Задача.
1. Произвести точечную и интервальную оценку истинного объема продаж в предположении его нормального распределения.
2. Проверить гипотезы о том, что истинный объем продаж: а) отсутствует, б) равен 5 условным денежным единицам.
3. Используя в качестве зависимой переменной данные последующего вариант определить выборочную ковариацию и выборочный коэффициент корреляции.
4. Проверить гипотезу об отсутствии связи между наборами данных.
5. Проверить гипотезу о том, что обе выборки произведены из одной совокупности.

Вариант 8.
Объем продаж, усл. ден. ед: 0, 2, 2, 4, 5, 5, 7, 7, 8, 10.

Решение. Составим простой вариационный ряд, подсчитав, сколько раз встречается каждая варианта. Получим:

x_i	n_i
0	1
2	2
4	1
5	2
7	2
8	1
10	1
Сумма	**10**

Найдем выборочное среднее:
$$\bar{x} = \frac{1}{n}\sum x_i n_i = \frac{1}{10}50 = 5.$$
Найдем дисперсию (исправленную, так как объем выборки очень маленький):
$$D_x = \frac{1}{n-1}\sum (x_i - \bar{x})^2 n_i = \frac{1}{9}86 \approx 9,556.$$
Найдем среднее квадратическое отклонение: $\sigma_x = \sqrt{D_x} \approx 3,091$.
Расчеты приведены в таблице ниже:

x_i	n_i	$x_i n_i$	$(x_i - \bar{x})^2 n_i$
0	1	0	25
2	2	4	18
4	1	4	1
5	2	10	0
7	2	14	8
8	1	8	9
10	1	10	25
Сумма	**10**	**50**	**86**

Тогда точечная оценка истинного объема продаж равна выборочному среднему:
$\bar{x} = 5$ усл. ден. ед.

Интервальную оценку (с доверительной вероятностью $\gamma = 95\% = 0,95$) найдем из формулы:

$$\overline{x} - t\frac{\sigma_x}{\sqrt{n}} < a < \overline{x} + t\frac{\sigma_x}{\sqrt{n}},$$

где t определяется по доверительной вероятности из условия $\Phi(t) = \gamma/2 = 0,475$, откуда $t = 1,96$.

Подставляем:

$$5 - 1,96\frac{3,091}{\sqrt{10}} < a < 5 + 1,96\frac{3,091}{\sqrt{10}},$$

$$3,084 < a < 6,916.$$

Таким образом, с вероятностью 95% истинный объем продаж попадает в интервал от 3,084 до 6,916 усл. ден. ед.

Проверим гипотезу $H_0 = \{a = 0\}$ о том, что истинный объем продаж отсутствует. Вычислим наблюдаемое значение критерия

$$T_{набл.} = \frac{\overline{x} - a_0}{\sigma_x}\sqrt{n} = \frac{5}{3,091}\sqrt{10} \approx 5,115.$$

По таблице критических точек распределения Стьюдента при уровне значимости $\alpha = 0,05$ и числе степеней свободы $k = n - 1 = 9$ найдем критическую точку $t_{кр.двуст.}(0,05;9) = 2,26$.

Так как $T_{набл.} = 5,115 > 2,26 = t_{кр.двуст.}$, то нулевую гипотезу следует отвергнуть, объем продаж присутствует (больше нуля).

Проверим гипотезу $H_0 = \{a = 5\}$ о том, что истинный объем продаж равен 5 условным денежным единицам. Вычислим наблюдаемое значение критерия

$$T_{набл.} = \frac{\overline{x} - a_0}{\sigma_x}\sqrt{n} = \frac{5-5}{3,091}\sqrt{10} = 0.$$

По таблице критических точек распределения Стьюдента при уровне значимости $\alpha = 0,05$ и числе степеней свободы $k = n - 1 = 9$ найдем критическую точку $t_{кр.двуст.}(0,05;9) = 2,26$.

Так как $T_{набл.} = 0 < 2,26 = t_{кр.двуст.}$, то нулевую гипотезу следует принять.

Используя в качестве зависимой переменной данные последующего варианта
(№9) 1, 0, 2, 4, 4, 6, 7, 8, 9, 10.
определим выборочную ковариацию и выборочный коэффициент корреляции.

Обозначим эту выборку за Y, а выборку своего варианта за X. Ковариация этих величин может быть найдена по формуле:

$$\text{cov}(X,Y) = M\left[(X - M(X))(Y - M(Y))\right] = M(XY) - M(X)M(Y).$$

Коэффициент корреляции найдем по формуле: $\rho(X,Y) = \dfrac{\text{cov}(X,Y)}{\sigma_x \cdot \sigma_y}$.

Сначала вычислим необходимые величины для подстановки в формулы:

Найдем выборочное среднее:

$$\overline{y}=\frac{1}{n}\sum y_i n_i=\frac{1}{10}51=5{,}1.$$

Найдем дисперсию (исправленную, так как объем выборки очень маленький):

$$D_x=\frac{1}{n-1}\sum (y_i-\overline{y})^2 n_i=\frac{1}{9}106{,}9\approx 11{,}878.$$

Найдем среднее квадратическое отклонение: $\sigma_y=\sqrt{D_y}\approx 3{,}446$.

Расчеты приведены в таблице ниже:

y_i	n_i	$y_i n_i$	$(y_i-\overline{y})^2 n_i$
0	1	0	26,01
1	1	1	16,81
2	1	2	9,61
4	2	8	2,42
6	1	6	0,81
7	1	7	3,61
8	1	8	8,41
9	1	9	15,21
10	1	10	24,01
Сумма	**10**	**51**	**106,9**

Осталось найти $M(XY)=\frac{1}{n}\sum x_i y_i=\frac{1}{10}347=34{,}7$:

x_i	y_i	$x_i y_i$
0	1	0
2	0	0
2	2	4
4	4	16
5	4	20
5	6	30
7	7	49
7	8	56
8	9	72
10	10	100

Сумма **347**

Теперь подставляем все найденные величины и находим:

Ковариацию: $\operatorname{cov}(X,Y)=34{,}7-5\cdot 5{,}1=9{,}2$.

Коэффициент корреляции: $\rho(X,Y)=\dfrac{9{,}2}{3{,}091\cdot 3{,}446}\approx 0{,}864$.

Проверим гипотезу $H_0=\{\rho=0\}$ об отсутствии связи между наборами данных. Вычислим наблюдаемое значение критерия:

$$T_{набл.} = \frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}} = \frac{0,864\sqrt{8}}{\sqrt{1-0,864^2}} \approx 4,854.$$

По таблице критических точек распределения Стьюдента при уровне значимости $\alpha = 0,05$ и числе степеней свободы $k = n-2 = 8$ найдем критическую точку $t_{кр.двуст.}(0,05;8) = 2,31$.

Так как $T_{набл.} = 4,854 > 2,31 = t_{кр.двуст.}$, то нулевую гипотезу следует отвергнуть, связь между наборами данных есть.

Проверить гипотезу о том, что обе выборки произведены из одной совокупности. Так как предполагается, что выборки сделаны из генеральной совокупности, распределенной по нормальному закону, достаточно проверить сначала гипотезу о равенстве дисперсий, а затем, если она окажется верной, о равенстве математических ожиданий.

Исправленные дисперсии различны, поэтому проверим гипотезу о равенстве дисперсий, используя критерий Фишера-Снедекора. Найдем отношение большей дисперсии к меньшей:

$$F_{набл} = \frac{11,878}{9,556} \approx 1,243.$$

В качестве конкурирующей гипотезы примем гипотезу $H_1 : D(X) < D(Y)$. По таблице при уровне значимости $\alpha = 0,05$ и числам степеней свободы $k_1 = m-1 = 9$ и $k_2 = n-1 = 9$ найдем критическую точку $F_{кр}(0,05,9,9) = 3,18$. Так как $F_{набл} = 1,243 < 3,18 = F_{кр}$, нет оснований отвергнуть нулевую гипотезу о равенстве генеральных дисперсий.

Итак, предположение о равенстве генеральных дисперсий выполняется, поэтому сравним средние.

Введем нулевую гипотезу: $H_0 : M(X) = M(Y)$. Вычислим наблюдаемое значение критерия по формуле:

$$T_{набл.} = \frac{\overline{x} - \overline{y}}{\sqrt{(n-1)\sigma_x^2 + (m-1)\sigma_y^2}} \cdot \sqrt{\frac{n \cdot m \cdot (n+m-2)}{n+m}}.$$

Получаем:

$$T_{набл.} = \frac{5-5,1}{\sqrt{9 \cdot 9,556 + 9 \cdot 11,878}} \cdot \sqrt{\frac{10 \cdot 10 \cdot 18}{20}} \approx -0,068.$$

Находим критическую точку (двусторонняя область) из таблицы Стьюдента при уровне значимости $\alpha = 0,05$ и числу степеней свободы $k = m+n-2 = 18$ $t_{дв.кр} = 2,1$. Так как модуль наблюдаемого значения критерия 0,068 меньше критического, нет оснований отвергнуть нулевую гипотезу.

Таким образом, можно считать, что выборки произведены из одной совокупности (распределенной по нормальному закону).