvo mejor rendimiento que el otro? ¿Por qué?

GRUPO 1					GRUPO 2				
Value	Freq	Percent	Cumulative		Value	Freq	Percent	Cumulative	
			Freq	Percent				Freq	Percent
4	4	20.0	4	20.0	5	3	20.0	3	20.0
5	5	25.0	9	45.0	7	4	26.7	7	46.7
6	5	25.0	14	70.0	8	3	20.0	10	66.7
7	4	20.0	18	90.0	9	3	20.0	13	86.7
8	1	5.0	19	95.0	10	2	13.3	15	100.0
10	1	5.0	20	100.0	Total	15	100.0		
Total	20	100.0							

- a) Sí; el grupo 2 porque tiene un porcentaje de aprobados del 26,7% en tanto que en el grupo 1 sólo fue del 20%.
- b) No puede saberse debido a que los grupos tienen diferente número de unidades de análisis.
- c) Sí, el grupo 2 porque tiene un porcentaje de aprobados del 80% en tanto que en el grupo 1 fue sólo del 30%.

11.- En un diagrama de dispersión, cada punto representa:

- a) Un elemento de la recta de regresión.
- b) La media y la desviación estándar de un test.
- c) Un par de observaciones de una unidad de análisis.

12.- El análisis de resultados de una tabla de contingencia de 4x3 para dos variables cualitativas dio indicios de que éstas eran independientes. ¿De cuál de los siguientes resultados se trata?

- a) $r = 0$
- b) $\chi^2_{obs} = 37$
- c) V de cramer = 0,02

REFERENCIAS

Bárbara Horrillo-Álvarez, Carolina Marín-Martín y Manuel R. Abuín (2019). La Adherencia al Entrenamiento en Meditación Mindfulness con Registro en Papel y en Aplicación Móvil, *Clínica y Salud* 30, 2, 99-108.

Maslach C, Jackson SE, Leiter MP, ed. Maslach (1996). Burnout Inventory Manual. 30 edición. Palo Alto, California: Consulting Psychologists Press, 19-26.

Martin Ramos, R., Jiménez Morales, R., Herrera Jiménez, L., Guerra Pérez, L., Martin Ramos, R. y Pérez Medinilla, Y. (2015). Influencia de la ansiedad y depresión sobre las funciones ejecutivas en pacientes con esclerosis múltiple. *Archivos de Neurociencias* 20, 4, 233-239.

Vargas Orozco, C; Villarreal Sotelo, K.; Guevara López, C.; Andrade Trillo, M. (2014). Depresión y ansiedad en estudiantes universitarios, *Revista de Psicología y Ciencias del Comportamiento* 5, 1, 36-44.

RESOLUCIÓN

1.- Los siguientes diagramas de tallo-hoja representan la distribución de los puntajes en una encuesta sobre la imagen positiva de dos precandidatos (A) y (B) de un mismo partido político entre los que se quiere decidir para proponer en las elecciones. Ambos diagramas corresponden a la misma muestra de sujetos encuestados, quienes expresaron su opinión en una escala de 0 a 100.

STATISTIX FOR WINDOWS

STEM AND LEAF PLOT OF A	STEM AND LEAF PLOT OF B																																																						
LEAF DIGIT UNIT = 1 4 2 REPRESENTS 42.	LEAF DIGIT UNIT = 1 3 2 REPRESENTS 32.																																																						
<table style="width: 100%; border: none;"> <thead> <tr> <th style="text-align: left;">STEM</th> <th style="text-align: left;">LEAVES</th> </tr> </thead> <tbody> <tr><td>2</td><td>4 24</td></tr> <tr><td>4</td><td>4 68</td></tr> <tr><td>7</td><td>5 234</td></tr> <tr><td>12</td><td>5 56789</td></tr> <tr><td>17</td><td>6 01234</td></tr> <tr><td>24</td><td>6 5566789</td></tr> <tr><td>(11)</td><td>7 01223334444</td></tr> <tr><td>25</td><td>7 555556666777889</td></tr> <tr><td>10</td><td>8 012344</td></tr> <tr><td>4</td><td>8 5678</td></tr> </tbody> </table>	STEM	LEAVES	2	4 24	4	4 68	7	5 234	12	5 56789	17	6 01234	24	6 5566789	(11)	7 01223334444	25	7 555556666777889	10	8 012344	4	8 5678	<table style="width: 100%; border: none;"> <thead> <tr> <th style="text-align: left;">STEM</th> <th style="text-align: left;">LEAVES</th> </tr> </thead> <tbody> <tr><td>3</td><td>3 234</td></tr> <tr><td>8</td><td>3 56789</td></tr> <tr><td>14</td><td>4 012344</td></tr> <tr><td>20</td><td>4 556789</td></tr> <tr><td>(11)</td><td>5 00112233344</td></tr> <tr><td>29</td><td>5 55556789</td></tr> <tr><td>21</td><td>6 01234</td></tr> <tr><td>16</td><td>6 5679</td></tr> <tr><td>12</td><td>7 0134</td></tr> <tr><td>8</td><td>7 56</td></tr> <tr><td>6</td><td>8 24</td></tr> <tr><td>4</td><td>8 5</td></tr> <tr><td>3</td><td>9 2</td></tr> <tr><td>2</td><td>9 5</td></tr> <tr><td>1</td><td>10 0</td></tr> </tbody> </table>	STEM	LEAVES	3	3 234	8	3 56789	14	4 012344	20	4 556789	(11)	5 00112233344	29	5 55556789	21	6 01234	16	6 5679	12	7 0134	8	7 56	6	8 24	4	8 5	3	9 2	2	9 5	1	10 0
STEM	LEAVES																																																						
2	4 24																																																						
4	4 68																																																						
7	5 234																																																						
12	5 56789																																																						
17	6 01234																																																						
24	6 5566789																																																						
(11)	7 01223334444																																																						
25	7 555556666777889																																																						
10	8 012344																																																						
4	8 5678																																																						
STEM	LEAVES																																																						
3	3 234																																																						
8	3 56789																																																						
14	4 012344																																																						
20	4 556789																																																						
(11)	5 00112233344																																																						
29	5 55556789																																																						
21	6 01234																																																						
16	6 5679																																																						
12	7 0134																																																						
8	7 56																																																						
6	8 24																																																						
4	8 5																																																						
3	9 2																																																						
2	9 5																																																						
1	10 0																																																						

a) ¿Cuál es el tamaño de la muestra de encuestados?

El tamaño de la muestra de encuestados coincide con la cantidad de hojas (LEAVES), dado que cada una se corresponde a la medición de la variable sobre cada sujeto. En este caso es 60. Sin necesidad de contar hoja por hoja, puede obtenerse fácilmente a partir de la primera columna de la izquierda de los tallo-hoja, donde se indican las frecuencias acumuladas: 24 hasta el segundo tallo con el valor 6, 11 dentro del primer tallo con el valor 7 y 25 a partir del segundo tallo con el valor 7; en total $24 + 11 + 25 = 60$. *Si lo entendió pruebe hacerlo con la distribución de la imagen de B. También se pueden contar las "hojas" de cualquiera de las dos distribuciones... ¡pero es más largo y engorroso!*

b) La siguiente tabla contiene en cada renglón los resúmenes estadísticos correspondientes a seis conjuntos de datos. **Sin hacer cálculos**, identifique qué renglón corresponde a la distribución A y qué renglón a la distribución B. Hágalo escribiendo A o B al lado del número de la fila elegida: 1), 2), 3), 4), 5) o 6). Sólo un renglón es para A y sólo un renglón es para B. En otras palabras, de las 5 opciones sólo dos son verdaderas.

Los diagramas de tallo-hoja muestran que la mediana de A (en el tallo 7) es mayor que la de B (en el tallo 5). En la columna MEDIAN de la tabla aparecen los valores 54 y 73, por lo que 54 es la mediana de B y 73 la de A. Por otra parte, también se aprecia que B es más variable que A. Como en la columna SD hay sólo dos valores diferentes: 15.822 y 11.208, el primero debe corresponder a B y el segundo a A. De los 6 renglones, el único que contiene una combinación donde la mediana es 73 y SD es 11.208 es el 4); por tanto éste corresponde a A. El renglón donde la mediana es 54 y SD es 15.822 es 1), que corresponde a B.

DESCRIPTIVE STATISTICS

	VARIABLE	MEAN	SD	1ST QUARTI	MEDIAN	3RD QUARTI
B	1)	56.583	15.822	45.000	54.000	65.750
	2)	69.750	15.822	62.250	73.000	77.000
	3)	56.583	11.208	45.000	54.000	65.750
A	4)	69.750	11.208	62.250	73.000	77.000
	5)	69.750	11.208	45.000	54.000	65.750
	6)	56.583	15.822	62.250	73.000	77.000

c) Señale con A y con B las cajas que corresponden a la distribución A y B respectivamente.

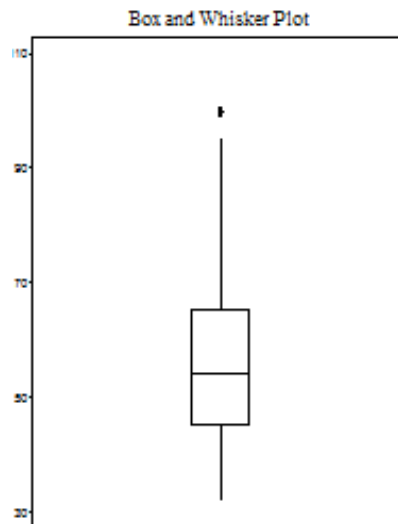
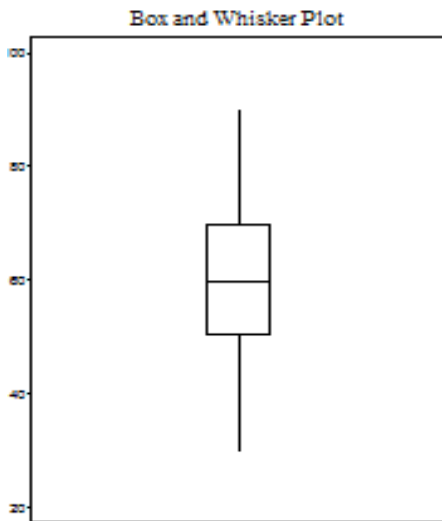
Recordemos los principios generales del diagrama caja y bigotes. La caja está delimitada por el primer y el tercer cuartil y los bigotes van desde dichos cuartiles hasta las observaciones más extremas dentro del rango de un intercuartil y medio. La longitud de cada bigote y de cada sección de la caja están dadas por la mayor o menor concentración de valores comprendidos en cada sección: del mínimo hasta la base de la caja (primer cuartil), desde esta base hasta la división de la caja (segundo cuartil o mediana), desde esta base hasta el techo de la caja (tercer cuartil) y desde este techo hasta la observación mayor.

Por eso es que el diagrama de caja y bigotes es útil para conocer rápidamente el tipo de asimetría de la distribución. Cuanto más centrada esté la mediana entre los dos cuartiles y cuanto más similares sean las longitudes de los bigotes, más simétrica es la distribución. Por el contrario, si un bigote se alarga más que otro y la mediana se acerca más a uno de los cuartiles, esto está hablando de mayor asimetría en alguna dirección.

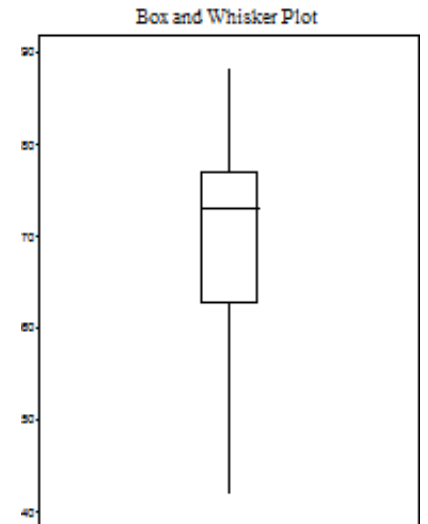
Una advertencia es que en estos diagramas de las cajas los valores van de menor a mayor desde abajo hacia arriba. En cambio, en los tallo-hoja los valores van de menor a mayor desde arriba hacia abajo.

En los diagramas de tallo-hoja se ve que ambas distribuciones A y B son asimétricas de distinto sentido: simetría negativa para A y positiva para B. De los tres diagramas de caja-bigotes, el primero corresponde a una distribución simétrica, por lo que queda descartado. En el diagrama central, el bigote más largo va para los valores más altos, por lo que se refiere a una asimetría positiva. Lo contrario ocurre con el tercer diagrama. Por tanto, el diagrama central corresponde a B y el tercero a A.

Nótese que el diagrama de caja-bigotes también sirve para detectar datos atípicos (outliers): los que están a más de un rango intercuartílico y medio hacia el lado del cuartil 1 o del cuartil 3. Eso se señala con un asterisco. En el diagrama del centro el asterisco corresponde al valor 100.



B



A

d) Sobre la base de la información obtenida ¿A cuál propondría como candidato? Fundamente su respuesta.

Propondría al candidato A, ya que los puntajes en su imagen son en general más altos, como muestran todas las medidas de tendencia central y la forma asimétrica negativa de su distribución de frecuencias.

2.- A continuación se exhiben dos conjuntos de datos. Corresponden a la cantidad de días que tardaron respectivamente los proveedores SALUTEM y VITAE en satisfacer el encargo del mismo producto a cierto hospital. A SALUTEM se le hicieron 10 encargos y a VITAE 20 encargos.

SALUTEM: 4 – 7 – 6 – 3 – 4 – 3 – 5 – 3 – 4 – 5

VITAE: 3 – 5 – 8 – 1 – 4 – 7 – 8 – 6 – 4 – 6 – 7 – 7 – 6 – 5 – 7 – 5 – 8 – 8 – 6 – 7

a) Mencione todas las variables que intervienen en el problema y clasifíquelas. Indique también el nivel de medición.

La variable clasificatoria es el proveedor. Es una variable cualitativa de nivel nominal.

La variable dependiente es la cantidad de días que tardó el proveedor en satisfacer el encargo al hospital. Es cuantitativa discreta (se consideran días enteros) de nivel de razón.

b) Organice la información en tablas de frecuencia.

La tabla de frecuencias tiene en la primera columna los valores de la variable, en la segunda las frecuencias absolutas (la cantidad de veces que se repite cada observación) y puede tener una tercera columna con las frecuencias porcentuales (f_i/n por 100).

Para realizarla es necesario ordenar los valores de la variable de menor a mayor y colocarlos en la primera columna. Y en las restantes completar las frecuencias respectivas.

Cantidad de días	PROVEEDOR			
	SALUTEM		VITAE	
	f_i	$f_i\%$	f_i	$f_i\%$
1	-	-	1	5
2	-	-	-	-
3	3	30	1	5
4	3	30	2	10
5	2	20	3	15
6	1	10	4	20
7	1	10	5	25
8	-	-	4	20
	10	100	20	100

c) Exhiba los datos en dos diagramas de tallo – hoja.

El diagrama de tallo-hoja organiza las observaciones partiendo en dos cada dato: si los números por representar tienen dos dígitos, se encolumnan en el tallo (stem) los dígitos correspondientes a las decenas y, para cada tallo, se colocan en el mismo renglón los dígitos correspondientes a las unidades, que son las hojas.

Si los datos son solo unidades, se encolumnan en el tallo de menor a mayor y en cada fila, se ponen tantos ceros como corresponda por cada unidad, su frecuencia absoluta (por ejemplo si el 3 se repite 4 veces, se pone 3 en el tallo y 0 0 0 0 en la fila).

Los obtendremos utilizando Statistix

Statistix 8.0

PrRepaso2, 25/04/2020, 2:35:38

Stem and Leaf Plot of x for y = SALUTEM

Leaf Digit Unit = 0.1
3 0 represents 3.0

Minimum 3.0000
Median 4.0000
Maximum 7.0000

Stem	Leaves
3	3 000
(3) 4	000
4	5 00
2	6 0
1	7 0

10 cases included 0 missing cases

Stem and Leaf Plot of x for y = VITAE

Leaf Digit Unit = 0.1
 1 0 represents 1.0

Minimum 1.0000
 Median 6.0000
 Maximum 8.0000

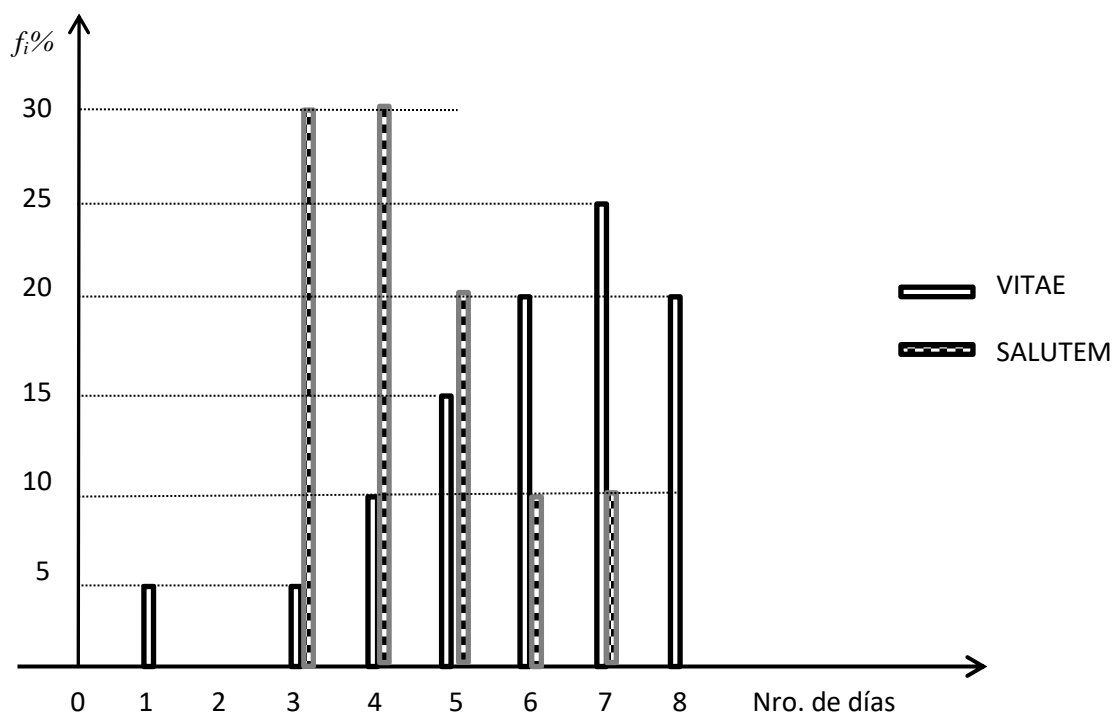
Stem	Leaves
1	1 0
1	2
2	3 0
4	4 00
7	5 000
(4)	6 0000
9	7 00000
4	8 0000

20 ases included 0 missing cases

d) Si quisiera comparar las distribuciones de frecuencia de los tiempos para ambas sucursales en un mismo gráfico ¿qué tipo de gráfico sería apropiado? Puede bosquejarlo o, si está familiarizado con una aplicación, hacerlo con precisión.

El gráfico pertinente para esta variable es el de bastones (barras delgadas). Este gráfico indica en el eje horizontal los valores de la variable discreta de menor a mayor y en el eje vertical las frecuencias.

En este caso se desean representar en un mismo gráfico dos distribuciones. Entonces se realiza un diagrama de bastones (o barras) adyacentes. Sobre cada valor se ubican dos bastones: uno para la primera distribución y otro para la segunda (en colores o trazos diferentes). Cuando se trabaja con una sola distribución se utilizan indistintamente las frecuencias absolutas, relativas o porcentuales pero si se desean comparar dos distribuciones que provienen de tamaños de muestra diferentes (como es el presente caso), deben utilizarse las frecuencias relativas o porcentuales para que puedan ser comparables. Veremos más adelante, que lo que se está comparando son las distribuciones condicionales de la cantidad de días dado el proveedor.



e) Calcule los resúmenes numéricos adecuados para comparar ambas distribuciones en todos los aspectos de las mismas.

	SALUTEM	VITAE
MODA	3,5	7
MEDIANA	4	6
MEDIA	4,4	5,9
AMPLITUD TOTAL	4	7
DESVIACIÓN ESTÁNDAR	1,3499	1,8610
COEFICIENTE DE VARIACIÓN	30,679	31,542
ASIMETRÍA	Positiva (0,6114)	Negativa (-0,9572)
CURTOSIS	-0,6163	0,5123

Como en este curso no hemos hecho hincapié en las fórmulas de asimetría y curtosis para su cálculo manual, estos índices se obtuvieron con un programa; los demás índices pueden obtenerse también con un programa o manualmente. Para los objetivos de este curso, es suficiente con detectar visualmente el tipo de asimetría.

f) Dé una conclusión breve pero completa de la comparación entre las dos distribuciones.

El Proveedor SALUTEM tarda, en general, menos días en satisfacer los encargos (como lo muestran las medidas de tendencia central y el sentido de la asimetría). También ofrece menor incertidumbre en cuanto a los tiempos de entrega, como expresan todas las medidas de variabilidad, que son menores en el caso de SALUTEM.

En efecto, las medidas de tendencia central de SALUTEM son menores que las de VITAE. Además, por la asimetría se aprecia que las mayores concentraciones de observaciones, en el caso de VITAE se dan en los valores mayores de la variable (ver el gráfico de barras y observar la mayor concentración hacia la derecha), mientras que en SALUTEM las mayores concentraciones de frecuencias están en los valores menores de la variable.

En cuanto a la curtosis, indica el mayor o menor amesetamiento de la distribución. El análisis de la curtosis es más relevante cuando se compara la distribución observada con el modelo normal (se verá más adelante); cuando es positiva es más espigada (leptocúrtica) que la normal y cuando es negativa es más aplanada (platicúrtica). En el caso de VITAE es más leptocúrtica y en el caso de SALUTEM, platicúrtica.

g) Sobre la base del análisis anterior, y suponiendo que ambos venden al mismo precio, ¿recomendaría a alguna de los dos proveedores por sobre el otro?

Elegiría a SALUTEM porque suele entregar antes que VITAE y es más previsible en cuanto al tiempo de entrega de los encargos.

3.- En un programa familiar de entretenimientos compiten un equipo de 6 niños con un equipo de 4 padres. El entretenimiento consiste en *cada una* de las 10 personas tiene que escribir cuantas palabras pueda en un minuto que comiencen con una letra dada. **Gana el equipo que en total haya producido la mayor cantidad de palabras** (es decir, se suma la cantidad de palabras producidas por cada miembro del equipo).

Se sabe que la cantidad promedio de palabras de las 10 personas fue 8 y que la media del equipo de niños fue 7. ¿Qué equipo ganó la competencia? Justifique con base en los cálculos adecuados.

Como la media del grupo de los niños es 7 y hay 6 niños, la suma de sus puntajes debió ser 42. Siendo 8 la media general, la suma de todos los puntajes fue 80. De modo que la suma de los puntajes de los adultos fue $80 - 42 = 38$. Por tanto, ganó el equipo de los niños.

4.- Leila es psicóloga y trabaja para las obras sociales A y B. Durante el último año llevó registro del tipo de trastorno que presentaba cada uno de los pacientes que era diagnosticado con un trastorno mental o de la conducta para cada obra social y organizó la información en la siguiente tabla.

Tipo de Trastorno	OBRA SOCIAL A	OBRA SOCIAL B	TOTAL
Ataques de pánico	5	3	8
Depresión	8	9	17
Ansiedad	14	4	18
Alteraciones en el sueño	6	1	7
Déficit de atención	7	3	10

a) Mencione las variables que intervienen en el problema, clasifíquelas e indique nivel de medición.

Las variables que intervienen son:

El Tipo de Trastorno. Cualitativa de nivel nominal.

Obra Social. Cualitativa de nivel nominal.

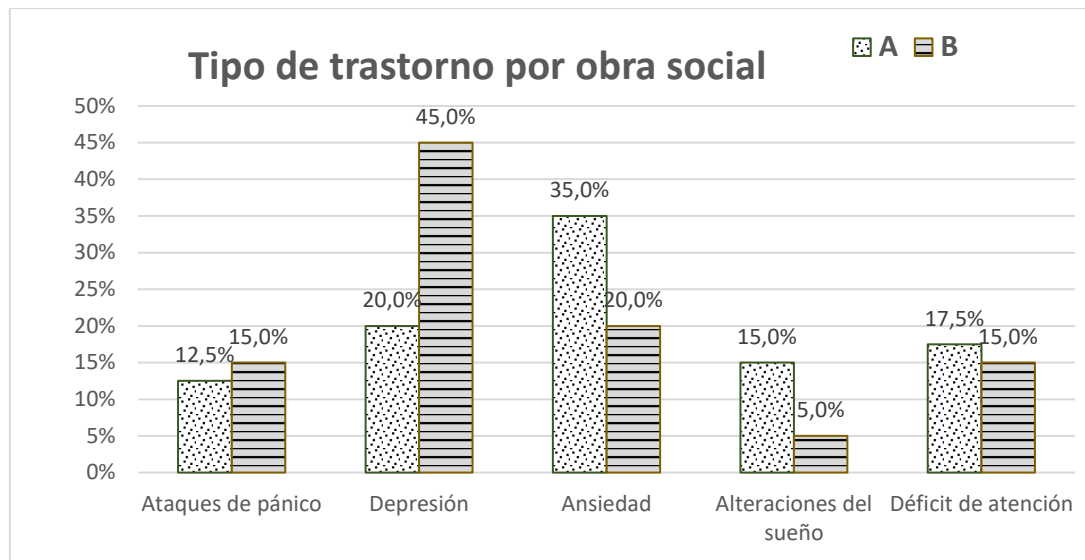
b) Indique qué diagrama sería adecuado para comparar las distribuciones de frecuencias de los tipos de trastornos en ambas obras sociales. Puede hacer un bosquejo manual del mismo o utilizar una aplicación para obtenerlo. Saque conclusiones de la comparación.

El diagrama apropiado para la variable “tipo de trastorno”, es el de barras y para comparar dicha variable en las dos obras sociales, corresponde hacer barras adyacentes. En otras palabras, lo que se comparan son las distribuciones condicionales del Tipo de Trastorno dada la Obra Social.

En el eje horizontal colocamos los distintos trastornos y en el vertical las frecuencias.

Si las dos obras sociales tuvieran la misma cantidad de pacientes, podríamos utilizar la frecuencia absoluta pero, como la obra social A tiene 40 y la B 20, corresponde utilizar las relativas o porcentuales para poder comparar ambas obras sociales. Se las diferencia con distintos tipos o colores de trazos y el ancho de las barras debe ser igual. Completaremos primero la tabla con las frecuencias porcentuales.

Tipo de Trastorno	OBRA SOCIAL A		OBRA SOCIAL B	
	f_i	$f_i\%$	f_i	$f_i\%$
Ataques de pánico	5	12,5	3	15,0
Depresión	8	20,0	9	45,0
Ansiedad	14	35,0	4	20,0
Alteraciones en el sueño	6	15,0	1	05,0
Déficit de atención	7	17,5	3	15,0
Total	40	100,0	20	100



En ambas obras sociales los pacientes consultan por los cinco tipos de trastornos. En A es más frecuente que consulten por *Ansiedad* y en B por *Depresión*. En la Obra Social B hay menos variabilidad que en A, ya que un 45% de los casos corresponden a *Depresión* mientras que en B, un 35% se concentra en *Ansiedad*. Las mayores diferencias entre ambas obras sociales se registran en los trastornos de *Depresión*, *Ansiedad* y *Alteraciones del sueño*.

c) Si se presenta un paciente de la Obra Social A

c1) ¿Qué trastorno mental o de la conducta es más probable que padezca?

Es más probable que padezca de *Ansiedad* (porque es la moda de la distribución). En efecto, por ser el trastorno de ansiedad el de mayor frecuencia en la Obra Social A, si se presenta un paciente de A, es más probable que padezca de ansiedad.

c2) ¿Y para la Obra Social B?

Es más probable que padezca de *Depresión*.

c3) ¿En cuál de los dos casos la incertidumbre es mayor?

Para conocer en cuál de los dos casos hay más incertidumbre con respecto al trastorno más probable se utiliza el concepto de entropía. Si se calculara, la obra social con mayor entropía sería en la cual hay mayor incertidumbre para prever cuál es el trastorno del paciente correspondiente, dado que la mayor entropía indica mayor incertidumbre.

Otra manera más sencilla para detectar en cuál existe mayor incertidumbre, sin necesidad de calcular la entropía, es considerar la relevancia que tiene la frecuencia porcentual de la moda en cada obra social: en la A, la moda (*Ansiedad*), alcanza un porcentaje del 35%; mientras que en la B, su moda (*Depresión*) alcanza un 45%. Por lo cual puede preverse que es más factible en la A encontrar un paciente con ansiedad, y en la B un paciente con depresión. Pero como el porcentaje de la moda de la A es menor que el de la B, se puede inferir que en la A la previsión tiene menos fuerza que en la B, en la cual el porcentaje es mayor. En otras palabras, puede observarse ocularmente que en la Obra Social A las frecuencias están más repartidas entre

todas las clases mientras que en B están más concentradas en menos clases: más concentración en *Depresión* y poca en *Alteraciones en el sueño*. Por tanto, hay mayor incertidumbre en la A.

d) ¿Qué resumen estadístico es adecuado para expresar la centralidad de los datos del tipo de trastorno? El único resumen adecuado para ello es la *Moda* porque es una variable de nivel nominal.

e) Si utilizara un diagrama circular para representar la distribución del tipo de trastorno sin distinguir Obra Social, ¿qué ángulo debería dar a cada sector?

Cada ángulo debe ser proporcional a la frecuencia de cada clase. Para ello se multiplica la frecuencia relativa por 360º como se muestra en la tabla.

Tipo de Trastorno	OBRA SOCIAL A	OBRA SOCIAL B	TOTAL		
			f_i	f_i'	Ángulo
Ataques de pánico	5	3	8	8/60	360º x 8/60 = 48º
Depresión	8	9	17	17/60	360º x 17/60 = 102º
Ansiedad	14	4	18	18/60	360º x 18/60 = 108º
Alteraciones en el sueño	6	1	7	7/60	360º x 7/60 = 42º
Déficit de atención	7	3	10	10/60	360º x 10/60 = 60º

f) Calcule un coeficiente adecuado para expresar el grado de asociación muestral entre la Obra Social y el Tipo de Trastorno.

Utilizaremos el coeficiente **V** de Cramer. Para ello debemos calcular previamente el coeficiente χ^2 de Pearson, lo cual requiere obtener la tabla con frecuencias esperadas bajo la hipótesis de independencia, a saber:

Tipo de Trastorno	OBRA SOCIAL A	OBRA SOCIAL B	TOTAL
Ataques de pánico	8x40/60 = 5,33	20x8/60 = 2,67	8
Depresión	40x17/60 = 11,33	20x17/60 = 5,67	17
Ansiedad	40x18/60 = 12	20x18/60 = 6	18
Alteraciones en el sueño	40x7/60 = 4,67	20x7/60 = 2,33	7
Déficit de atención	40x10/60 = 6,67	20x10/60 = 3,33	10
TOTAL	40	20	60

$$\chi^2 = (5-5,33)^2/5,33 + (3-2,67)^2/2,67 + (8-11,33)^2/11,33 + (9-5,67)^2/5,67 + (14-12)^2/12 + (4-6)^2/6 + (6-4,67)^2/4,67 + (1-2,33)^2/2,33 + (7-6,67)^2/6,67 + (3-3,33)^2/3,33 = 5,1826$$

Cálculo del Coeficiente V de Cramer

El mínimo entre número de columnas menos 1: (2-1) y número de filas menos 1: (5-1) es 2-1=1.

$$\text{Coeficiente V de Cramer} = \sqrt{\frac{\chi^2}{n \cdot \min(f-1, c-1)}} = \sqrt{\frac{5,1826}{60 \times 1}} = 0,2939$$

5.- Treinta y cinco personas fueron encuestadas con respecto a la penalización en años de un homicida antes y después de ver un video sobre su vida. Se les preguntó con cuántos años penalizarían al homicida. La distribución de frecuencias se halla en la tabla. Sobre la base de las medidas de tendencia central adecuadas, analice los datos y concluya si el video tuvo efecto en la decisión de los espectadores y cómo incidió.

IMAGEN	ANTES	DESPUÉS
PERPETUA	10	6
LARGA	15	11
MODERADA	7	10
BREVE	3	7
NULA		1

La variable en cuestión es la opinión sobre la penalización en años de un homicida, pero medida en una escala ordinal, con valores de menor a mayor: nula, breve, moderada, larga o perpetua.

Dicha variable fue medida en treinta y cinco personas antes de ver un video sobre la vida del homicida y también fue medida sobre las mismas personas después de haber visto un video de la vida del homicida (datos pareados).

Para apreciar si la opinión de los encuestados fue modificada por haber visto el video, teniendo en cuenta el tipo de variable (cualitativa ordinal), se pueden utilizar la *Mediana* y la *Moda*.

La *Moda* no es tan clara para representar la opinión después del video porque hay dos frecuencias similares más altas que las demás: 11 y 10. No obstante podría decirse que antes del video la duración de la pena más elegida (moda) fue *Larga* (elegida por casi el 43%) mientras que después la frecuencia de esta categoría cayó a un 31% y la categoría *Moderada* pasó a concentrar una frecuencia de casi el 28%, que la acerca a la *Moda*. En cuanto a la *Mediana*, habiendo 35 observaciones, corresponde a la observación nro. 18, que antes del video está en la categoría *Larga* y después del video en *Moderada*. Por tanto vemos, que tanto con la *Moda* como con la *Mediana* hubo un desplazamiento de las opiniones hacia relajar la longitud de la pena. Por eso puede decirse que el video parece haber tenido un efecto en la decisión mitigando la duración de la pena.

6.- A continuación se muestra el resumen y parte de los resultados de la investigación de Vargas Orozco et al (2014).

En esta investigación presentamos los índices de Depresión y Ansiedad de los estudiantes de la Universidad Autónoma de Tamaulipas campus Reynosa-Aztlán. La muestra estuvo conformada por 500 estudiantes clasificados por género y por programa académico, se aplicaron dos instrumentos estandarizados para medir depresión, se utilizó la Escala de Depresión de Beck y la Escala de Ansiedad de Hamilton, los resultados nos muestran que existe un mayor puntaje en los cuadros depresivos en el sexo femenino así como en los índices de ansiedad.

Sexo	NIVELES DE DEPRESIÓN			
	Normal	Leve	Moderada	Severa
Femenino	176	33	8	2
Masculino	259	15	6	1

Sexo	ANSIEDAD	
	Con ansiedad	Sin ansiedad
Femenino	169	50
Masculino	101	180

- a) Mencione y clasifique todas las variables a las que se alude en el resumen y cuyos resultados de distribuciones de frecuencias se exhiben en las tablas.

Variabes: Nivel de depresi3n a partir del Puntaje en la escala de Beck.

Por los valores que especifica la tabla (normal, leve, moderada, severa) es una variable cualitativa ordinal

Presencia de (alto nivel de) ansiedad a partir del puntaje en la escala de Hamilton.

Por los valores que especifica la tabla (con o sin) es una variable cualitativa nominal.

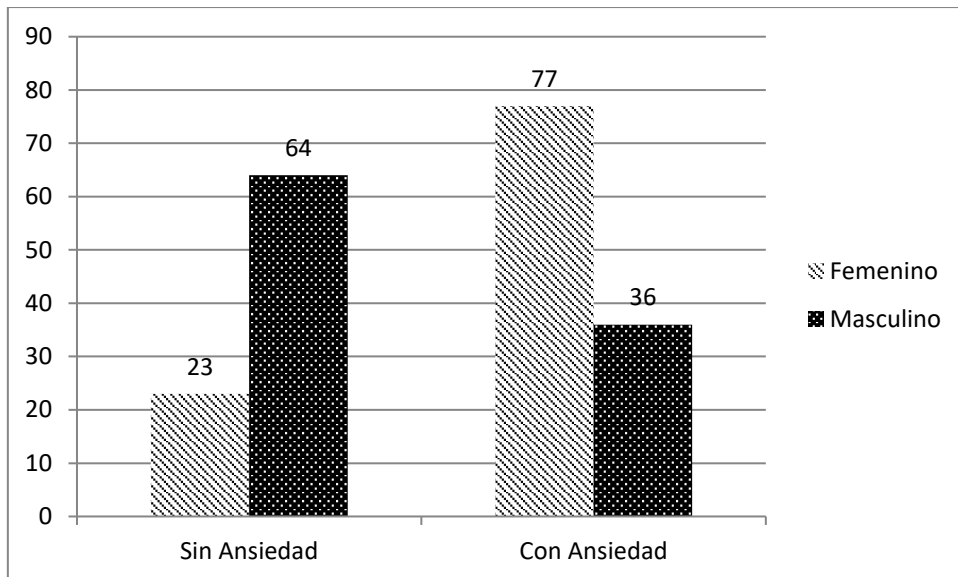
Sexo: cualitativa nominal.

- b) b1) Obtenga las tablas de frecuencias porcentuales para las distribuciones condicionales de la depresi3n segun el sexo. ¿Cuál presenta mayor variabilidad? Fundamente.

Sexo	Normal	Leve	Moderada	Severa	Total
Femenino	176	33	8	2	219
	80,37%	15,07%	3,65%	0,91%	100%
Masculino	259	15	6	1	281
	92,17%	5,34%	2,14%	0,36%	100%
	435	48	14	3	500
	87,00%	9,60%	2,80%	0,60%	100%

La mayor variabilidad se da en las personas de sexo femenino porque, si bien en general hay concentraci3n de los casos en valor *normal* y frecuencias menores a medida que nos alejamos de él, en las personas de sexo femenino, la frecuencia se reparte algo más que en el caso de los varones entre el resto de las clases, especialmente la *leve*.

- b2) Grafique las distribuciones condicionales de la ansiedad segun el sexo en un mismo gráfico de modo que puedan ser comparadas.



b3) Sobre la base de b1) y de b2) saque una conclusión de la comparación entre sexos.

En ambos sexos los niveles predominantes de depresión corresponden a la normalidad. Las dos distribuciones de los niveles de depresión según el sexo son similares aunque hay algo más de variabilidad en el caso de las mujeres donde los niveles de depresión leve y moderado concentran un porcentaje mayor de observaciones que en el grupo de los varones. Con respecto a la Ansiedad, la distribución no sólo es diferente entre los sexos sino opuesta: las mujeres se presentan mayoritariamente *Con Ansiedad* y los varones *Sin Ansiedad*. En el caso de los varones hay algo más de variabilidad, ya que la moda concentra menos frecuencia (64%) que en el caso de las mujeres (77%).

b4) Obtenga las distribuciones marginales de la Depresión y de la Ansiedad. ¿Qué medidas de tendencia central son adecuadas para resumir estos datos? Calcúlelas e interprete.

Depresión	f_i	$f_i \%$
Normal	435	87,00%
Leve	48	9,60%
Moderada	14	2,80%
Severa	3	0,60%
Total	500	100,00%

Ansiedad	f_i	$f_i \%$
Con ansiedad	270	54%
Sin ansiedad	230	46%
	500	100%

Para Depresión es adecuada la mediana, puesto que su nivel de medición es ordinal, y su valor es *Normal*.

Para Ansiedad, la única posible por ser de nivel de medición nominal, es la moda que es *Con Ansiedad*.

b5) ¿En qué medida, a nivel muestral, el nivel de depresión, por un lado, y la ansiedad, por el otro, están asociados al sexo?

Calcularemos los coeficientes de asociación de Cramer y de Kendall-Yule respectivamente.

Tabla con las frecuencias esperadas de los niveles de depresión bajo la hipótesis de independencia entre sexos.

Sexo	NIVELES DE DEPRESIÓN				Total
	Normal	Leve	Moderada	Severa	
Femenino	$435 \times 219 / 500 = 190,53$	$48 \times 219 / 500 = 21,02$	$14 \times 219 / 500 = 6,13$	$3 \times 219 / 500 = 1,31$	219
Masculino	$435 \times 281 / 500 = 244,47$	$48 \times 281 / 500 = 26,98$	$14 \times 281 / 500 = 7,87$	$3 \times 281 / 500 = 1,69$	281
Total	435	48	14	3	500

$$\chi^2 = (176-190,53)^2/190,53 + (33-21,02)^2/21,02 + (8-6,13)^2/6,13 + (2-1,31)^2/1,31 + (259-244,47)^2/244,47 + (15-26,98)^2/26,98 + (6-7,87)^2/7,87 + (1-1,69)^2/1,69 = 15,78$$

$$\text{Coeficiente V de Cramer} = \sqrt{\frac{\chi^2}{n \cdot \min(f-1, c-1)}} = \sqrt{\frac{15,78}{500 \times 1}} = 0,1777$$

Coeficiente Q de Kendall-Yule

Sexo	ANSIEDAD	
	Con ansiedad	Sin ansiedad
Femenino	169	50
Masculino	101	180

$$Q = (169 \times 180 - 50 \times 101) / (169 \times 180 + 50 \times 101) = 0,7153$$

De acuerdo con lo que ya se observaba gráficamente, el Sexo está asociado con la Ansiedad mucho más que con la Depresión, en esta muestra.

- c) Si la edad promedio de las mujeres fue de 22,8 años y la de los varones de 23,2 ¿Cuál es fue la edad promedio del total de estudiantes?

$$\text{Media general para la edad} = (22,8 \times 219 + 23,2 \times 281) / (219 + 281) = 23,0248$$

7.- A continuación se muestra el resumen e instrumentos de la investigación de Horrillo-Álvarez, C. et al (2019).

Los efectos beneficiosos de la meditación mindfulness (MM) se han observado en numerosos estudios, pudiendo ser la atención y conciencia en el momento presente una variable esencial. Sin embargo, la adherencia al entrenamiento de la práctica meditativa es en ocasiones baja. Se investigaron los efectos del entrenamiento en MM sobre la atención y conciencia y si el formato de registro de la práctica de MM, formato papel (PP) (n = 27) o aplicación móvil (APP) (n = 21), produce diferencias en la adherencia al entrenamiento y en la atención y conciencia. Los resultados mostraron un aumento de la atención y

conciencia tras el entrenamiento, independientemente del formato de registro de la práctica. La APP redujo el abandono al entrenamiento y aumentó el tiempo de práctica de MM cada día que se practicó.

Instrumentos

Escala de atención y conciencia en el momento presente (Mindfulness Attention Awareness Scale - MAAS). Compuesta por 15 ítems permite obtener una puntuación entre 15 y 90 puntos.

Aplicación de móvil Meditate. Diseñada ad hoc, registra la fecha y los minutos meditados cada vez que se practica MM durante las 8 semanas de entrenamiento. El tiempo máximo que permite practicar la aplicación son 30 minutos diarios.

Hoja de registro de la práctica de MM. En formato papel y creada ad hoc, recoge la fecha y los minutos de MM cada día que se practica.

I)

a) Mencione las variables bajo estudio, clasifíquelas según sus valores y nivel de medición e indique cuáles son las dependientes y las independientes.

1. Formato del registro: papel/digital. Cualitativa, nominal.
2. Entrenamiento en MM: sí/no. Cualitativa, nominal.
3. Puntaje en la Escala de atención y conciencia en el momento presente (MAAS). Cuantitativa continua de nivel intervalar (no está detallada la construcción de la escala para interpretar su origen).
4. Fecha en que se meditó (registro papel o APP): Se registra día, mes y año. En este caso no interesa para su tratamiento estadístico sino sólo para control y seguimiento del sujeto en su experiencia. Puede ser considerada cualitativa o cuantitativa, según el contexto. Por ejemplo, si interesara cada cuántos días se medita dentro de un mes (frecuencia de meditación mensual) se estaría considerando la fecha como cuantitativa. Como variable cuantitativa, la fecha es discreta de nivel intervalar.
5. Tiempo de práctica medida en minutos diarios meditados, durante las 8 semanas de entrenamiento (registro papel o APP). Cuantitativa continua. Nivel de medición de razón.
6. Adherencia al Entrenamiento. Cualitativa, nominal.

Las variables independientes son las mencionadas en 1 y 2.

Las variables dependientes son las mencionadas en 3, 5 y 6.

b) ¿Qué resúmenes estadísticos numéricos elegiría y cómo lo utilizaría para:

b1) clasificar a los sujetos en dos polos opuestos de menor MM y mayor MM.

Elegiría los cuartiles 1 y 3 para determinar dos polos de mayor y menor MM; es decir, considerar la cuarta parte de las personas con mayores niveles de MM y la cuarta parte de las personas con menores niveles de MM. También podrían elegirse otros percentiles; por ejemplo el 30 y el 70, siempre dejando un porcentaje en la parte central de la distribución de modo que los dos polos queden bien diferenciados y dependiendo de qué tanto se los quiera diferenciar. En una investigación más amplia y con mayores tamaños de muestra, estos grupos podrían utilizarse para explorar y comparar; por ejemplo, los hábitos y estrategias de cada uno de estos grupos.

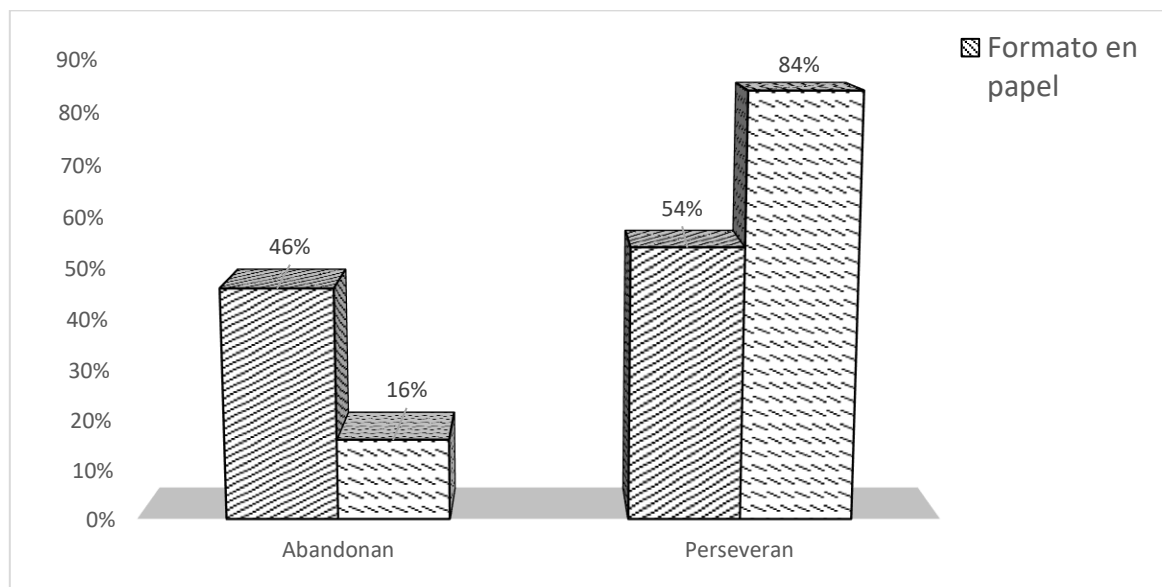
b2) decidir si el formato de registro de la práctica incide en la atención y conciencia.

Compararía las medias de los puntajes de MM entre los dos grupos. El tipo de registro con una media superior indicaría mayor efectividad de la práctica.

II)

Adherencia en el entrenamiento		
Tipo de registro	Abandonan	Perseveran
Formato en papel	23	27
Aplicación móvil	4	21

a) Represente con un diagrama adecuado la distribución de la adherencia según el tipo de registro de modo que pueda indagar si éste tiene incidencia sobre la misma.



El gráfico muestra que son más los que perseveran que los que no; pero que la adherencia se da en mayor medida en los que hicieron la meditación utilizando la aplicación móvil.

b) Obtenga una medida de tendencia central adecuada para comparar los dos grupos. ¿En cuál de ellos se registra mayor variabilidad?

La *Moda* es para ambos grupos: *Perseveran*. Hay mayor variabilidad en el grupo que hizo la meditación con el soporte en papel.

c) ¿Proporcionan estos datos indicios de que el tipo de registro esté asociado a la adherencia en algún sentido? Fundamente su respuesta.

Sí; resulta claro del gráfico que la práctica de MM en APP está asociada a mayor adherencia.

III)

Tiempo de práctica durante las 8 semanas de entrenamiento según el tipo de registro

	Formato en papel (n=27)		Aplicación móvil (n=21)	
	Media	Desv. Estándar	Media	Desv. Estándar
Nro. de días de práctica	46,78	8,059	38,10	8,882
Nro. de minutos por día de práctica	8,55	2,299	12,64	6,076
Total hs práctica en los 56 días	7,98	2,145	7,93	4,210

a) A partir de la información precedente dé una conclusión sintética pero completa de la comparación de ambos grupos en cuanto al tiempo de práctica.

En general, mirando las medias se puede concluir que, con aplicaciones móviles: la media de días de práctica es menor, pero los minutos diarios de práctica son mayores que con papel. Pero si se cuentan las horas en total resultan equivalentes.

b) Sobre la base del cálculo de un resumen estadístico adecuado, decida cuál de los dos grupos es más homogéneo en cuanto al número de días de práctica.

Siendo el tiempo una variable de nivel de razón, lo más apropiado para comparar la homogeneidad de los grupos es hacerlo calculando el coeficiente de variación (CV).

	Formato en papel	Aplicación móvil
CV	$(8,059/46,78) \times 100 = 17,23$	$(8,882/38,10) \times 100 = 23,31$

El grupo que utilizó el registro en papel fue más homogéneo en cuanto a la cantidad de días de ejercitación.

c) Calcule el promedio de días de práctica de la muestra del total de los participantes.

$$\text{Media general: } (27 \times 46,78 + 21 \times 38,10) / (27 + 21) = 42,9825$$

8.- El Inventario Burnout o «síndrome de quemarse por el trabajo» de Maslach-Escala General de Schaufeli et al. (1996) es un constructo psicológico en el que se diferencian tres subescalas. Las medias y desviaciones estándar de los puntajes de dichas subescalas en el grupo normativo se muestran en la siguiente tabla:

	Agotamiento (Fatiga física y emocional)	Cinismo (Indiferencia o actitudes de distanciamiento hacia el trabajo)	Eficacia Profesional (Sentido de autocompetencia al desempeñar el trabajo)
Media	2,15	1,55	4,83
Desv.estándar	1,26	1,16	1

Ordene de modo creciente en una escala común los puntajes obtenidos por un trabajador a quien se le administró este instrumento, según la siguiente información:

Obtuvo un puntaje en *Agotamiento* que está a un desvío por encima de la media.
Obtuvo un puntaje directo de 0,57 en *Cinismo*.
Obtuvo un puntaje en T igual a 40 en *Eficacia Profesional*.

Para ordenar los puntajes en una escala común es necesario estandarizarlos. Utilizaremos el puntaje z.

Agotamiento: $z = 1$ (es dato directo del enunciado: un desvío por encima de la media).

Cinismo: $z = (0,57 - 1,55) / 1,16 = -0,84$

Eficacia Profesional: $z = (40 - 50) / 10 = -1$

Al ordenar los puntajes z de menor a mayor: -1, -0,84, 1 se concluye que el trabajador tiene niveles eficacia y de cinismo menores que la media del grupo normativo, estando más lejos aún en eficacia que en cinismo. Por otra parte, tiene más agotamiento que la media del grupo normativo.

9.- Ubique el artículo de Martín-Ramos, R. et al (2015) en el material de la Cátedra o en la web:

<https://www.medigraphic.com/pdfs/arcneu/ane-2015/ane154a.pdf>. Lea los párrafos de Objetivo (p.234), *Instrumentos de evaluación psicológica* (p. 235), Tabla 2 (p. 235), Tabla 3 (p. 236) y Tabla 4 (p. 236).

a) En la Tabla 3, encuentre los dos casos de variables más correlacionadas y los dos casos de variables menos correlacionadas e indique cuáles son dichas correlaciones.

Las dos correlaciones más intensas se dieron entre el puntaje en el Inventario de Depresión de Beck (BDI) con el puntaje en el subtest de diseño con cubos de WAISS III, la cual fue de -0,547 y con el puntaje en Fluidez Semántica: -0,494.

Las variables menos correlacionadas, de correlación casi nula, fueron el puntaje en el subtest de IDARE de Ansiedad Estado y el subtest de clasificación de tarjetas de Wisconsin de Categorías alcanzadas: 0,007 y con el de Errores Perseverativos: 0,03.

b) La correlación entre los puntajes en el test de Errores Perseverativos y en el Inventario de Depresión de Beck (BDI) fue 0,398.

b1) Interprete el signo positivo de esta correlación.

El signo positivo significa que hay una relación creciente entre los puntajes en los tests de Depresión y de Errores Perseverativos; lo que se puede entender como que a hay una cierta tendencia a cometer más errores perseverativos cuando los niveles de depresión son más altos, y viceversa: menos errores en los niveles más bajos de depresión.

b2) ¿Qué porcentaje de la variabilidad de los puntajes en el test de Errores Perseverativos es explicado por el Puntaje en Depresión de Beck? Verifique la correspondencia entre el resultado obtenido en la Tabla 3 con el que se muestra en la Tabla 4.

La proporción de variabilidad explicada en el modelo lineal es el cuadrado del coeficiente de correlación lineal r de Pearson; esto es: $0,398^2 = 0,1584$. Por tanto puede decirse que el 15,84% de los Errores

Perseverativos es explicado por el puntaje en Depresión y viceversa. Esta misma información se halla en la Tabla 4, al ajustar las rectas de regresión, y corresponde al coeficiente de determinación R^2 que se halla justo debajo del título Errores Perseverativos, en su cruce la tercera columna BDI.

c) La correlación entre los puntajes en el test de Fluidez Verbal Semántica (COWA) y en Ansiedad Rasgo (IDARE) fue -0,414.

c1) Interprete el signo negativo de esta correlación.

El signo negativo indica que la relación entre dichas variables es decreciente; es decir, que a mayor Fluidez Verbal Semántica le corresponde menor Ansiedad Rasgo y viceversa.

c2) ¿Qué porcentaje de la variabilidad de los puntajes en el test de Fluidez verbal semántica es explicado por el Puntaje en Ansiedad rasgo? Verifique la correspondencia entre el resultado obtenido en la Tabla 3 con el que se muestra en la Tabla 4.

$R^2 = (-0,414)^2 = 0,1714$. Hacia el final de la Tabla 4, debajo del título Semántica y en la segunda columna, aparece el valor de $R^2 = 0,171$ (redondeado a 3 decimales). Por tanto el porcentaje de varianza explicada en el modelo lineal para estas variables es 17,14%.

d) ¿De cuántos análisis de regresión dan cuenta los resultados de la Tabla 4? y en cuáles dos de ellos se logró un mejor ajuste?

Se realizaron 24 análisis de regresión. Los mejores ajustes se evidencian por los valores de R^2 más altos: 0,299 y 0,244; que se refieren a la relación de los puntajes en el Inventario de Depresión (BDI) con los de los subtests de Diseño con Cubos y Fluidez Semántica respectivamente.

e) ¿Cuáles son las dos rectas para las que se obtuvieron los mejores ajustes? Justifique.

Llamando y = Puntaje en BDI, x_1 = Puntaje en Diseño con Cubos, x_2 = Puntaje en Fluidez Semántica, las dos rectas para las que se obtuvieron mejores ajustes fueron la de y sobre x_1 y la de y sobre x_2 , como lo indican los coeficientes de determinación mencionados en d).

f) Utilice la información sobre las medias de la Tabla 2 y la información sobre las pendientes de la Tabla 4 para deducir las ecuaciones de las dos rectas de regresión mencionadas en e).

Como las rectas se obtienen para las personas que padecen esclerosis múltiple (EM), en la Tabla 2 hay que mirar sólo esa columna para hallar las medias en las variables en cuestión:

Medias de y : 20,43 x_1 : 30,83 x_2 : 41,30

En la Tabla 4 se informan las pendientes de la rectas (β^*) que, para la regresión del BDI sobre el puntaje en Diseño con Cubos fue -1,56 y sobre el de Fluidez Semántica fue -0,861.

Por tanto las ecuaciones de las correspondientes rectas son:

$$y = -1,56 x_1 + b_1 \quad \text{e} \quad y = -0,861 x_2 + b_2$$

donde b_1 y b_2 son las ordenadas al origen de cada recta que hay que determinar. Para ello se utilizan las medias; ya que en las rectas de regresión el par de medias de (x,y) pertenecen a la recta.

Luego, reemplazando y y x_1 en la primera recta tenemos:

$$20,43 = -1,56 \cdot 30,83 + b_1. \text{ Despejando } b_1 \text{ resulta: } b_1 = 20,43 + 48,0948 = 68,5248 \approx 68,52$$

Análogamente, para la segunda recta:

$$20,43 = -0,861 \cdot 41,30 + b_2. \text{ Despejando } b_2 \text{ resulta: } b_2 = 20,43 + 35,5593 = 55,9893 \approx 55,99$$

Por tanto, las dos rectas de regresión son:

$$y = -1,56 x_1 + 68,52 \quad \text{e} \quad y = -0,861 x_2 + 55,99$$

g) Si una persona puntúa 18 en el BDI y obtiene 36 puntos en Diseño con Cubos, ¿cuál es el error de predicción?

Reemplazamos $x_1 = 36$ en la recta correspondiente para obtener el valor predicho en BDI:

$$y = -1,56 \cdot 36 + 68,52 = 12,36$$

Si el valor observado en el BDI para una persona con 36 puntos en Diseño con Cubos fue 18, entonces el error de predicción (residuo) es $18 - 12,36 = 5,64$.

10.- Las tablas muestran salidas de Statistix referidas a las notas de dos grupos de alumnos en un examen de inglés. Tomando como criterio de comparación el porcentaje de aprobados y siendo 7 la nota mínima de aprobación ¿podría decirse que alguno de los grupos tuvo mejor rendimiento que el otro? ¿Por qué?

GRUPO 1					GRUPO 2				
Value	Freq	Percent	Cumulative		Value	Freq	Percent	Cumulative	
			Freq	Percent				Freq	Percent
4	4	20.0	4	20.0	5	3	20.0	3	20.0
5	5	25.0	9	45.0	7	4	26.7	7	46.7
6	5	25.0	14	70.0	8	3	20.0	10	66.7
7	4	20.0	18	90.0	9	3	20.0	13	86.7
8	1	5.0	19	95.0	10	2	13.3	15	100.0
10	1	5.0	20	100.0	Total	15	100.0		
Total	20	100.0							

- a) Sí; el grupo 2 porque tiene un porcentaje de aprobados del 26,7% en tanto que en el grupo 1 sólo fue del 20%.
- b) No puede saberse debido a que los grupos tienen diferente número de unidades de análisis.
- c) Sí, el grupo 2 porque tiene un porcentaje de aprobados del 80% en tanto que en el grupo 1 fue sólo del 30%.

11.- En un diagrama de dispersión, cada punto representa:

- a) la media y la desviación estándar de un test.
- b) un par de observaciones de una unidad de análisis.
- c) un elemento de la recta de regresión.

12.- El análisis de resultados de una tabla de contingencia de 4x3 para dos variables cualitativas dio indicios de que éstas eran independientes. ¿De cuál de los siguientes resultados se trata?

- a) $r = 0$
- b) $\chi^2_{\text{obs}} = 37$
- c) $V \text{ de cramer} = 0,02$