

Problemario EP1-U1

Contents

0.1	Tablas de frecuencia	1
0.2	Distribuciones de frecuencia	2
0.3	Diagrama de tallo y hojas	2
0.4	Análisis de regresión	4

Los siguientes ejercicios deben ser realizados en dos formas, primero de manera manual y posteriormente por medio de software. Al ser realizados a mano, deben ser hechos con limpieza y orden, para posteriormente ser digitalizados por el método que tengan a su alcance. Al ser realizados por medio de software se recomienda el uso de RStudio para la resolución de los problemas y la generación del documento a entregar. La combinación de ambos documentos debe ser enviada en una carpeta comprimida a la plataforma habilitada para tal uso. No se admiten evidencias en físico.

0.1 Tablas de frecuencia

Enseguida se presentan algunas tablas. Las columnas son arbitrarias y no implican variables diferentes ni nada por el estilo, es decir, los datos se dividieron en columnas solo para una mejor presentación en pantalla. En realidad se refieren a una sola columna.

1. Los cinco automóviles más vendidos en Estados Unidos durante 2003 fueron la camioneta Chevrolet Silverado/C/K, la camioneta Dodge Ram, la camioneta Ford F-Series, el Honda Accord y el Toyota Camry (Motor Trend, 2003). En la siguiente tabla se presenta una muestra de 50 compras de automóviles.

##	V1	V2	V3	V4	V5
## 1	Silverado	Ram	Accord	Camry	Camry
## 2	Silverado	Silverado	Camry	Ram	F-Series
## 3	Ram	F-Series	Accord	Ram	Ram
## 4	Silverado	F-Series	F-Series	Silverado	Ram
## 5	Ram	Ram	Accord	Silverado	Camry
## 6	F-Series	Ram	Silverado	Accord	Silverado
## 7	Camry	F-Series	F-Series	F-Series	Silverado
## 8	F-Series	Silverado	F-Series	F-Series	Ram
## 9	Silverado	Silverado	Camry	Camry	F-Series
## 10	Silverado	F-Series	F-Series	Accord	Accord

- a) Elabore una distribución de frecuencia y otra de frecuencia porcentual (relativa).
- b) ¿Cuál es la camioneta y el automóvil de pasajeros más vendidos?
- c) Haga una gráfica de pastel.

2. El Higher Education Research Institute de UCLA cuenta con estadísticas sobre las áreas que son más elegidas por los estudiantes de nuevo ingreso. Las cinco más elegidas son arte y humanidades (A), administración de negocios (B), ingeniería (E), política (P) y ciencias sociales (S) (The New York Times Almanac, 2006). Otras áreas (O), entre las que se encuentran biología, física, ciencias de la computación y educación se agruparon todas en una sola categoría. Las siguientes fueron las áreas elegidas por 64 estudiantes de recién ingreso de una muestra.

##	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16
## 1	S	P	P	O	B	E	O	E	P	O	O	B	O	O	O	A
## 2	O	E	E	B	S	O	B	O	A	O	E	O	E	O	B	P
## 3	B	A	S	O	E	A	B	O	S	S	O	O	E	B	O	B
## 4	A	E	B	E	A	A	P	O	O	E	O	B	B	O	P	B

- a) Dé una distribución de frecuencia y otra de frecuencia porcentual.

- b) Elabore una gráfica de barras.
- c) ¿Que porcentaje de los estudiantes de nuevo ingreso elige una de las cinco áreas más elegidas?
- d) ¿Cuál es el área más elegida por los estudiantes de nuevo ingreso? ¿Qué porcentaje de los estudiantes de nuevo ingreso elige esta área?

0.2 Distribuciones de frecuencia

3. Cada año en Estados Unidos, aproximadamente 1.5 millones de los estudiantes de educación superior presentan un examen de aptitud escolar (SAT, por sus siglas en inglés). Cerca de 80% de las universidades e instituciones de educación superior emplean las puntuaciones obtenidas por los estudiantes en este examen como criterio de admisión (College Board, marzo de 2006). A continuación se presentan las puntuaciones obtenidas en las áreas de matemáticas y expresión verbal por una muestra de estudiantes.

##		V1	V2	V3	V4	V5
## 1		1025	1042	1195	880	945
## 2		1102	845	1095	936	790
## 3		1097	913	1245	1040	998
## 4		998	940	1043	1048	1130
## 5		1017	1140	1030	1171	1035

- a) Presente una distribución de frecuencia y un histograma de estas puntuaciones.
- b) Elabore otra distribución e histograma con la primera clase comenzando en la puntuación 750 y la amplitud de clase deberá ser 100.
- c) Dé un comentario sobre la forma de la distribución.
- d) ¿Qué otras observaciones puede hacer acerca de estas puntuaciones con base en los resúmenes tabulares y gráficos?

0.3 Diagrama de tallo y hojas

4. Drug Store News (septiembre de 2002) proporciona datos sobre ventas de medicamentos de las principales farmacias de Estados Unidos. Los datos siguientes son ventas anuales en millones.

##	Farmacia	Ventas
## 1	Ahold-USA	1700
## 2	CVS	12700
## 3	Eckerd	7739
## 4	Kmart	1863
## 5	Kroger	3400
## 6	MedicineShope	1757
## 7	Rite-Aid	8637
## 8	Safeway	2150
## 9	Walgreens	11660
## 10	Wal-Mart	7250

- a) Elabore un diagrama de tallo y hojas.
- b) Indique cuáles son las ventas anuales menores, mayores e intermedias.
- c) ¿Cuáles son las dos farmacias mayores?

5. A continuación se presentan las temperaturas diarias más altas y más bajas registradas en 20 ciudades de Estados Unidos (USA Today, 3 de marzo 2006).

##	Ciudad	Alta	Baja
## 1	Albuquerque	66	39
## 2	Atlanta	61	35
## 3	Baltimore	42	26
## 4	Charlotte	60	29
## 5	Cincinnati	41	21

## 6	Dallas	62	47
## 7	Denver	60	31
## 8	Houston	70	54
## 9	Indianapolis	42	22
## 10	LasVegas	65	43
## 11	LosAngeles	60	46
## 12	Miami	84	65
## 13	Minneapolis	30	11
## 14	NewOrleans	68	50
## 15	OklahomaCity	62	40
## 16	Phoenix	77	50
## 17	Portland	54	38
## 18	St.Louis	45	27
## 19	SanFrancisco	55	43
## 20	Seattle	52	36

- Con las temperaturas altas elabore un diagrama de tallo y hojas.
 - Con las temperaturas bajas elabore un diagrama de tallo y hojas.
 - Compare los dos diagramas y haga comentarios acerca de las diferencias entre las temperaturas más altas y las más bajas.
 - Proporcione una distribución de frecuencia de las temperaturas más altas y de las más bajas.
6. De acuerdo con 2003 Annual Consumer Spending Survey, el cargo promedio mensual a una tarjeta de crédito Bank of America Visa fue de \$1,838 (U.S. Airways Attaché Magazine, diciembre de 2003). En una muestra de cargos mensuales a tarjetas de crédito los datos obtenidos son los siguientes.

##	V1	V2	V3	V4	V5
## 1	236	1710	1351	825	7450
## 2	316	4135	1333	1584	387
## 3	991	3396	170	1428	1688

- Calcule la media y la mediana.
- Calcule el primero y tercer cuartil.
- Calcule el rango y el rango intercuartílico.
- Calcule la varianza y la desviación estándar.
- Comente la forma de la distribución. ¿Esta es la forma que esperaría? ¿Por qué sí o por qué no?
- ¿Hay observaciones atípicas en estos datos?
- Trace el gráfico de caja pertinente.

7. La oficina de censos de Estados Unidos proporciona estadísticas sobre las familias en ese país, informaciones como edad al contraer el primer matrimonio, estado civil actual y tamaño de la casa (www.census.gov, 20 de marzo de 2006). Los datos siguientes son edades al contraer el primer matrimonio en una muestra de hombres y en una muestra de mujeres.

##	V1	V2
## 1	HOMBRES	MUJERES
## 2	26	20
## 3	23	28
## 4	28	23
## 5	25	30
## 6	27	24
## 7	30	29
## 8	26	26
## 9	35	25
## 10	28	22
## 11	21	22
## 12	24	25

## 13	27	23
## 14	29	27
## 15	30	26
## 16	27	19
## 17	32	<NA>
## 18	27	<NA>
## 19	25	<NA>

- Determine la mediana en la edad de hombres y mujeres al contraer el primer matrimonio.
- Calcule el primer y tercer cuartil tanto en los hombres como en las mujeres.
- Hace 30 años la mediana en la edad al contraer el primer matrimonio era 25 años entre los hombres y 22 años entre las mujeres. ¿Qué indica esta información acerca de la edad a la que deciden contraer matrimonio los jóvenes de hoy en día?
- Trace la gráfica de caja para este problema.

0.4 Análisis de regresión

Vuelva al conjunto de datos del problema 5.

- Elabore un diagrama de dispersión que muestre la relación entre las dos variables, temperatura más alta y temperatura más baja.