

Informe Ciencia de Datos - LA ConGA

Carlos Andres Pinzòn Osorio

UNIVERSIDAD ANTONIO NARIÑO

February 26, 2021

Content

- 1 Introduccion
- 2 Datos y Procesamiento
- 3 Analisis por Pais
- 4 Analisis Global
- 5 Conclusiones

Introducción

La obtención de los datos se dio mediante el uso del repositorio de ciencia de datos perteneciente a la clase 7 en Git.

El objetivo principal del presente informe pretende aplicar los conocimientos que se obtuvieron en el modulo 1 concernientes al uso de herramientas tales como Python o ROOT para el procesamiento y análisis de datos, así como la acción de estos mediante los diferentes ambientes de colaboración y aplicación de herramientas que permiten la transmisión de la información a su vez que su manejo. En un archivo .ipnyb adjunto se muestra el proceso para convertir el archivo de extension .CSV a .ROOT. El trabajo se desarrollo en tres secciones como se indica a continuación:

- 1 Herramientas y procesamiento
- 2 Análisis
- 3 Conclusiones y observaciones

Datos y Procesamiento

	individuo	origen	peso	tamano
0	cam01	Alemania	18	26.0
1	cam02	Alemania	38	62.0
2	cam03	Alemania	39	59.0
3	cam04	Alemania	28	41.0
4	cam05	Alemania	7	13.5
5	cam06	Alemania	29	46.5
6	cam07	Alemania	19	25.5
7	cam08	Alemania	25	40.5
8	cam09	Alemania	40	59.0
9	cam10	Alemania	16	19.0
10	tur01	Inglaterra	26	39.0
11	tur02	Inglaterra	17	21.5
12	tur03	Inglaterra	32	49.0
13	tur04	Inglaterra	21	31.5
14	tur05	Inglaterra	21	27.5
15	pet01	Inglaterra	6	11.0
16	pet02	Italia	8	12.0
17	pet03	Italia	28	40.0
18	pet04	Italia	10	15.0
19	pet05	Italia	19	28.5
20	pet06	Italia	13	17.5
21	pet07	Italia	16	29.0
22	pet08	Italia	15	20.5
23	pet09	Italia	36	54.0
24	pet10	Francia	39	53.5
25	tur06	Francia	33	49.5
26	tur07	Francia	31	43.5
27	tur08	Francia	9	15.5
28	tur09	Francia	10	17.0
29	tur10	Francia	5	11.5
30	tur11	Francia	20	26.0
31	tur12	Francia	27	43.5

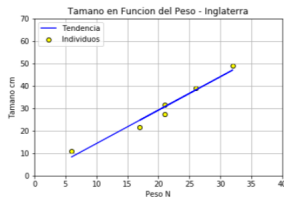
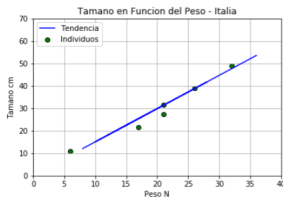
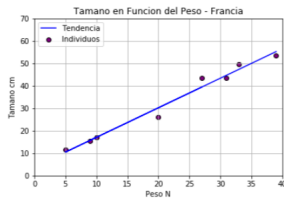
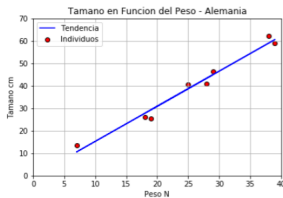
Archivo : analisis.CSV

Lenguaje : PYTHON 3

- **Pandas**
Manejo de dataframe
- **Numpy**
Conversión y extracción de datos
- **Matplotlib**
Gráficas
- **statistics**
Cálculos de las medidas de dispersión
- **sklearn**
Calculo de lineas de tendencia

Manejo de 4 variables donde se relacionan peso , tamaño , origen y se etiqueta a un conjunto de 32 individuos.

Analisis Por Pais



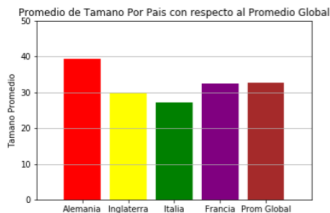
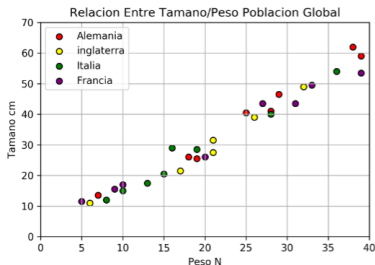
Se estableció una correlación de la información mediante el peso (P) y tamaño (T) de los individuos de cada país. La media y la desviación estándar se calcularon usando:

`statistics.mean()` --> μ
`statistics.stdev()` --> σ

Las funciones lineales se encuentran en el archivo ipynb.

País	μP	μT	σP	σT
Ale	25.37	39.25	10.67	16.87
Ing	20.50	29.91	8.78	13.29
Ita	18.12	27.06	9.46	14.17
Fra	21.75	32.5	12.65	16.83

Analisis Global



País	Promedio Peso	Promedio Tamaño
Ale	25.37	39.25
Ing	20.50	29.91
Ita	18.12	27.06
Fra	21.75	32.5
Global	21.90	32.75

Conclusiones y Observaciones

- 1 La población analizada corresponde a 10 alemanes, 8 franceses, 6 ingleses y 8 italianos.
- 2 Las gráficas indican que la mejor relación y homogeneidad entre tamaño/peso de individuos se encuentra en Francia.
- 3 La mayor homogeneidad de tamaños se encuentra en Inglaterra, donde se pudo evidenciar que su desviación es la mas pequeña con respecto a su media local.
- 4 En cuanto al peso la mayor homogeneidad se encuentra en Inglaterra, donde de nuevo su desviación es la mas baja con respecto a la su media.
- 5 Italia posee los individuos mas heterogéneos en cuanto su tamaño y peso de acuerdo a sus desviaciones estándar.

Conclusiones y Observaciones

- 1 La poblacion global de individuos representada en la grafica " relacion entre tamano/peso poblacion global" indica que los italianos poseen el mayor porcentaje de individuos (75 por ciento) con menor tamano y peso al ubicarse por debajo de los 25N y 40 cm)
- 2 Los Alemanes poseen la poblacion de mayor tamano (62.5 por ciento) ubicada en peso suepriores a los 25 N y 40 cm de tamano.
- 3 Cabe mencionar que existe una franja donde no se tienen datos de individuos , esa franaj esta entre los 22N a 24 N y 32 cm a 38cm.