«Исследование статистических связей»

Шарифуллин Ринас Рамилевич гр. 09-131

Вариант – 13

Задание 4_1

1. Говорят, что уровень успеваемости студентов по математической статистики зависит от их роста.
2. В выборку вошло n = 103 человек. В данных, представленных ниже, горизонтальный столбец таблицы показывает рост студента (в см.). Вертикальный – баллы учащегося за экзамен по математической статистике.
3. Рост студента есть случайная величина с функцией распределения F_1, а F_2 – функция распределения баллов.
4. Ожидается, что F_1 и F_2 зависимые случайные величины. Т.о., нулевая гипотеза $\mathbf{H}_0$: F_1 и F_2 – независимые сл. вел. при альтернативе $\mathbf{H}_1$: F_1 и F_2 – зависимые сл. вел.
5. Уровень значимости α = 0.03.
6. Применим критерий сопряженности хи-квадрат. Области признаков разобьём на r = 3 и s = 5 интервалы. Ожидания будут подтверждены, если X^2 примет маленькое значение, т.е. критическая область имеет вид $\{X^2 > C\}$.

$$\mathbb{X}^2 = n\sum_{k=1}^{r}\sum_{j=1}^{s}\frac{\left(\frac{n_{kj}}{n}-\frac{n_{*j}}{n}\frac{n_{k*}}{n}\right)^2}{\frac{n_{*j}}{n}\frac{n_{k*}}{n}} = \frac{1}{n}\sum_{k=1}^{r}\sum_{j=1}^{s}\frac{\left(n\,n_{kj}-n_{*j}\,n_{k*}\right)^2}{n_{*j}\,n_{k*}}.$$

7. При справедливости нулевой гипотезы функцию распределения статистики X^2 можно приблизить функцией хи-квадрат распределения $\mathbb{Khi}(x \mid (r - 1) * (s - 1)) = \mathbf{P}_{H0}\{X^2 \le x\}$ с $(r$ - 1) * (s - 1) = 8 степенями свободы.
8. Критическая константа C_α находится как решение неравенства

$$\mathbf{P}_{H0}\,\{X^2 \ge C_\alpha\} = 1 - \mathbb{Khi}(C_\alpha \mid 8) = 0.03,$$

т.е. C_α равна квантили порядка 0.97 хи-квадрат распределения с 8 ст. св. По таблице хи-квадрат распределения (процедурой Excel), находим C_α = 17,01.

a. Вид критерия: гипотеза однородности отвергается, если $\{X^2 \ge$ 17,01}.

9.

a. По представленным данным найдено

	172,45	182,83	>182,83	Σ
>59	0	3	7	10
59	8	13	9	30
54,05	6	8	8	22
49,1	10	5	5	20
44,15	11	9	1	21
Σ	35	38	30	$n = 103$
Статистика X^2			19,74	
степени свободы			8	
3%-я критическая область			$X^2 \geq 17{,}01$	
Гипотеза независимости			отвергается	
с критическим уровнем значимости			*p-val* < 0,0041	

b. Критический уровень значимости p-value вычисляется по формуле *p-val* = 1 − $\mathbb{Khi}$(19,74 | 8) = 0,0041. Есть все основания считать, что успеваемость студента по математической статистике зависит от его роста.

Задание 4_2

1. Говорят, что уровень успеваемости студентов по математической статистики зависит от их роста.
2. В выборку вошло n = 103 человек. В данных, представленных ниже, горизонтальный столбец таблицы показывает рост студента (в см.). Вертикальный – баллы учащегося за экзамен по математической статистике.
3. Рост студента есть случайная величина с функцией распределения F_1, а F_2 – функция распределения баллов.
4. Ожидается, что F_1 и F_2 зависимые случайные величины. Т.о., нулевая гипотеза $\mathbf{H}_0$: F_1 и F_2 – независимые сл. вел. при альтернативе $\mathbf{H}_1$: F_1 и F_2 – зависимые сл. вел.
5. Уровень значимости α = 0.01.
6. Применим критерий независимости компонент двумерного случайного вектора. Статистика Стьюдента вычисляется по следующей формуле:

$$t = \sqrt{n-2}\frac{R}{\sqrt{1-R^2}}.$$

   Критическая область будет принимать вид: $|t| \geq C$
7. Функция распределения тестовой статистики совпадает с функцией распределения Стьюдента $\mathbb{St}_{(n-2)}$ с n − 2 = 101 степенями свободы.
8. Критическая константа C_α находится из уравнения

$$2(1 - \mathbb{St}_{101}(C)) = 0.01,$$

   т.е. – равна верхней 0.005-квантили распределения Стьюдента с 101 степенями свободы.
    a) Воспользовавшись таблицей (пакетом Excel), нашли, что C_α = 2,87.
    b) Окончательный вид критической области |t| ≥ 2,87
9.
    a. По представленным данным найдено

	x	y
Среднее,	177,12	50,56
Дисперсия, s^2	65,99	40,91
Станд. отклонение, s	8,12	6,39
Объем выборки, n	103	103
Коэффициент корреляции, R		0,39
Преобразование Стьюдента, t		4,28
1%-я критическая область		$\|t\| > 2,87$

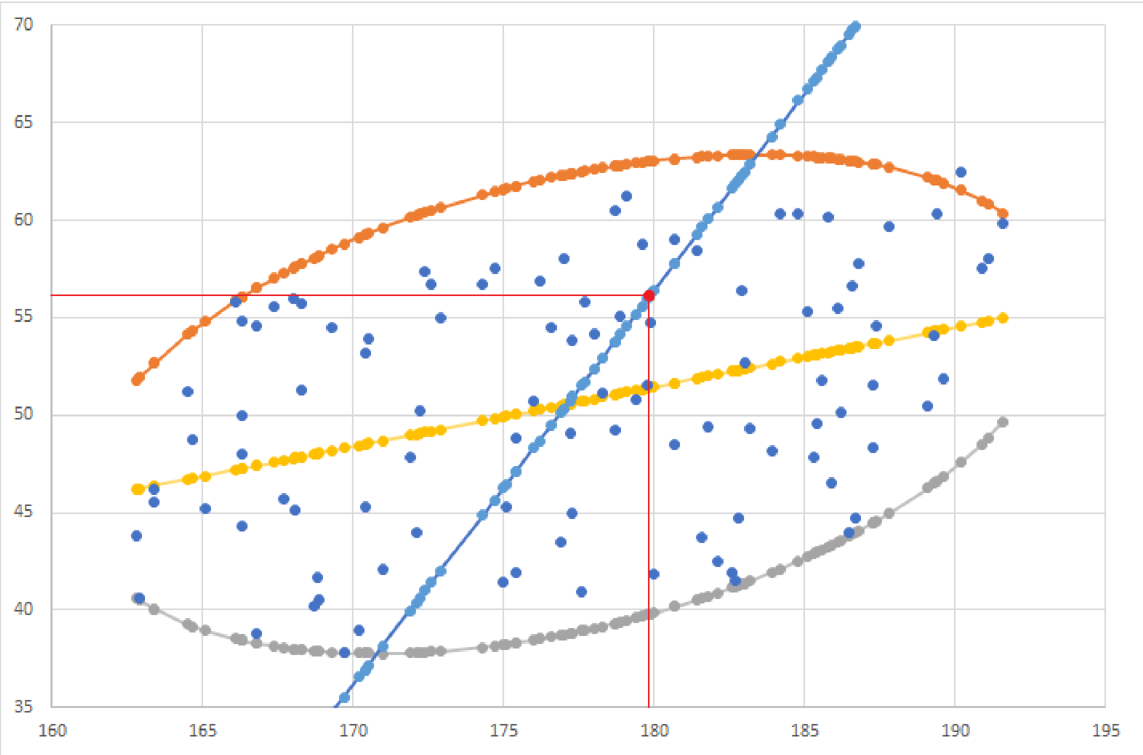
Гипотеза независимости	отвергается
с критическим уровнем значимости	p-val < 0,000043

b. Критический уровень значимости p-value вычисляется по формуле *p-val* = $2(1 - \mathbb{St}_{101}(4{,}28)) = 0{,}000043$. Есть все основания считать, что успеваемость студента по математической статистике зависит от его роста.

Задание 4_3

1. По результатам независимых измерений значения баллов за экзамен по математической статистике и роста n = 103 студентов найти оценки коэффициентов линейной среднеквадратической регрессии роста учащегося (ξ) на полученную оценку (η); представить график линии регрессии в поле всех данных; найти прогноз значения ξ при значении η = 56; дать оценку точности прогноза, изобразить эллипс рассеяния.
2. Для проверки предположения в эксперименте приняло участие n = 103 студента, у которых были измерены рост и уровень успеваемости по математической статистике.
3. По представленным данным найдено

Коэффициент линейной регрессии	$\beta = 0{,}49$
Уравнение регрессии η на ξ	$y = 0{,}49x + 152{,}16$
Прогноз при y = 56	$x = 179{,}8$
Коэфф.корреляции	$R = 0{,}39$
Стандарт.отклонение наблюдений	$S_x = 8{,}12$
Оценка стандарт. ошибки прогноза	7,49



Вывод. При таком невысоком значении коэффициента корреляции (R = 0,39; стандартная ошибка прогноза равна 7,49) прогностические качества линии регрессии очень низкие.