Министерство науки и высшего образования РФ

Федеральное государственное бюджетное образовательное учреждение высшего образования

«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ» (ТУСУР)

Кафедра автоматизированной обработки информации (АОИ)

Отчет по контрольной работе №2 по дисциплине
«Теория вероятностей и математическая статистика»
Вариант №17

Студент ФДО ТУСУР
Гр. з-422П8-5
____ Курьянович В.А.
«___»___________202_ г.

Руководитель:
Старший преподаватель
каф. АОИ
____Ю. В. Синчинова
«___»___________202_ г.

Томск 2023

Задача 1. Тема: «Нормальное распределение»

Предположим, что в течение года цена на акции некоторой компании есть случайная величина, распределенная по нормальному закону с математическим ожиданием, равным 48 у.е., и стандартным отклонением, равным 6. Определите вероятность того, что в случайно выбранный день обсуждаемого периода цена отклонится не больше чем на 7 у.е. за акцию.

Решение:

Для решения задачи необходимо использовать формулу вероятности нормального распределения:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-¿¿¿}$$

где μ - математическое ожидание, σ - стандартное отклонение.
В нашем случае μ=48, σ=6.

Так как мы ищем вероятность того, что цена отклонится не больше чем на 7 у.е., то нам нужно найти вероятность того, что цена будет находиться в интервале

$$[48-7;48+7]=[41;55]$$

Для этого необходимо вычислить две вероятности:

$$P(X\leq 55)$$

$$P(X\leq 41)$$

Затем найдем разность этих вероятностей:

$$P(41\leq X\leq 55)=P(X\leq 55)-P(X\leq 41)$$

$$P(X\leq 55)=\int_{-\infty}^{55}\square f(x)\,dx=\int_{-\infty}^{55}\square\frac{1}{6\sqrt{2\pi}}e^{-¿¿¿}$$

$$P(X\leq 41)=\int_{-\infty}^{41}\square f(x)\,dx=\int_{-\infty}^{41}\square\frac{1}{6\sqrt{2\pi}}e^{-¿¿¿}$$

Тогда искомая вероятность равна:

$$P(41\leq X\leq 55)=P(X\leq 55)-P(X\leq 41)=0.9332-0.0668=0.8664$$

Таким образом, вероятность того, что в случайно выбранный день обсуждаемого периода цена отклонится не больше чем на 7 у.е. за акцию равна 0.8664.

Задача 2. Тема: «Интервальные оценки».

Задание:

При выборочном опросе 1200 телезрителей оказалось, что 456 из них регулярно смотрят программы телевидения НТВ. Постройте 99%-ый доверительный интервал, оценивающий долю всех телезрителей, предпочитающих программы телеканала НТВ.

Решение:

Для построения доверительного интервала для доли всех телезрителей, предпочитающих программы телеканала НТВ, можно воспользоваться методом Уилсона.

Метод Уилсона:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\frac{\alpha}{2}}^2}{4n^2}}$$

Где $\hat{p}$ выборочная доля, n - объем выборки, $z_{\alpha/2}$ - квантиль стандартного нормального распределения уровня $\alpha/2$.

Подставим значения из условия задачи:

$$\hat{p} = \frac{456}{1200} = 0.38$$

$$n = 1200$$

$$\alpha = 0.01$$

$$0.38 \pm 2.58 \sqrt{\frac{0.38(1-0.38)}{1200}} = [0.34; 0.42]$$

Таким образом, с вероятностью 99% можно утверждать, что доля всех телезрителей, предпочитающих программы телеканала НТВ находится в интервале от 34% до 42%.

Задача 3. Тема: «Проверка статистических гипотез»

Задание:

Крупный коммерческий банк заказал маркетинговое исследование по выявлению эффекта «премирования» (калькулятор, набор ручек и др.) как стимула для открытия счета в банке. Для проверки случайным образом было отобрано 200 «премированных» посетителей и 200 «непремированных». В результате выяснилось, что 79% посетителей, которым не предлагалась премия, и 89% посетителей, которым премия предлагалась, открыли счет в банке в течение 6 мес. Используя эти данные, проверьте гипотезу о том, что доля «премированных» посетителей, открывших счет в банке, существенно отличается от удельного веса «непремированных», открывших счет. Уровень значимости $\alpha = 0.05$.

Решение:

Для проверки гипотезы о том, что доля «премированных» посетителей, открывших счет в банке, существенно отличается от удельного веса «непремированных», открывших счет, используется критерий хи-квадрат (χ^2).

Пусть p1 - доля «непремированных» посетителей, открывших счет в банке, а p2 - доля «премированных» посетителей, открывших счет в банке. Тогда гипотезы будут следующими:

H0: $p1 = p2$
H1: $p1 \neq p2$

Уровень значимости $\alpha = 0.05$.

Рассчитаем значения χ^2 и найдем критическое значение $\chi^2\alpha(1)$ для уровня значимости $\alpha = 0.05$ и степеней свободы df = 1 (так как имеем две выборки).

$\chi^2 = ((0.79 * 200 - 0.89 * 200)^2 / (0.79 * 200 + 0.89 * 200)) + ((0.21 * 200 - 0.11 * 200)^2 / (0.21 * 200 + 0.11 * 200)) = 10.24$

Критическое значение $\chi^2\alpha(1) = 3.84$.

Так как $\chi^2 > \chi^2\alpha(1)$, то гипотеза H0 отвергается на уровне значимости $\alpha = 0.05$ в пользу альтернативной гипотезы H1.

Таким образом, можно сделать вывод о том, что доля «премированных» посетителей, открывших счет в банке, существенно отличается от удельного веса «непремированных», открывших счет.

Задача 4. Тема: «Критерий согласия Пирсона»

Задание: С помощью критерия согласия Пирсона на уровне значимости $\alpha = 0{,}05$ выяснить, можно ли считать случайную величину X, заданную в виде сгруппированного статистического ряда, нормально распределенной с параметрами

$\vec{x}$ и s, рассчитанными по выборке.

$(x_j; x_{j+1})$	[1.2; 1.6)	[1.6; 2.0)	[2.0; 2.4)	[2.4; 2.8)	[2.8; 3.2)	[3.2; 3.6)
n_j	3	5	10	8	4	2

Рассчитаем значения $\overline{x}\, u\, \overline{s}$:

n = 32 (общее число наблюдений)

Выборочное среднее - x_i равно

$(1.4 * 3+ 1.8 * 5 + 2.2 * 10 + 2.6 * 8 + 3 * 4 + 3.4 * 2)/32 = 2.3375$

Дисперсия равна

$((1.4 - 2.3375)^2 * 3 + (1.8 - 2.3375)^2 *5 + (2.2 - 2.3375)^2 * 10 + (2.6 - 2.3375)^2 * 8 + (3 - 2.3375)^2 * 4 + (3.4 - 2.3375)^2 * 2) /32 \approx 0.276$

Стандартное отклонение равно $\sqrt{0.276} = 0.525$

Рассчитаем наблюдаемое значение $K_{набл}$ статистики Пирсона по формуле:

$$K=\sum_{j=1}^{k}\frac{\left(n_j-np_j\right)^2}{np_j}$$

Эмпирические частоты n_j уже известны, а для вычисления вероятностей p_j, в предположении, что гипотеза справедлива, применим формулу

$$p_j=P\left(x_j<X<x_{j+1}\right)=\Phi\left(\frac{x_{j+1}-\bar{x}}{\bar{s}}\right)-\Phi\left(\frac{x_j-\bar{x}}{\bar{s}}\right),\ j=1,2\ ...\ k$$

$$p_1=\Phi\left(\frac{1.6-2.3375}{0.525}\right)-\Phi\left(\frac{1.2-2.3375}{0.525}\right)=\Phi(-1.4047)-\Phi(-2.1666)=0.0654$$
$$p_2=\Phi\left(\frac{2.0-2.3375}{0.525}\right)-\Phi\left(\frac{1.6-2.3375}{0.525}\right)=\Phi(-0.6428)-\Phi(-1.4047)=0.1803$$
$$p_3=\Phi\left(\frac{2.4-2.3375}{0.525}\right)-\Phi\left(\frac{2.0-2.3375}{0.525}\right)=\Phi(0.1190)-\Phi(-0.6428)=0.2827$$
$$p_4=\Phi\left(\frac{2.8-2.3375}{0.525}\right)-\Phi\left(\frac{2.4-2.3375}{0.525}\right)=\Phi(0.8809)-\Phi(0.1190)=0.2628$$
$$p_5=\Phi\left(\frac{3.2-2.3375}{0.525}\right)-\Phi\left(\frac{2.8-2.3375}{0.525}\right)=\Phi(1.6428)-\Phi(0.8809)=0.1389$$
$$p_6=\Phi\left(\frac{3.6-2.3375}{0.525}\right)-\Phi\left(\frac{3.2-2.3375}{0.525}\right)=\Phi(2.4047)-\Phi(1.6428)=0{,}0423$$

$(x_j;x_{j=1})$	Выборочное среднее x_i	Наблюдаемая частота n_j	Вероятность p_j попадания в j-й интервал	Ожидаемая частота np_j	Слагаемые статистики Пирсона $\frac{\left(n_j-np_j\right)^2}{np_j}$
[1.2; 1.6)	1.4	3	0.0654	2,0928	0.39325
[1.6; 2.0)	1.8	5	0.1803	5,7696	0.10265
[2.0; 2.4)	2.2	10	0.2827	9,0464	0.10052
[2.4; 2.8)	2.6	8	0.2628	8,4096	0.01995
[2.8; 3.2)	3	4	0.1389	4,4448	0.04451
[3.2; 3.6)	3.4	2	0.0423	1,3536	0.30868
$\sum$ ❑		32	1,2404	31,1168	$К_{набл}$=0.96956

Наблюдаемое значение статистики Пирсона равно $К_{набл}$=0.96956

Определим границу критической области. Так как статистика Пирсона измеряет разницу между эмпирическим и теоретическим распределениями, то, чем больше ее наблюдаемое значение $К_{набл}$, тем сильнее довод против основной гипотезы. Поэтому критическая область для этой статистики всегда правосторонняя: [$К_{кр}$; +∞). Ее границу $К_{кр}$= $x^2_{(k-r-1;\alpha)}$ находим по таблицам распределения «хи-квадрат» и заданным значениям α=0,05, число интервалов k=6, r=2 (так как параметры $\bar{x}\ u\ \bar{s}$ оценены по выборке): $К_{кр}$= $x^2(6-2-1;0{,}05)=7{,}8$

Наблюдаемое значение статистики Пирсона попадает в критическую область: $К_{набл}$< $К_{кр}$, поэтому есть основания не отвергать основную гипотезу.

Вывод: на уровне значимости 0,05 справедливо предположение о том, что данная случайная величина X имеет нормальное распределение.

Задача 5. Тема: «Ранговая корреляция»

Задание: По заданной таблице рангов найти выборочный коэффициент ранговой корреляции Спирмена и проверить значимость полученного результата при α = 0.05.

Десять предприятий проранжированы по двум признакам: X — коэффициент механизации работ, Y — производительность труда.

Ранг 1	1	2	3	4	5	6	7	8	9	10
Ранг 2	2	1	5	4	3	6	8	7	10	9

Решение:

Для нахождения выборочного коэффициента ранговой корреляции Спирмена (rS) по данной таблице рангов воспользуемся следующей формулой:

$$r_s = 1 - \frac{6S}{n\left(n^2-1\right)},$$

где d - разность рангов по соответствующим признакам для каждого предприятия, n - общее число предприятий.

Построим ранговую таблицу.

i	1	2	3	4	5	6	7	8	9	10
r_i	1	2	3	4	5	6	7	8	9	10
s_i	2	1	5	4	3	6	8	7	10	9
$r_i - s_i$	-1	1	-2	0	2	0	-1	1	-1	1

Величина S равна

$$S = \sum_{i=1}^{10} \left(r_i - s_i\right)^2 = (-1)^2 + (1)^2 + (-2)^2 + (0)^2 + (2)^2 + (0)^2 + (-1)^2 + (1)^2 + ¿$$
$$+(-1)^2 + (1)^2 = 14$$

Коэффициент корреляции Спирмена соответственно равен

$$r_s = 1 - \frac{6 \cdot 14}{10\left(10^2-1\right)} = 0.92$$

Значимость коэффициента корреляции проверяется с помощью статистики, имеющей распределение Стьюдента с (n-2) степенями свободы:

$$K = r_s \sqrt{\frac{n-2}{1-r_s^2}}$$

$$K_{набл} = 0.92\sqrt{\frac{10-2}{1-(0.92)^2}} = 6.63$$

$T_{крит}$ - критическая точка двусторонней критической области, которую находим по таблице критических точек распределения Стьюдента по уровню значимости и числу степеней свободы k=n-2=8

$T_{крит}(0,05; 8) = 2,31$

Так как эмпирическое значение коэффициента корреляции Спирмена меньше критического:

$|T_{набл}| = 6,63 > T_{крит} = 2,31$

Значение коэффициента корреляции Спирмена близко к 1, поэтому между признаками X и У есть сильная корреляционная связь. Этот результат согласуется с результатом применения коэффициента корреляции Пирсона.

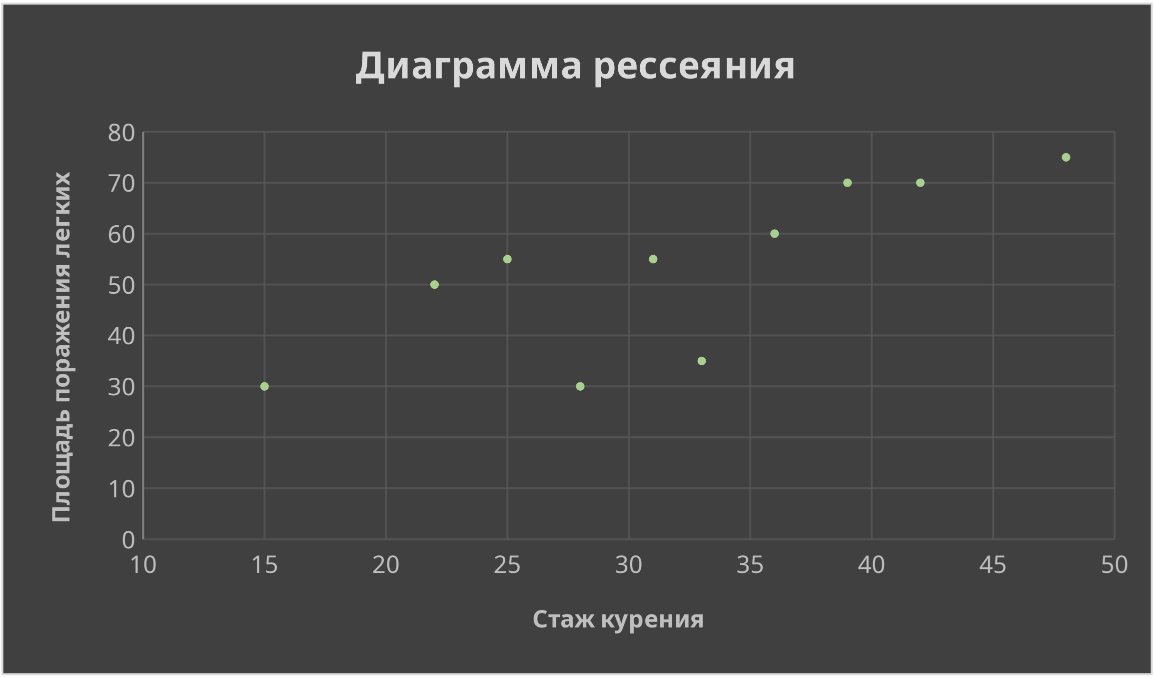
Задача 6. Тема: «Линейная корреляция и регрессия»

Задание: Для приведенных исходных данных постройте диаграмму рассеяния и определите по ней характер зависимости. Рассчитайте выборочный коэффициент корреляции Пирсона, проверьте его значимость при $\alpha = 0.05$. Запишите уравнение регрессии и дайте интерпретацию полученных результатов.

Врач-исследователь выясняет зависимость площади пораженной части легких (Y, %) у людей, заболевших эмфиземой легких, от числа лет курения (X, лет). Исследуйте зависимость по выборочным данным.

X	25	36	22	15	48	39	42	31	28	33
Y	55	60	50	30	75	70	70	55	30	35

Решение:
Для начала построим диаграмму рассеяния:



По виду диаграммы есть основания предполагать линейную зависимость площади пораженной части легких от стажа курения . Теперь рассчитаем выборочный коэффициент корреляции Пирсона.

Для удобства проведем все необходимые предварительные расчеты в таблице.

x	y	x^2	y^2	xy
25	55	625	3025	1375
36	60	1296	3600	2160
22	50	484	2500	1100
15	30	225	900	450
48	75	2304	5625	3600
39	70	1521	4900	2730
42	70	1764	4900	2940
31	55	961	3025	1705
28	30	784	900	840
33	35	1089	1225	1155
319	**530**	**11053**	**30600**	**18055**

Рассчитаем числовые характеристики выборки, используя итоговую строку расчетной таблицы и учитывая, что объем выборки n=7.

Выборочные средние:

$$\bar{x}=\frac{1}{n}\sum x_i=\frac{1}{10}\cdot 319=31,9$$

$$\bar{y}=\frac{1}{n}\sum y_i=\frac{1}{10}\cdot 530=53$$

$$\overline{xy}=\frac{1}{n}\sum x_i y_i=\frac{1}{10}\cdot 18055=1805,5$$

Выборочные дисперсии:

$$\sigma_x^2=\overline{x^2}-\bar{x}^2=\frac{1}{10}\cdot 11053-31,9^2=87.69$$

$$\sigma_y^2=\overline{y^2}-\bar{y}^2=\frac{1}{10}\cdot 30600-53^2=251$$

Выборочные среднеквадратические отклонения:

$$\sigma_x=\sqrt{87.69}=9.36$$

$$\sigma_y=\sqrt{251}=15.84$$

Вычислим выборочный коэффициент корреляции по формуле:

$$r_{xy}=\frac{\overline{xy}-\bar{x}\cdot\bar{y}}{\sigma_x\sigma_y}=\frac{1805.5-31.9\cdot 53}{9.36\cdot 15.84}=0.774$$

Рассчитаем теперь оценки коэффициентов регрессии:

$$\hat{b}_1=0.774\cdot\frac{\sqrt{87.69}}{\sqrt{251}}=0.457$$

$$\hat{b}_0=53-0.457\cdot 31.9=38.421$$

Таким образом, уравнение регрессии Y на X имеет вид:

$$y=0.457x+38.421$$

Наблюдаемое значение критерия t находим по формуле:

$$t_{набл}=r_s\sqrt{\frac{n-2}{1-r_s^2}}=4.605$$

По таблице находим критическое значение $t_{0,05;8}=2.31$

Так как $t_{набл}$>$t_{крит}$, то полученное значение коэффициента корреляции признается значимым. Коэффициент корреляции признается статистически значимым или существенно отличным от нуля в генеральной совокупности. Таким образом, стаж курения оказывает статистически существенное влияние на площадь поражения легких.