|  |
| --- |
| Description: http://virtualcampus.pupr.edu/images/ci/sets/set01/lesson_on.gif**Modulo#7: Descriptiva II: Medidas de Posición Relativa y Diagrama de Cajas.** |
|  |

|  |
| --- |
| **ENGI 2270 Probability and Statistics for Engineers** |
| **Modulo 07: Datos Cuantitativos: Medidas de Posición Relativa y Diagrama de Cajas.** |
| 🗒Introduccion |
| En este modulo se presentara el teorema de Tchebyshev y la regla empírica. Además del diagrama de cajas. También se presentara el cómputo de los números atípicos y valor Z.  Por último se presentara el uso de programas de computadora para la generación de gráficos. |
| **☑ Learning Goals /Outcomes** (What you need to know) |
| Upon completion of this module, you will be able to:   * Medidas de posición relativa entre µ y σ   + Teorema Tchebyshev   + Regla Empírica * Diagrama de Caja (Box and Wisker Diagram) * Tipos de sesgos * Números Atípicos (Outliers) * Valor Zo * Estimación de la desviación estándar * Uso de software para generar diagramas   Leccion#7.1 Medidas de posición relativa  Leccion#7.2 Box and Wisker y Outliers  Leccion#7.3 Sesgos, y Valor Z Uso de STATDISK para generar box plot. |
| **🗍 Lección # 7.1: Medidas de Posición Relativa entre µ y σ**  En las medidas de posición relativa se estudiara la relación entre desviación estándar y su promedio. Además de la cantidad de datos entre las medidas asociadas al promedio y a la cantidad de desviaciones estándares. También son medidas que podemos utilizar para comparar los conjuntos de datos o varios experimentos , conociendo como se distribuyen los datos de cada conjunto. En lo que respecta a esta lección estudiaremos: Teorema Tchebyshev y Regla Empírica.  Nota: Otras medidas de posición relativa son las ya estudiadas en modulo anterior están :  Cuartiles y Percentiles y Rango Intercuartilico (IQR). Como también la que se presentara más adelante, valor Zo pueden ser de mucha utilidad al comparar experimentos.  **TEOREMA DE TCHEVYSHEV**  Matemático ruso que determino la cantidad de datos que en cualquier experimento deben haber a 2 o más desviaciones estándares del promedio.  Sin importar la forma de la distribución, Tchevyshev prueba de forma cuantitativa que alrededor o en la vecindad del promedio de cualquier experimento se encuentran la mayoría de los datos. ¿Cuánto es alrededor? Esa es la pregunta que contesto Tchevyshev.; Es a +/- 2 desviaciones estándares, estará no menos del 75% de los datos de cualquier experimento, puede haber más pero nunca menos. Se expresa como a K desviaciones estará 1-(1/k2) del área. Donde k>=2, no puede ser 1.  chevy0001.jpg  Imágenes extraidas y manipuladas utilizando Paint; del libreo : William Menderhall, Robert Beaver y Barbara Beave; Introducción a la probabilidad y estadística ; 1ra. Edición 2002 International Thomson Editores, USA, capitulo 2, pagina 64  Donde K es el número de desviaciones estándares a cada lado del promedio.  , donde k>= 2   |  |  | | --- | --- | | **Intervalo** | **Área o % de Datos** | |  | **75%** | |  | **89%** | |
| **REGLA EMPIRICA**  Solo aplica si los datos al presentarlos gráficamente nos da una campana o distribución Normal o Gaussiana. Entonces la aplica siguiente Regla Empírica:     |  |  | | --- | --- | | **Intervalo** | **Area o % de Datos** | |  | **68%** | |  | **95%** | |  | **99%** |   **normal0001.jpg**  Imágenes extraidas y manipuladas utilizando Paint; del libreo : William Menderhall, Robert Beaver y Barbara Beave; Introducción a la probabilidad y estadística ; 1ra. Edición 2002 International Thomson Editores, USA, capitulo 2, pagina 74    Imagen extraida y manipulada por Paint del libro :David Doane and Lori Seward; Applied Statistics in Business and Economics; McGraw-Hill Irwin, NY 2007, USA |
| **🗍 Lección # 7.2 DIAGRAMA DE CAJAS**  **También se conoce como: Diagrama de Bloques y Líneas**  **BOX AND WISKER DIAGRAM y BOX PLOT**  **Utilizando las 5 medidas del resumen**: Con estas medidas podemos definir la distribución de los datos del experimento   * Xmin, Q1, Q2, Q3, Xmax   Se construye un diagrama de cajas que representa la forma de una distribución.  El Diagrama de Cajas ( box-plot ) se construye al calcular Xmin, Q1,Q2,Q3 y Xmax colocándoles en línea en forma proporcional,    Imagen extraida y manipulada por Paint del libro: Mario Triola; Elementary Statistics ; 10th ed.,2006, Pearson Educational Inc. USA  Ejemplo1: Para los siguientes n=25 datos determine Box and Wisker  12 12 13 13 13 14 14 14 15 15 15 16 16 17 18 20 20 21 21 24 25 30 30 36 40  Utilizando STADISK software  Minimum: 12  1st Quartile: 14  2nd Quartile: 16  3rd Quartile: 21  Maximum: 40    **Boxplot**    Imagenes generadas por programa STATDISK  **EJEMPLO#2**  Para los siguentes datos desarrolle el box-plot (box and wisker):   |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Data1 | 12 | 12 | 20 | 21 | 22 | 24 | 25 | 25 | 30 | 30 | 40 | | Data2 | 8 | 8 | 10 | 10 | 15 | 15 | 16 | 20 | 20 | 23 | 25 |     Imagenes generadas por programa STATDISK  **NUMEROS ATIPICOS (OUTLIERS**)  **OUTLIER- NUMEROS ATIPICOS**-son valores extremos. O son muy grandes o muy pequeños relativos a la mayoría de los números del experimento.  El propósito de determinar los “outliers” es primeramente la protección de la integridad de los datos. Para detectar errores al entrar los datos. Para determinar “outliers” se deben determinar las siguientes cercas: (Fronteras o limites)   * Valor atípico (Grande) (Upper Outlier) = Q3 + 1.5(IQR) * Valor atípico (Pequeño)(Lower Outlier) = Q1 - 1.5(IQR)   Cualquier valor mayor al Upper Outlier o menor a Lower Outlier son números atípicos o outliers. La mayoría de los programas estadísticos presentan los outliers como asteriscos al dibujar el Box Plot.    Imágenes extraídas y manipuladas utilizando Microsoft Paint, de la presentación en Power Point del CD del libro de Montgomery , Douglas C; Applied Statistics and Probability for Engineers, 2003, John Wiley & Sons Inc. USA. Chapter2  En la figura anterior se presenta un box plot (box and wisker) y a la vez las escalas o cercas (fronteras o limites) que representan los números atípicos (outliers)  A Q3+1.5(IQR) outlier a Q3+3(IQR) extream outlier por grande.  A Q1-1.5(IQR) outlier a Q1-3(IQR) extream outlier por pequeño.  EJEMPLO3: Utilizando los datos del siguiente ejemplo:  12 12 13 13 13 14 14 14 15 15 15 16 16 17 18 20 20 21 21 24 25 30 30 36 40  Minimum: 12  1nd Quartile: 14  2nd Quartile: 16  3rd Quartile: 21  Maximum: 40    IQR = 21-14 =7  1.5(IQR) = 1.5\*7=10.5  Upper Outlier= Q3+1.5(IQR)=21+10.5=31.5 Los valores 35 y 40 son outlies, porque son mayores a 31.5  Lower Outlier= Q1-1.5(IQR)= 14-10.5=3.5 No hay outliers por pequeños    **Boxplot**  ---------------  ----I + I------------------ \* \*  ---------------  ------+---------+---------+---------+---------+---------+C1  15.0 20.0 25.0 30.0 35.0 40.0  Imagen generada por programa MIN-TAB  12 12 13 13 13 14 14 14 15 15 15 16 16 17 18 20 20 21 21 24 25 30 30 36 40 |
|  |
| **🗍 Lección #7.3 Sesgos y Medidas de Posición Relativa (valor Z) Uso de STATDISK.**   * **Medidas de Forma y Sesgo (skewness)**     Imagen extraida y manipulada por Paint del libro :David Doane and Lori Seward; Applied Statistics in Business and Economics; McGraw-Hill Irwin, NY 2007, USA  **SESGO POSITIVO o A LA DERECHA – COLA A LA DERECHA**  Si una distribución tiene una cola a la derecha se dice que esta sesgada a la derecha o sesgo positivo.(promedio > mediana > moda)  **SESGO NEGATIVO o A LA IZQUIERDA – COLA A LA IZQUIERDA**  Si una distribución tiene una cola hacia la izquierda se dice que esta sesgada ala izquierda o sesgo negativo. (promedio < mediana<moda)  **AUSENUCIA DE SESGO = SIMETRIA**  Si una distribución tiene la misma cola a la derecha y ala izquierda es simétrica . Simetría es ausencia de sesgo. (promedio = mediana= moda)     * **VALOR Zo**   El valor Zo nos indica a cuantas desviaciones estándares esta un numero X de su propio promedio. Es una medida que nos relaciona la desviación estándar con el promedio. (Al igual que las medidas de posición relativa ( Regla empírica y Teorema de Tchevyshev) ).  Mediante la siguiente ecuación:   * **POBLACION** * **MUESTRA**   Zo = nos presenta el numero de desviaciones estándares a que un valor X esta de su promedio  Ejemplo: Para los siguientes datos : 1, 1, 2, 2, 2, 5, 5, 7, 8, 8,10 . Determine a cuantas desviaciones estándares esta 7 de su promedio.  Este problema requiere el computo del promedio de la desviación estándar.   |  |  | | --- | --- | | X | X^2 | | 1 | 1 | | 1 | 1 | | 2 | 4 | | 2 | 4 | | 2 | 4 | | 5 | 25 | | 5 | 25 | | 7 | 49 | | 8 | 64 | | 8 | 64 | | 10 | 100 | | 51 | 341 |   ∑X ∑X^2  PROMEDIO =  = 51/11 = 4.6363  VARIANZA =  = ( 341-(51)2/11)/ (11-1)=10.4545  DESVIACION ESTANDAR= S = SQRT( 10.4545)=3.2333  = (7-4.6363)/ 3.2333 = 0.731 desviaciones estándares     * .**ESTIMACION DE LA DESVIACION ESTANDAR**   En casos de emergencia que se requiera el valor de la desviación estándar y no se tenga tiempo para computo por falta de todos los datos o acceso a todos los datos. Podemos estimar la varianza utilizando el siguiente computo:  donde R=rango=Xmax - Xmin  Para los siguientes datos: 1, 1, 2, 2, 2, 5, 5, 7, 8, 8,10  S = (10-1)/4 = 9/4 =2.25 es un estimado de la desviación estándar.  La verdadera desviación estándar es 3.23. El estimado tiene un error de (3.23-2.25)/3.23=30% con relación al valor verdadero. Los estimados no son el valor verdadero, peor es no tener nada. Nos dan una idea para salir de un apuro o emergencia.  **🗁 USO DE STATDISK (**Uso de software para generar diagramas) En nuestro curso se utilizan programas como Mini-Tab, Excel y Statdisk.  El programa STATDISK lo supple el professor Mario Triola para el uso de los estudiantes de estadistica.( <http://wps.aw.com/aw_triola_stats_series/> http://www.statdisk.org/)    Se accesa por [www.statdisk](http://www.statdisk) luego te registraras luego, y te permitirá bajarlo.  Como hace un Box Plot en STATDISK  1. Llamar al programa STATDISK, sale hoja de entrada de datos, parecida a hoja electrónica. Favor de entrar los datos, en columna escogida.  2. Ir a menú de opciones DATA, escoger BOX PLOT    3. Luego escoger columna apropiada en este caso Col 1 y StatDISk presentara BOX PLOT    4. Si se desea información de estadística descriptiva. En menú, escoger DATA, luego escoger Explore Data e informar columna deseada. Ver próximas figuras.      Tambien me pueden contactar para darles una copia.  <http://freestatistics.altervista.org/?p=stat> |
| **Conclusion** (Closing statement paragraph) |
| Durante este modulo se estudiaron las medidas de posición relativa entre µ y σ (Teorema Tchebyshev y la Regla Empírica) . Ademas de cómo crear Diagrama de Caja (Box and Wisker Diagram) (Box Plot Diagram). Tambien se presentaron los números atípicos (outliers), tipos de sesgos y valor Zo. La estimación de la desviación estándar fue también presentada asi como el uso de software para generar diagramas de cajas. |
| **🗁 Learning Activities** (What you need to do) |
| Activities for this lesson: |
| **🗁 Learning Activities** (What you need to do) |
| Activities for this lesson:  Universidad Politécnica de Puerto Rico  **EJERCISIOS DE PRACTICA**  ENGI 2210/2270 Probability and Statistics for Engineers  Nombre:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Num:Est:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Prof. Ing. José Raúl Díaz, PE;    **I. Realicé el siguiente problema**: (Continuación del mismo problema en Modulo#3)  a- Para los siguientes datos: (**Ordenados)**  3 4 5 8 8 10 12 12 20 22  a) Calcule  1) la media o promedio 6) el rango  2) la mediana 7) el rango intercuartilico  3) la moda 8) la varianza  4) el rango medio 9) la desviación estándar  5) el eje medio 10) el coeficiente de variación  b) Prepare un grafico Box Plot (Box and Wisker)  c) Determine los valores de la frontera o cerca de los números atípicos ( outliers).  Conteste si hay o no números atípicos. De haberlos ¿cuáles son?  d) Determine a cuantas desviaciones estándares esta el numero 8 de su promedio.  e) Si se tuviera que estimar la desviación estándar ¿Cuál sería su estimado?  f) De acuerdo al Teorema de Tchevyshev entre que par de valores deben estar por lo menos el 75% de los datos. |
| **📋 Assignment #1** |
| (Include due dates, time frames and point possible).  1. Para los siguientes datos: Continuación del mismo problema en Modulo#3)  1, 1, 2, 3, 3, 3, 3, 3, 4, 4, 5, 5, 5  a) Calcule  1) la media o promedio  2) cuartil-1, cuartil-2, cuartil-3  3) Rango Intercuartilico  4) varianza (use formula computacional)  5) coeficiente de variación  b) Prepare un grafico Box Plot (Box and Wisker)  c) Determine los valores de la frontera o cerca de los números atípicos ( outliers).  Conteste si hay o no números atípicos. De haberlos ¿cuáles son?  d) Determine a cuantas desviaciones estándares esta el numero 4 de su promedio.  e) Si se tuviera que estimar la desviación estándar ¿Cuál sería su estimado?  **📋 Assignment #2**  Universidad Politécnica de Puerto Rico  **ASIGNACION**  ENGI 2210/2270 Probability and Statistics for Engineers  Nombre:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Num:Est:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Prof. Ing. José Raúl Díaz, PE;    **I. Realicé el siguiente problema**:  Para los siguientes datos: (**Ordenados) ∑X=880 ; ∑X2=39648**  30 34 37 38 38 40 42 42 42 42  44 44 47 48 48 49 50 52 56 57  a) Realice Tabla de clases, Intervalos ó fronteras , frecuencias relativa que se requiere para un Dibuje un Diagrama de Frecuencia Relativa. **Utilice 6 Clases**  b) Calcule  1) la media o promedio 5) Rango  2) quartil-1 6) Varianza  3) quartil-2 7) desviación estándar  4) quartil-3 8) coeficiente de variación  **c**) Prepare un grafico Box Plot (Box and Wisker)  d) Determine los valores de la frontera o cerca de los números atípicos ( outliers).  Conteste si hay o no números atípicos. De haberlos ¿cuáles son?  e) Determine a cuantas desviaciones estándares esta el numero 42 de su promedio.  f) Si se tuviera que estimar la desviación estándar ¿Cuál sería su estimado?  g) De acuerdo al Teorema de Tchevyshev entre que par de valores deben estar por lo menos el 75% de los datos.  **II Determine :** El coeficiente de variación (CV) de los dos conjuntos de datos.   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Data1 | 11 | 12 | 12 | 14 | 15 | 20 | | Data2 | 8 | 8 | 10 | 10 | 15 | 15 |     Para los dos sets :  a) Prepare un grafico Box Plot (Box and Wisker)  b) Determine los valores de la frontera o cerca de los números atípicos ( outliers).  Conteste si hay o no números atípicos. De haberlos ¿cuáles son? |
| **🗎 Self Assessment** (How to Know if you’re ready for grading) |
| Check your understanding: |
| **🗔 Assessment** (How to show me what you’ve learned)  Select the assessment tool to used: (Test or SafeAssignment) |
| (Include due dates, time frames and point possible). |
| **⭯ Virtual Group Activity**  Select the tool to used: (Blog, Discussion Board, Journal, Wiki, Group, Chat or Virtual Classroom) |
| (Include due dates, time frames and point possible). |
| **⏭ Optional Activity** |
|  |
| **Learning Resources** (Tools to help you learn) |
| **References**  David Doane and Lori Seward; Applied Statistics in Business and Economics; McGraw-Hill Irwin, NY 2007, USA  Mario Triola; Elementary Statistics ; 10th ed.,2006, Pearson Educational Inc. USA  Mark Berenson, David Levine and Timothy Krehbiel; Basic Business Statistics Concepts and Applications; 9th edition, 2004; Pearson Prentice-Hall, USA  Ken Black, David Eldredge3; Business & Economics Statistics; Using Microsoft Excel; South-Western; Thomson Learning, 2002; USA  William Menderhall, Robert Beaver y Barbara Beave; Introducción a la probabilidad y estadística ; 1ra. Edición 2002 International Thomson Editores, México DF  Ken Black; Business Statistics Comtemporary Decision Making; 3rd edition;2001; South-Western College Publishing Thomson Learning, USA |