

Семинарска работа по предметот Бизнис статистика

Важен индикатор за развојот на човештвото, и воопшто опстанокот на едно човечко битие, е среќата. Затоа како многу интригантна тема, предмет на анализа на оваа семинарска работа се податоци од светскиот извештај за среќа (World Happiness Report) во 157 земји, за 2016 година, кои можат да се најдат на следниот линк: <https://www.kaggle.com/unsdsn/world-happiness>.

За World Happiness Report

World Happiness Report или Светски извештај за среќа, е годишно издание на организацијата за одржлив развој на Обединетите Нации кое содржи разни рангирања и артикли поврзани со нивото на среќа ширум светот и индикациите кои тоа ги носи со себе.

Постоењето на овој извештај за нивото на среќа во земјите во светот, ја рефлектира зголемената интерес за тоа како подобро да се искористи среќата како важен фактор за квалитетот на човечкиот живот и развој, од разни здруженија и организации, и најважно од владите на овие земји.

Во табелата со податоци, ги среќаваме следните колони:

- Ниво на среќа – на кое испитаниците одговарале со број од 0 до 10 , каде 10 е најсреќно.
- Слобода
- Дарежливост
- Доверба (корупција на Владите)
- Здравје (очекувања за здрав живот, животен век)
- Дистопија (имагинарна земја која ја има одликата на земја каде што живеат најнесреќните луѓе, има најмали приходи, најмалку дарежливост, најмногу корупција, најмалку социјална поддршка.) Целта на дистопијата е споредба на другите земји, каде што ниедна не би имала полоши резултати од дистопијата.

Прв дел

(станува збор за колоната Happiness Score)

1. Да се изработи табела со распределба на честоти. Да се определат средните точки на интервалите, релативните фреквенции и кумулативните фреквенции. Потоа, добиените вредности графички да се претстават со хистограми и полигони.

Кога се работи со непрекинати променливи, има посебна постапка по која се делат податоците во групи или интервали.

Чекор 1: Наоѓање број на интервали. Оптимален број на интервали е помеѓу 5 и 15. Еден начин да се пресмета бројот на интервали е да се пресмета како корен од бројот на податоци. Бројот на единки е 157, чиј корен е **12.529**. **Заокружуваме на 12 интервали**.

Чекор 2: Наоѓање ширина на интервали.

(Најголема вредност – најмала вредност) / бројот на интервали.

$(7526 - 2905) / 12 = 0.385$, **заокружуваме на 0.4**.

Со помош на функции, ја добиваме следната табела, каде што:

Freq се честоти.

Rfreq се релативни честоти.

Cumfreq се кумулативни честоти.

Pfreq се проценти на релативните честоти.

P_Cumfreq се проценти на кумулативните честоти.

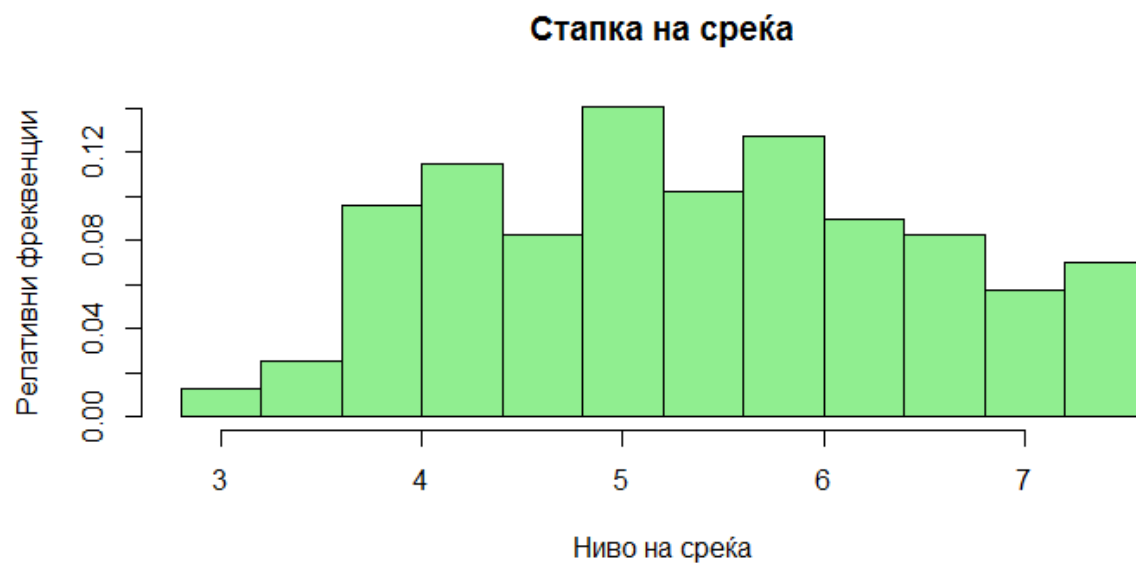
	freq	Rfreq	Cumfreq	Pfreq	P_Cumfreq
[2.8,3.2)	2	0.01273885	2	1.273885	1.273885
[3.2,3.6)	4	0.02547771	6	2.547771	3.821656
[3.6,4)	15	0.09554140	21	9.554140	13.375796
[4,4.4)	18	0.11464968	39	11.464968	24.840764
[4.4,4.8)	13	0.08280255	52	8.280255	33.121019
[4.8,5.2)	22	0.14012739	74	14.012739	47.133758
[5.2,5.6)	16	0.10191083	90	10.191083	57.324841
[5.6,6)	20	0.12738854	110	12.738854	70.063694
[6,6.4)	14	0.08917197	124	8.917197	78.980892
[6.4,6.8)	13	0.08280255	137	8.280255	87.261146
[6.8,7.2)	9	0.05732484	146	5.732484	92.993631
[7.2,7.6)	11	0.07006369	157	7.006369	100.000000

Слика 1: Табела на распределба на честоти за обележјето Happiness Score

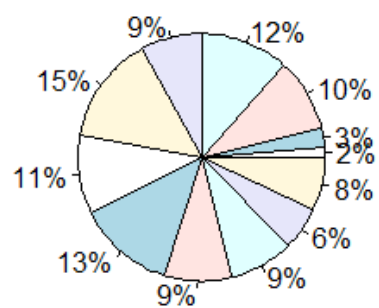
Слика 2: Хистограм на честоти за обележјето Happiness Score, изразен преку средни точки на интервали



Слика 2: Хистограм на релативни честоти за обележјето Happiness Score



Честоти во проценти за стапката на среќа



2. Конструирање стебло-лист дијаграм.

2 | 90

3 | 7 30 36 48 52 61 62 67 69 72 74 74 76 83 86 87 91 92 96 97

4 | 3 7 12 14 16 19 20 22 22 24 25 27 28 32 36 36 36 39 40 42 46 51 51 57 58 63 64 66 75 79 80 81
87 88 88 91 100

5 | 3 4 6 6 12 12 13 13 14 15 16 16 16 18 18 20 25 28 29 30 31 39 40 44 46 49 51 52 53 54 55 56 62
65 66 74 77 77 80 81 82 84 84 86 90 92 92 96 98 98 99 99

6 | 0 7 8 8 17 22 24 27 32 36 36 38 38 38 47 48 48 49 54 57 60 65 70 70 72 74 78 87 91 93 95 99

7 | 4 9 10 12 27 29 31 33 34 40 41 50 50 51 53.

Можеме да заклучиме дека најголем број од податоците започнуваат со цифрата 5.

3. Графика на расејување за податоци од обележјето Happiness Score и Health Life Expectancy



Според графикот на расејување можеме да забележиме дека постои силна линеарна врска помеѓу двете обележја.

4. Мода, медијана, просек

ПРОСЕК	5.382185
МЕДИЈАНА	5.314
МОДА	3.739

5. Квартали, опсег, интерквартален распон

Квартал Q1	Квартал Q2	Квартал Q3
4.404	5.314	6.269

Опсег : $7.526 - 2.905 = 4.621$

Интерквартален распон : 1,865 (разлика меѓу третиот и првиот квартал)

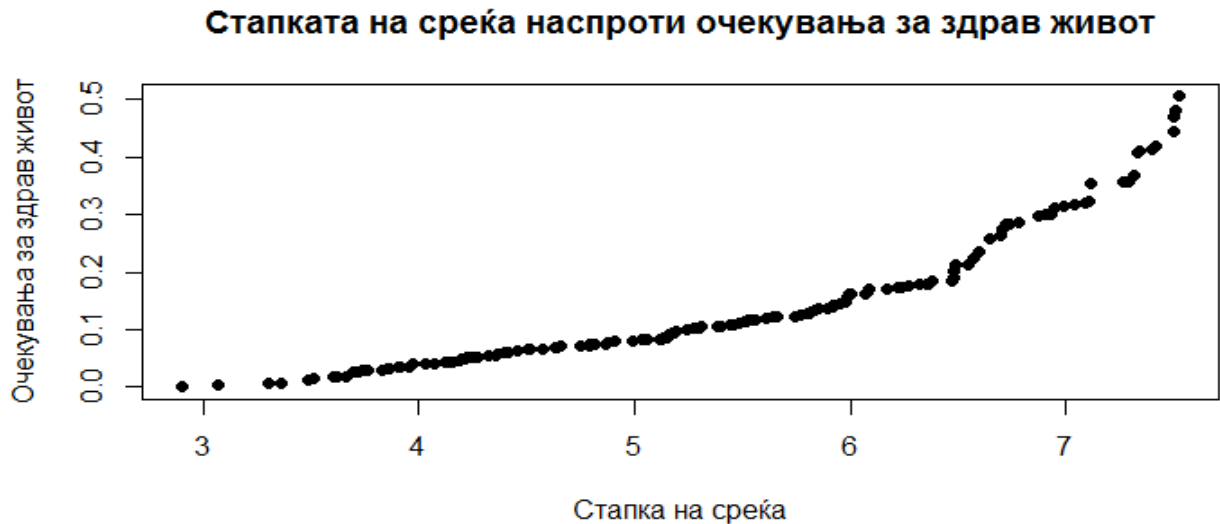
6. Дисперзија и стандардна девијација

Дисперзија	Стандардна девијација
1.303418	1.141674

Дисперзијата е пресметана како квадрат од девијацијата, со помош на готови функции.

7. Коефициент на корелација

Очекувања за здрав живот – просечен број на години што се очекува дека еден човек ќе ги живее во “полно здравје”.



Коефициентот на корелација е **0.9215576** и значи дека има многу силна линеарна позитивна поврзаност помеѓу нивото на среќа и очекувањата за здрав живот.

Втор дел

(станува збор за колоната Happiness Score)

1. Интервал на доверба

Се тестира 95% интервал на доверба за истото обележје (колona) Happiness Score, односно за просекот. Со помош на готови функции добиваме дека 95% од просеците за нивото на среќа, за земени примероци од овие земји би биле во интервал 5.178301 - 5.586068.

2. Тестирање хипотези за просекот

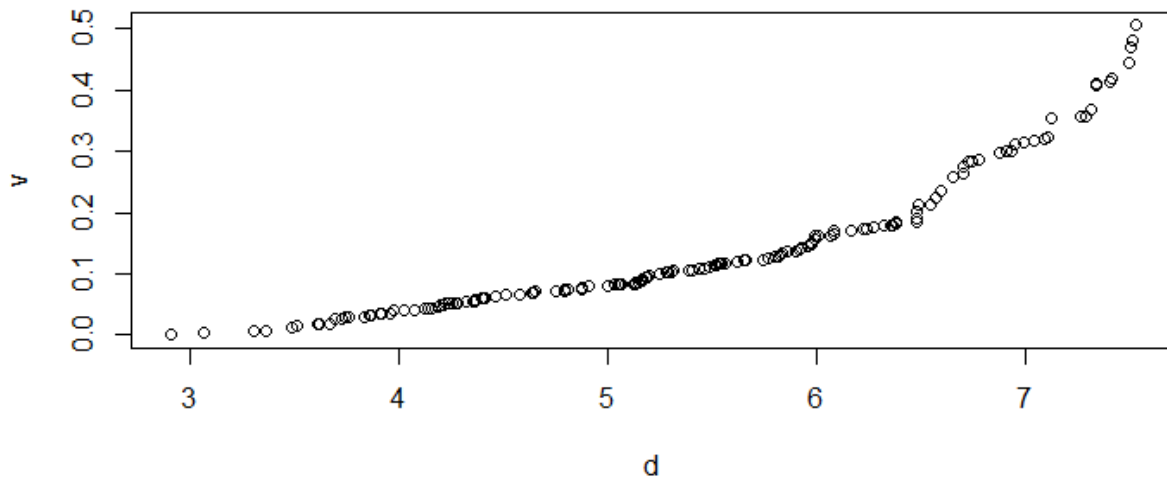
Велиме дека нултата хипотеза $H_0 - \mu \geq 4$,

алтернативната хипотеза $H_A - \mu < 4$.

Со помош на готови функции за тестирање на хипотези, добиваме резултат дека нултата хипотеза е точна, односно просекот е поголем или еднаков на 4, што се потврдува и преку пресметаниот интервал на доверба кој вели дека 95% од просеците на нивото на среќа ќе се наоѓаат помеѓу 5.178301 и 5.586068.

5. Регресиона анализа

(всушност коефициентот на корелација помеѓу нивото на среќа и очекувањата за здрав живот од првиот дел, барање 7, пресметани на поинаков начин)



Емилија Георгиевска 183044, група 6 ПИТ