

Aplicacion de tecnicas de estimacion y prueba de hipotesis

Caso: tendencias socio-economicas de algunas lineas de carrea de Ingenieria de sistemas

Alvarez Bautista Burga Casanova Cuyate

Facultad de Ingenieria Industrial y de Sistemas
Universidad Nacional de Ingenieria

Octubre 2022

Tabla de Contenido

- 1 Problema
- 2 Objetivos
- 3 Importancia de los objetivos
- 4 Resultados y Conclusiones

Tabla de Contenido

1 Problema

2 Objetivos

3 Importancia de los objetivos

4 Resultados y Conclusiones

- Empiricamente se observa que en el mundo la precarizacion del trabajo se acrecenta cada vez mas, de igual modo con la evolucion de personas casadas y los salarios promedio de los jovenes
- Con motivo de generar informacion util para la prediccion de estas tendencias socio-economicas se ha procedio a realizar un analisis estadistico sobre 9 hipotesis planteadas

Tabla de Contenido

1 Problema

2 **Objetivos**

3 Importancia de los objetivos

4 Resultados y Conclusiones

Objetivos del trabajo

General

Generar informacion relevante para la prediccion de tendencias socio-economicas en el mundo tomando como referencia datos provenientes de distintos paises.

Hipotesis especificas

- La distribucion de ingresos de sigue la ley normal
- La eleccion de especialidad es causa de la formacion de 2 grupos en la poblacion.
- Las personas que trabajan una cantidad de horas superior a la media tienen una mejor distribucion de ingresos que aquellas que no lo hacen
- En paises desarrollados existe una mayor cantidad de mujeres con puestos de trabajos relacionados a ingenieria que en paises en via de desarrollo
- Los cientificos de datos poseen un mejor distribucion de ingresos que los ingenieros de datos
- El sector (*publico / privado*) al que pertenece un trabajador es causa de la diferencia de salarios

Hipotesis especificas

- El **promedio de ingresos** de las personas que trabajan una cantidad de horas superior a la mediana es mayor al promedio de ingreso de personas que laburan una cantidad menor de horas que la mediana
- Las personas de mediana edad poseen una mejor distribucion de ingreso que las personas jovenes
- El promedio de ingresos de la poblacion mexicana es mayor que la peruana

Cada hipotesis tiene asociado el objetivo de comprobar o rechazar la suposicion

Tabla de Contenido

- 1 Problema
- 2 Objetivos
- 3 Importancia de los objetivos
- 4 Resultados y Conclusiones

Acerca de la normalidad de la distribucion de ingresos

Ever no te olvides del texto

Sobre la eleccion de especialidad

Se desea analizar principalmente:

- 1 Si la eleccion de la especializacion es causa de la formacion de 2 clusters en la poblacion
- 2 Que tan confiable resulta usar **ANOVA** ¹

¹*Analysis of Variance*

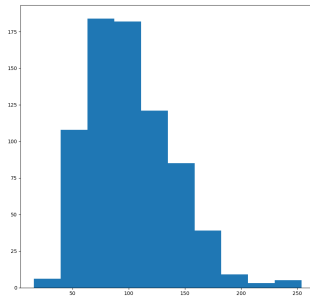
Sobre los demas objetivos

Se consideraron las hipotesis de tal forma que brinde información relevante para el análisis del mercado laboral para los egresados de la carrera de *Ingenieria de sistemas*

Tabla de Contenido

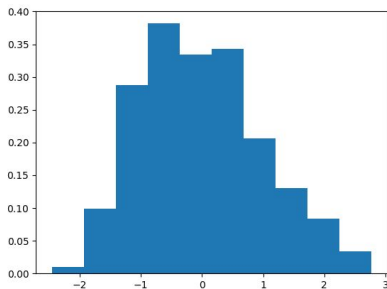
- 1 Problema
- 2 Objetivos
- 3 Importancia de los objetivos
- 4 Resultados y Conclusiones

Figura: Data sin estandarizar



Hipotesis 1

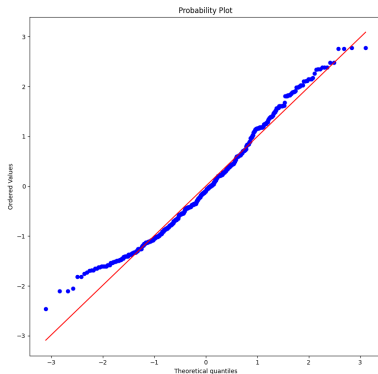
Figura: Data estandarizada y sin outliers



Puede parecer una distribucion Normal

Hipotesis 1

Figura: Grafica Q-Q



Se aleja de la distribución normal

Se aplicó el test de *Jarque-Bera*, para comprobar si la muestra presenta una **curtosis** y **asimetría** correspondientes a una ley normal.

El estadístico de *Jarque Bera* es asintóticamente un estimador de una *Chi-Cuadrado* (χ_n^2) y toma como hipótesis nula que los datos de la muestra siguen la ley normal

Test de Jarque-Bera

$$\mathbf{JB} = \frac{n}{6}(S^2 + \frac{1}{4}(K - 3)^3)$$

Siendo n los grados de libertad

Estimadores de momentos centrales

- Tercer Momento Central

$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3}$$

- Cuarto Momento Central

$$K = \frac{\hat{\mu}_4}{\hat{\sigma}^4}$$

Adicionalmente, se usara el test de *Kolmogorov-Smirnov*, donde se plantea que la distribucion de ingresos en la poblacion de ciencia de datos no sigue la ley normal y se comparará con la funcion acumulada teoria de esta

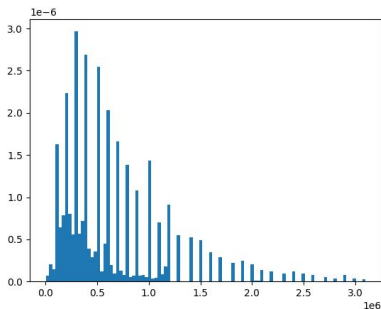
Conclusiones hipotesis 1

El test K-S y el de Jarque-Bera muestran los siguientes p-values.

Hipotesis 2

Antes de realizar cualquier tecnica de inferencia es necesario conocer la forma de las distribuciones, incluso antes de analizar la varianza

Figura: Distribución de ingresos de ingenieros de software en la India

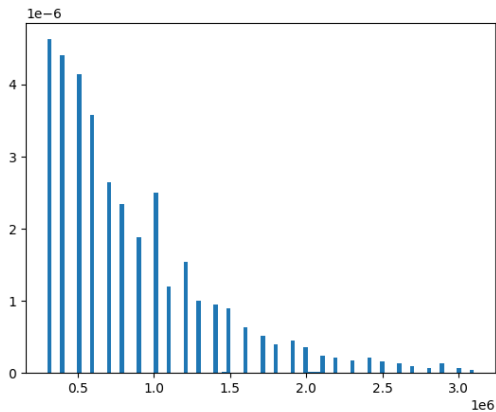


se puede notar como existen *2 grupos en la poblacion*

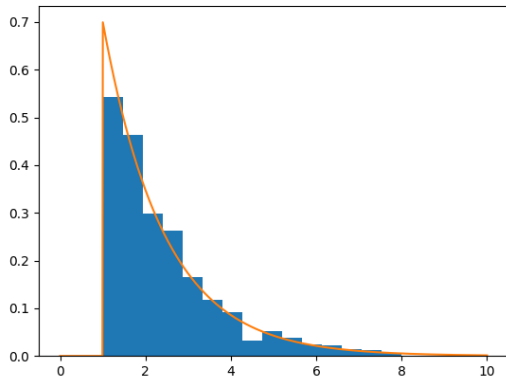
Aplicacion del test **Kolmogórov-Smirnov**

En este caso se va a comprar la funcion de distribucion acumulada observada con la de la distribucion teoria de una exponencial

Separando grupos aparentes



Ajustando Curva



Funciones acumuladas

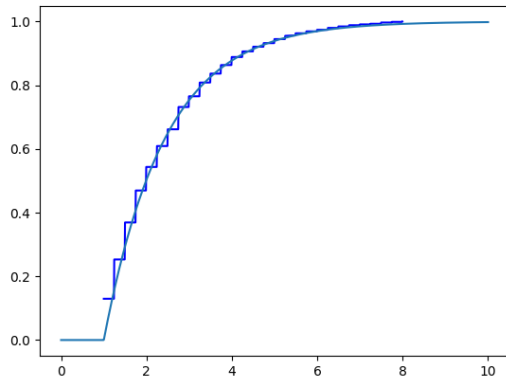
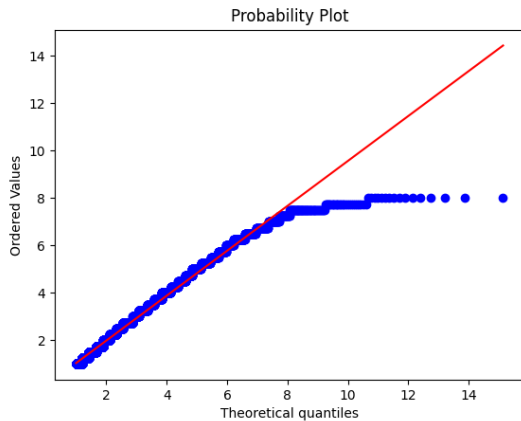


Grafico P-P



De acuerdo al p-value obtenido no se puede rechazar la hipótesis nula

```
Length of list: 11128
```

```
KstestResult(statistic=0.129449368008136, pvalue=1.4083532442224077e-201)
```