Условие задачи.

Фирма предлагает автоматы по розливу напитков. При выборке n-16 найдена средняя величина x =182 г дозы, наливаемой в стакан автоматом №1. При выборке m-9 найдена средняя величина y =185 дозы, наливаемой в стакан автоматом №2. По утверждению изготовителя, случайная величина наливаемой дозы имеет нормальный закон распределения с дисперсией, равной $\sigma_x^2 = \sigma_y^2 = 25$. Можно ли считать отличия выборочных средних случайной ошибкой при уровне значимости $\alpha = 0.01$?

Поставлены задачи:

1) Проверить гипотезу: $H_0:\ \mathrm{a}_1 = a_2$

2) Проверить гипотезу: $H_1:\ \mathrm{a}_1 \succ a_2$

3) Проверить гипотезу: $H_1:\ \mathrm{a}_1 \prec a_2$.

Решение

Пусть a_x и a_y – математические ожидания доз, наливаемых автоматом №1 и автоматом №2. Нулевая гипотеза в данном случае $H_0: a_x = a_y$ при альтернативных $H_1: a_x \neq a_y$, $H_1: a_x \succ a_y$ и $H_1: a_x < a_y$. Дисперсия известна: σ^2=25. В качестве критерия справедливости статистической гипотезы выбирается функция

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}},$$

распределенная по нормальному закону с параметрами (0, 1).

1) Рассмотрим гипотезу $H_0: a_x = a_y$ при альтернативной $H_1: a_x \neq a_y$. В этом случае критическая область двусторонняя и имеет вид $(-\infty; z_{кр}^{л}) \cup (z_{кр}^{п}; \infty)$. Величины $z_{кр}^{л}$ и $z_{кр}^{п}$ рассчитываются из условий

$$P(Z < z_{кр}^{л}) = \alpha/2 \text{ и } P(Z > z_{кр}^{п}) = \alpha/2.$$

Воспользовавшись таблицей функции Лапласа, имеем

$$\Phi(z_{кр}^{п}) = 0{,}5 - \alpha/2 = 0{,}495,$$

$$z_{кр}^{п} = 2{,}57.$$

Критическая область имеет вид $(-\infty; -2{,}57) \cup (2{,}57; \infty)$. Значение z_r= –1,44 не попадает в критическую область, поэтому нулевая гипотеза принимается.

2) Рассмотрим гипотезу $H_0: a_x = a_y$ для альтернативной $H_1: a_x < a_y$. В этом случае критическая область имеет вид $(-\infty; z_{кр}^{л})$, где $z_{кр}^{л}$ определяется из условия $P(Z < z_{кр}^{л}) = \alpha$.

Так как функция Лапласа – нечетная функция, т.е. $\Phi(-z) = -\Phi(z)$, а таблица этой функции содержит только положительные значения, то найдем вначале $z_{кр}^{п}$.

Для этого вычислим значение функции Лапласа в критической точке: $\Phi(z_{кр}^{п}) = 0{,}5 - \alpha = 0{,}49$. Откуда $z_{кр}^{п} = 2{,}33$. Значит, левосторонняя критическая область будет $(-\infty; -2{,}33)$.

Рассчитаем z_r:

$$z_r = \frac{182 - 185}{\sqrt{\frac{25}{16} + \frac{25}{9}}} = \frac{-3 \cdot 12}{25} = -1{,}44.$$

Полученное значение z_r= –1,44 не входит в критическую область $(-\infty; -2{,}33)$, поэтому нулевая гипотеза принимается.

3) Рассмотрим гипотезу $H_0 : a_x = a_y$ для альтернативной $H_1 : a_x \succ a_y$. В этом случае критическая область имеет вид $(z_{кр}^{л};\infty)$, где $z_{кр}^{л}$ определяется из условия $P(Z < z_{кр}^{л}) = \alpha$.

Для этого вычислим значение функции Лапласа в критической точке: $\Phi(z_{кр}^{п}) = 0{,}5 - \alpha = 0{,}49$. Откуда $z_{кр}^{п} = 2{,}33$. Значит, правосторонняя критическая область будет $(2{,}33;\infty)$.

Рассчитаем z_r:

$$z_r = \left| \frac{182 - 185}{\sqrt{\frac{25}{16} + \frac{25}{9}}} \right| = \frac{3 \cdot 12}{25} = 1{,}44.$$

Полученное значение z_r= 1,44 не входит в критическую область $(2{,}33;\infty)$, поэтому нулевая гипотеза принимается.

Ответ: результаты приведены выше.

Условие задачи.

Масса (в граммах) 30 пачек полуфабриката «Геркулес» такова:

503,509,495,493,489,485,507,511,487,485,506,504,507,511,499,491,494,518,506,515, 487,509,507,488,495,490,498,497,492,495.

Поставлены задачи:

4) Построить статистический ряд распределения относительных частот;

5) Построить гистограмму выборочной оценки плотности вероятности;

6) Найти выборочную среднюю и выборочную дисперсию по распределению выборки

7) Найти несмещенную оценку математического ожидания дисперсии

8) Найти доверительный интервал с надежностью 0,95 для оценки математического ожидания нормально распределенной случайной величины;

9-10) Проверить гипотезу о нормальном законе распределения случайной величины на уровне значимости $\alpha = 0.05$.

Решение

4) Число групп возьмем равным 5. Ширина интервала составит: $h = \frac{x_{\max} - x_{\min}}{n} = \frac{518 - 485}{5} = 6.6$.

$x_{\max}$ - максимальное значение группировочного признака в совокупности.

$x_{\min}$ - минимальное значение группировочного признака.

Определим границы группы.

Номер группы	Нижняя граница	Верхняя граница
1	485	491.6
2	491.6	498.2
3	498.2	504.8
4	504.8	511.4
5	511.4	518

Для каждого значения ряда подсчитаем, какое количество раз оно попадает в тот или иной интервал. Для этого сортируем ряд по возрастанию.

485	485 – 491.6	1
485	485 – 491.6	2
487	485 – 491.6	3
487	485 – 491.6	4
488	485 – 491.6	5
489	485 – 491.6	6
490	485 – 491.6	7
491	485 – 491.6	8
492	491.6 – 498.2	1
493	491.6 – 498.2	2
494	491.6 – 498.2	3
495	491.6 – 498.2	4
495	491.6 – 498.2	5
495	491.6 – 498.2	6
497	491.6 – 498.2	7
498	491.6 – 498.2	8
499	498.2 – 504.8	1
503	498.2 – 504.8	2
504	498.2 – 504.8	3
506	504.8 – 511.4	1
506	504.8 – 511.4	2
507	504.8 – 511.4	3
507	504.8 – 511.4	4
507	504.8 – 511.4	5
509	504.8 – 511.4	6
509	504.8 – 511.4	7
511	504.8 – 511.4	8

511	504.8 – 511.4	9
515	511.4 – 518	1
518	511.4 – 518	2

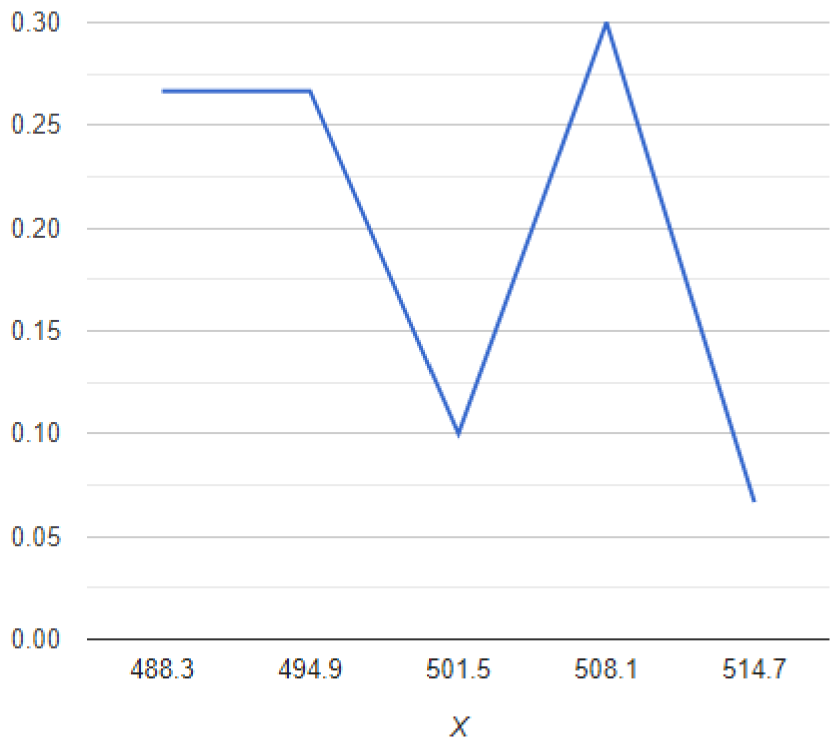
Результаты группировки оформим в виде таблицы:

Группы	№ совокупности	Частота fi
485 – 491.6	6,10,9,21,24,5,26,16	8
491.6 – 498.2	29,4,17,3,25,30,28,27	8
498.2 – 504.8	15,1,12	3
504.8 – 511.4	11,19,7,13,23,2,22,8,14	9
511.4 – 518	20,18	2

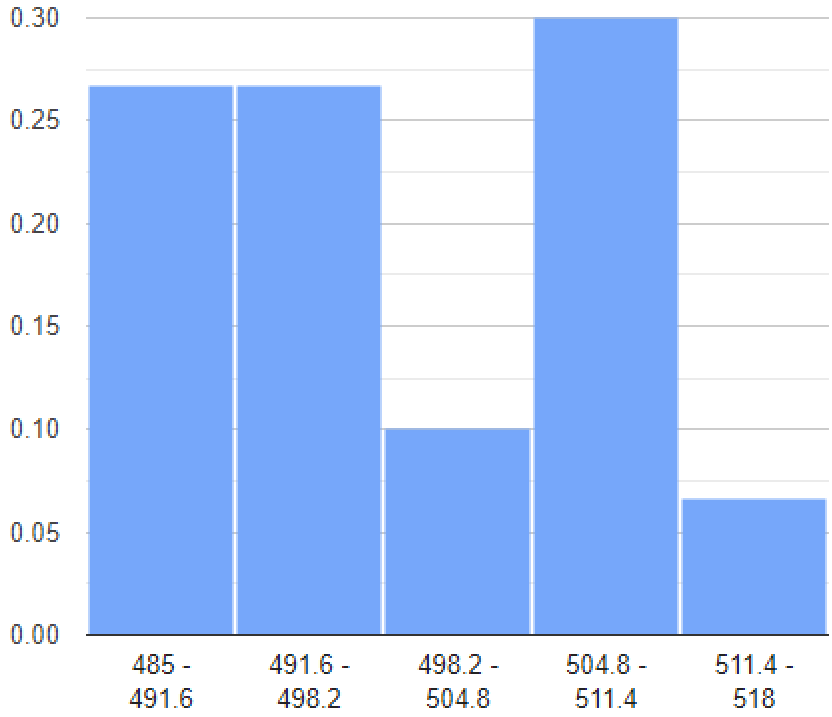
Составляем статистическую таблицу относительных частот:

X	485 – 491.6	491.6 – 498.2	498.2 – 504.8	504.8 – 511.4	511.4 – 518
$p_i^* = \frac{m_i}{n}$	$\frac{8}{30}$	$\frac{8}{30}$	$\frac{3}{30}$	$\frac{9}{30}$	$\frac{2}{30}$

Построим полигон относительных частот:



5) Строим гистограмму выборочной оценки плотности вероятности:



6) Таблица для расчета показателей.

Группы	Середина интервала, $x_{центр}$	Кол-во, f_i	$x_i \cdot f_i$	Накопленная частота, S	$\|x-x_{ср}\|\cdot f_i$	$(x-x_{ср})^2 \cdot f_i$	Относительная частота, f_i/f
485 - 491.6	488.3	8	3906.4	8	86.24	929.667	0.267
491.6 - 498.2	494.9	8	3959.2	16	33.44	139.779	0.267
498.2 - 504.8	501.5	3	1504.5	19	7.26	17.569	0.1
504.8 - 511.4	508.1	9	4572.9	28	81.18	732.244	0.3
511.4 - 518	514.7	2	1029.4	30	31.24	487.969	0.0667
Итого		30	14972.4		239.36	2307.228	1

Найдем выборочную среднюю: $\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i} = \frac{14972.4}{30} \approx 499.08$.

Найдем выборочную дисперсию: $D = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i} = \frac{2307.228}{30} \approx 76.908$.

Среднеквадратическое отклонение: $\sigma = \sqrt{D} = \sqrt{76.908} \approx 8.77$.

7) Несмещенная оценка дисперсии: $S^2 = \dfrac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i - 1} = \dfrac{2307.228}{29} \approx 79.56$.

Оценка среднеквадратического отклонения: $s = \sqrt{S^2} = \sqrt{79.56} \approx 8.92$.

8) Доверительный интервал для математического ожидания: $\left(\bar{x} - \dfrac{t_{kp} s}{\sqrt{n}};\ \bar{x} + \dfrac{t_{kp} s}{\sqrt{n}}\right)$.

В нашем случае $2\Phi(t_{kp}) = \gamma \Rightarrow \Phi(t_{kp}) = \gamma/2 = 0.95/2 = 0.475$.

По таблице функции Лапласа найдем, при каком t_{kp} значение $\Phi(t_{kp}) = 0.475$.

$t_{kp}(\gamma) = (0.475) = 1.96$, тогда $\varepsilon = t_{kp} \dfrac{s}{\sqrt{n}} = 1.96 \cdot \dfrac{8.92}{\sqrt{30}} \approx 3.192$.

Следовательно, доверительный интервал для математического ожидания
(499.08 - 3.192; 499.08 + 3.192) = (495.888; 502.272).

С вероятностью 0.95 можно утверждать, что среднее значение при выборке большего объема не выйдет за пределы найденного интервала.

9) Проверим гипотезу о том, что данные выборки распределены по нормальному закону с помощью критерия согласия Пирсона: $K = \sum \dfrac{(f_i - f \cdot p_i)^2}{f \cdot p_i}$, где p_i — вероятность попадания в i-й интервал случайной величины, распределенной по гипотетическому закону.

Для вычисления вероятностей p_i применим формулу и таблицу функции Лапласа $\Phi\left(\dfrac{x_{i+1} - \bar{x}}{\sigma}\right) - \Phi\left(\dfrac{x_i - \bar{x}}{\sigma}\right)$, где $\sigma = 8.92$, $x_{cp} = 499.08$.

Теоретическая (ожидаемая) частота равна $f_i = f \cdot p_i$, где $f = 30$.

Интервалы группировки	Наблюдаемая частота f_i	$x_1 = \dfrac{x_i - x_{cp}}{\sigma}$	$x_1 = \dfrac{x_{i+1} - x_{cp}}{\sigma}$	$\Phi(x_1)$	$\Phi(x_2)$	Вероятность попадания в i-й интервал, $p_i = \Phi(x_2) - \Phi(x_1)$	Ожидаемая частота, $30 p_i$	Слагаемые статистики Пирсона, K_i
485 - 491.6	8	-1.5785	-0.8386	-0.4429	-0.2995	0.1434	4.302	3.1788
491.6 - 498.2	8	-0.8386	-0.09866	-0.2995	-0.0398	0.2597	7.791	0.00561
498.2 - 504.8	3	-0.09866	0.6413	-0.0398	0.2422	0.282	8.46	3.5238
504.8 - 511.4	9	0.6413	1.3812	0.2422	0.4177	0.1755	5.265	2.6496
511.4 - 518	2	1.3812	2.1212	0.4177	0.4838	0.0661	1.983	0.000146
	30							9.358

Границу критической области $K_{kp} = \chi^2(k - r - 1; \alpha)$ находим по таблицам распределения χ^2 и заданным значениям s, k (число интервалов), r=2 (параметры x_{cp} и s оценены по выборке).

Для $\alpha = 0.05$: $K_{kp} = \chi^2(5 - 2 - 1; 0.05) = 5.99146$, $K_{набл} = 9.358$

Наблюдаемое значение статистики Пирсона попадает в критическую область: $K_{набл} = 9.358 \succ K_{kp} = 5.99146$, поэтому есть основания отвергать основную гипотезу. Данные выборки распределены не по нормальному закону.

Ответ: результаты расчетов приведены в соответствующих пунктах.

Условие задачи

С целью анализа взаимного влияния зарплаты и текучести рабочей силы на пяти однотипных фирмах с одинаковым числом работников проведены измерения уровня месячной зарплаты *X* и числа уволившихся за год рабочих *Y*.

Поставлены задачи:

11) Построить поле корреляции

12) Найти выборочное уравнение линейной регрессии *Y* на *X*

13) Найти выборочное уравнение линейной регрессии *X* на *Y*

14) Найти выборочный коэффициент корреляции

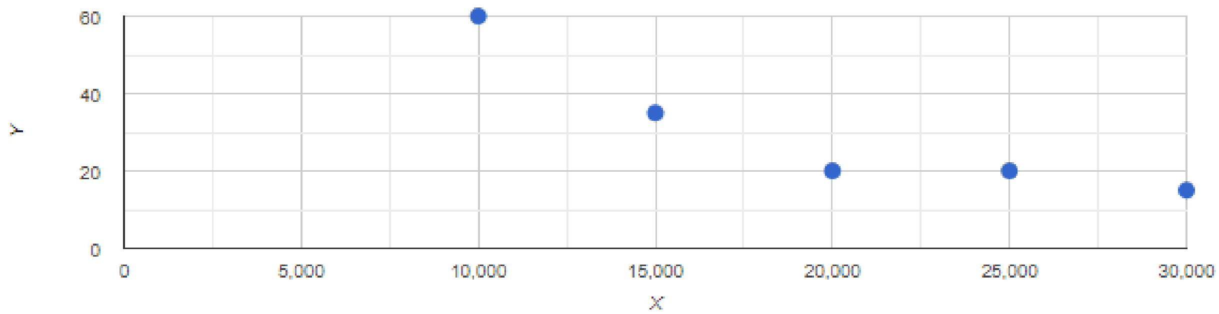
15) Построить линии регрессии в поле корреляции.

Решение

Возьмем следующие данные:

X	10000	15000	20000	25000	30000
Y	60	35	20	20	15

11) Построим поле корреляции.



12) Составим расчетную таблицу:

x	y	x^2	y^2	$x \cdot y$
10000	60	100000000	3600	600000
15000	35	225000000	1225	525000
20000	20	400000000	400	400000
25000	20	625000000	400	500000
30000	15	900000000	225	450000
100000	150	2250000000	5850	2475000

Выборочные средние.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{100000}{5} = 20000, \qquad \bar{y} = \frac{\sum y_i}{n} = \frac{150}{5} = 30$$ и

$$\overline{xy} = \frac{\sum x_i y_i}{n} = \frac{2475000}{10} = 495000.$$

Выборочные дисперсии:

$$S^2(x) = \frac{\sum x_i^2}{n} - \bar{x}^2 = \frac{2250000000}{5} - 20000^2 = 50000000.$$

$$S^2(y) = \frac{\sum y_i^2}{n} - \bar{y}^2 = \frac{5850}{5} - 30^2 = 270.$$

Среднеквадратическое отклонение:

$$S(x) = \sqrt{S^2(x)} = \sqrt{50000000} \approx 7071.068.$$

$$S(y) = \sqrt{S^2(y)} = \sqrt{270} \approx 16.432.$$

Ковариация: $\operatorname{cov}(x, y) = \overline{x \cdot y} - \bar{x} \cdot \bar{y} = 495000 - 20000 \cdot 30 = -105000$.

Рассчитываем показатель тесноты связи. Таким показателем является выборочный линейный коэффициент корреляции, который рассчитывается по формуле:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{S(x)S(y)} = \frac{495000 - 20000 \cdot 30}{7071.068 \cdot 16.432} \approx -0.904.$$

В нашем примере связь между признаком Y фактором X высокая и обратная.

Уравнение регрессии Y на X.

$$y_x = r_{xy} \frac{x - \bar{x}}{S(x)} S(y) + \bar{y} = -0.904 \cdot \frac{x - 20000}{7071.068} \cdot 16.432 + 30 = -0.021 \cdot x + 72.$$

Линейное уравнение регрессии имеет вид $y_x = -0.021 \cdot x + 72$.

13) Уравнение регрессии X на Y.

$$x_y = r_{xy} \frac{y - \bar{y}}{S(y)} S(x) + \bar{x} = -0.904 \cdot \frac{y - 30}{16.432} \cdot 7071.068 + 20000 = -389.012 \cdot y + 31670.36.$$

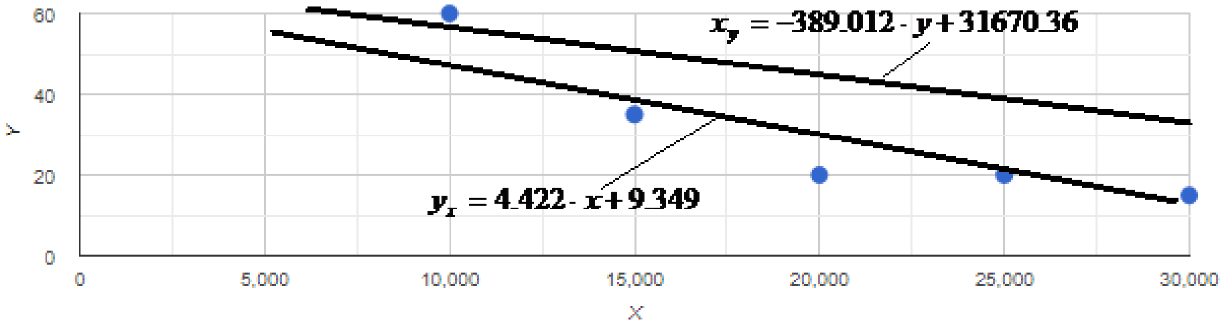
Линейное уравнение регрессии имеет вид $x_y = -389.012 \cdot y + 31670.36$.

14) Рассчитываем показатель тесноты связи. Таким показателем является выборочный линейный коэффициент корреляции, который рассчитывается по формуле:

$$r_{xy} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{S(x)S(y)} = \frac{495000 - 20000 \cdot 30}{7071.068 \cdot 16.432} \approx -0.904.$$

В нашем примере связь между признаком Y фактором X высокая и обратная.

15) Построим линии регрессии в поле корреляции.



Ответ: $y_x = 4.422 \cdot x + 9.349$, $x_y = -389.012 \cdot y + 31670.36$.