

Семинарска работа по предметот Бизнис статистика

Стефан Везенкоски 233152

Извор на податочното множество: <https://www.kaggle.com/datasets/gchan357/human-glaucoma>

(го скартив податочното множество и во прилог ќе ви испратам excel документ со точната бројка на вредности со која работев)

Работам со податочното множество каде се прикажуваат податоци за пациенти кои имаат глауком.

Се состои од следите осум обележја:

- Glaucoma
- Age
- ocular_pressure
- MD
- PSD: promenlivost na standardot na defektot
- GHT
- cornea_thickness: debelina na roznica
- RNFL4

Јас работам и правам детални статистички анализи за следниве две обележја:

- PSD
- cornea_thickness

1. ПРВ ДЕЛ

Табела со распределба на честоти за PSD:

```
> print(freq_table_df_PSD)
```

Интервали	Средини_точки	фреквенција	Релативна_фреквенција	Релативна_фреквенција_проценти	Кумулативна_фреквенција	Кумулативна_фреквенција_проценти
1 [1.18, 2.02)	1.599474	57	0.163323782	16.3323782	57	16.33238
2 [2.02, 2.86)	2.438421	62	0.177650430	17.7650430	119	34.09742
3 [2.86, 3.7)	3.277368	39	0.111747851	11.1747851	158	45.27221
4 [3.7, 4.54)	4.116316	15	0.042979943	4.2979943	173	49.57020
5 [4.54, 5.37)	4.955263	14	0.040114613	4.0114613	187	53.58166
6 [5.37, 6.21)	5.794211	13	0.037249284	3.7249284	200	57.30659
7 [6.21, 7.05)	6.633158	12	0.034383954	3.4383954	212	60.74499
8 [7.05, 7.89)	7.472105	11	0.031518625	3.1518625	223	63.89685
9 [7.89, 8.73)	8.311053	10	0.028653295	2.8653295	233	66.76218
10 [8.73, 9.57)	9.150000	9	0.025787966	2.5787966	242	69.34097
11 [9.57, 10.4)	9.988947	16	0.045845272	4.5845272	258	73.92550
12 [10.4, 11.2)	10.827895	26	0.074498567	7.4498567	284	81.37536
13 [11.2, 12.1)	11.666842	16	0.045845272	4.5845272	300	85.95989
14 [12.1, 12.9)	12.505789	17	0.048710602	4.8710602	317	90.83095
15 [12.9, 13.8)	13.344737	14	0.040114613	4.0114613	331	94.84241
16 [13.8, 14.6)	14.183684	9	0.025787966	2.5787966	340	97.42120
17 [14.6, 15.4)	15.022632	5	0.014326648	1.4326648	345	98.85387
18 [15.4, 16.3)	15.861579	2	0.005730659	0.5730659	347	99.42693
19 [16.3, 17.1]	16.700526	2	0.005730659	0.5730659	349	100.00000

1. ТАБЕЛИ ЗА РАСПРЕДЕЛБА НА ЧЕСТОТИ ЗА PSD и cornea_thickness

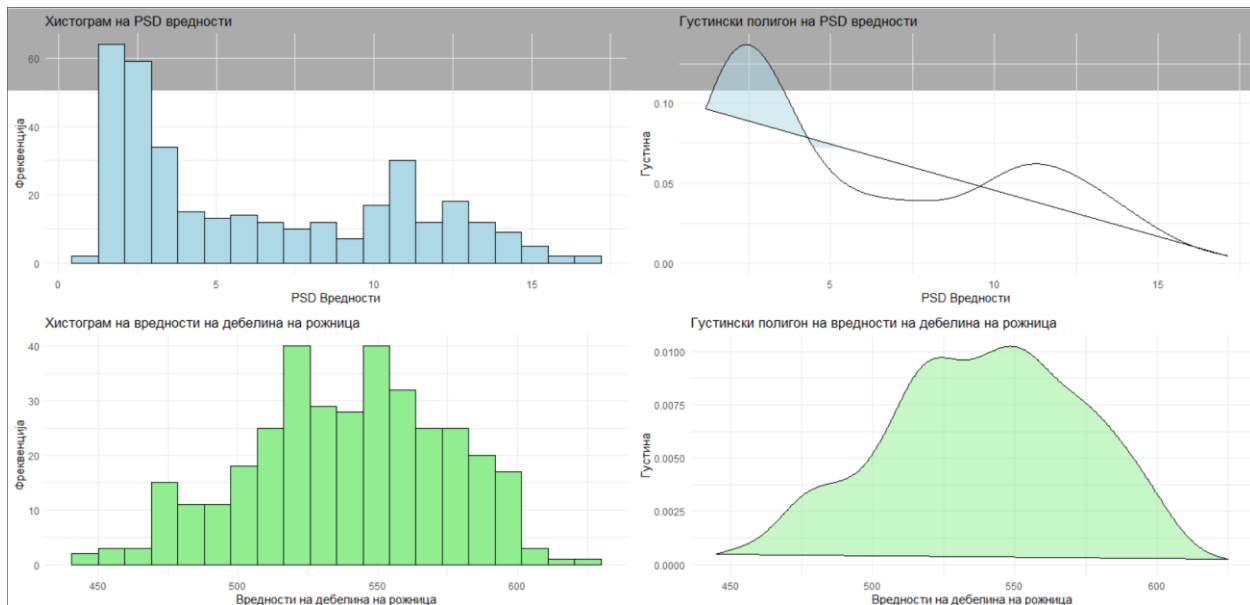
-
- Интервали каде што соодветните податоци се поделени на одреден број на интервали и соодветна ширина на интервалите.
- пресметани се средните точки на секој од интервалите(во децимали)
- фреквенција за секој од интервалите
- Релативна честота за секој од интервалите (и релативна честота во %)
- Кумулативна честота за секој од интервалите (и кумулативна честота во %)

Табела со распределба на честоти за cornea_thickness:

```
> print(freq_table_df_cornea)
```

	Интервали	Средни точки	Фреквенција	Релативна_фреквенција	Релативна_фреквенција_проценти	Кумулативна_фреквенција	Кумулативна_фреквенција_проценти
1	[445,454)	449.7368	4	0.011461318	1.1461318	4	1.146132
2	[454,464)	459.2105	1	0.002865330	0.2865330	5	1.432665
3	[464,473)	468.6842	5	0.014326648	1.4326648	10	2.865330
4	[473,483)	478.1579	17	0.048710602	4.8710602	27	7.736390
5	[483,492)	487.6316	14	0.040114613	4.0114613	41	11.747851
6	[492,502)	497.1053	9	0.025787966	2.5787966	50	14.326648
7	[502,511)	506.5789	23	0.065902579	6.5902579	73	20.916905
8	[511,521)	516.0526	38	0.108882521	10.8882521	111	31.805158
9	[521,530)	525.5263	31	0.088825215	8.8825215	142	40.687679
10	[530,540)	535.0000	27	0.077363897	7.7363897	169	48.424069
11	[540,549)	544.4737	38	0.108882521	10.8882521	207	59.312321
12	[549,559)	553.9474	37	0.106017192	10.6017192	244	69.914040
13	[559,568)	563.4211	26	0.074498567	7.4498567	270	77.363897
14	[568,578)	572.8947	29	0.083094556	8.3094556	299	85.673352
15	[578,587)	582.3684	19	0.054441261	5.4441261	318	91.117479
16	[587,597)	591.8421	15	0.042979943	4.2979943	333	95.415473
17	[597,606)	601.3158	13	0.037249284	3.7249284	346	99.140401
18	[606,616)	610.7895	2	0.005730659	0.5730659	348	99.713467
19	[616,625]	620.2632	1	0.002865330	0.2865330	349	100.000000

Исто така податоците се претставени и со соодветни хистограми и полигони.



```
Стебло-лист дијаграм за PSD:
> myStem(PSD_values, 0, 2)
1 | 18 25 28 29 33 36 38 39 40 41 41 42 43 44 46 46 47 48 48 53 55 57 59 60 62 62 63 63 66 67 68 68 69 70 70 72 77 77 78 79 81 81 81 82 82 83 83 84 85 86 88 92 94 96
2 | 0 0 1 2 3 3 4 4 5 6 7 8 10 14 14 14 14 14 16 16 17 18 18 18 19 23 23 24 27 28 28 29 29 29 29 31 31 32 33 33 36 40 42 44 48 51 52 52 55 56 56 57 58 60 61 63 64 67 71 75 80 80 81 83 88
3 | 0 2 2 5 6 6 6 8 11 13 16 21 22 23 23 23 28 28 31 35 36 41 48 50 56 60 67 69 76 78 88 96 97 97 98 99
12 | 0 0 1 2 6 6 11 28 32 34 35 41 51 54 55 56 63 64 70 75 76 79 83 94 96
4 | 3 4 12 17 27 34 35 60 65 71 74 77 80 83 85 95
11 | 1 6 8 12 13 14 17 18 18 20 23 25 25 26 27 29 46 58 64 66 74
10 | 2 6 8 9 19 24 28 32 33 37 40 40 40 48 51 56 57 60 62 67 73 73 86 90 93 93 96 96
14 | 29 35 51 51 51 55 69 71 90 96
8 | 6 8 21 27 33 37 38 55 70 73 77 79 82 91 94
5 | 5 11 15 19 30 46 50 53 54 78 86 90 93 99
13 | 2 2 11 20 29 31 41 63 66 67 68 73 88 97 99
7 | 6 28 40 48 52 59 69 69 78 78 83
6 | 1 14 15 19 28 50 57 65 65 77 85 85 91 92 96 97
16 | 17 90
9 | 11 12 35 55 91 94 99
15 | 7 59
17 | 12
> #
>
```

Стебло-лист дијаграм за PSD.

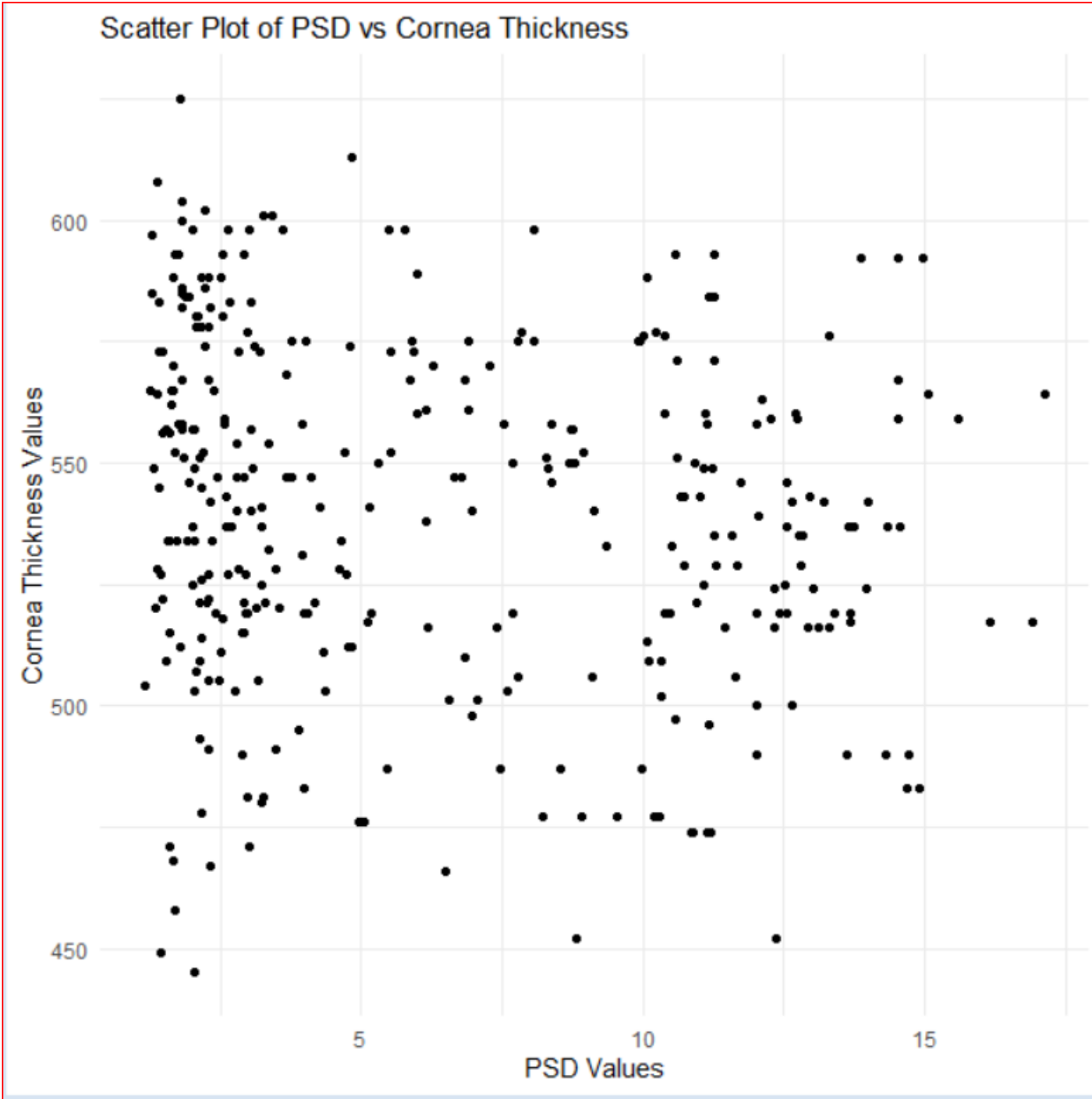


График на расејување за податоците од PSD и Cornea Thichness

Дискусија за врската помеѓу двете обележја: од графикот може да се забележи дека постои негативна линеарна корелација помеѓу двете обележја. Коефициентот на корелација r е -0.1501 , што укажува на слаба негативна корелација помеѓу двете обележја.

Мода, медијана и просек за cornea thickness:

Мода: **519 (се повторува 15 пати)**

Медијана: **541**

Просек: **539.002865329513**

Квартали, опсег и интерквартален распон за cornea thickness:

Квартили: **25% 50% 75%**

516 541 565

Опсег: **180**

Интерквартален распон(IQR): 49 (КВАРТИЛ 3(75%) – КВАРТИЛ 1(25%))

565-516 = 49

Дисперзија и стандардна девијација за cornea thickness:

Дисперзија: **1245.30746303066**

(пресметана со формулата која е користена на факултет)

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

Стандардна девијација: sqrt од 1245.30746303066 = **35.2889141662174**

Мода, медијана, просек, интерквартален распон и кватили за PSD:

Мода: **2.14 (се повторува 5 пати)**

Медијана: **4.65**

Просек: **6.29962750716332**

Опсег: **15.94**

Кватили: **25% 50% 75%**

2.29 4.65 10.57

Интерквартален распон: **8.28**

Дисперзија и стандардна девијација за PSD:

Дисперзија: **19.5440053206205**

Стандардна девијација: **4.42086024667377"**

Коефициент на корелација помеѓу PSD и cornea_thickness:

-0.1500741 (слаба негативна врска)

2. ВТОР ДЕЛ

1. Интервал на доверба за PSD:

95% интервал на доверба

$$\text{Алфа} = 1 - 0.95 = 0.05$$

Пресметка на интервалот: (min.interval, max.interval)

(5.835815, 6.763440) : интервалот на доверба, што значи дека со **95% сигурност**, просекот на PSD_values е помеѓу **5.835815** и **6.763440**

6.299628: Ова е просечната вредност на примерокот што е избран.

2. Хипотези за тестирањето параметар

$$E_X = 5$$

H₀: E_X = 5 (просекот е 5)

H_a: E_X ≠ 5 (просекот не е 5)

$$\text{Алфа} = 0.05$$

$$\text{Ниво на значајност} = 1 - \alpha/2 = 0,025$$

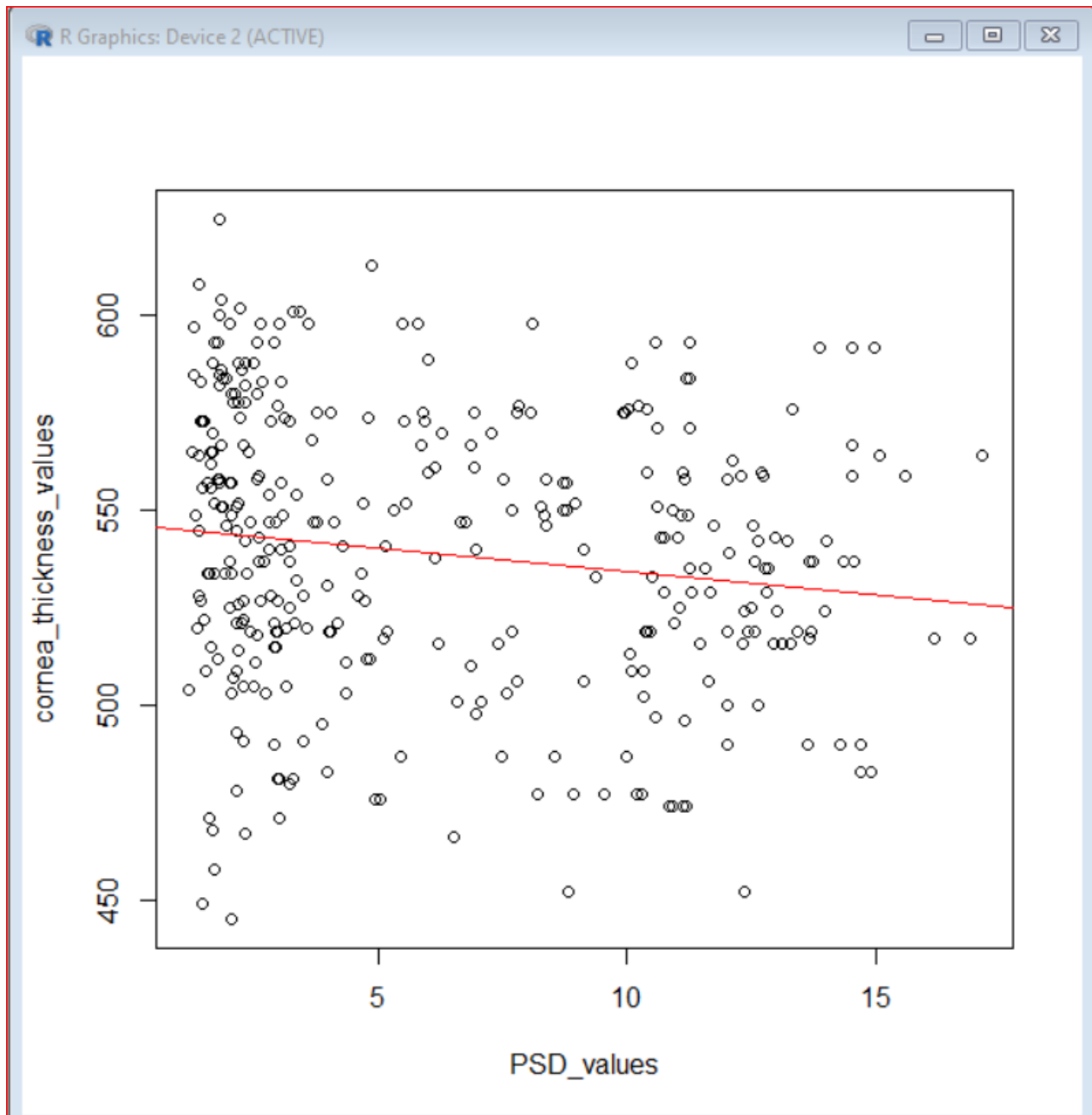
Критична вредност за Z тестот: **1.96**

(-1.96, 1.96) – критичен домен

Пресметана вредност за Z: **5.491928**

Z не припаѓа во критичниот домен со што H₀ отфрла и се прифаќа H_a.

Регресиона анализа



Права на регресија за двете обележја

мануелно да се внесе X и со моделот да се предвиди вредноста на Y

X = 15

$X(15) = 528.5803$: Оваа вредност (528.5803) е предвидената вредност за дебелината на роговица (Y) кога PSD е 15. Тоа е просечната предвидена вредност од моделот.

