****

**МАНДАХ ИХ СУРГУУЛЬ**

**Тэнхмийн нэр**

**SW2**

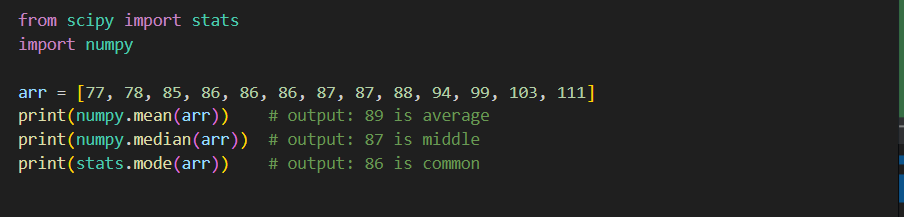
Machine Learning

Гүйцэтгэсэн : Болд Эрдэнэ , О.Лхагвабаатар

**УЛААНБААТАР ХОТ**

**2020ОН**

Mean median and mode



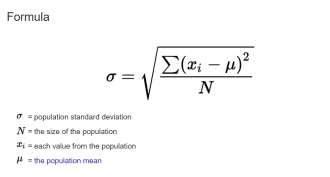
Mean: гараас авсан тооны дундажыг гаргаж ирдэг функц

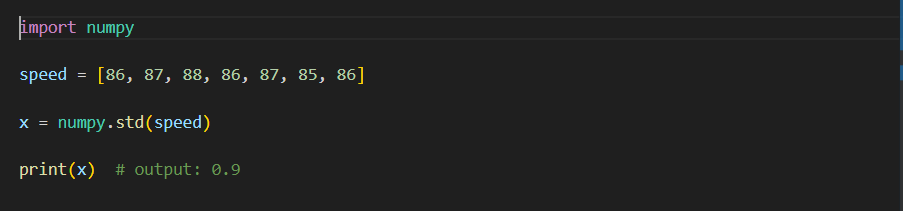
Median- гараас авсан тооны яг голын тоог гаргаж ирдэг функц

Mode- гараас авсан тооны хамгийн их давтагтаж буй тоог гаргаж ирдэг функц

Standard deviation буюу

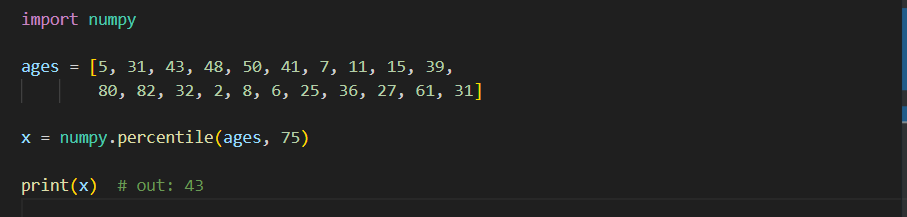
Стандарт хэлбэлзэл

 нь тайлбарлах статистикт дундаж утгатай харьцуулахад өгөгдлийн цэгүүдийн тархалтын зэрэг юм. Энэ нь өгөгдлийн түүврийн утгууд хэрхэн тархаж байгааг хэлдэг бөгөөд энэ нь өгөгдлийн цэгүүдийн дунджаас ялгаатай байдлын хэмжүүр юм. Өгөгдлийн багц, түүвэр, статистик популяци, санамсаргүй хэмжигдэхүүн эсвэл магадлалын тархалтын стандарт хазайлт нь түүний дисперсийн квадрат язгуур юм.

page4image26562496

Энэ функц нь дээрх Стандарт хэлбэлзэл томъеог ашиглаж гараас өгсөн тоон утгын тархалтыг харуулдаг

Percentiles буюу хувь

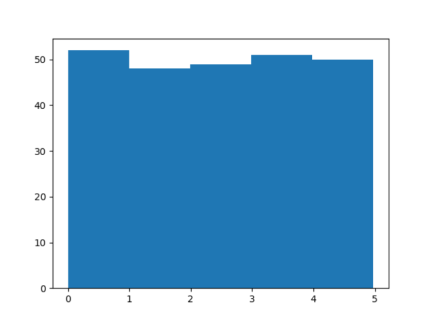
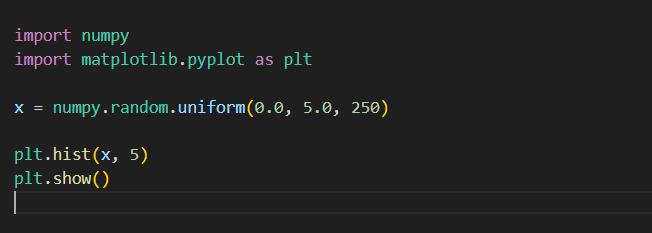


гараас авсан тооны хэдэн % ямар утга эзэлж байгааг харуулдаг функц

жишээн дээрх кодын 75% нь 43 наснаас залуу байна.

Data Distribution (histogram)

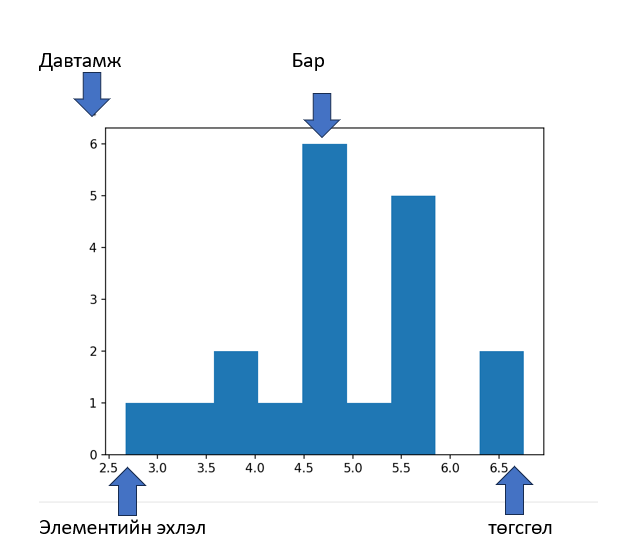
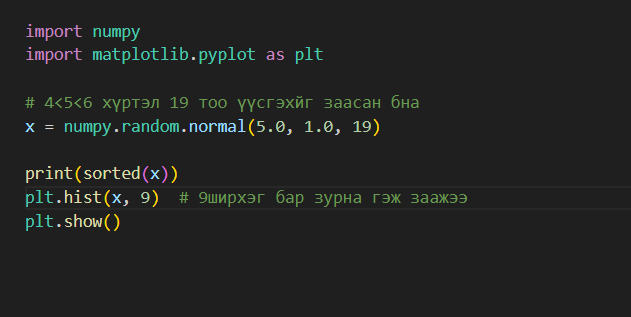
буюу Өгөгдлийн хуваарилалт (гистограм)

page6image26559616

Энэ код нь - гистограмм нь тоон хувьсагчийн утгуудын тархалтыг зураасаар харуулсан диаграмм юм.

Normal Data Distribution

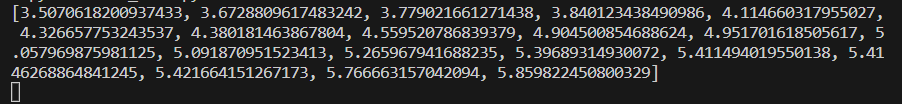
Ердийн өгөгдөл түгээлт

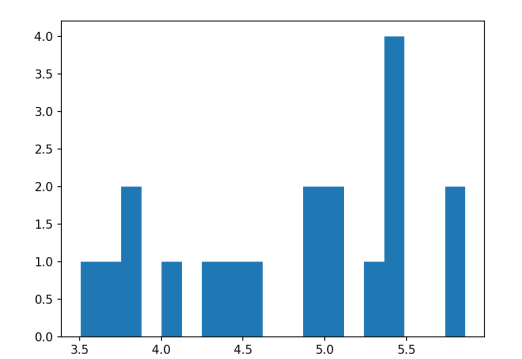
page7image26480576

Энэ код нь өгөгдөл нь хазайлтгүйгээр тэгш хэмтэй тархсан. Ихэнх үнэт зүйлс нь төвийн бүсийг тойрон бөөгнөрөх ба төвөөс холдох тусам үнэ цэнэ нь багасдаг.

Array to histogram

Гистограм руу массив





массив шиг

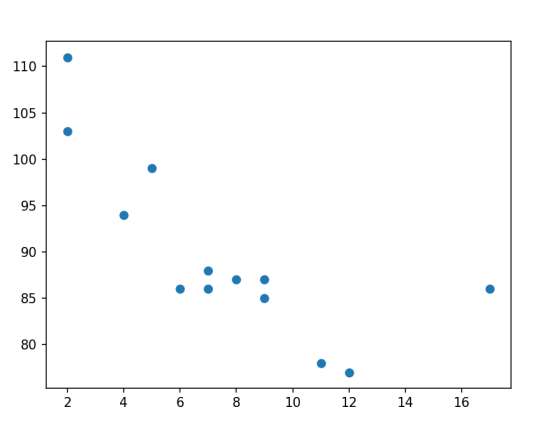
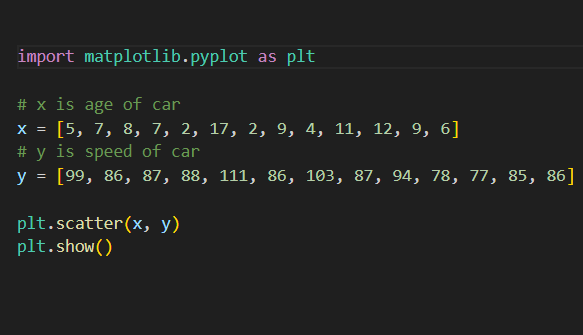
Мэдээлэл оруулах. Гистограммыг хавтгайруулсан массив дээр тооцоолно.

binsint эсвэл скалярын дараалал эсвэл str, сонголттой

Хэрэв bins нь int бол өгөгдсөн муж дахь ижил өргөнтэй савны тоог тодорхойлно

Scatter Plot (diagram)

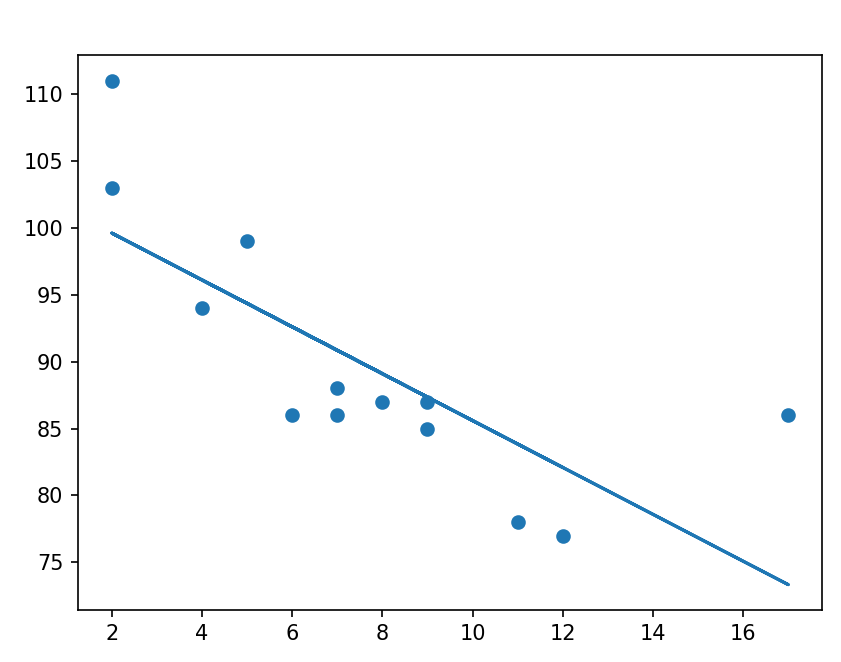
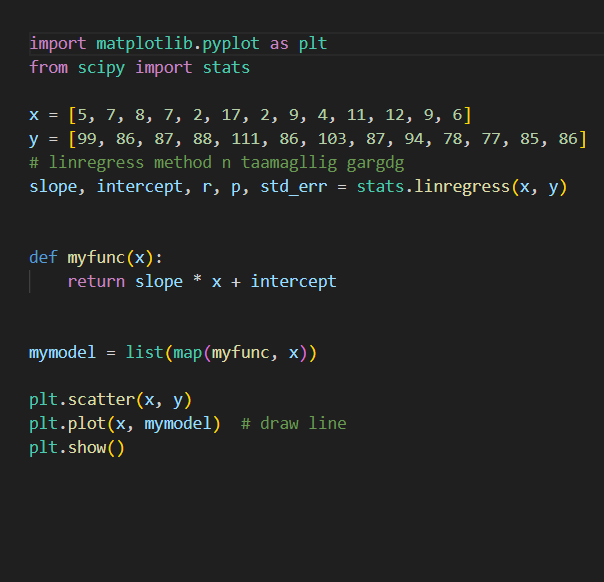
Тархалтын график (диаграмм)

page9image26521408

Тархалтын график нь өгөгдлийн багц дахь утга бүрийг цэгээр дүрсэлсэн диаграмм юм. Matplotlib модуль нь тархалтын график зурах аргатай бөгөөд түүнд ижил урттай хоёр массив хэрэгтэй, нэг нь x тэнхлэгийн утгууд, нөгөө нь у тэнхлэгийн утгууд юм.

Linear Regression(to predict outcome)

Шугаман регресс (үр дүнг урьдчилан таамаглах)

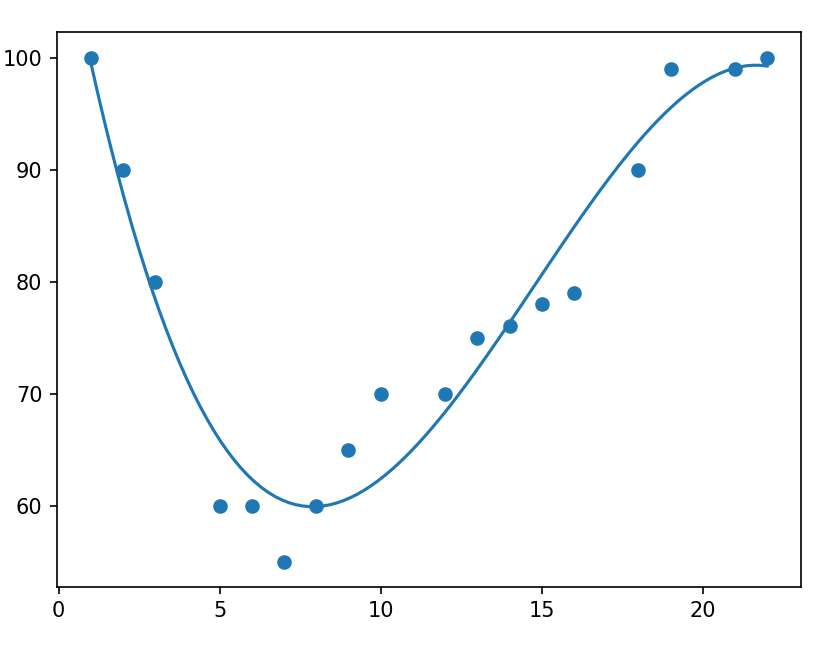
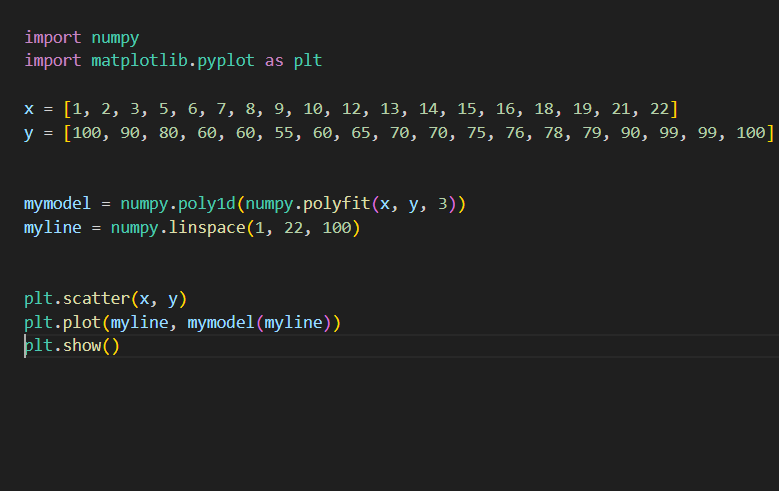
page10image26539136

Хувьсагчдын хоорондын хамаарлыг олох гэж оролдох үед регресс гэдэг нэр томъёог ашигладаг.

Machine Learning болон статистик загварчлалд энэ хамаарлыг ирээдүйн үйл явдлын үр дүнг урьдчилан таамаглахад ашигладаг.

Polynomial Regression

Олон гишүүнт регресс

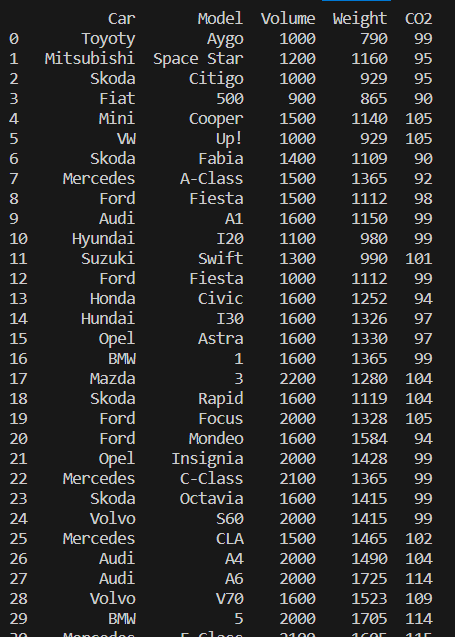
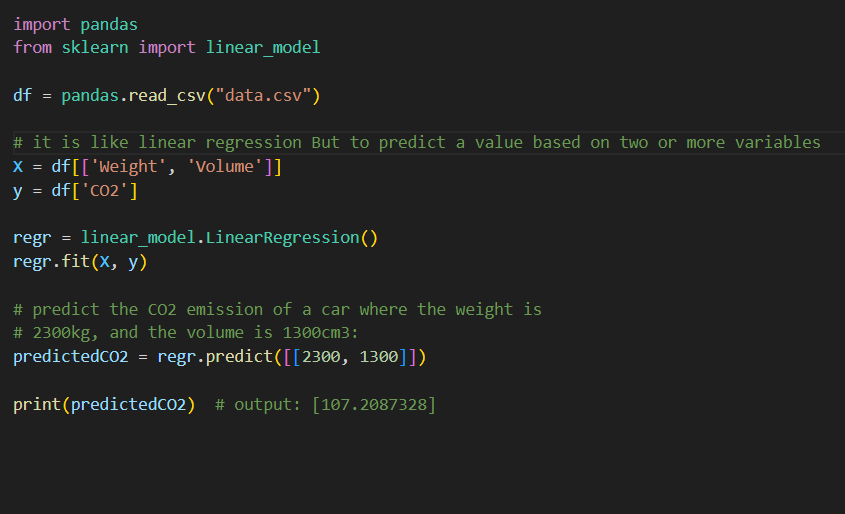
page11image26539520

Хэрэв таны өгөгдлийн цэгүүд шугаман регресстэй (бүх өгөгдлийн цэгээр дамжин өнгөрөх шулуун шугам) тохирохгүй байвал энэ нь олон гишүүнт регрессийн хувьд тохиромжтой байж болох юм.

Шугаман регрессийн нэгэн адил олон гишүүнт регресс нь өгөгдлийн цэгүүдээр шугам татах хамгийн сайн аргыг олохын тулд x ба y хувьсагчдын хоорондын хамаарлыг ашигладаг.

Multiple Regression

Олон регресс

page12image26479232

эцсийн регресс нь шугаман регресстэй адил боловч нэгээс олон бие даасан утгатай, өөрөөр хэлбэл бид хоёр ба түүнээс дээш хувьсагч дээр үндэслэн утгыг таамаглахыг оролддог гэсэн үг юм.

Доорх өгөгдлийн багцыг хараарай, энэ нь автомашины талаархи зарим мэдээллийг агуулдаг.

Train/Test

import numpy  
import matplotlib.pyplot as plt numpy.random.seed(2)

x = numpy.random.normal(3, 1, 100)  
y = numpy.random.normal(150, 40, 100) / x

plt.scatter(x, y) plt.show() page14image26414464

