

Отчет о Лабораторной №1

Введение

В рамках данной лабораторной работы проводится первичный статистический анализ двух наборов данных - "Airport" и "babyboom".

Ключевые задачи исследования:

1. Визуальный анализ распределений переменных с использованием гистограмм и ящиков с усами
2. Определение типа распределений и идентификация выбросов
3. Расчет основных описательных статистик:
4. Исследование взаимосвязей между переменными с помощью корреляционного анали

Цель - освоение базовых методов описательной статистики и визуализации данных

Задание

По датасету "Airport"

- Построить гистограммы и ящики с усами для переменных: Scheduled departures, Performed departures, Enplaned passengers, Enplaned revenue tons of freight, Enplaned revenue tons of mail, сделать предположение о виде распределения (по каждой переменной), объяснить полученные значения в ящике с усами

Результат

Предположение по всем переменным: Логнормальные распределения

На boxplot медиана меньше среднего, узковатый график, медиана почти вплотную к Q1, что указывает на логнормальное распределение, сильная ассиметрия

Задание

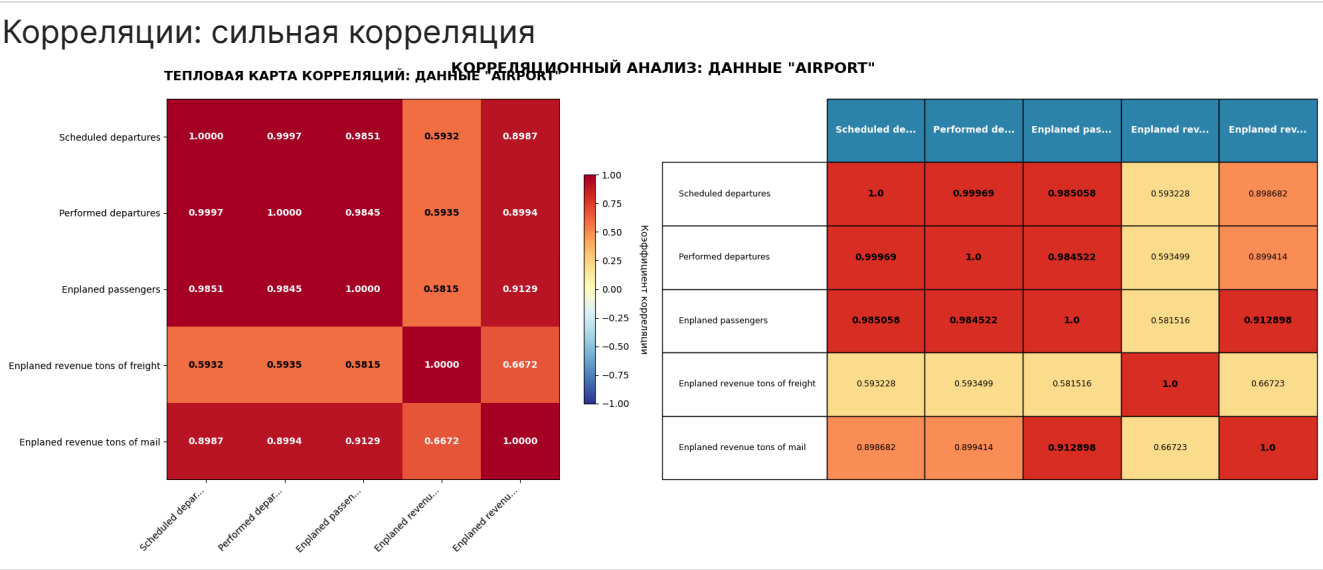
Рассчитать числовые характеристики этих переменных, описательную статистику

Результат

Введение

ОПИСАТЕЛЬНАЯ СТАТИСТИКА: ДАННЫЕ "AIRPORT"									
Variable	Mean	Median	Mode	Variance	Range	Std	IQR	Skewness	Kurtosis
Scheduled departures	45562.43	22864.5	1188.0	3202940000.0	321242.0	56594.53	54594.75	2.44	7.25
Performed departures	46304.96	23027.5	1253.0	3331110000.0	331085.0	57715.76	55001.75	2.46	7.37
Enplaned passengers	3136192.0	1210722.0	0.0	21202700000000.0	25636383.0	4604642.0	3441284.0	2.67	8.35
Enplaned revenue tons of freight	33858.16	6195.27	7.95	6575910000.0	614215.65	81091.95	19018.79	4.29	22.69
Enplaned revenue tons of mail	11461.87	2919.54	0.0	423492000.0	140359.38	20578.91	11091.14	3.36	14.18

* Все денежные показатели в соответствующих единицах измерения



Задание

По датасету "babyboom"
(Birth weight in grams, Number of minutes after midnight of each birth): построить гистограммы для каждой изучаемой переменной и ящички с усами. Сделать предположение о виде распределения , объяснить полученные значения в в ящичке с усами.

Результат

Предположение о Birth Weight: нормальное усеченное или нормальное сдвинутое распределение
На boxplot видно, умеренная левосторонняя асимметрия, медиана ближе к Q3 чем к Q1, небольшое количество выбросов слева
Предположение о minutes since midnight: Мультимодальное распределение - очень много пиков
boxplot в мультимодальном распределении говорит о симметрии модальностей - медиана близко к центру, примерно равные усы

Задание

Введение

Рассчитайте числовые характеристики этих переменных, описательную статистику

Результат

ОПИСАТЕЛЬНАЯ СТАТИСТИКА: ДАННЫЕ "BABYBOOM"

Variable	Mean	Median	Mode	Variance	Range	Std	IQR	Skewness	Kurtosis
Birth Weight (grams)	3275.95	3404.0	3428.0	278818.28	2417.0	528.03	430.25	-1.16	1.06
Minutes since midnight	788.73	846.5	5.0	173111.69	1430.0	416.07	675.75	-0.33	-1.08

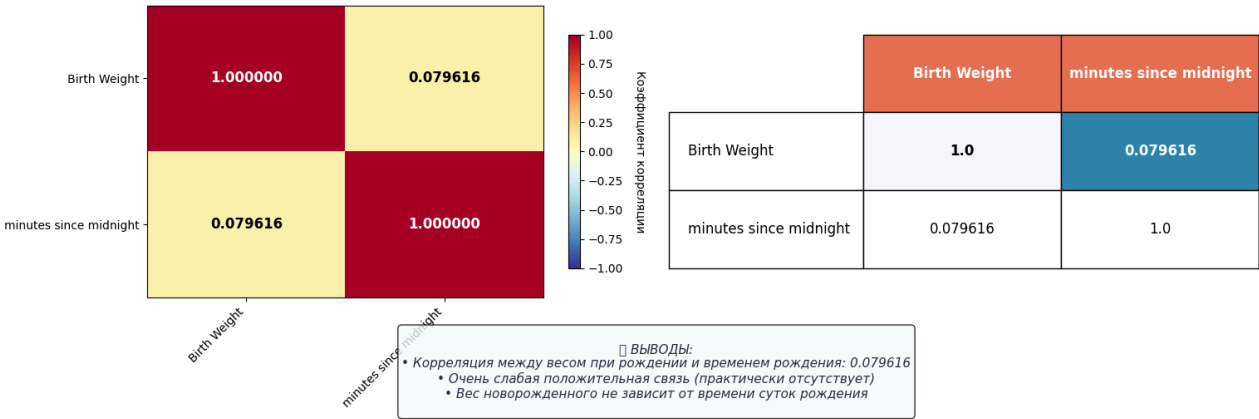
• Вес при рождении: отрицательная асимметрия (левостороннее распределение)

• Время рождения: близко к симметричному распределению

• Мода времени рождения: 00:05 (первые минуты после полуночи)

Корреляции:

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ: ДАННЫЕ "BABYBOOM"
ТЕПЛОВАЯ КАРТА КОРРЕЛЯЦИЙ: ДАННЫЕ "BABYBOOM"



Выводы: