

TALLER EN CLASE No 1

ESTADÍSTICA DESCRIPTIVA

Felipe Alejandro Avilés Cedeño - 201520281  
Juliana Ramos González - 201520745  
Mauricio Sánchez Parra - 201522503

UNIVERSIDAD DE LA SABANA

PROBABILIDAD Y ESTADISTICA

GRUPO 1

15 de Febrero del 2017

**INTRODUCCION**

El presente informe tiene por objetivo organizar y analizar dos grupos de datos relacionados con diferentes características. En primer lugar, se analizarán datos que representan las medidas de longitud (en mm) producidas por una máquina de tornillos de acero en un día laboral, por lo cual pertenecen a una variable continua. En segundo lugar, se analizarán datos que representan el número de ciclos transcurridos hasta que se presenta una falla en una prueba de piezas de aluminio sujetas a un esfuerzo de piezas alternante repetido de 21 000 psi, a 10 ciclos por segundo, siendo estos datos parte de una variable discreta.

En ambos casos se realizará una tabla de frecuencias agrupando los datos, al igual que se analizarán los resultados más detalladamente a través de tres representaciones gráficas: histogramas de frecuencias, polígono de frecuencias y ojiva de frecuencias. Finalmente se sacarán conclusiones que aporten información detallada sobre cada muestra.

**JUSTIFICACIÓN**

Es importante tener conocimiento sobre estadística descriptiva ya que nos ayuda no solo a organizar datos sino a sacar conclusiones sobre la muestra y en la mayoría de los casos a dar solución a un problema. Por otro lado la estadística descriptiva se puede aplicar a cualquier campo laboras que maneje una base de datos y es necesario para la resolución de problemas aplicados en ingeniería, programa que los tres estudiantes que realizan este informe estudian.

**MARCO TEORICO**

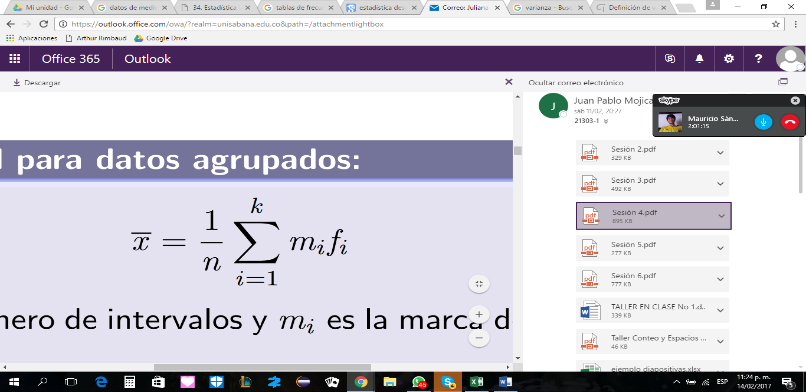
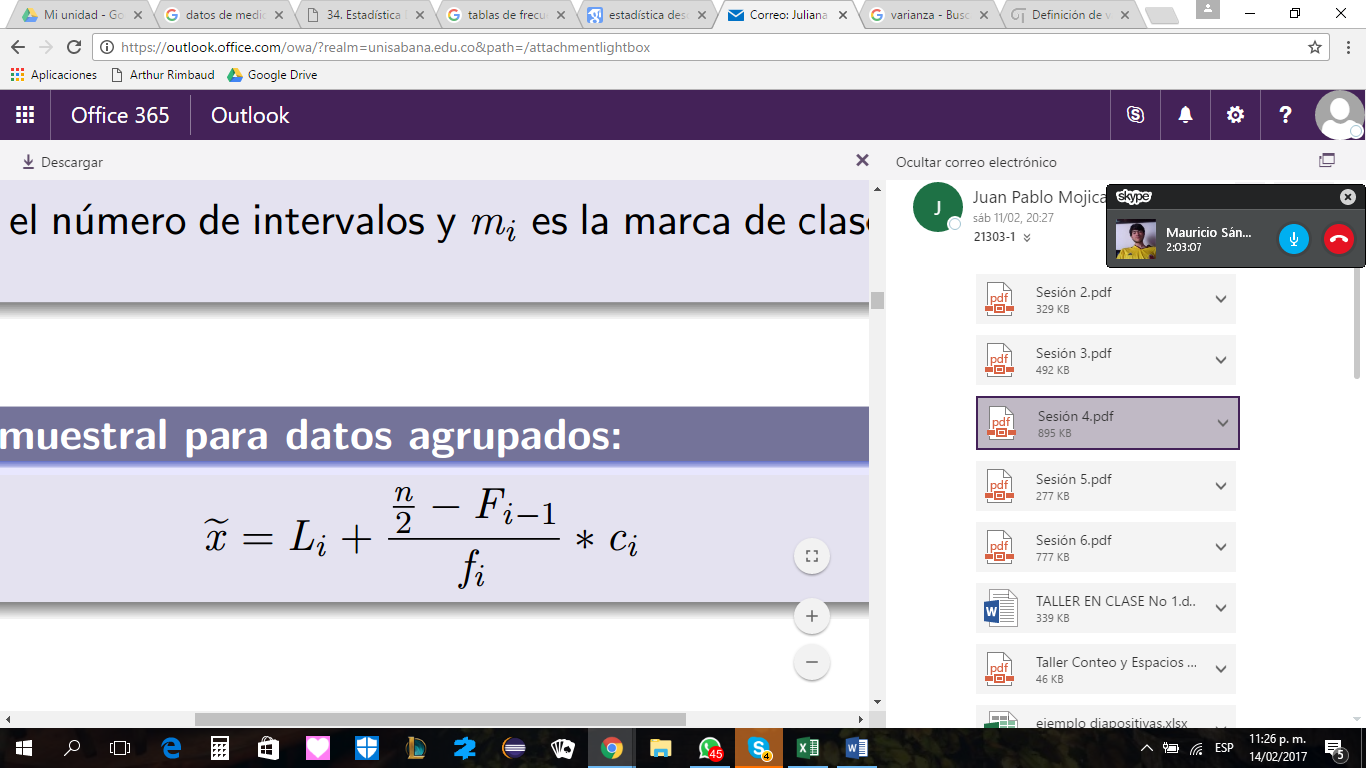
**Estadística**: Es el conjunto de métodos y técnicas que permiten recopilar, presentar, analizar y tomar decisiones respecto de un conjunto de datos. Existen métodos descriptivos e inferenciales.

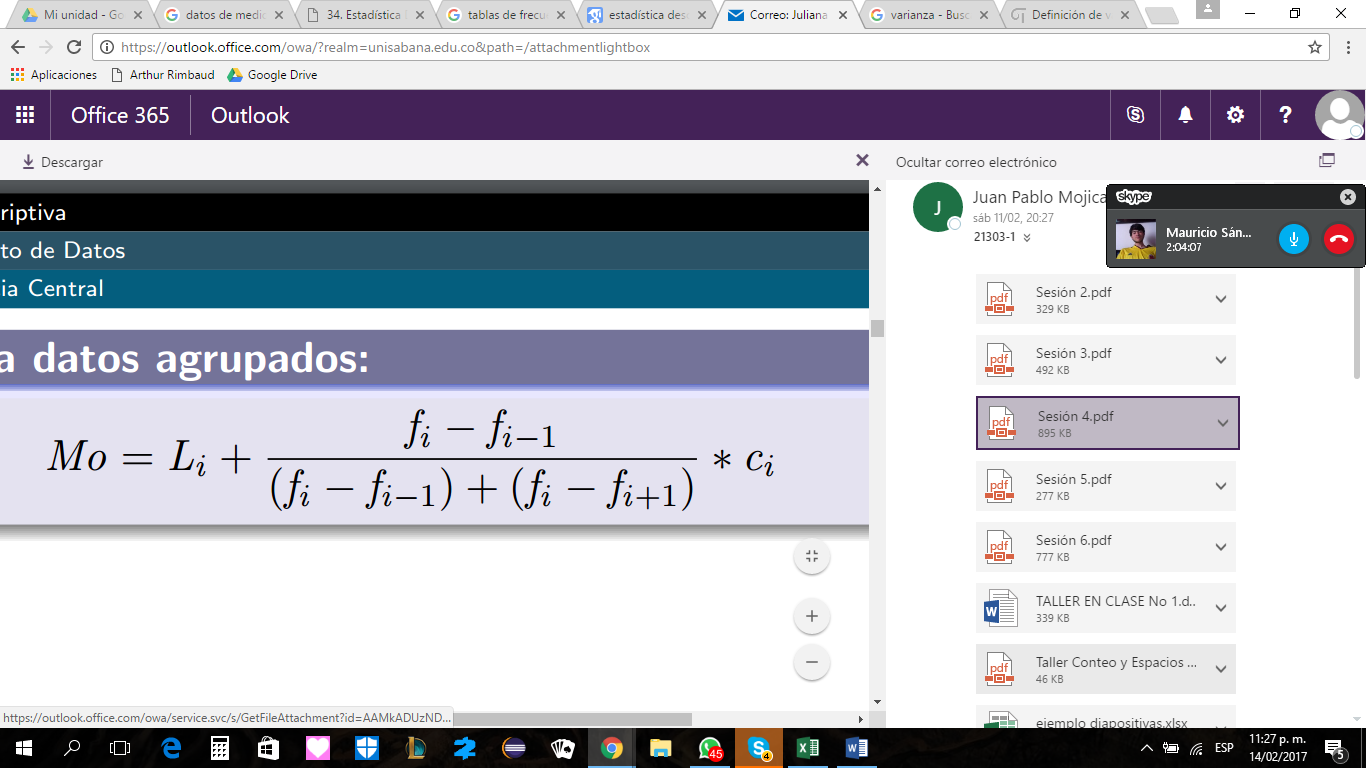
**Métodos descriptivos:** Estos métodos permiten conocer, representar y cuantificar el comportamiento de un conjunto de datos.

**Tabla de frecuencias para datos agrupados:** Son herramientas de estadística donde los datos son colocados en columnas, los cuales representan los diferentes valores obtenidos en la muestra, con sus respectivas.

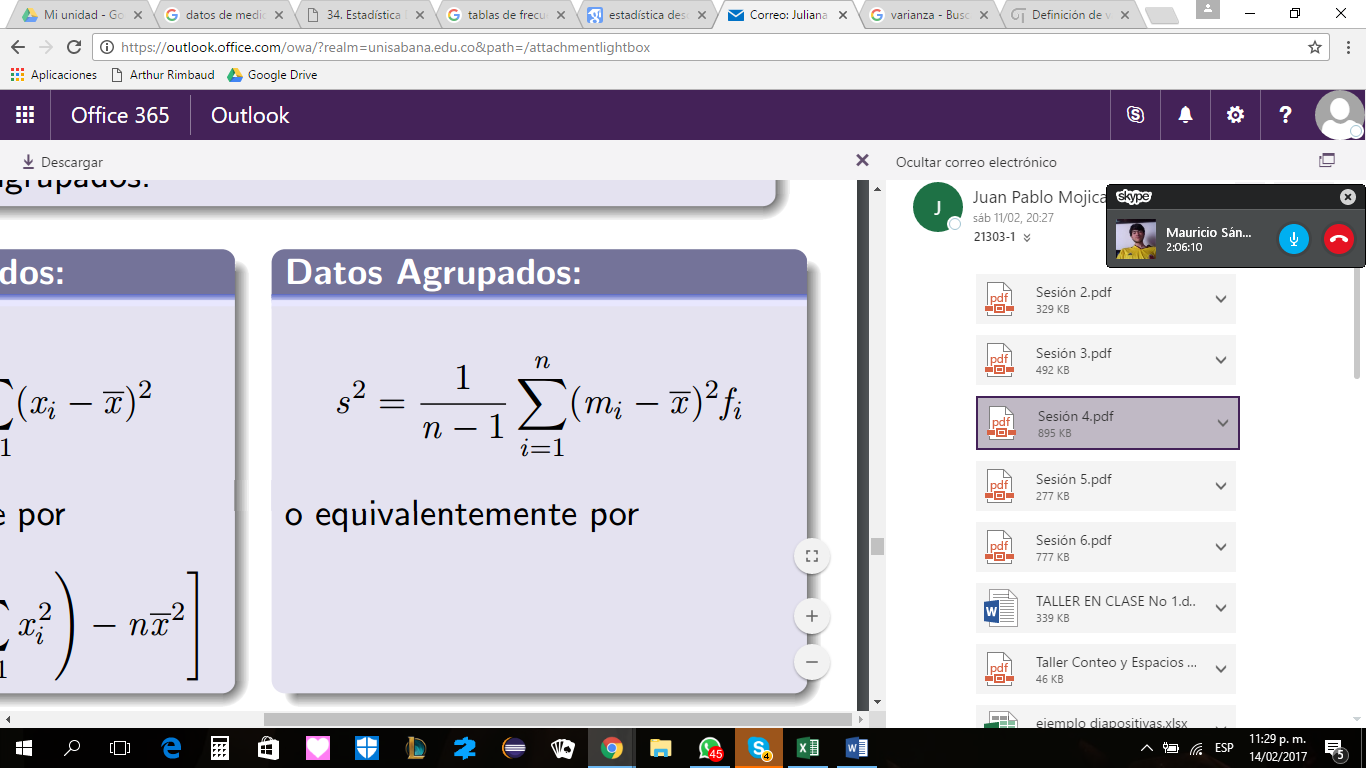
* **Datos(n):** son los valores obtenidos durante el estudio estadístico.
* **Rango (rg):** Es la diferencia entre el mayor y menor valor del dato. .
* **Clases (K):** Número de intervalos ; .
* **Amplitud (A):** Es la cantidad de datos que corresponden a cada clase. .
* **Marca de clase (Mi):** es el punto medio de cada intervalo. .
* **Frecuencia absoluta (Fa):** es la cantidad de datos que cuentan en cada intervalo, se debe tener en cuenta que la suma de todas las frecuencias absolutas debe ser igual al número de datos estudiados.
* **Frecuencia relativa (Fr):** Es la probabilidad de encontrar datos en el respectivo intervalo, el porcentaje de participación en cada clase, la suma total debe ser igual a 1. .
* **Frecuencia absoluta acumulada:** Es ir acumulando clase a clase la frecuencia absoluta.
* **Frecuencia relativa acumulada:** Se procede de manera igual que en el caso anterior pero con la frecuencia relativa.

**Medidas de tendencia central:** Son las cantidades que indican el punto central del conjunto de datos.

* **Media muestral:** Usada cuando la variable es numérica y poco dispersa.
* 
* **Mediana:** Es la cantidad que esta justamente en el centro de un conjunto de datos
* **Moda:** Es la cantidad que más frecuencia tiene dentro de un conjunto de datos, se debe tener en cuenta si n es par o impar.



* **Varianza:** Sera un valor positivo o cero

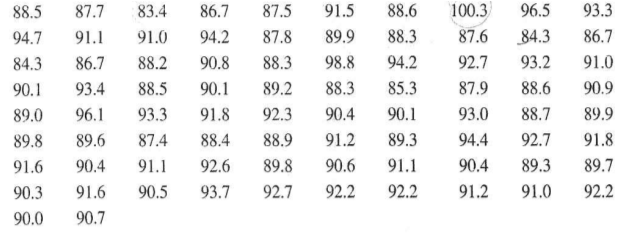


* **Desviación estándar:**

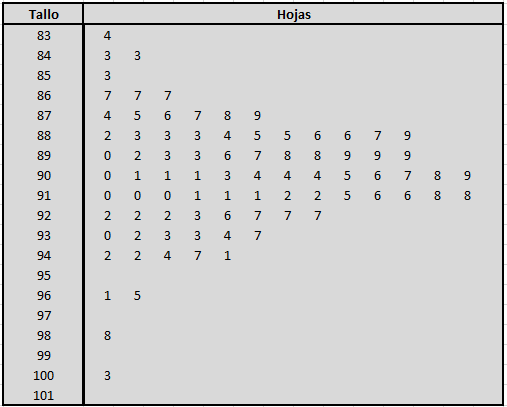
**RESULTADOS**

Muestra 1: Los datos siguientes representan las medidas de longitud (en mm) producidas por una máquina de tornillos de acero en un día laboral.

Datos:



1. Construya un diagrama de tallos y hojas para estos datos.



b) Elabore una tabla de frecuencias de datos agrupados para estos datos.

|  |  |
| --- | --- |
| **n** | 82 |
| **xmin** | 83,4 |
| **xmax** | 100,3 |
| **ran** | 16,9 |
| **k** | 7 |
| **amp** | 2,00 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **clase** | **L inf** | **L sup** | **mi** | **fi** | **Fi** | **ri** | **Ri** | **%ri** | **%Ri** |
| 1 | 83 | 85,00 | 84,0 | 3 | 3 | 0,03658537 | 0,03658537 | 3,66 | 3,66 |
| 2 | 85,00 | 87,00 | 86,0 | 4 | 7 | 0,04878049 | 0,08536585 | 4,88 | 8,54 |
| 3 | 87,00 | 89,00 | 88,0 | 18 | 25 | 0,2195122 | 0,30487805 | 21,95 | 30,49 |
| 4 | 89,00 | 91,00 | 90,0 | 25 | 50 | 0,30487805 | 0,6097561 | 30,49 | 60,98 |
| 5 | 91,00 | 93,00 | 92,0 | 19 | 69 | 0,23170732 | 0,84146341 | 23,17 | 84,15 |
| 6 | 93,00 | 95,00 | 94,0 | 9 | 78 | 0,1097561 | 0,95121951 | 10,98 | 95,12 |
| 7 | 95,00 | 97,00 | 96,0 | 2 | 80 | 0,02439024 | 0,97560976 | 2,44 | 97,56 |
| 8 | 97,00 | 99,00 | 98,0 | 1 | 81 | 0,01219512 | 0,98780488 | 1,22 | 98,78 |
| 9 | 99,00 | 101,00 | 100,0 | 1 | 82 | 0,01219512 | 1 | 1,22 | 100 |
|  |  |  |  |  |  |  |  |  |  |

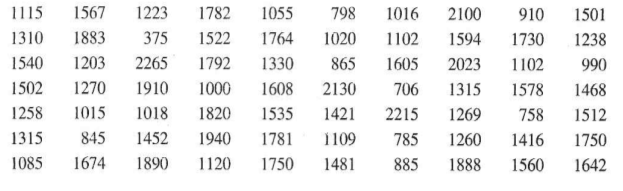
Cálculo de medidas de tendencia central:

|  |  |
| --- | --- |
| **media** | 90,4146341 |
| **mediana** | 90,28 |
| **moda** | 90,0769231 |
| **varianza** | 3,24222795 |
| **desviación estándar** | 1,80061877 |

c) Elabore la representación gráfica (histogramas de frecuencias, polígono de frecuencias y ojiva de frecuencias.

Aunque el histograma y el polígono de frecuencias son iguales según sus ejes, es mucho más claro ver la tendencia de los datos de la base de datos con el histograma.

Muestra 2: Los datos representan el número de ciclos transcurridos hasta que se presenta una falla en una prueba de piezas de aluminio sujetas a un esfuerzo de piezas alternante repetido de 21 000 psi, a 10 ciclos por segundo.



a) Elabore una tabla de frecuencias de datos agrupados para estos datos.

|  |  |
| --- | --- |
| **n** | 70 |
| **xmin** | 375 |
| **xmax** | 2265 |
| **ran** | 1890 |
| **k** | 7 |
| **amp** | 270,00 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **clase** | **L inf** | **L sup** | **mi** | **fi** | **Fi** | **ri** | **Ri** | **%ri** | **%Ri** |
| 1 | 375 | 645,00 | 510,0 | 1 | 1 | 0,01428571 | 0,01428571 | 1,43 | 1,43 |
| 2 | 645,00 | 915,00 | 780,0 | 8 | 9 | 0,11428571 | 0,12857143 | 11,43 | 12,86 |
| 3 | 915,00 | 1185,00 | 1050,0 | 13 | 22 | 0,18571429 | 0,31428571 | 18,57 | 31,43 |
| 4 | 1185,00 | 1455,00 | 1320,0 | 14 | 36 | 0,2 | 0,51428571 | 20 | 51,43 |
| 5 | 1455,00 | 1725,00 | 1590,0 | 16 | 52 | 0,22857143 | 0,74285714 | 22,86 | 74,29 |
| 6 | 1725,00 | 1995,00 | 1860,0 | 13 | 65 | 0,18571429 | 0,92857143 | 18,57 | 92,86 |
| 7 | 1995,00 | 2265,00 | 2130,0 | 5 | 70 | 0,07142857 | 1 | 7,14 | 100 |

Cálculo de medidas de tendencia central:

|  |  |
| --- | --- |
| **media** | 1416,42857 |
| **mediana** | 1435,71429 |
| **moda** | 1563 |
| **varianza** | 30525,9317 |
| **desviación**  **estándar** | 174,716718 |

1. Elabore la representación gráfica (histogramas de frecuencias, polígono de frecuencias y ojiva de frecuencias.

**CONCLUSIONES**

-Se puede llegar a que en la primera base de datos la gran parte de la fabricación de tornillos tiende a ser entre las longitudes de 90 a 91 mm de longitud, y esto se puede llegar a concluir, gracias al diagrama de tallo y hojas sin aún revisar la tabla de frecuencia.

-Los cálculos de medida central de moda, media y mediana en la primera base de datos, se tiene que los tres datos son muy semejantes al valor de 90 mm, con lo cual se puede llegar a que la fabricación normal de estos tornillos llega a los 90 mm con muy pocos errores, y también con muy poca variabilidad entre las medidas de los tornillos.

-El cálculo de la media en ambos estudios se realizó con la fórmula de datos agrupados, que no resulta ser muy exacto a comparación de hacer la media como datos no agrupados, sin embargo, su diferencia entre un cálculo y otro, no es muy significativo en sus cifras decimales.

-De los 82 datos de la primera base de datos, solo 4 tornillos excedieron de los 95 mm de longitud, con lo cual también se puede observar, que con el valor de la desviación estándar nos permite también llegar a concluir lo mismo sin tener que observar toda la muestra.

-En la segunda base de datos los datos son mucho más variados y se puede saber por su valor máximo y mínimo, por los limites inferiores y superiores de cada intervalo, y a su vez el cálculo de varianza y desviación estándar presentadas anteriormente.

-Normalmente en la segunda base de datos, los valores de ciclos hasta antes presentarse una falla se muestran más entre el número 915 ciclos hasta 1590 ciclos. Aunque son 43 datos de los 70, no es muy certero cual es el verdadero valor con mayor repetición (diciendo esto solo viendo la tabla de frecuencias).

-Los cálculos de moda, media y mediana en la segunda base de datos no son semejantes a comparación de la primera base de datos (según similitud), pero esto se puede deber al valor de la marca de clase en cada intervalo

- Con la ayuda de la ojiva de frecuencias, se puede observar que el crecimiento de la frecuencia absoluta acumulada en la primera base de datos es mayor a comparación de la segunda base de datos, aunque no tiene que ver una base de datos esté relacionada con la otra, sino que, con este gráfico, se puede obtener una mejor visualización de los datos y su frecuencia absoluta acumulada.

BIBLIOGRAFIA

Parra, J. M. (1995). ESTADÍSTICA DESCRIPTIVA E INFERENCIAL I.

Fernández, S. F., Sánchez, J. M. C., Córdoba, A., Cordero, J. M., & Largo, A. C. (2002). *Estadística descriptiva*. ESIC Editorial.

Gómez, M. (1998). Elementos de estadística descriptiva. *Universidad Estatal a Distancia, San José, Costa Rica*.

Barrantes, M. G. (1998). *Elementos de estadística descriptiva*. EUNED.