Отчёт
По индивидуальному заданию
Вариант 23, 24
Выполнил: Радько В. А.

Первичная обработка статистических данных

Была дана выборка значений случайных величин «Число вызовов, поступивших на АТС за некоторый промежуток времени», обозначим её ξ, и «Продолжительность разговора (мин.)» обозначим её η.

Рассмотрим первую выборку $\{x_i\}$, её объём $n = 100$. Сначала определим размах выборки, для этого находим $\min(x_i) = 1$ и $\max(x_i) = 21$. Разобьём интервал $[1, 21]$ на $k_1 = 10$ частей и построим вариационный ряд.

$[a_i, a_{i+1})$	$[1,3)$	$[3,5)$	$[5,7)$	$[7,9)$	$[9,11)$	$[11,13)$	$[13,15)$	$[15,17)$	$[17,18)$	$[19,21]$
$\frac{m_i}{n}$	0.06	0.23	0.35	0.2	0.1	0.04	0.01	0.0	0.0	0.1

Таблица 1: Вариационный ряд $\{x_i\}$

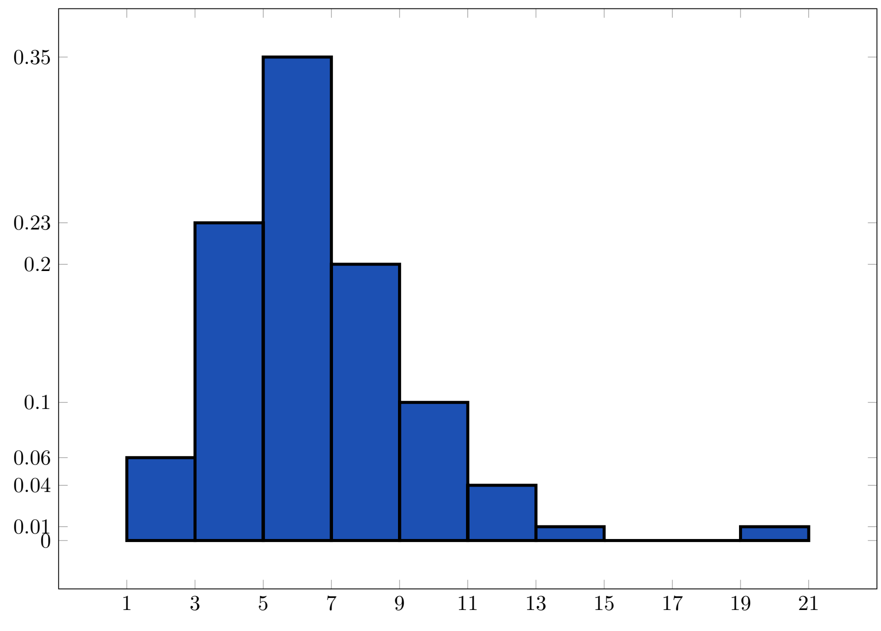


Рис. 1: Гистограмма $\{x_i\}$

Оценим $M\xi$ средним арифметическим:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \approx 6.08 \tag{1}$$

Дисперсию $D\xi$ будем оценивать формулой:

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\bar{x}^2 \approx 8.2562 \tag{2}$$

Оценка среднеквадратического отклонения:

$$s_x = \sqrt{s_x^2} \approx 2.8734 \tag{3}$$

Интервальные оценки числовых характеристик

Так как случайная величина

$$\frac{s^2}{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}\frac{\left(\xi_i - \bar{\xi}\right)^2}{\sigma^2} \in \frac{1}{n-1}\chi^2_{n-1}, \tag{4}$$

тогда получим, что

$$\frac{\bar{\xi} - m}{s}\sqrt{n} = \frac{\frac{\bar{\xi}-m}{\sigma}\sqrt{n}}{\frac{s}{\sigma}} = \frac{\frac{\bar{\xi}-m}{\sigma}\sqrt{n}}{\sqrt{\frac{s^2}{\sigma^2}}} \in t_{n-1}. \tag{5}$$

Выберем уровень доверия $\gamma = 0.95$. Тогда для выбранного уровня доверия определим величину t_γ такую, что будет выполняться $P\left(|t_{n-1}| < t_\gamma\right) = \gamma$. Воспользуемся функцией, обратной функции распределения, `tinv` из пакета Matlab.

```
tinv(       (1 - 0.95)/2, 99) = -1.98;
tinv(0.95 + (1 - 0.95)/2, 99) = 1.98;
```

Преобразуем левую часть уравнения (5):

$$\begin{gathered}\bar{x} - \frac{s_x}{\sqrt{n}}t_\gamma < M\xi < \bar{x} + \frac{s_x}{\sqrt{n}}t_\gamma \\ \frac{s_x}{\sqrt{n}}t_\gamma = \frac{2.8734}{10} \cdot 1.98 = 0.5689 \\ \bar{x} - 0.5689 < M\xi < \bar{x} + 0.5689 \\ 5.5111 < M\xi < 6.6489\end{gathered}$$

Следовательно, $M\xi \in (5.5111, 6.6489)$ с уровнем доверия $\gamma = 0.95$.

Теперь будем искать доверительный интервал для $D\xi$. Рассмотрим случайную величину $U = (n-1)\frac{s^2}{\sigma^2_\xi}$. Из уравнения (4) видно, что $U \in \chi^2_{n-1}$. Выберем интервал (U_1, U_2) так, чтобы вероятность попадания U левее и правее него была одинаковой и равной, следовательно, $(1-\gamma)/2$. При доверительной вероятности $\gamma = 0.95$ имеем $(1-0.95)/2 = 0.025$. Для вычисления значения U_1 и U_2 воспользуемся функцией `chi2inv` из пакета Matlab.

```
chi2inv(    0.025, 99) = 73.36
chi2inv(1 - 0.025, 99) = 128.42
```

Тогда получаем:

$$\begin{gathered}73.36 < 99 \cdot \frac{8.2562}{\sigma^2_\xi} < 128.42 \\ 6.36 < D\xi < 11.14\end{gathered} \tag{6}$$

Следовательно $D\xi \in (6.36, 11.14)$ с уровнем доверия $\gamma = 0.95$.

Проверка гипотезы о равенстве значения числовой характеристики фиксированному числу

Проверим гипотезу о равенстве значения $M\xi$ некоторому фиксированному числу. Сформулируем простую основную гипотезу и сложную альтернативную:

$$H_0 : M\xi = \mu_0 = [\bar{x}] = 6$$
$$H_1 : M\xi \neq \mu_0 = 6$$

Запишем критерий проверки справедливости гипотезы:

$$T = t_{n-1} = \frac{\bar{x} - \mu_0}{s}\sqrt{n} \tag{7}$$

Выберем уровень значимости $\alpha = 0.01$, тогда поскольку критерий подчиняется закону распределения Стьюдента с параметром $n - 1$, тогда определим $t_{\text{кр.}}$, исходя из того, что $P\left(|t_{n-1}| > t_{\text{кр.}}\right) = \alpha$.

Используя функцию, обратную функции распределения `tinv`, получим:

```
tinv(            0.01/2, 99) = -2.62
tinv((1 - 0.01) + 0.01/2, 99) =  2.62
```

Вычислим $T_{\text{набл.}}$:

$$T_{\text{набл.}} = \frac{\bar{x} - \mu_0}{s_x} \cdot \sqrt{n} = \frac{6.08 - 6}{2.8734} \cdot 10 = 0.27842 \tag{8}$$

Так как $T_{\text{набл.}} \in Q_{\text{доп.}} \Rightarrow$ оснований отвергать гипотезу H_0 у нас нет, и мы её принимаем. Вычислим вероятности ошибок второго рода полагая: $H_1' : \mu_1 = 5.5$ и $H_2'' : \mu_2 = 6.5$. Согласно теореме Леви, будем считать, что распределение вероятностей t_{n-1} мало отличается от $N\left(\bar{y} - \mu_1; \frac{s_x}{\sqrt{n}}\right)$. тогда вероятность ошибки второго рода будет:

$$\beta = \Phi\left(\frac{\mu_0 - \mu_1}{s}\sqrt{n} + t_{\text{кр.}}\right) - \Phi\left(\frac{\mu_0 - \mu_1}{s}\sqrt{n} - t_{\text{кр.}}\right).$$

Получим $\beta' = \beta'' = 0.81054$.

Проверим гипотезу равенства дисперсии определённому числу. Запишем гипотезы, критерий и уровень значимости:

$$H_0 : D\xi = [s_x^2] = 8$$
$$H_1 : D\xi \neq 8$$
$$T = \chi^2_{n-1} = (n-1)\frac{s_x^2}{\sigma_0^2}$$
$$\alpha = 0.01$$

С помощью `chi2inv` вычислим $\chi^2_{\text{крит.}}$, такое что $P\left(\chi^2_{n-1} > \chi^2_{\text{крит.}}\right) = \alpha$.

```
chi2inv(1 - 0.01, 99) = 134.64
```

Вычислим $T_{\text{набл.}}$:

$$T_{\text{набл.}} = (n-1)\frac{s_x^2}{\sigma_0^2} = 99 \cdot \frac{8.2562}{8} = 102.17 \tag{9}$$

Поскольку $T_{\text{набл.}} \in Q_{\text{доп.}} \Rightarrow$ оснований не принимать гипотезу H_0 нет, мы её принимаем.

Проверка гипотезы о совпадении значений одноимённых числовых характеристик двух различных случайных величин

Проверим гипотезу о совпадении дисперсий двух случайных величин. Рассмотрим дисперсии рассматриваемой нами случайной величины $D\xi$ и другой $D\xi'$. Полученные точечные оценки дисперсий: $s_x^2 = 8.2562$ и $s_x'^2 = 9.4016$. Сформулируем гипотезы:

$$H_0 : \frac{D\xi'}{D\xi} = 1$$
$$H_1 : \frac{D\xi'}{D\xi} > 1$$
$$T = f_{n'-1,n-1} = \frac{\chi^2_{n'-1}/(n'-1)}{\chi^2_{n-1}/(n-1)}$$

Критерием проверки справедливости гипотезы будет отношение $T \in F(n'-1, n-1)$ - подчиняющееся распределению Фишера-Снедекора (F-распределению).

Выберем уровень значимости $\alpha = 0.01$, тогда:

```
finv(1 - 0.01, 99, 99) = 1.6015
```

Вычислим $f_{\text{набл.}}$:

$$f_{\text{набл.}} = \frac{s_x'^2}{s_x^2} = \frac{9.4016}{8.2562} = 1.1387 \tag{10}$$

Поскольку $1.1387 < 1.6015 \Rightarrow T \in Q_{\text{доп.}}$, а значит мы принимаем гипотезу $H_0 : D\xi = D\xi'$. Используем этот результат в проверке гипотезы о равенстве математических ожиданий. Запишем гипотезы и критерий:

$$H_0 : M\xi = M\xi'$$
$$H_1 : M\xi \neq M\xi'$$
$$T = t_{n_1+n_2-2} = \frac{\bar{x} - \bar{x}'}{s} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}},$$
$$s^2 = \frac{1}{n_1 + n_2 - 2} \left((n_1 - 1)\, s_x^2 + (n_2 - 1) s_{x'}^2 \right)$$

```
tinv(1 - 0.01, 100 + 100 - 2) = 2.345
```

Следовательно, при уровне значимости $\alpha = 0.01$, $|t_{\text{крит.}}| = 2.345$. Вычислим $T_{\text{набл.}}$:

$$s^2 = \frac{1}{198} (99 \cdot 8.2562 + 99 \cdot 9.4016) = 8.8289$$
$$T_{\text{набл.}} = \frac{6.08 - 4.82}{\sqrt{8.8289}} \cdot \sqrt{\frac{100 \cdot 100}{2 \cdot 100}} = 2.9985$$

Поскольку $T_{\text{набл.}} > t_{\text{крит.}} \Rightarrow$ отклоняем гипотезу H_0 и принимаем H_1, при этом вероятность ошибки первого рода равна $\alpha = 0.01$.

Проверка гипотезы о виде закона распределения

В нашем случае гистограмма похожа на плотность вероятности гамма-распределения. Плотность вероянтости гамма-распределения:

$$p(x) = \begin{cases} 0, & x \leqslant 0 \\ \frac{1}{\beta^\alpha \Gamma(\alpha)} \cdot x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)}, & x > 0 \end{cases},$$

где α и β – числовые параметры распределения. Найдём эти параметры:

$$M\xi = \alpha \cdot \beta, D\xi = \alpha \cdot \beta^2 \Rightarrow$$
$$\Rightarrow \alpha = \frac{M^2\xi}{D\xi}, \beta = \frac{D\xi}{M\xi}$$
$$M\xi = 6, D\xi = 8 \Rightarrow \alpha = \frac{36}{8} = 4.5, \beta = \frac{8}{6} = 1.33$$

Критерием будет случайная величина:

$$T = \chi^2_{k-l-1} = \sum_{j=1}^{k} \frac{(m_j - np_j)^2}{np_j}$$

где $k = 6$ – количество интервалов вариационного ряда, а $l = 2$ – количество параметров распределения, которые заменяются их точечными оценками, тогда при урове значимости $\alpha = 0.01$, $\chi^2_{\text{крит.}} = 11.345$.

```
chi2inv(1 - 0.01, 6 - 2 - 1) = 11.345
```

	$(a_{j_1}; a_j]$	m_j	p_j	np_j	$(m_j - np_j)^2/np_j$
1	$[1;3)$	6	0.12245	12.245	2.6187
2	$[3;5)$	23	0.2914	29.14	1.2937
3	$[5;7)$	35	0.27369	27.369	2.1277
4	$[7;9)$	20	0.16968	12.968	3.8132
5	$[9;11)$	10	0.083499	8.3499	0.32609
6	$[11;21]$	6	0.056166	5.6166	0.026172
Σ		100	0.99689		10.206

Получили $\chi^2_{\text{набл.}} = 10.206 < 11.345 \Rightarrow$ у нас нет оснований отклонять гипотезу $H_0 : \xi \in \Gamma(\alpha, \beta)$.

Проверка гипотезы о совпадении законов распределения двух случайных величин

Имеем вариационный ряд, построенный по выборке x'_i значений случайной величины ξ' и подогнанный к разбиению вариационного ряда ξ, чтобы $a_i = a'_i$.

$[a_i, a_{i+1})$	$[1,3)$	$[3,5)$	$[5,7)$	$[7,9)$	$[9,11)$	$[11,13)$	$[13,15)$	$[15,17)$	$[17,18)$	$[19,21]$
$\frac{m_i}{n}$	0.17	0.35	0.31	0.11	0.02	0.01	0.0	0.01	0.01	0.01

Таблица 2: Вариационный ряд $\{x'_i\}$

По этой выборке так же имеем: $\bar{x}' = 4.82, s'^2 = 9.4016, s' = 3.0662$. Формулируем гипотезу и критерий:

$$H_0 : F_\xi(x) = F_{\xi'}(x)$$
$$T = \chi^2_{k-1} = n_1 \cdot n_2 \sum_{j=1}^{k} \frac{1}{m_j + m'_j} \left(\frac{m_j}{n_1} - \frac{m'_j}{n_2} \right)^2 \overset{(n_1=n_2=100)}{=} \sum_{j=1}^{k} \frac{\left(m_j - m'_j\right)^2}{m_j + m'_j}.$$

Выбираем уровень значимости $\alpha = 0.01$, тогда получаем $\chi^2_{\text{крит.}} = 15.086$. Вычислим $\chi^2_{\text{набл.}}$:

$$\chi^2_{\text{набл.}} = 16.332$$

Так как $16.332 > 15.086 \Rightarrow T_{\text{набл.}} \in S_{\text{крит.}}$, значит мы отклоняем гипотезу H_0 с уровнем значимости α и принимаем гипотезу H_1. При этом вероятность ошибки первого рода $\alpha = 0.01$.

Первичная обработка статистических данных

Вторая выборка $\{y_i\}$ объёма $n = 100$. Размах выборки: $\min(y_i) = 1.8$ и $\max(y_i) = 68.9$. Интервал $[1.8, 68.9]$ так же делим на $k_2 = 10$ частей.

$[a_i, a_{i+1})$	$[1.8, 8.51)$	$[8.51, 15.22)$	$[15.22, 21.93)$	$[21.93, 28.64)$	$[28.64, 35.35)$
$\frac{m_i}{n}$	0.08	0.08	0.12	0.13	0.14
	$[35.35, 42.06)$	$[42.06, 48.77)$	$[48.77, 55.48)$	$[55.48, 62.19)$	$[62.19, 68.9]$
	0.08	0.18	0.04	0.07	0.08

Таблица 3: Вариационный ряд $\{y_i\}$

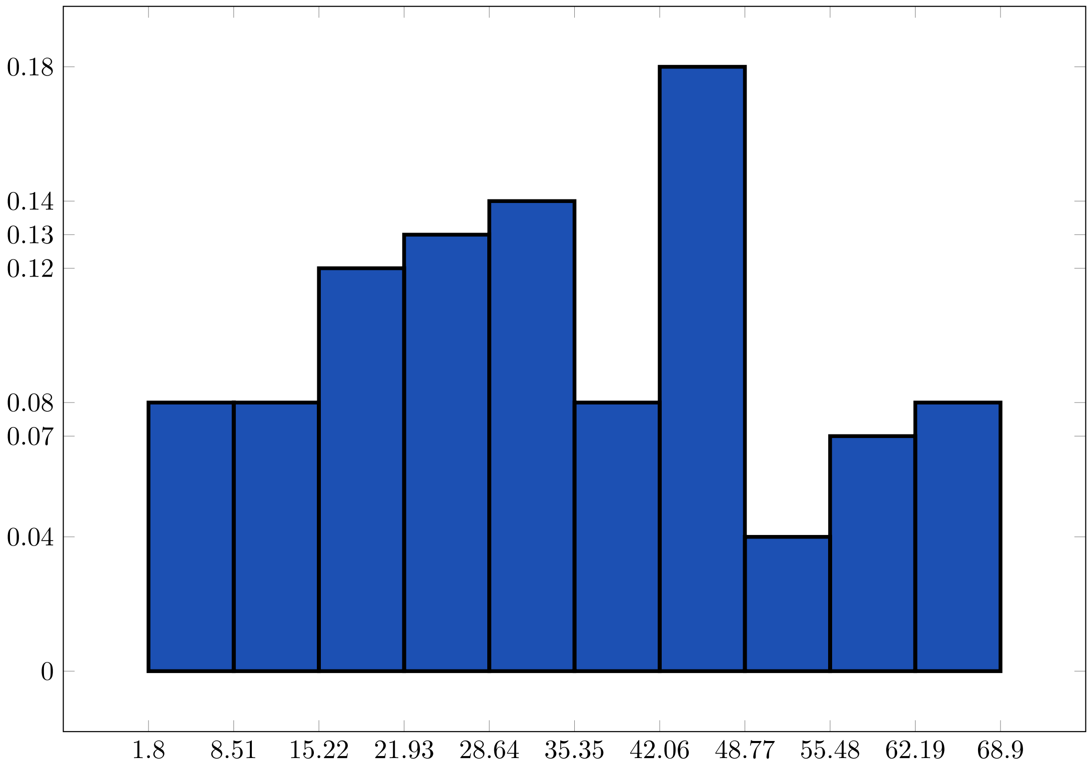


Рис. 2: Гистограмма $\{y_i\}$

Оценка мат. ожидания $M\eta$:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \approx 33.923 \tag{11}$$

Оценка дисперсии $D\eta$:

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n} y_i^2 - \frac{n}{n-1}\bar{y}^2 \approx 323.4 \tag{12}$$

Оценка среднеквадратического отклонения:

$$s_y = \sqrt{s_y^2} \approx 17.983 \tag{13}$$

Интервальные оценки числовых характеристик

Найдём доверительный интервал для $M\eta$. Для выбранного уровня доверия $\gamma = 0.95$, $t_\gamma = 1.98$. Тогда:

$$\bar{y} - \frac{s_y}{\sqrt{n}} t_\gamma < M\eta < \bar{y} + \frac{s_y}{\sqrt{n}} t_\gamma$$

$$\frac{s_y}{\sqrt{n}} t_\gamma = \frac{17.983}{10} \cdot 1.98 = 3.56$$

$$\bar{y} - 3.56 < M\eta < \bar{y} + 3.56$$

$$30.363 < M\eta < 37.483$$

Следовательно, $M\eta \in (30.363, 37.483)$ с уровнем доверия $\gamma = 0.95$.

Теперь будем искать доверительный интервал для $D\eta$:

$$\chi_1^2 < (n-1) \cdot \frac{s_y^2}{\sigma_\eta^2} < \chi_2^2$$

$$73.36 < 99 \cdot \frac{323.4}{\sigma_\eta^2} < 128.42$$

$$249.31 < D\eta < 436.43$$

Следовательно $D\eta \in (249.31, 436.43)$ с уровнем доверия $\gamma = 0.95$.

Проверка гипотезы о равенстве значения числовой характеристики фиксированному числу

$$H_0 : M\eta = \mu_0 = [\bar{y}] = 34$$

$$H_1 : M\eta \neq \mu_1 = 34$$

$$\alpha = 0.01$$

$$T = \frac{\bar{y} - \mu_0}{s} \sqrt{n}$$

Найдём $t_{\text{крит.}}$, исходя из того, что $P\left(|t_{n-1}| > t_{\text{кр.}}\right) = \alpha$:

```
tinv(            0.01/2, 99) = -2.62
tinv((1 - 0.01) + 0.01/2, 99) =  2.62
```

Вычислим $T_{\text{набл.}}$:

$$T_{\text{набл.}} = \frac{33.923 - 34}{17.983} \cdot 10 = -0.042818 \tag{14}$$

Так как $0.042818 < 2.62 \Rightarrow T_{\text{набл.}} \in Q_{\text{доп.}} \Rightarrow$ оснований отвергать гипотезу H_0 нет, мы её принимаем.

Теперь проверим гипотезу о равенстве $D\eta$:

$$H_0 : D\eta = [s_y^2] = 323$$

$$H_1 : D\eta \neq = 323$$

$$\alpha = 0.01$$

$$T = (n-1)\, \frac{s_y^2}{\sigma_0^2}$$

Найдём $\chi^2_{\text{крит.}}$, такое что $P\left(\chi^2_{n-1} > \chi^2_{\text{крит.}}\right) = \alpha$.

```
chi2inv(1 - 0.01, 99) = 134.64
```

Вычислим $T_{\text{набл.}}$:

$$T_{\text{набл.}} = (n-1)\, \frac{s_y^2}{\sigma_0^2} = 99 \cdot \frac{323.4}{323} = 99.123 \tag{15}$$

Поскольку $T_{\text{набл.}} \in Q_{\text{доп.}} \Rightarrow$ оснований не принимать гипотезу H_0 нет, мы её принимаем.

Проверка гипотезы о совпадении значений одноимённых числовых характеристик двух различных случайных величин

Проверим гипотезу о совпадении дисперсий рассматриваемой нами случайной величины $D\eta$ и $D\eta'$. Полученные точечные оценки дисперсий: $s_y^2 = 323.4$ и $s_y'^2 = 815.59$. Запишем гипотезы, критерий и уровень значимости:

$$
\begin{aligned}
&H_0 : \frac{D\eta'}{D\eta} = 1 \\
&H_1 : \frac{D\eta'}{D\eta} > 1 \\
&\alpha = 0.01 \\
&T = f_{n'-1,n-1} = \frac{\chi^2_{n'-1} / \left(n' - 1\right)}{\chi^2_{n-1} / \left(n - 1\right)}
\end{aligned}
$$

Для уровня значимости $\alpha = 0.01$:

```
finv(1 - 0.01, 99, 99) = 1.6015
```

Вычислим $T_{\text{набл.}}$:

$$T_{\text{набл.}} = \frac{s_y'^2}{s_y^2} = \frac{815.59}{323.4} = 2.5219 \tag{16}$$

Поскольку $2.5219 > 1.6015 \Rightarrow T \in S_{\text{крит.}}$, а значит мы отклоняем гипотезу H_0 и принимаем H_1. При этом вероятность ошибки первого рода равна $\alpha = 0.01$.

Запишем гипотезы:

$$
\begin{aligned}
&H_0 : M\eta = M\eta' \\
&H_1 : M\eta \neq M\eta'
\end{aligned}
$$

Поскольку гипотеза о равенстве дисперсий не подтвердилась, будем использовать другой критерий:

$$T = u = \left(\bar{y} - \bar{y}'\right) / \sqrt{\frac{s_y^2}{n} + \frac{s_y'^2}{n'}} \tag{17}$$

Этот критерий, согласно теореме Леви, асимптотически нормален, тогда $P\left(|u| > u_{\text{кр.}}\right) = 1 - 2\Phi\left(u_{\text{кр.}}\right) = \alpha$, или $u_{\text{кр.}} = \Phi^{-1}\left(\left(1 - \alpha\right)/2\right)$.

```
norminv(    0.01/2) = -2.5758
norminv(1 - 0.01/2) =  2.5758
```

Вычислим $T_{\text{набл.}}$:

$$T_{\text{набл.}} = (33.923 - 51.184) / \sqrt{\frac{323.4}{100} + \frac{815.59}{100}} = -17.261 \cdot \sqrt{11.390} = -58.254 \tag{18}$$

Поскольку $|T_{\text{набл.}}| > u_{\text{кр.}} \Rightarrow$ мы отклоняем гипотезу H_0 и принимаем H_1.

Проверка гипотезы о виде закона распределения

В нашем случае гистограмма похожа на плотность вероятности равномерного распределения. Плотность вероятности равномерного распределения:

$$p(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a \leqslant x < b \\ 0, & b \leqslant x \end{cases},$$

где a и b – начало и конец отрезка, на котором η принимает не нулевые значения. Возьмём в качестве этих параметров $x_{min} = 1.8$ и $x_{max} = 68.9$. Критерий:

$$T = \chi^2_{k-l-1} = \sum_{j=1}^{k} \frac{(m_j - np_j)^2}{np_j} \tag{19}$$

где $k = 10$ – количество интервалов вариационного ряда, а $l = 2$ – количество параметров распределения, которые заменяются их точечными оценками. Тогда при уровне значимости $\alpha = 0.01$, $\chi^2_{\text{крит.}} = 18.475$.

```
chi2inv(1 - 0.01, 10 - 2 - 1) = 18.475
```

	$(a_{j_1}; a_j]$	m_j	p_j	np_j	$(m_j - np_j)^2/np_j$
1	$[1.8; 8.51)$	8	0.1	10	0.39996
2	$[8.51; 15.22)$	8	0.1	10	0.39996
3	$[15.22; 21.93)$	12	0.1	10	0.40004
4	$[21.93; 28.64)$	13	0.1	10	0.90007
5	$[28.64; 35.35)$	14	0.1	10	1.6001
6	$[35.35; 42.06)$	8	0.1	10	0.39996
7	$[42.06; 48.77)$	18	0.1	10	6.4002
8	$[48.77; 55.48)$	4	0.1	10	3.5999
9	$[55.48; 62.19)$	7	0.1	10	0.89995
10	$[62.19; 68.9]$	8	0.1	10	0.39996
Σ		100	1		15.400

Получили $\chi^2_{\text{набл.}} = 15.4 < 18.475 \Rightarrow$ у нас нет оснований отклонять гипотезу H_0.

Проверка гипотезы о совпадении законов распределения двух случайных величин

Перепишем оба вариационных ряда так, чтобы разбиения у них совадали ($a_i = a'_i$).

$[a_i, a_{i+1})$	$[0, 10)$	$[10, 20)$	$[20, 30)$	$[30, 40)$	$[40, 50)$
$\frac{m_i}{n}$	8	19	19	14	21
	$[50, 60)$	$[60, 70)$	$[70, 80)$	$[80, 90)$	$[90, 100]$
	9	10	0	0	0

Таблица 3: Вариационный ряд $\{y_i\}$

$[a_i, a_{i+1})$	$[0, 10)$	$[10, 20)$	$[20, 30)$	$[30, 40)$	$[40, 50)$
$\frac{m_i}{n}$	11	8	8	10	13
	$[50, 60)$	$[60, 70)$	$[70, 80)$	$[80, 90)$	$[90, 100]$
	8	14	7	8	13

Таблица 3: Вариационный ряд $\{y'_i\}$

Мы видим, что в первом вариационном ряду значения на отрезках $[70, 80), [80, 90)$ и $[90, 100)$ равны нулю. Мы можем сделать проверку двумя способами: 1) можно выбросить часть вариационного ряда y'_i, и 2) объединим интервалы в $[60, 100)$ в обоих вариационных рядах.

Рассмотрим вариант 1):

Имеем гипотезу: $H_0 : F_\eta(x) = F_{\eta'}(x)$;

Тогда, поскольку мы выбросили часть выборки $\{y'_i\}$, тогда $n_2 = 100 - 13 - 8 - 8 = 71$ и критерий будет иметь вид:

$$T = \chi^2_{k-1} = n_1 \cdot n_2 \sum_{j=1}^{k} \frac{1}{m_j + m'_j} \left(\frac{m_j}{n_1} - \frac{m'_j}{n_2} \right)^2 \tag{20}$$

Для уровня значимости $\alpha = 0.01$ получим $\chi^2_{\text{крит.}} = 16.812$

Вычислим $T_{\text{набл.}} = 8.3750 < 16.812 \Rightarrow H_0$ принимается.

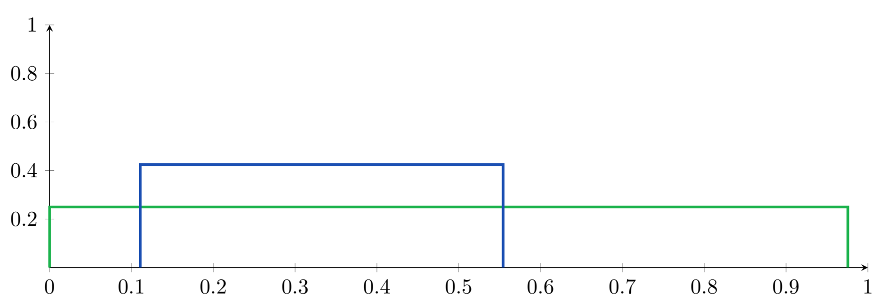
Рассмотрим вариант 2):

Гипотеза H_0 - та же. Поскольку $n_1 = n_2 = 100$, то воспользуемся более простым критерием:

$$T = \sum_{j=1}^{k} \frac{\left(m_j - m'_j\right)^2}{m_j + m'_j}. \tag{21}$$

Тогда при тех же $\alpha = 0.01$ и $\chi^2_{\text{крит.}} = 16.812$ получим $\chi^2_{\text{набл.}} = 31.737 > 16.812 \Rightarrow H_0$ отклоняется.

Какой из способов выбрать? Мы знаем, что оба распределения - равномерны, но их графики, видимо, выглядят так:



Корреляционный анализ

По элементам двумерной выборки $\{(x_i; y_i) : y \in \overline{1, n}\}$ составим корреляционную таблицу:

$\eta \backslash \xi$	$[1,3)$	$[3,5)$	$[5,7)$	$[7,9)$	$[9,11)$	$[11,13)$	$[13,15)$	$[15,17)$	$[17,19)$	$[19,21]$	
$[1.8; 8.51)$		1		2	2	3					8
$[8.51; 15.22)$	1		3	2	1	1					8
$[15.22; 21.93)$			3	3	5		1				12
$[21.93; 28.64)$			6	5	1					1	13
$[28.64; 35.35)$		2	10	1	1						14
$[35.35; 42.06)$		3	3	2							8
$[42.06; 48.77)$		4	10	4							18
$[48.77; 55.48)$	1	2		1							4
$[55.48; 62.19)$	1	6									7
$[62.19; 68.9]$	3	5									8
	6	23	35	20	10	4	1			1	100

Сила статистической связи оценивается с помощью коэффициента линейной корреляции:

$$\rho = \frac{\alpha_{11} - m_\xi m\eta}{\sigma_\xi \sigma_\eta} \tag{22}$$

Точечная оценка коэффициента линейной корреляции:

$$\begin{gathered} r = \frac{a_{11} - \bar{x}\bar{y}}{s_x s_y}, \text{где} \\ a_{11} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i \end{gathered} \tag{23}$$

В нашем случае $r = -0.011603$. Получнное близкое к нулю значение позволяет сделать вывод о почти отсутствующей связи между случайными величинами.

Найдём функции регрессии: $f(x) = M\left[\eta / \{\xi = x\}\right]$. По корреляционной таблице видно, что обе функции регрессии - линейные, то есть $f(x) = a_1 x + b_1$. Теоретическое уравнение линейной функции регрессии η на ξ имеет вид:

$$y - m_\eta = \rho \frac{\sigma_\eta}{\sigma_\xi}\left(x - m_\xi\right) \tag{24}$$

Заменяя в этом уравнении теоретические числовые характеристики их точечными оценками получим:

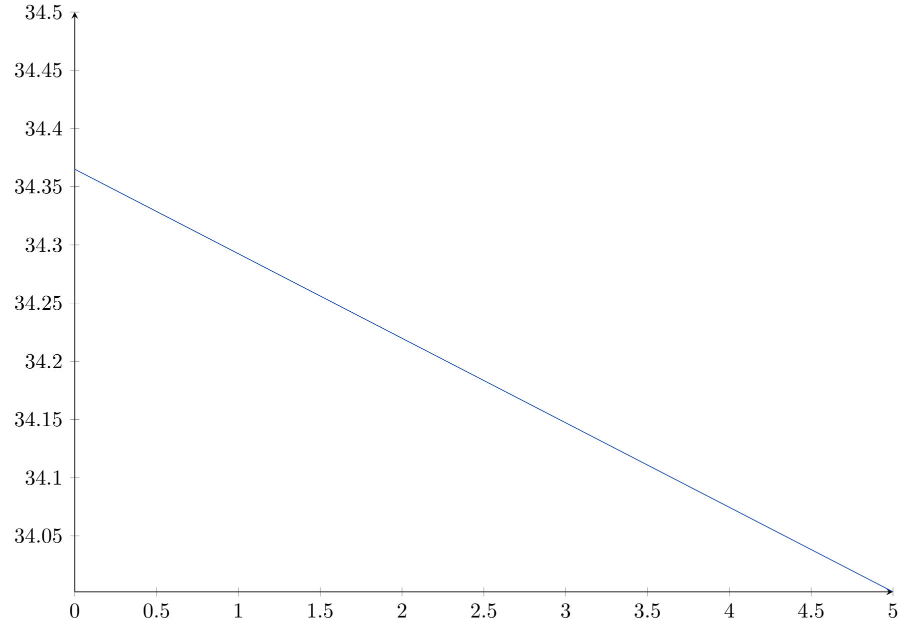
$$y - \bar{y} = r\frac{s_y}{s_x}\left(x - \bar{x}\right) \tag{25}$$

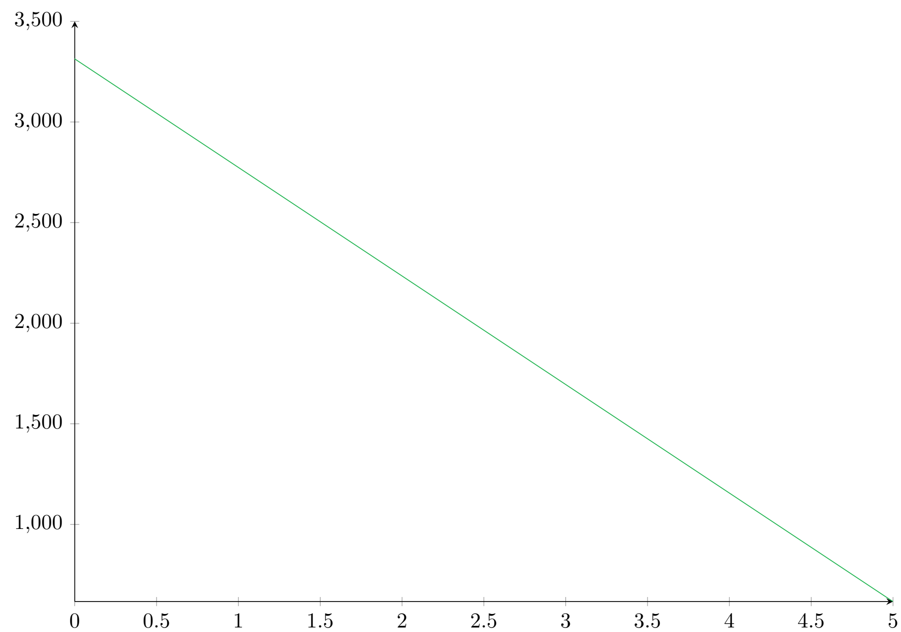
В нашем случае уравнение принимает вид:

$$\begin{aligned} & y - 33.923 = -0.011603 \cdot \frac{17.983}{2.8734}\left(x - 6.08\right) \\ & y - 33.923 = -0.072617 \cdot \left(x - 6.08\right) \\ & y = -0.072617x + 34.365 \end{aligned} \tag{26}$$

Аналогично находим регрессию ξ на η:

$$\begin{aligned} & x - m_\xi = \rho\frac{\sigma_\xi}{\sigma_\eta}\left(y - m_\eta\right) \\ & x - \bar{x} = r\frac{s_x}{s_y}\left(y - \bar{y}\right) \\ & x - 6.08 = -0.011603 \cdot \frac{2.8734}{17.983}\left(y - 33.923\right) \\ & x = -0.0018539y + 6.1429 \end{aligned} \tag{27}$$


34.5
34.45
34.4
34.35
34.3
34.25
34.2
34.15
34.1
34.05
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5


3,500
3,000
2,500
2,000
1,500
1,000
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5