

# Домашнее задание к семинару 20 ноября

Ластовецкий Дмитрий Александрович

20 ноября 2020 г.

## Задание 1.

$$1. P(\chi_1^2 \geq 4.45) = 1 - P(\chi_1^2 < 4.45) = 1 - (\Phi(\sqrt{4.45}) - \Phi(-\sqrt{4.45})) = 1 - (\Phi(\sqrt{4.45}) - (1 - \Phi(\sqrt{4.45}))) = 2 - 2\Phi(\sqrt{4.45}) \approx 2 - 2\Phi(2.11) = 0.0348$$

$$2. P(94 \geq \chi_{110}^2 \geq 98) = P(\chi_{110}^2 \leq 98) - P(\chi_{110}^2 \leq 94) \rightarrow \Phi\left(\frac{98-110}{\sqrt{110}}\right) - \Phi\left(\frac{94-110}{\sqrt{110}}\right) \approx \Phi(-1.14) - \Phi(-1.53) = 0.1271 - 0.0630 = 0.0641$$

$$3. P(\chi_{250}^2 \leq X) = 0.47 \rightarrow \Phi\left(\frac{X-250}{\sqrt{500}}\right) = 0.47 \Leftrightarrow \Phi\left(\frac{X-250}{\sqrt{500}}\right) = \Phi(-0.08) \Leftrightarrow \frac{X-250}{\sqrt{500}} = -0.08 \Leftrightarrow X = (-0.08 * \sqrt{500} + 250) \Leftrightarrow X \approx 248.21$$

## Задание 2.

Дана следующая выборка из распределения сл.в. X: -2.24, 7.69, 4.19, -7.07, -2.14, 16.37, 17.79. Найдите 98%-ый доверительный интервал для дисперсии данной случайной величины в предположении о нормальном законе распределения. Запишите общий вид доверительного интервала. Укажите уровень квантилей и количество степеней свободы, которые вы использовали для построения доверительного интервала. Проинтерпретируйте полученный доверительный интервал.

Общий вид доверительного интервала для дисперсии при неизвестном среднем согласно теореме Фишера для нормальных выборок:

$$P\left(\frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}, n-1}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}, n-1}^2}\right) = 1 - \alpha$$

$$\alpha = 0.02$$

Выборочное среднее равно  $\bar{x} \approx 4.94$

Для расчета выборочной дисперсии составим следующую таблицу:

Значение сл.в. $x$	$x - \bar{x}$	$(x - \bar{x})^2$
-2.24	-7.18	51.5524
7.69	2.75	7.5625
4.19	-0.75	0.5625
-7.07	-12.01	144.2401
-2.14	-7.11	50.5521
16.37	11.43	130.6449
17.79	12.85	165.1225

$$\sum_1^n (x - \bar{x})^2 = 550.237$$

$$S^2 = \frac{\sum_1^n (x - \bar{x})^2}{n-1} = \frac{550.237}{6} = 91.7$$

Количество степеней свободы равно  $n - 1 = 7 - 1 = 6$

Уровень первого квантиля для распределения хи-квадрат  $1 - \frac{\alpha}{2} = 0.99$

Уровень второго квантиля для распределения хи-квадрат  $\frac{\alpha}{2} = 0.01$

Соответственно, доверительный интервал имеет вид:

$$\left[ \frac{6*91.7}{\chi_{0.99}^2(6)}, \frac{6*91.7}{\chi_{0.01}^2(6)} \right],$$

При упрощении и подстановке:

$$[32.72683; 630.8979]$$

Интерпретация: с уверенностью в 98% мы можем говорить о том, что дисперсия генеральной совокупности (истинная дисперсия) находится в интервале  $[32.72683; 630.8979]$ . Доверительный интервал широкий, что может означать неоднородность выборки.

### Задание 3.

Без использования R проверьте гипотезу о независимости признаков A и B на основе следующей таблицы сопряженности:

A/B	B1	B2	sum
A1	105	46	151
A2	30	200	230
sum	135	246	total sum = 381

Проинтерпретируйте полученное значение p-value. Сделайте вывод об отвержении или не отвержении нулевой гипотезы о независимости признаков. Рассчитайте и проинтерпретируйте стандартизированные остатки.

$$H_0 : p_{ij} = p_i * p_j$$

$$H_1 : p_{ij} \neq p_i * p_j$$

Ожидаемые значения при независимости признаков:

$$\hat{n}_{11} = 151 * 135 : 381 \approx 53.5$$

$$\hat{n}_{12} = 151 * 246 : 381 \approx 97.5$$

$$\hat{n}_{21} = 230 * 135 : 381 \approx 81.5$$

$$\hat{n}_{22} = 230 * 246 : 381 \approx 148.5$$

$$\chi_1^2 = \frac{(105-53.5)^2}{53.5} + \frac{(46-97.5)^2}{97.5} + \frac{(30-81.5)^2}{81.5} + \frac{(200-148.5)^2}{148.5} \approx 127.1801$$

$$p_{value} = P(\chi_1^2 \geq 127.1801) \rightarrow 0$$

$p_{value} < 0.05$ , мы можем отвергнуть нулевую гипотезу.

Расчет стандартизированных остатков:

$$\frac{(105-53.5)}{\sqrt{53.5}} = 7.04$$

$$\frac{(46-97.5)}{\sqrt{97.5}} = -5.1$$

$$\frac{(30-81.5)}{\sqrt{81.5}} = -5.7$$

$$\frac{(200-148.5)}{\sqrt{148.5}} = 4.2$$