

Unit 2 Caregiver Support

Unit Overview + Narrative Connections

In this unit, students will explore univariate and bivariate data sets. They will investigate, calculate, and interpret descriptive statistics like measures of center, variability, and association. Students will connect the learning in this unit to climate change, considering how statistics can give a more accurate and complete picture. However, they will also examine how statistics can be manipulated, distorted, and misinterpreted!

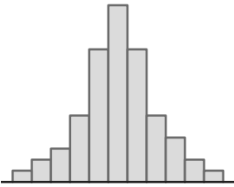
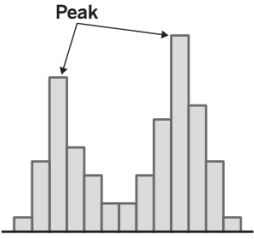
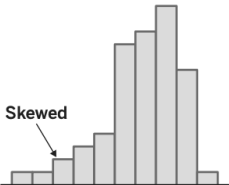
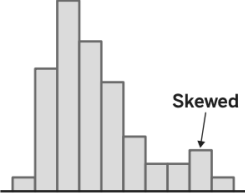

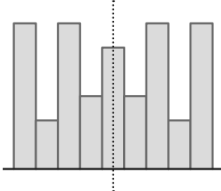
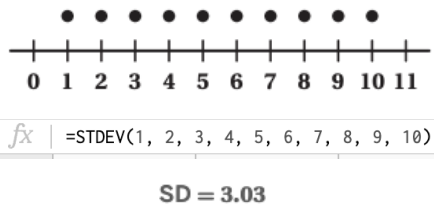


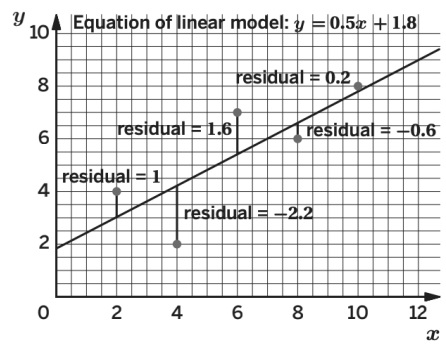
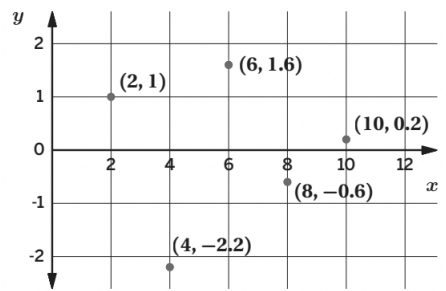
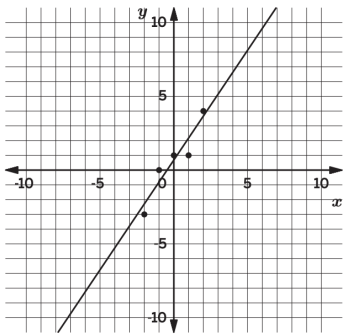
Prior Learning	Current Learning	Future Learning
<ul style="list-style-type: none"> • Dot plots, histograms, and box plots • Measures of center • Measures of variation • Fit a line to bivariate data 	<ul style="list-style-type: none"> • Data distributions • Standard deviation • Determining outliers • Fit lines to data • Correlation 	<ul style="list-style-type: none"> • Estimate population characteristics • Normal distribution • Margin of error • Randomization distributions

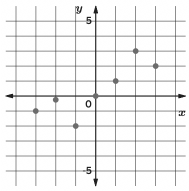
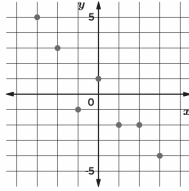
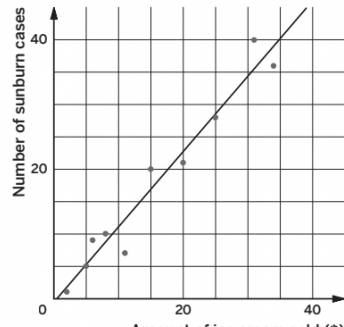
Key Ideas

- Data distributions can be described as *bell-shaped*, *bimodal*, *skewed*, or *uniform*.
- Spreadsheet technology can be used to create data distributions.
- The *standard deviation* of a data set is another measure of variability.
- *Residuals* can be used to judge whether a linear model is a good fit for a data set and to determine the *line of best fit*.
- Data can be organized in *two-way* and *relative frequency tables*.
- The *correlation coefficient* can be used to describe *associations* of bivariate data.

Vocabulary

<p>bell-shaped</p>	<p>A distribution that looks like a bell, with most of the data near the center and fewer points farther from the center.</p>	
<p>bimodal</p>	<p>A distribution with two distinct peaks.</p>	
<p>skewed left</p>	<p>A distribution with a long left tail, where data extends far away from the center.</p>	
<p>skewed right</p>	<p>A distribution with a long right tail, where data extends far away from the center.</p>	
<p>uniform</p>	<p>A distribution in which data is evenly distributed throughout the range.</p>	
<p>symmetric</p>	<p>A distribution that has a vertical line of symmetry and the mean is equal to the median.</p>	
<p>standard deviation</p>	<p>A commonly used measure of variability. It is the square root of the average of the squares of the distances between data values and the mean. The spreadsheet function “=STDEV()” can be used to calculate the standard deviation.</p>	

<p>residual</p>	<p>The difference between the actual y-coordinate and the y-coordinate predicted by a model, given the x-coordinate.</p>										
<p>residual plot</p>	<p>Shows the residuals on the vertical axis, with the independent variable (x) on the horizontal axis.</p>										
<p>line of best fit</p>	<p>The linear model that has the smallest possible sum of the squares of the residuals.</p>										
<p>categorical variable</p>	<p>A variable that can be partitioned into groups or categories.</p>	<table border="1" data-bbox="1068 1241 1511 1318"> <tr> <td>Takes a car to school</td> <td>Uses another form of transportation</td> </tr> </table>	Takes a car to school	Uses another form of transportation							
Takes a car to school	Uses another form of transportation										
<p>two-way table</p>	<p>A table that organizes categorical data into cells. The categories do not overlap, so that each data value is recorded in exactly one cell.</p>	<table border="1" data-bbox="1084 1375 1507 1606"> <thead> <tr> <th></th> <th>Flood insurance</th> <th>No flood insurance</th> </tr> </thead> <tbody> <tr> <th>Income below poverty line</th> <td>200</td> <td>700</td> </tr> <tr> <th>Income above poverty line</th> <td>2,000</td> <td>500</td> </tr> </tbody> </table>		Flood insurance	No flood insurance	Income below poverty line	200	700	Income above poverty line	2,000	500
	Flood insurance	No flood insurance									
Income below poverty line	200	700									
Income above poverty line	2,000	500									

<p>relative frequency table</p>	<p>A two-way table that shows the proportion of each value — expressed as fractions, decimals, or percentages — compared to the total in each row, column, or in the entire table.</p>	<table border="1"> <thead> <tr> <th></th> <th>Flood insurance</th> <th>No flood insurance</th> </tr> </thead> <tbody> <tr> <th>Income below poverty line</th> <td>22%</td> <td>78%</td> </tr> <tr> <th>Income above poverty line</th> <td>80%</td> <td>20%</td> </tr> </tbody> </table>		Flood insurance	No flood insurance	Income below poverty line	22%	78%	Income above poverty line	80%	20%									
	Flood insurance	No flood insurance																		
Income below poverty line	22%	78%																		
Income above poverty line	80%	20%																		
<p>association</p>	<p>When a change in one variable suggests another may change as well, the variables have an association and are said to be associated with one another.</p>	<p>Likely association:</p> <table border="1"> <thead> <tr> <th></th> <th>Likes school</th> <th>Dislikes school</th> </tr> </thead> <tbody> <tr> <th>Part of a club</th> <td>23 (71.875%)</td> <td>5 (≈ 20.8%)</td> </tr> <tr> <th>Not part of a club</th> <td>9 (28.125%)</td> <td>19 (≈ 79.2%)</td> </tr> </tbody> </table> <p>Unlikely association:</p> <table border="1"> <thead> <tr> <th></th> <th>Left handed</th> <th>Right handed</th> </tr> </thead> <tbody> <tr> <th>Composts food waste</th> <td>10 (10%)</td> <td>100 (10%)</td> </tr> <tr> <th>Does not compost food waste</th> <td>90 (90%)</td> <td>900 (90%)</td> </tr> </tbody> </table>		Likes school	Dislikes school	Part of a club	23 (71.875%)	5 (≈ 20.8%)	Not part of a club	9 (28.125%)	19 (≈ 79.2%)		Left handed	Right handed	Composts food waste	10 (10%)	100 (10%)	Does not compost food waste	90 (90%)	900 (90%)
	Likes school	Dislikes school																		
Part of a club	23 (71.875%)	5 (≈ 20.8%)																		
Not part of a club	9 (28.125%)	19 (≈ 79.2%)																		
	Left handed	Right handed																		
Composts food waste	10 (10%)	100 (10%)																		
Does not compost food waste	90 (90%)	900 (90%)																		
<p>correlation coefficient</p>	<p>A value that describes the strength and direction of a linear association between two variables. Strong positive associations have correlation coefficients close to 1, strong negative associations have correlation coefficients close to -1, and weak associations have correlation coefficients close to 0.</p>	<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>$r = 0.8$</p>  </div> <div style="text-align: center;"> <p>$r = -0.9$</p>  </div> </div>																		
<p>causation</p>	<p>When a change in one variable is shown, through careful experimentation, to cause a change in another variable.</p>	<p>Ice cream sold and sunburn cases are correlated but are <i>not</i> causal.</p>  <p style="text-align: center;">Amount of ice cream sold (\$)</p> <p>$y = 1.2x - 0.6$ $r = 0.98$</p>																		

Example Problems + Discussion Prompts

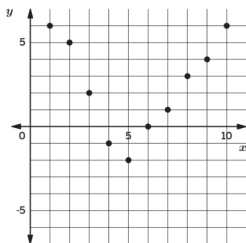
Sub-Unit 1

Problem	Sample Solution
<p style="text-align: center;">Lesson 5</p> <p>The data set represents the scores of Bard's assignments.</p> <p style="text-align: center;">0, 40, 60, 70, 75, 80, 85, 95, 95, 100</p> <p>a. Is 0 an outlier?</p> <p>b. The teacher is considering dropping the lowest score. What effect does eliminating the lowest value, 0, from the data set have on the mean and median?</p>	<p>a. A value is an outlier if it is less than $Q1 - 1.5 \cdot IQR$ or greater than $Q3 + 1.5 \cdot IQR$.</p> <p>$Q1$ is the median of the first 50% of the data and $Q3$ is the median of the second 50% of the data. The IQR is the range of the middle 50% of the data and can be calculated by subtracting $Q3 - Q1$.</p> <p>In this data set, $Q1$ is 60 and $Q3$ is 95. The IQR is 35 because $95 - 60 = 35$.</p> <p style="text-align: center;"> $\overset{Q1}{0, 40, 60, 70, 75}, [\overset{Q3}{80, 85, 95, 95, 100}]$ </p> <p>To determine whether a value is an outlier, it must either be less than $Q1 - 1.5 \cdot IQR$ or greater than $Q3 + 1.5 \cdot IQR$.</p> <p>Applying the formula $Q1 - 1.5 \cdot IQR$ to this data set results in $60 - 1.5 \cdot 35 = 7.5$, and 0 is less than 7.5.</p> <p>Yes, 0 is an outlier.</p> <p>b. By removing the lowest value, both the mean and the median will increase, but outliers have a greater effect on the mean than the median.</p> <p>Both the mean and median will increase but the mean will increase more than the median because the value eliminated is an outlier.</p>
<p>Discuss this question with your student:</p> <ul style="list-style-type: none"> How could you use the shape of the data distribution to determine whether the mean or the median is greater in this data set? 	

Sub-Unit 2

Problem	Sample Solution
<p>Lesson 7</p> <p>The mean of Data set A is 43.5 and the standard deviation is 3.7. The mean of Data set B is 12.8 and the standard deviation is 4.1.</p> <p>a. Which data set shows greater variability?</p> <p>b. What differences would you expect to see when comparing the dot plots of the two data sets?</p>	<p>a. The standard deviation is a measure of variability. The greater the standard deviation, the greater the variation in the data.</p> <p>Data Set B has a standard deviation greater than Data Set A.</p> <p>Data Set B shows greater variability.</p> <p>b. In a dot plot, most of the data centers around the mean, and the standard deviation determines the average number of units the rest of the data will appear above or below the mean.</p> <p>Data set A's dot plot will have most of the data centered around 43.5 and, on average, 3.7 units above or below 43.5.</p> <p>Data set B's dot plot will have most of the data centered around 12.8 and, on average, 4.1 units above or below 12.8.</p>
<p>Discuss this question with your student:</p> <ul style="list-style-type: none">• How are the standard deviation and MAD similar and different?	

Sub-Unit 3

Problem	Sample Solution
<p>Lesson 13</p> <p>Consider the residual plot shown. Is the line of fit for the data from which it was calculated a good fit?</p> 	<p>When the residuals are close to zero and appear to be randomly distributed above and below the x-axis on a residual plot, it indicates that a linear model is a good fit for the data. If the residuals follow a pattern, it indicates that a linear model may not be a good fit.</p> <p>The residual plot shows a pattern in the data, so a nonlinear model best represents the data.</p> <p>No, the line of fit for the data is not a good fit.</p>

Discuss these questions with your student:

- How can you use the values of residuals to determine whether a linear model is a good fit for a data set?
- How can you use the residual plot to determine whether a linear model is a good fit for a data set?

Sub-Unit 4

Problem	Sample Solution																
<p style="text-align: center;">Lesson 15</p> <p>Mai conducts a survey in her community asking if the homeowners think they have safe drinking water and if they think more efforts should be made to provide clean water. The results of her survey are shown.</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th></th> <th>More can be done to provide clean water</th> <th>More cannot be done to provide clean water</th> <th>Total</th> </tr> </thead> <tbody> <tr> <th>Have safe drinking water</th> <td style="text-align: center;">32</td> <td style="text-align: center;">23</td> <td style="text-align: center;">55</td> </tr> <tr> <th>Do not have safe drinking water</th> <td style="text-align: center;">61</td> <td style="text-align: center;">11</td> <td style="text-align: center;">72</td> </tr> <tr> <th>Total</th> <td style="text-align: center;">93</td> <td style="text-align: center;">34</td> <td style="text-align: center;">127</td> </tr> </tbody> </table> <p>a. How many members of the community participated in Mai’s survey?</p> <p>b. How many members of the community think more cannot be done to provide clean water?</p> <p>c. Of the members of the community who do not have safe drinking water, how many think more can be done to provide clean water?</p>		More can be done to provide clean water	More cannot be done to provide clean water	Total	Have safe drinking water	32	23	55	Do not have safe drinking water	61	11	72	Total	93	34	127	<p>a. The total number of participants in Mai’s survey can be determined from the intersection of the row total and the column total.</p> <p>127 members.</p> <p>b. The number of members of the community that think more cannot be done to provide clean water can be determined from the column “More cannot be done to provide clean water” total.</p> <p>34 members.</p> <p>c. First, only consider the number of members of the community who do not have safe drinking water. This can be determined by looking at the cells in the row “Do not have safe drinking water.” Then, find the cell that is under the column header “More can be done to provide clean water.”</p> <p>61 members.</p>
	More can be done to provide clean water	More cannot be done to provide clean water	Total														
Have safe drinking water	32	23	55														
Do not have safe drinking water	61	11	72														
Total	93	34	127														

Discuss these questions with your student:

- What does the value 55 represent in the two-way table?
- Why is a two-way table called a two-way table?

Lesson 17

The relative frequency table shows the percentage of each type of art (painting or sculpture) in a museum that would be classified in the different styles (modern or classical).

	Modern	Classical
Painting	41%	59%
Sculpture	38%	62%

Based on these percentages, is there evidence to suggest an association between the variables?

An association between two variables means that the two variables are statistically related to each other.

When row or column relative frequencies are vastly different, association is likely. When row or column relative frequencies are similar, association is unlikely.

Because the two types of art have similar percentages of each style, there is no reason to believe there is an association.

No, there is not enough evidence to suggest an association.

Discuss this question with your student:

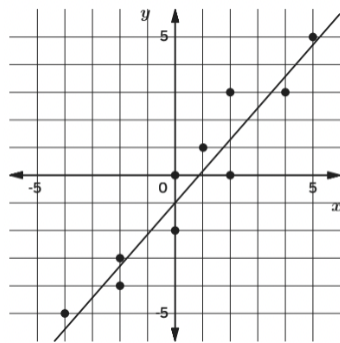
- What adjustments could you make to the relative frequency table to illustrate an association between the variables? Why?

Sub-Unit 5

Problem

Lesson 19

Which of the following is the best estimate for the correlation coefficient of the scatter plot and line of fit shown?



- A. $r = 0.96$
- B. $r = -0.96$
- C. $r = 0.15$
- D. $r = -0.15$

Sample Solution

The correlation coefficient in B is negative and the slope of the line of fit shown is positive.

The correlation coefficient in C is very close to 0, indicating no linear association but the data shown can be modeled with a line of fit.

The correlation coefficient in D is both negative, and the slope of the line of fit shown is positive, and very close to 0, indicating no linear association but the data shown can be modeled with a line of fit.

The correct answer is **A**.

Discuss this question with your student:

- What might a scatterplot with the correlation coefficient in B, C, or D look like? Why?

Sample Answers to Discussion Questions

Answers may vary.

- How could you use the shape of the data distribution to determine whether the mean or the median is greater in this data set?
 - *The data appears left skewed in a dot plot. Distributions that are skewed left typically have a mean that is less than the median.*
- How are the standard deviation and MAD similar and different?
 - *Both are measures of variability calculated from the mean. While the MAD is the average distance to the mean, the standard deviation involves the average of the squares of these distances. The standard deviation is more commonly used to measure variability than MAD.*
- How can you use the values of residuals to determine whether a linear model is a good fit for a data set?
 - *The smaller the absolute value of the residuals, the closer they are to zero and the better the fit of the linear model.*
- How can you use the residual plot to determine whether a linear model is a good fit for a data set?
 - *When the residuals are close to zero and appear to be randomly distributed above and below the x-axis on a residual plot, it indicates that a linear model is a good fit for the data. If the residuals follow a pattern, it indicates that a linear model may not be a good fit.*
- What does the value 55 represent in the two-way table?
 - *The 55 represents the total number of members of the community that have safe drinking water.*
- Why is a two-table called a two-way table?
 - *The table is examining two different categories.*
- What adjustments could you make to the relative frequency table to illustrate an association between the variables? Why?
 - *If the table had a higher percentage for modern paintings, like 65%, and a lower percentage for classical paintings, like 35%. This would illustrate a likely association because the column relative frequencies are vastly different.*
- What might a scatterplot with the correlation coefficient in B, C, or D look like? Why?
 - *A scatterplot with the correlation coefficient in B, $r = -0.96$, would look very similar to the current scatter plot just reflected across the y-axis to have a negative slope. A scatterplot with the correlation coefficient in C, $r = 0.15$, would look like mostly randomly distributed data points with a very low likelihood that*

they would fit to a positive line. A scatterplot with the correlation coefficient in C, $r = -0.15$, would look like mostly randomly distributed data points with a very low likelihood that they would fit to a negative line.

Apoyo para cuidadores/as, Unidad 2

Vista general de la unidad + Conexiones narrativas

En esta unidad los/as estudiantes explorarán conjuntos de datos con una o más variables. Investigarán, calcularán e interpretarán estadísticas descriptivas como medidas de centro, variabilidad y asociación. Los/as estudiantes relacionarán lo aprendido en esta unidad con el cambio climático, considerando cómo las estadísticas pueden ofrecer una imagen más precisa y completa. Sin embargo, también examinarán cómo las estadísticas pueden ser manipuladas, distorsionadas y malinterpretadas.

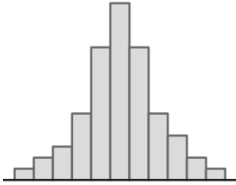
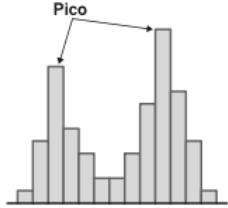
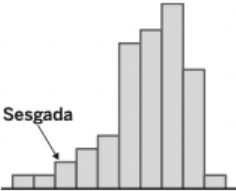
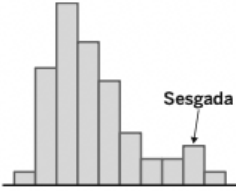

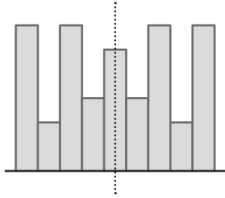
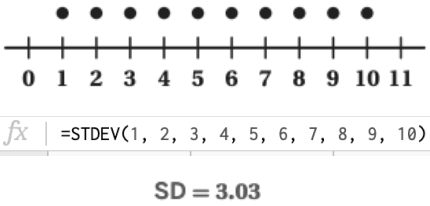


Aprendizaje previo	Aprendizaje actual	Aprendizaje futuro
<ul style="list-style-type: none"> • Diagramas de puntos, histogramas y diagrama de cajas • Medidas del centro • Medidas de variación • Ajustar una línea a los datos bivariados 	<ul style="list-style-type: none"> • Distribución de datos • Desviación estándar • Determinación de los valores atípicos • Ajustar las líneas a los datos • Correlación 	<ul style="list-style-type: none"> • Estimar las características de la población • Distribución normal • Margen de error • Distribuciones de aleatoriedad

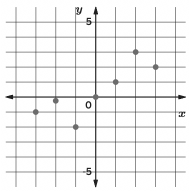
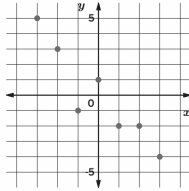
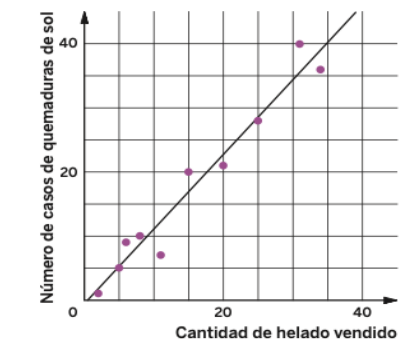
Ideas clave

- Las distribuciones de datos pueden describirse como acampanadas, bimodales, sesgadas o uniformes.
- La tecnología de las hojas de cálculo puede utilizarse para crear distribuciones de datos.
- La *desviación estándar* de un conjunto de datos es otra medida de la variabilidad.
- Los *residuos* pueden utilizarse para juzgar si un modelo lineal se ajusta bien a un conjunto de datos y para determinar la *línea de ajuste óptimo*.
- Los datos pueden organizarse en tablas de *frecuencias relativas* y *bidireccionales*.
- El *coeficiente de correlación* puede utilizarse para describir *asociaciones* de datos bivariados.

Vocabulario

<p>acampanada</p>	<p>Una distribución que asemeja a una campana, con la mayoría de los datos cerca del centro y una menor cantidad de puntos más lejos del centro.</p>	
<p>bimodal</p>	<p>Una distribución con dos picos distintos.</p>	
<p>sesgada a la izquierda</p>	<p>Una distribución con una larga cola izquierda, en la que los datos se extienden lejos del centro.</p>	
<p>sesgada a la derecha</p>	<p>Una distribución con una larga cola derecha, en la que los datos se extienden lejos del centro.</p>	
<p>uniforme</p>	<p>Distribución en la cual los datos son distribuidos de manera regular a través del rango.</p>	
<p>simétrica</p>	<p>Una distribución que tiene una línea de simetría vertical y la media es igual a la mediana.</p>	
<p>desviación estándar</p>	<p>Medida de variabilidad de uso común. Se trata de la raíz cuadrada del promedio de las distancias elevadas al cuadrado entre los valores de los datos y la media. La función de la hoja de cálculo</p>	 <p style="text-align: center;">SD = 3.03</p>

	"=STDEV()" puede utilizarse para calcular la desviación estándar.										
residuo	Diferencia entre la coordenada y real y la coordenada y pronosticada por un modelo, dada la coordenada x.										
diagrama de residuos	Muestra los residuos en el eje vertical, con la variable independiente (x) en el eje horizontal.										
línea de ajuste óptimo	Modelo lineal que tiene la menor suma posible de los cuadrados de los residuos.										
variable categórica	Variable que puede partirse en grupos o categorías.	<table border="1"> <tr> <td>Va en automóvil a la escuela</td> <td>Usa otra forma de transporte</td> </tr> </table>	Va en automóvil a la escuela	Usa otra forma de transporte							
Va en automóvil a la escuela	Usa otra forma de transporte										
tabla de doble entrada	Tabla que organiza datos categóricos en celdas. Las categorías no se superponen, de manera que el valor de cada dato es registrado exactamente en una sola celda.	<table border="1"> <thead> <tr> <th></th> <th>Seguro contra inundaciones</th> <th>Sin seguro contra inundaciones</th> </tr> </thead> <tbody> <tr> <th>Ingresos por debajo del umbral de pobreza</th> <td>200</td> <td>700</td> </tr> <tr> <th>Ingresos por encima del umbral de pobreza</th> <td>2,000</td> <td>500</td> </tr> </tbody> </table>		Seguro contra inundaciones	Sin seguro contra inundaciones	Ingresos por debajo del umbral de pobreza	200	700	Ingresos por encima del umbral de pobreza	2,000	500
	Seguro contra inundaciones	Sin seguro contra inundaciones									
Ingresos por debajo del umbral de pobreza	200	700									
Ingresos por encima del umbral de pobreza	2,000	500									

<p>tabla de frecuencia relativa</p>	<p>Tabla de doble entrada que muestra la proporción de cada valor (expresada como fracciones, decimales o porcentajes), en comparación con el total de cada fila, columna o con toda la tabla.</p>	<table border="1"> <thead> <tr> <th></th> <th>Seguro contra inundaciones</th> <th>Sin seguro contra inundaciones</th> </tr> </thead> <tbody> <tr> <th>Ingresos por debajo del umbral de pobreza</th> <td>22%</td> <td>78%</td> </tr> <tr> <th>Ingresos por encima del umbral de pobreza</th> <td>80%</td> <td>20%</td> </tr> </tbody> </table>		Seguro contra inundaciones	Sin seguro contra inundaciones	Ingresos por debajo del umbral de pobreza	22%	78%	Ingresos por encima del umbral de pobreza	80%	20%									
	Seguro contra inundaciones	Sin seguro contra inundaciones																		
Ingresos por debajo del umbral de pobreza	22%	78%																		
Ingresos por encima del umbral de pobreza	80%	20%																		
<p>asociación</p>	<p>Cuando un cambio en una variable sugiere que otra también podría cambiar, las variables tienen una asociación y están asociadas entre sí.</p>	<p>Asociación probable:</p> <table border="1"> <thead> <tr> <th></th> <th>Le gusta la escuela</th> <th>No le gusta la escuela</th> </tr> </thead> <tbody> <tr> <th>Pertenece a un club</th> <td>23 (71.875%)</td> <td>5 (≈ 20.8%)</td> </tr> <tr> <th>No pertenece a un club</th> <td>9 (28.125%)</td> <td>19 (≈ 79.2%)</td> </tr> </tbody> </table> <p>Asociación improbable:</p> <table border="1"> <thead> <tr> <th></th> <th>Zurdo</th> <th>Derecho</th> </tr> </thead> <tbody> <tr> <th>Convierte en abono los restos de comida</th> <td>10 (10%)</td> <td>100 (10%)</td> </tr> <tr> <th>No convierte en abono los restos de comida</th> <td>90 (90%)</td> <td>900 (90%)</td> </tr> </tbody> </table>		Le gusta la escuela	No le gusta la escuela	Pertenece a un club	23 (71.875%)	5 (≈ 20.8%)	No pertenece a un club	9 (28.125%)	19 (≈ 79.2%)		Zurdo	Derecho	Convierte en abono los restos de comida	10 (10%)	100 (10%)	No convierte en abono los restos de comida	90 (90%)	900 (90%)
	Le gusta la escuela	No le gusta la escuela																		
Pertenece a un club	23 (71.875%)	5 (≈ 20.8%)																		
No pertenece a un club	9 (28.125%)	19 (≈ 79.2%)																		
	Zurdo	Derecho																		
Convierte en abono los restos de comida	10 (10%)	100 (10%)																		
No convierte en abono los restos de comida	90 (90%)	900 (90%)																		
<p>coeficiente de correlación</p>	<p>Valor que describe la fuerza y dirección de una asociación lineal entre dos variables. Asociaciones positivas fuertes tienen coeficientes de correlación cercanos a 1, mientras que asociaciones negativas fuertes tienen coeficientes de correlación cercanos a -1, y asociaciones débiles tienen coeficientes de correlación cercanos a 0.</p>	<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>$r = 0.8$</p>  </div> <div style="text-align: center;"> <p>$r = -0.9$</p>  </div> </div>																		
<p>causalidad</p>	<p>Cuando se muestra que un cambio en una variable causa un cambio en otra variable, a través de cuidadosa experimentación.</p>	<p>Los helados vendidos y los casos de quemaduras solares están correlacionados pero no son causales.</p>  <p>$y = 1.2x - 0.6$ $r = 0.98$</p>																		

Problemas de ejemplo + Temas de discusión

Subunidad 1

Problema	Solución de ejemplo
<p style="text-align: center;">Lección 5</p> <p>El conjunto de datos representa las calificaciones de las tareas de Bard.</p> <p style="text-align: center;">0, 40, 60, 70, 75, 80, 85, 95, 95, 100</p> <p>a. ¿Es 0 un valor atípico?</p> <p>b. El/la maestro/a está considerando no tomar en cuenta la calificación más baja. ¿Qué efecto tiene eliminar el valor más bajo, 0, del conjunto de datos sobre la media y la mediana?</p>	<p>a. A value is an outlier if it is less than $C1 - 1.5 \cdot RIC$ o mayor que $C3 + 1.5 \cdot RIC$.</p> <p>C1 es la mediana del primer 50% de los datos y C3 es la mediana del segundo 50% de los datos. El RIC es el rango del 50% del medio de los datos y puede calcularse restando $C3 - C1$.</p> <p>En este conjunto de datos, C1 es 60 y C3 es 95. El RIC es 35 porque $95 - 60 = 35$.</p> <p style="text-align: center;"> $\overset{C1}{\cancel{0}, \cancel{40}, 60, 70, 75}, [\cancel{80}, \cancel{85}, 95, 95, \cancel{100}]^{\overset{C3}$ </p> <p>Para determinar si un valor es un valor atípico, debe ser menor que $C1 - 1,5 \cdot RIC$ o mayor que $C3 + 1,5 \cdot RIC$.</p> <p>Aplicando la fórmula $C1 - 1,5 \cdot RIC$ a este conjunto de datos se obtiene que $60 - 1,5 \cdot 35 = 7,5$, y 0 es menor que 7,5.</p> <p>Sí, 0 es un valor atípico.</p> <p>b. Al eliminar el valor más bajo, tanto la media como la mediana aumentarán, pero los valores atípicos tienen un efecto mayor sobre la media que la mediana.</p> <p>Tanto la media como la mediana aumentarán. Sin embargo, la media aumentará más que la mediana porque el valor que se elimina es un valor atípico.</p>
<p>Comente esta pregunta con su estudiante:</p>	

- ¿Cómo podrías utilizar la forma de la distribución de datos para determinar si la media o la mediana es mayor en este conjunto de datos?

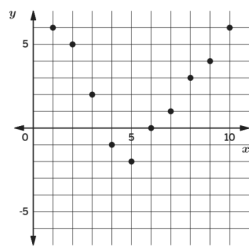
Subunidad 2

Problema	Solución de ejemplo
<p style="text-align: center;">Lección 7</p> <p>La media del Conjunto de datos A es 43.5 y la desviación estándar es 3.7. La media del Conjunto de datos B es 12.8 y la desviación estándar es 4.1.</p> <p>a. ¿Cuál conjunto de datos muestra mayor variabilidad?</p> <p>b. ¿Qué diferencias esperarías ver al comparar los diagramas de puntos de los dos conjuntos de datos?</p>	<p>a. La desviación estándar es una medida de variabilidad. Cuanto mayor sea la desviación estándar, mayor será la variación de los datos.</p> <p>El Conjunto de datos B muestra mayor variabilidad que el Conjunto de datos A.</p> <p>El Conjunto de datos B muestra mayor variabilidad.</p> <p>b. En un diagrama de puntos, la mayoría de los datos se centran en la media, y la desviación estándar determina el número promedio de unidades que el resto de los datos aparecerán por encima o por debajo de la media.</p> <p>El diagrama de puntos del Conjunto de datos A tendrá la mayoría de los datos centrados alrededor de 43.5 y, en promedio, 3.7 unidades por encima o por debajo de 43.5.</p> <p>El diagrama de puntos del Conjunto de datos B tendrá la mayoría de los datos centrados alrededor de 12.8 y, en promedio, 4.1 unidades por encima o por debajo de 12.8.</p>
<p>Comente esta pregunta con su estudiante:</p> <ul style="list-style-type: none"> • ¿En qué se parecen y en qué se diferencian la desviación estándar y la DAM? 	

Subunidad 3

Problema	Solución de ejemplo
Lección 13	Cuando los residuos se acercan a cero y parecen

Consulta el diagrama de residuo que se muestra. ¿Es la línea de ajuste para los datos a partir de los cuales se calculó un buen ajuste?



estar distribuidos aleatoriamente por encima y por debajo del eje x en un diagrama de residuos, indica que un modelo lineal se ajusta bien a los datos. Si los residuos siguen un patrón, indica que un modelo lineal puede no ajustarse bien.

El diagrama de residuos muestra un patrón en los datos, por lo que un modelo no lineal es el que mejor representa los datos.

No, esta no representa una buena línea de ajuste.

Comente estas preguntas con su estudiante:

- ¿Cómo puedes utilizar los valores de los residuos para determinar si un modelo lineal se ajusta bien a un conjunto de datos?
- ¿Cómo puedes utilizar el diagrama de residuos para determinar si un modelo lineal se ajusta bien a un conjunto de datos?

Subunidad 4

Problema		Solución de ejemplo																	
<p>Lección 15</p> <p>Mai realiza una encuesta en su comunidad que pregunta si los propietarios de vivienda creen que tienen agua potable segura y si creen que deberían realizarse más esfuerzos para proporcionar agua limpia. Se muestran los resultados de su encuesta.</p> <table border="1"> <thead> <tr> <th></th> <th>Se puede hacer más para proporcionar agua limpia</th> <th>No se puede hacer más para proporcionar agua limpia</th> <th>Total</th> </tr> </thead> <tbody> <tr> <th>Tienen agua potable segura</th> <td>32</td> <td>23</td> <td>55</td> </tr> <tr> <th>No tienen agua potable segura</th> <td>61</td> <td>11</td> <td>72</td> </tr> <tr> <th>Total</th> <td>93</td> <td>34</td> <td>127</td> </tr> </tbody> </table> <p>a. ¿Cuántos miembros de la comunidad participaron en la encuesta de Mai?</p>			Se puede hacer más para proporcionar agua limpia	No se puede hacer más para proporcionar agua limpia	Total	Tienen agua potable segura	32	23	55	No tienen agua potable segura	61	11	72	Total	93	34	127	<p>a. El número total de participantes en la encuesta de Mai puede determinarse a partir de la intersección del total de filas y el total de columnas.</p> <p>127 miembros.</p> <p>b. El número de miembros de la comunidad que piensan que no se puede hacer más para proporcionar agua limpia se puede determinar a partir del total de la columna "No se puede hacer más para proporcionar agua limpia".</p> <p>34 miembros.</p> <p>c. En primer lugar, sólo hay que tener en cuenta el número de miembros de la comunidad que no disponen de agua potable. Esto se puede</p>	
	Se puede hacer más para proporcionar agua limpia	No se puede hacer más para proporcionar agua limpia	Total																
Tienen agua potable segura	32	23	55																
No tienen agua potable segura	61	11	72																
Total	93	34	127																

b. ¿Cuántos miembros de la comunidad creen que no se puede hacer más para proporcionar agua limpia?

c. De los miembros de la comunidad que no tienen agua potable segura, ¿cuántos creen que se puede hacer más para proporcionar agua limpia?

determinar mirando las celdas de la fila "No tienen agua potable segura". Luego, busque la celda que está bajo el encabezado de la columna "Se puede hacer más para proporcionar agua limpia".

61 miembros.

Comente estas preguntas con su estudiante:

- ¿Qué representa el valor 55 en la tabla de doble entrada?
- ¿Por qué una tabla de doble entrada se llama tabla de doble entrada?

Lección 17

La tabla de frecuencia relativa muestra el porcentaje de cada tipo de arte (pintura o escultura) en un museo que se clasificaría en los diferentes estilos (moderno o clásico).

	Moderno	Clásico
Pintura	41%	59%
Escultura	38%	62%

Con base en estos porcentajes, ¿hay evidencia para sugerir una asociación entre las variables?

Una asociación entre dos variables significa que las dos variables están estadísticamente relacionadas entre sí.

Cuando las frecuencias relativas de filas o columnas son muy diferentes, la asociación es probable. Cuando las frecuencias relativas de filas o columnas son similares, la asociación es improbable.

Puesto que los dos tipos de arte tienen porcentajes similares de cada estilo, no hay razón para creer que haya una asociación.

No, no hay evidencia suficiente para sugerir una asociación.

Comente esta pregunta con su estudiante:

- ¿Qué ajustes podrías hacer en la tabla de frecuencias relativas para ilustrar una asociación entre las variables? ¿Por qué?

Subunidad 4

Problema	Solución de ejemplo
Lección 19	El coeficiente de correlación en B es negativo y la pendiente de la línea de ajuste mostrada es

¿Cuál de las siguientes es la mejor estimación para el coeficiente de correlación del diagrama de dispersión y la línea de ajuste mostrados?

- A. $r = 0.96$
- B. $r = -0.96$
- C. $r = 0.15$
- D. $r = -0.15$

positiva.

El coeficiente de correlación en C es muy cercano a 0, lo que indica que no hay asociación lineal, pero los datos mostrados pueden modelarse con una línea de ajuste.

El coeficiente de correlación en D es negativo, y la pendiente de la línea de ajuste mostrada es positiva, y muy cercana a 0, lo que indica que no hay asociación lineal, pero los datos mostrados pueden ser modelados con una línea de ajuste.

La respuesta correcta es **A**.

Comente esta pregunta con su estudiante:

- ¿Qué aspecto podría tener un diagrama de dispersión con el coeficiente de correlación en B, C o D?
¿Por qué?

Respuestas de ejemplo a las preguntas de discusión

Puede haber varias respuestas.

- ¿Cómo podrías utilizar la forma de la distribución de datos para determinar si la media o la mediana es mayor en este conjunto de datos?
 - *Los datos aparecen sesgados a la izquierda en un diagrama de puntos. Las distribuciones sesgadas a la izquierda suelen tener una media inferior a la mediana.*
- ¿En qué se parecen y en qué se diferencian la desviación estándar y la DAM?
 - *Ambas son medidas de variabilidad calculadas a partir de la media. Mientras que la DAM es la distancia promedio a la media, la desviación estándar implica el promedio de los cuadrados de estas distancias. La desviación estándar se utiliza más comúnmente para medir la variabilidad que la DAM.*
- ¿Cómo puedes utilizar los valores de los residuos para determinar si un modelo lineal se ajusta bien a un conjunto de datos?
 - *Cuanto menor sea el valor absoluto de los residuos, más cerca estarán de cero y mejor será el ajuste del modelo lineal.*
- ¿Cómo puedes utilizar el diagrama de residuos para determinar si un modelo lineal se ajusta bien a un conjunto de datos?
 - *Cuando los residuos se acercan a cero y parecen estar distribuidos aleatoriamente por encima y por debajo del eje x en un diagrama de residuos, indica que un modelo lineal se ajusta bien a los datos. Si los residuos siguen un patrón, indica que un modelo lineal puede no ajustarse bien.*
- ¿Qué representa el valor 55 en la tabla de doble entrada?
 - *El 55 representa el número total de miembros de la comunidad que disponen de agua potable.*
- ¿Por qué una tabla de doble entrada se llama tabla de doble entrada?
 - *La tabla examina dos categorías diferentes.*
- ¿Qué ajustes podrías hacer en la tabla de frecuencias relativas para ilustrar una asociación entre las variables? ¿Por qué?
 - *Si la tabla tuviera un porcentaje mayor para las pinturas modernas, como el 65%, y un porcentaje menor para las pinturas clásicas, como el 35%. Esto ilustraría una asociación probable porque las frecuencias relativas de las columnas son muy diferentes.*
- ¿Qué aspecto podría tener un diagrama de dispersión con el coeficiente de correlación en B, C o D? ¿Por qué?

- *Un diagrama de dispersión con el coeficiente de correlación en B, $r = -0,96$, tendría un aspecto muy similar al diagrama de dispersión actual, simplemente reflejado en el eje Y para tener una pendiente negativa. Un diagrama de dispersión con el coeficiente de correlación en C, $r = 0,15$, se vería como puntos de datos distribuidos al azar con una probabilidad muy baja de que se ajusten a una línea positiva. Un diagrama de dispersión con el coeficiente de correlación en C, $r = -0,15$, se vería como puntos de datos distribuidos al azar con una probabilidad muy baja de que se ajusten a una línea negativa.*