

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра МОЭВМ

ОТЧЕТ
по лабораторной работе №3
по дисциплине «Методы статистической обработки данных»
Тема: Обработка выборочных данных. Нахождение интервальных
оценок параметров распределения. Проверка статистической гипотезы
о нормальном законе распределения.

Студент гр. 5381

Лянгузов А. А.

Преподаватель

Середа В.И.

Санкт-Петербург

2019

Цель работы.

Получение практических навыков вычисления интервальных статистических оценок параметров распределения выборочных данных и проверки «справедливости» статистических гипотез.

Основные теоретические положения.

Доверительный интервал – интервал, который с заданной надежностью γ покрывает заданный параметр.

Доверительный интервал для математического ожидания при неизвестном СКО σ вычисляется по формуле (1):

$$\bar{x}_g - \frac{S}{\sqrt{n}} t_\gamma < a < \bar{x}_g + \frac{S}{\sqrt{n}} t_\gamma, \text{ где} \quad (1)$$

a – оцениваемый параметр (математическое ожидание), \bar{x}_g – выборочное среднее, S – исправленное СКО, n – объем выборки, t_γ – табличное значение.

Точность оценки вычисляется по формуле (2):

$$\delta = \frac{S}{\sqrt{n}} t_\gamma \quad (2)$$

Доверительный интервал для среднеквадратического отклонения вычисляется по формуле (3):

$$S(1 - q) < \sigma < S(1 + q), \text{ где} \quad (3)$$

σ – интервальная оценка среднеквадратического отклонения, S – исправленное СКО, q – табличное значение.

Критерий согласия Пирсона (χ^2) применяют для проверки гипотезы о соответствии эмпирического распределения предполагаемому теоретическому распределению.

Критерий Пирсона представлен в формуле (4):

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - n'_i)^2}{n'_i}, \text{ где} \quad (4)$$

k – номер интервала, n_i – интервальные частоты, n'_i – теоритические частоты для нормального распределения.

Если $\chi^2 > \chi_{кр}^2$, где $\chi_{кр}^2$ – табличное значение, то гипотеза отвергнута.

Примечание: при выполнении работы, некоторые из вышеприведенных формул не были использованы. Вместо них использовались альтернативные подходы, ввиду особенностей (или личного опыта) работы с языком R.

Задание.

Для заданной надежности определить (на основании выборочных данных и результатов выполнения лабораторной работы №2) границы доверительных интервалов для математического ожидания и среднеквадратического отклонения случайной величины. Проверить гипотезу о нормальном распределении исследуемой случайной величины с помощью критерия Пирсона χ^2 . Дать содержательную интерпретацию полученным результатам.

Ход выполнения.

I. Интервальные оценки для мат.ожидания.

Зададим надежность оценки $\gamma = 0.95$.

Далее вычислим значение t_γ (точность) как квантиль распределения Стьюдента с $N-1$ степенью свободы, где N – объем выборки:

$$\mathbb{P} \left(\left| \frac{(\bar{x}_B - a)\sqrt{N}}{S} \right| \leq t_\gamma \right) = \gamma$$

Поскольку распределение Стьюдента симметричное и оценка двусторонняя, вычисляем $(1+\gamma)/2$ (или $-(1-\gamma)/2$) квантиль.

При объеме выборки в 107 измерений, $t_\gamma = 1.983$.

Далее доверительные интервалы строятся по формулам:

$$a \in \left(\bar{x}_B - \frac{S}{\sqrt{N}} t_\gamma; \bar{x}_B + \frac{S}{\sqrt{N}} t_\gamma \right).$$

1. Для величины v .

В ходе предыдущей лабораторной работы были получены точечные оценки мат.ожидания \bar{x}_B и СКВО S для величины v :

$$\bar{x}_v = 449.74299065.$$

$$S = 54.04939061.$$

Тогда доверительный интервал:

$$\bar{x}_v \in [439; 460].$$

2. Для величины E .

Аналогично.

$$\bar{x}_e = 127.20133511$$

$$S = 21.88787065$$

Тогда доверительный интервал:

$$\bar{x}_E \in [123; 131].$$

II. Интервальные оценки для среднеквадратического отклонения.

Определим надежность оценки $\gamma = 0.95$.

Формулы для вычисления доверительного интервала для СКВО:

$$a \in \left(\sqrt{\frac{(N-1)D_g}{\chi_{N-1}^2\left(\frac{1+\gamma}{2}\right)}}; \sqrt{\frac{(N-1)D_g}{\chi_{N-1}^2\left(\frac{1-\gamma}{2}\right)}} \right).$$

$\chi_{N-1}^2(r)$ – это r -квантиль распределения хи-квадрат со степенью свободы $N-1$;

D_g – исправленная выборочная дисперсия;

N – объем выборки.

(Замечание: сделано так, потому что это в языке R есть поддержка автоматизации вычисления r -квантили)

$$\chi_{N-1}^2\left(\frac{1+\gamma}{2}\right) = 136.4;$$

$$\chi_{N-1}^2\left(\frac{1-\gamma}{2}\right) = 79.4.$$

1. Для величины v .

В ходе предыдущей лабораторной работы была получена точечная оценка исправленной дисперсии для величины v :

$$D_g = 2921.33662493;$$

$$S = 54.04939061.$$

Тогда доверительный интервал:

$$S_v \in [47.7; 62.4].$$

2. Для величины E .

Аналогично для величины E :

$$D_g = 479.07888153;$$

$$S = 21.88787065.$$

Тогда доверительный интервал:

$$\bar{x}_E \in [19.3; 25.3].$$

Примечание: доверительные интервалы представляют собой замкнутые интервалы так как использовалось определение функции вероятности с нестрогим неравенством: $\mathbb{P}(X \leq x) = \alpha$.

III. Проверка простой гипотезы о нормальном распределении с помощью критерия Пирсона χ^2 .

Нулевая гипотеза H_0 : $v_1, v_2, \dots, v_N \sim \mathcal{N}(\bar{x}_e, D_e)$ – для величины v .

Нулевая гипотеза H_0 : $e_1, e_2, \dots, e_N \sim \mathcal{N}(\bar{x}_e, D_e)$ – для величины E .

Необходимо вычислить значение хи-квадрат для имеющегося интервального ряда и сравнить его с критическим значением.

Критическое значение полагается равным квантили распределения хи-квадрат с $(K-3)$ степенями свободы, где K – количество интервалов в интервальном ряду.

Если вычисленное по интервальному ряду значение превосходит критическое, то гипотеза отклоняется на данном уровне значимости.

Выберем уровень значимости $\alpha = 0.05$.

Тогда критическое значение:

$$\chi_{кр}^2 = \chi_{K-3}^2(1 - \alpha) = \chi_4^2(0.95) = 9.487729$$

Для вычисления интервального хи-квадрат составим таблицу.

1. Для величины v .

F_{th} – функция распределения для нормального закона с мат. ожиданием $\bar{x}_e = 449.74299065$ и исправленной дисперсией $D_e = 2921.33662493$.

Пятый и шестой столбец в таблице содержат соответствующие значения функции распределения в точках левой (leftB) и правой (rightB) границ интервалов (значения были получены с помощью встроенной функции `pnorm` в языке R).

Седьмой столбец таблицы содержит разницу между значениями шестой и пятой столбцов, а именно вероятность попадания в данный интервал.

Последний столбец вычисляется по формуле

$$n'_i = \left(F_{th}(\text{right}B_i) - F_{th}(\text{left}B_i) \right) * N,$$

где N – объем выборки, i – номер интервала.

Данный столбец содержит искомые теоретические частоты, то есть показывает, сколько «попаданий» (абсолютная частота) элементов выборки объема N должно быть в данном интервале,

если элементы выборки действительно распределены по нормальному закону с заданными параметрами.

Интервалы	Средины	Частоты абс.	Частоты отн.	F _{thLB}	F _{thRB}	F _{thRB} - F _{thLB}	Теор. частоты
[340; 375)	357,5	12	0.11214 953	0.00000 000	0.08335 319	0.08335 319	8.91879 1
[375; 410)	392,5	12	0.11214 953	0.08335 319	0.23107 574	0.14772 255	15.8063 13
[410; 445)	427,5	27	0.25233 645	0.23107 574	0.46503 654	0.23396 080	25.0338 06
[445; 480)	462,5	23	0.21495 327	0.46503 654	0.71219 308	0.24715 654	26.4457 50
[480; 515)	497,5	21	0.19626 168	0.71219 308	0.88635 301	0.17415 993	18.6351 12
[515; 550)	532,5	9	0.08411 215	0.88635 301	0.96819 580	0.08184 279	8.75717 8
[550; 585)	567,5	3	0.02803 738	0.96819 580	1.00000 000	0.03180 420	3.40305 0

Тогда значение хи-квадрат:

$$\chi^2 = \sum_i^K \frac{(freq_i - n'_i)^2}{n'_i}$$

$$\chi^2 = 2.939051$$

Полученное значение хи-квадрат меньше критического ($X^2_{кр} = 9.487729$), следовательно нулевая гипотеза принимается при уровне значимости $\alpha = 0.05$.

2. Для величины E.

Интервалы	Средины	Частоты абс.	Частоты отн.	F _{thLB}	F _{thRB}	F _{thRB} - F _{thLB}	Теор. частоты
[71,9,87)	79.45714	5	0.04672897	0.00000000	0.03317572	0.03317572	3.549802
[87,102)	94.57143	8	0.07476636	0.03317572	0.12599908	0.09282336	9.932099
[102,117)	109.68571	19	0.17757009	0.12599908	0.32456289	0.19856381	21.246327
[117,132)	124.80000	32	0.29906542	0.32456289	0.59311119	0.26854830	28.734668
[132,147)	139.91429	26	0.24299065	0.59311119	0.82279986	0.22968868	24.576689
[147,163)	155.02857	10	0.09345794	0.82279986	0.94701988	0.12422002	13.291542
[163,178]	170.14286	7	0.06542056	0.94701988	1.00000000	0.05298012	5.668873

Тогда значение хи-квадрат:

$$\chi^2 = \sum_i^K \frac{(freq_i - n'_i)^2}{n'_i}$$

$$\chi^2 = 2.786982$$

Полученное значение хи-квадрат меньше критического ($\chi^2_{кр} = 9.487729$), следовательно нулевая гипотеза принимается при уровне значимости $\alpha = 0.05$.

Выводы.

В работе были найдены границы доверительных интервалов для математического ожидания и среднеквадратического отклонения случайной величины для надежности $\gamma = 0,95$. Доверительный интервал для математического ожидания $\bar{x}_v \in [439; 460]$ и $\bar{x}_E \in [123; 131]$ покрывает значение математического ожидания $\bar{x}_v = 449.74299065$ и $\bar{x}_E = 127.20133511$ соответственно, те же выводы можно сделать и для доверительного интервала СКО. Проведенная оценка статистической гипотезы посредством критерия Пирсона принята для обеих исследуемых величин при уровне значимости $\alpha = 0,05$