

Лабораторная работа №3.

Проверка статистических гипотез о виде и параметрах распределения случайных величин.

Часть 1.

1. Используя данные Росстата, приведенные в файле данных CHISLO_DOCTORS.xlsx, найти значения случайного показателя X : **Число врачей на 10 тысяч человек населения** – в указанном в Вашем варианте Федеральном Округе в каждый год периода T : 2005, 2010, 2015, 2019, 2010, 2021 гг.
2. Визуализировать данные показателя X в каждый год периода T по указанному в Вашем варианте Федеральному Округу с помощью графиков и боксплотов.
3. Вычислить описательную статистику: среднее, стандартное отклонение, квартили, минимальное и максимальное значения показателя X в каждый год периода T по указанному в Вашем варианте Федеральному Округу.
4. Проверить, можно ли считать, что распределение случайной величины X в указанном Федеральном Округе в каждый год периода T подчинено нормальному закону распределения. Использовать для проверки тест Шапиро-Уилка (уровень значимости α указан в Вашем варианте).

Для дальнейшего исследования использовать только те года, для которых распределение случайного показателя X в Федеральном Округе можно считать нормальным.

5. Выделить те года t_1 - t_m с нормально распределенными значениями рядов данных X_{t_1} - X_{t_m} , где X_{t_1} - X_{t_m} имеют одинаковую дисперсию (уровень значимости взять равным α). Использовать для проверки нулевой гипотезы о равенстве дисперсий тесты Бартлетта и Левена.

Для выполнения п. 6-8 использовать ряды данных X_{t_1} - X_{t_m} из этой группы.

6. Проверить, можно ли считать, что среднее значение показателя X по данному Федеральному Округу в каждый год периода t_1 - t_m значимо выше (ниже) общероссийского значения показателя X (уровень значимости взять равным α). Общероссийские значения показателя X найти в файле CHISLO_DOCTORS.xlsx. Использовать для проверки гипотезы о равенстве средних t -тест для одной выборки.
7. Проверить, можно ли считать, что различия между средними значениями показателя X по данному Федеральному Округу в какие-то два года из периода t_1 - t_m незначимы, появились случайно (уровень значимости взять равным α). Использовать для проверки гипотезы о равенстве средних t -тест для двух выборок.
8. Проверить значимость отличий средних в выбранной группе (уровень значимости взять равным α). Использовать для проверки гипотезы о равенстве средних групп тест Тьюки и односторонний тест ANOVA.

Часть 2.

1. Проверить, можно ли считать, что распределение случайной величины X в указанном Федеральном Округе за весь период T подчинено нормальному закону распределения. Использовать для проверки следующие три критерия:

Хи-квадрат, Шапиро-Уилка и критерий Д'Агостино (уровень значимости α указан в Вашем варианте).

2. Смоделировать M выборок объемом n из значений случайной величины X , имеющей нормальное распределение с параметрами, указанными в Вашем варианте. На уровне значимости α проверить для каждой выборки гипотезу о нормальном законе распределения с помощью критериев Шапиро-Уилка и Д'Агостино. По результатам моделирования M выборок вычислить оценку вероятности совершить ошибку первого рода.

3. Смоделировать M выборок объемом n из значений случайной величины Y , имеющей указанное в Вашем варианте распределение. На уровне значимости α проверить для каждой выборки гипотезу о нормальном законе распределения с помощью критериев Шапиро-Уилка и Д'Агостино. Вычислить оценку вероятности не допустить ошибку второго рода. Какой из критериев при данной альтернативе является более мощным?

УКАЗАНИЯ.

Полную информацию о статистических критериях от команды разработчиков см.

<https://docs.scipy.org/doc/scipy/reference/stats.html>

Информацию о t-тестах см.

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_1samp.html#scipy.stats.ttest_1samp

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind_from_stats.html#scipy.stats.ttest_ind_from_stats

Информацию о тестах для проверки на нормальность см.

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.normaltest.html#scipy.stats.normaltest>

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.shapiro.html#scipy.stats.shapiro>

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chisquare.html#scipy.stats.chisquare>

Информацию о тестах для проверки выборок на равенство дисперсий см.

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.bartlett.html>

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.levene.html#scipy.stats.levene>

Информацию о тестах для проверки групп на равенство средних (ожидаемых) значений групп см.

Краткая информация по параметрическим гипотезам

Предполагается, что случайная величина X имеет нормальный закон распределения.

Таблица 1. Основные параметрические гипотезы для одной выборки.

H_0	Предположения	Статистика критерия	H_1	Область принятия H_0
$a = a_0$	σ^2 известно	$U = \frac{\bar{x} - a_0}{\sigma} \sqrt{n}$	$a > a_0$ $a < a_0$ $a \neq a_0$	$U < u_{\text{кр}}, \Phi_0(u_{\text{кр}}) = 1/2 - \alpha$ $U > -u_{\text{кр}}, \Phi_0(u_{\text{кр}}) = 1/2 - \alpha$ $ U < u_{\text{кр}}, \Phi_0(u_{\text{кр}}) = (1-\alpha)/2$
	σ^2 неизвестно	$T = \frac{\bar{x} - a_0}{s} \sqrt{n}$	$a > a_0$ $a < a_0$ $a \neq a_0$	$T < t_{\text{кр}}(\alpha, n-1)$ для односторонней области $T > -t_{\text{кр}}(\alpha, n-1)$ для односторонней области $ T < t_{\text{кр}}(\alpha, n-1)$ для двусторонней области
$\sigma^2 = \sigma_0^2$	a неизвестно	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$ $\sigma^2 \neq \sigma_0^2$	$\chi^2 < \chi_{\alpha; n-1}^2$ $\chi^2 > \chi_{1-\alpha; n-1}^2$ $\chi_{1-\alpha/2; n-1}^2 < \chi^2 < \chi_{\alpha/2; n-1}^2$
$p = p_0$	n порядка нескольких десятков или сотен	$U = \frac{w - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n}$, где $w = m/n$	$p > p_0$ $p < p_0$ $p \neq p_0$	$U < u_{\text{кр}}, \Phi_0(u_{\text{кр}}) = 1/2 - \alpha$ $U > -u_{\text{кр}}, \Phi_0(u_{\text{кр}}) = 1/2 - \alpha$ $ U < u_{\text{кр}}, \Phi_0(u_{\text{кр}}) = (1-\alpha)/2$

Предполагается, что случайные величины X и Y являются независимыми и имеют нормальный закон распределения.

Таблица 2. Основные параметрические гипотезы для двух выборок.

H_0	Предположения	Статистика критерия	H_1	Область принятия H_0
$\sigma_x^2 = \sigma_y^2$	σ_x^2 и σ_y^2 известны	$U = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$	$a_x > a_y$ $a_x < a_y$ $a_x \neq a_y$	$U < u_{кр}, \Phi_0(u_{кр}) = 1/2 - \alpha$ $U > -u_{кр}, \Phi_0(u_{кр}) = 1/2 - \alpha$ $ U < u_{кр}, \Phi_0(u_{кр}) = (1 - \alpha)/2$
	σ_x^2 и σ_y^2 неизвестны, но равны	$T = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{n} + \frac{1}{m}}}, \text{ где } s^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$	$a_x > a_y$ $a_x < a_y$ $a_x \neq a_y$	$T < t_{кр}(\alpha, n+m-2)$ для односторонней области $T > -t_{кр}(\alpha, n+m-2)$ для односторонней области $ T < t_{кр}(\alpha, n+m-2)$ для двусторонней области
$\sigma_x^2 \neq \sigma_y^2$	a_x и a_y неизвестны	$F = \frac{s_x^2}{s_y^2}, \text{ где } s_x^2 > s_y^2$	$\sigma_x^2 > \sigma_y^2$ $\sigma_x^2 \neq \sigma_y^2$	$F < F_{кр}(\alpha, n-1, m-1)$ $F < F_{кр}(\alpha/2, n-1, m-1)$
$p_1 = p_2$	n_1 и n_2 порядка нескольких десятков или сотен	$U = \frac{w_1 - w_2}{\sqrt{w(1-w)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \text{ где } w = \frac{m_1 + m_2}{n_1 + n_2}$	$p_1 > p_2$ $p_1 < p_2$ $p_1 \neq p_2$	$U < u_{кр}, \Phi_0(u_{кр}) = 1/2 - \alpha$ $U > -u_{кр}, \Phi_0(u_{кр}) = 1/2 - \alpha$ $ U < u_{кр}, \Phi_0(u_{кр}) = (1 - \alpha)/2$

См. также критерии Колмогорова-Смирнова, Хи-квадрат, Шапиро-Уилка, Бартлета, Тьюки, Левена и др..

ВАРИАНТЫ ЗАДАНИЙ

Вариант	Федеральный округ (ФО)	α	M	(a, sigma)	n	Закон распределения случайной величины Y
<u>25</u>	Центральный ФО	0,025	3000	(1, 2)	55	Хи-квадрат распределение с числом степеней свободы $k=19$
<u>24</u>	Южный ФО	0,01	4000	(2, 3)	65	Распределение Стьюдента с числом степеней свободы $k=13$
<u>23</u>	Приволжский ФО	0,015	5000	(-1, 2)	60	Релея с модой, равной 16.
<u>22</u>	Уральский ФО	0,02	5500	(-2, 3)	70	F-распределение с числом степеней свободы $k_1=k_2=13$
<u>21</u>	Сибирский ФО	0,03	6000	(3, 2)	75	Хи-квадрат с числом степеней свободы $k=15$
<u>20</u>	Дальневосточный ФО	0,035	2500	(- 3, 2)	80	Распределение Стьюдента с числом степеней свободы $k=14$
<u>19</u>	Северо-Кавказский ФО	0,01	3000	(1, 3)	85	Треугольное на отрезке (1, 4) и модой, равной 2
<u>18</u>	Центральный ФО	0,02	4000	(-1, 3)	90	Релея с модой, равной 7.
<u>17</u>	Южный ФО	0,015	5000	(-3, 4)	95	Логистическое с параметрами масштаба и сдвига 9 и 4 соответственно
<u>16</u>	Приволжский ФО	0,025	5500	(-2, 4)	100	Хи-квадрат распределение с числом степеней свободы $k=7$
<u>15</u>	Уральский ФО	0,03	2700	(3, 4)	55	F-распределение с числом степеней свободы $k_1=16$ и $k_2=17$
<u>14</u>	Сибирский ФО	0,035	3200	(3, 4)	66	Логистическое с параметрами масштаба и сдвига 5 и 2 соответственно.
<u>13</u>	Дальневосточный ФО	0,01	4400	(0, 3)	100	F-распределение с числом степеней свободы $k_1=5$ и $k_2=22$
<u>12</u>	Северо-Кавказский ФО	0,02	5600	(-1,5, 3)	70	Распределение Стьюдента с

						числом степеней свободы $k = 9$
<u>11</u>	Центральный ФО	0,015	5900	(1,5, 3)	80	Релея с модой, равной 10.
<u>10</u>	Южный ФО	0,025	3300	(-2, 4)	60	Хи-квадрат распределение с числом степеней свободы $k = 19$
<u>9</u>	Приволжский ФО	0,03	4600	(0, 4)	58	F-распределение с числом степеней свободы $k_1 = 11$ и $k_2 = 10$.
<u>8</u>	Уральский ФО	0,035	5400	(0,2)	68	Показательное распределение с математическим ожиданием, равным 100.
<u>7</u>	Сибирский ФО	0,01	6000	(2, 1)	73	Равномерное распределение на отрезке (1,2)
<u>6</u>	Дальневосточный ФО	0,02	6500	(-2, 1)	45	Распределение Стьюдента с числом степеней свободы $k = 16$
<u>5</u>	Северо-Кавказский ФО	0,015	2700	(3, 1)	56	Релея с модой, равной 12.
<u>4</u>	Центральный ФО	0,025	3600	(-3, 1)	67	Логистическое с параметрами масштаба и сдвига 6 и 2 соответственно.
<u>3</u>	Южный ФО	0,009	4800	(4, 2)	78	Хи-квадрат распределение с числом степеней свободы $k = 13$
<u>2</u>	Приволжский ФО	0,01	5100	(-4, 2)	87	F-распределение с числом степеней свободы $k_1 = 8$ и $k_2 = 22$
<u>1</u>	Сибирский ФО	0,015	4800	(3, 5)	66	Релея с модой, равной 17.