УДК 658.382

Н. А. Сорокин, канд. техн. наук, доцент, БГТУ

МАТЕМАТИКО-СТАТИСТИЧЕСКИЙ АНАЛИЗ ТРАВМАТИЗМА – РЕШЕНИЕ ЗАДАЧ

Simple comparison of indicators of a traumatism frequently does not give the answer to a question on the importance of their distinction.

The decision of problems of the analysis of a traumatism methods of mathematical statistics - an estimation of influence of age and the work experience on level of a traumatism, a deviation of number of accidents from their average value (at various o'clock of change, change, number of months, days of week, days of biological rhythms, months, seasons), comparison of several averages (at the various enterprises, for kinds of works, etc.) is offered.

Such analysis allows to receive with high degree of reliability authentic results about the importance of distinctions of average indexes of an industrial traumatism to various signs and it is more rational to plan an expenditure of means for its preventive maintenance.

Введение. При анализе травматизма возникают задачи сравнения нескольких средних показателей для оценки их различия. Пути решения таких задач методами математической статистики сформулированы в [1].

В качестве количественной характеристики уровня производственного травматизма используются коэффициенты частоты, тяжести и другие, а также (в тех случаях, когда число работников по сравниваемым признакам одинаково) число несчастных случаев.

Основная часть. Отдельные члены выборки называются вариантами x_i, а число членов, входящих в выборку, носит название объема выборки *n*.

Важнейшей характеристикой выборки являются среднее арифметическое или, сокращенно, среднее

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1)$$

и выборочная дисперсия, которая при $n \le 30$ определяется по формуле

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2, , \qquad (2)$$

а при $n > 30$ по формуле

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 . \qquad (3)$$

Задачи решаются при уровне значимости $\alpha = 0{,}05$ и мощности критерия $\beta = 0{,}1$.

Для оценки отклонения показателей травматизма (у различных предприятий, по видам работ и т. п.) или числа несчастных случаев от их среднего значения (в различные месяцы, числа месяцев, дни недели, часы смены, дни биологических ритмов) вычисляется среднее значение выборки $\bar{x}$. Формулируется нулевая гипотеза $H_0 : \bar{x} = x_i$ о том, что отклонения незначимы, и альтернативная $H_1 : \bar{x} \neq x_i$ о том, что они существенны.

Вычисляется статистика

$$X^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\bar{x}}. \qquad (4)$$

Определяется число степеней свободы

$$\nu = n - m , \qquad (5)$$

где m – число параметров, которое вычисляется при выполняемой проверке (в данном случае $m = 1$).

Далее по статистическим таблицам, например [2], определяется значение критерия Пирсона $\chi^2_{\nu;1-\alpha}$.

Нулевая гипотеза принимается, если

$$X^2 \le \chi^2_{\nu;\,0,95} \qquad (6)$$

(отклонения числа несчастных случаев от их среднего значения носят случайный характер). В противном случае она отвергается в пользу альтернативной, то есть отклонения значимы.

Выбор критерия при сравнении средних выборочных совокупностей у различных предприятий, по видам работ, профессиям, оборудованию и т. п. определяется видом распределения исходных данных, вследствие чего необходима предварительная проверка каждой выборки на нормальность распределения.

При объеме выборки $n > 50$ для проверки принадлежности ее членов к нормальному распределению предпочтительным является один из критериев: χ^2 Пирсона или ω^2 Мизеса – Смирнова.

При объеме выборки $50 \ge n > 15$ для проверки принадлежности ее членов к нормальному распределению предпочтительным является составной критерий [3].

При объеме выборки $n \le 15$ проверку нормальности распределения выборочных данных можно проводить по оценке асимметрии и эксцесса. Такой объем выборки сравнительно часто встречается при анализе данных о травматизме. Порядок проверки при этом следующий.

Выдвигается нулевая гипотеза о том, что выборочная совокупность подчиняется нормальному закону распределения, и альтернативная – о том, что члены выборки к нему не принадлежат.

Вычисляется асимметрия

$$A = \frac{1}{n\,S^3} \sum_{i=1}^{n} (x_i - \bar{x})^3 , \qquad (7)$$

где $S = \sqrt{S^2}$ – среднеквадратичное отклонение выборочной дисперсии; эксцесс

$$E = \frac{1}{n\,S^4} \sum_{i=1}^{n} (x_i - \bar{x})^3 - 3 . \qquad (8)$$

Затем определяются математические ожидания дисперсий этих величин:

$$D\,(A) = \frac{6\,(n-1)}{(n+1)\,(n+3)} ; \qquad (9)$$

$$D\,(E) = \frac{24\,n\,(n-2)\,(n-3)}{(n+1)^2\,(n+3)\,(n+5)} . \qquad (10)$$

И в заключение проверяются неравенства

$$|A| \le 3\sqrt{D\,(A)} ; \qquad (11)$$

$$|E| \le 5\sqrt{D\,(E)} . \qquad (12)$$

Если эти неравенства выполняются, то распределение данных соответствует нормальному закону; если хотя бы одно из них не выполняется, то распределение данных отличается от нормального.

При сравнении нескольких средних показателей травматизма важно не только установить равенство или различие нескольких средних совокупностей, но и определить, по каким дескрипторам средние показатели значимо больше или меньше общего среднего показателя.

При этом желательно применять параметрический критерий [1]. Однако область его использования ограничена условиями нормальности распределения выборочных совокупностей и равенством их дисперсий.

В других случаях следует применять непараметрические критерии.

Нулевая гипотеза о равенстве дисперсий нескольких средних нормально распределенных совокупностей проверяется следующим образом.

Сумма чисел степеней свободы

$$\nu = \sum_{j=1}^{k} \nu_j , \qquad (13)$$

где k – число выборок; $\nu_j = n_j - 1$ – число степеней свободы, относящееся к j-й выборке, n_j – объем j-й выборки.

Средневзвешенная дисперсия

$$S_p^2 = \frac{\sum_{j=1}^{k} \nu_j\, S_j^2}{\nu} , \qquad (14)$$

где S_j^2 – дисперсия j-й выборки.

Статистика критерия

$$\chi^{2*} = \frac{\nu \ln S_p^2 - \sum_{j=1}^{k} \nu_j \ln S_p^2}{1 + \dfrac{\sum_{j=1}^{k} \dfrac{1}{\nu_j} - \dfrac{1}{\nu}}{3\,(k-1)}} . \qquad (15)$$

Эта статистика – предложенная Бартлеттом модификация критерия, разработанная Нейманом и Пирсоном.

Далее по статистическим таблицам определяется табличное значение критерия Пирсона $\chi^2_{k-1;p}$ и проверяется неравенство вида

$$\chi^{2*} \le \chi^2_{k-1;\,0{,}95} . \qquad (16)$$

Если неравенство выполняется, то нулевая гипотеза принимается (дисперсии равны), в противном случае – отклоняется.

Метод сравнения нескольких средних при нормальном распределении выборочных данных и равных дисперсиях зависит от того, равны объемы выборок или нет. От этого зависит, в свою очередь, способ определения так называемых границ регулирования.

Верхняя и нижняя границы регулирования определяются относительно общего среднего показателя травматизма. В случае сравнения нескольких средних при равных объемах выборки для каждой из них определяется размах. Размах j-й выборки

$$R_j = x_{j\,\max} - x_{j\,\min} , \qquad (17)$$

где $x_{j\,\max}$ – максимальное значение показателя в j-й выборке; $x_{j\,\min}$ – минимальное значение показателя в j-й выборке.

Средний размах

$$\bar{R} = \frac{\sum_{j=1}^{k} R_j}{k} . \qquad (18)$$

Общее среднее показателей

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k} \bar{x}_j}{k} , \qquad (19)$$

где $\bar{x}_j$ – среднее значение показателя по j-й выборке.

Верхняя граница регулирования

$$\text{ВГР} = \bar{\bar{x}} + 0{,}65\ \text{A}_2\ \bar{R}\,, \qquad (20)$$

где A_2 – величина, зависящая от объема выборок [2].

Нижняя граница регулирования

$$\text{НГР} = \bar{\bar{x}} - 0{,}65\ \text{A}_2\ \bar{R}\,. \qquad (21)$$

Для сравнения нескольких средних при неравных объемах выборок сначала необходимо вычислить общее среднее

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k} n_j\ \bar{x}_j}{\sum_{j=1}^{k} n_j}\,. \qquad (22)$$

Средневзвешенная дисперсия

$$\text{S}_p^2 = \frac{\sum_{j=1}^{k} (n_j - 1)\ \text{S}_j^2}{\sum_{j=1}^{k} (n_j - 1)}\,. \qquad (23)$$

Границы регулирования в данном случае меняются от выборки к выборке.

Верхняя граница регулирования для 5%-го уровня значимости

$$\text{ВГР}_j = \bar{\bar{x}} + 0{,}65\ \text{A}_j\ S_p\,, \qquad (24)$$

где A_j – величина А для j-й выборки [3].

Нижняя граница регулирования

$$\text{НГР}_j = \bar{\bar{x}} - 0{,}65\ \text{A}_j\ S_p\,. \qquad (25)$$

Средние показатели, которые оказываются внутри границ регулирования, признаются незначимо отличающимися от общего среднего.

Сравнение нескольких средних при распределении выборочных данных, отличном от нормального, или при неравных дисперсиях производится следующим образом.

Варианты всех выборок сводятся в одну совокупность, где располагаются в порядке убывания и соответственно подчеркиваются: относящиеся к первой выборке – одной чертой, ко второй – двумя и т. д. Затем элементы полученной совокупности заменяются по порядку так называемыми рангами: 1, 2, …, N и таким же образом подчеркиваются. Если встречаются одинаковые значения членов совокупности, то им присваиваются среднеарифметические значения рангов. Затем определяются средние значения рангов по каждой выборке:

$$\bar{x}_{ji} = \frac{\sum_{i=1}^{n_j} x_{ji}}{n_j}\,, \qquad (26)$$

где x_{ji} – i-й ранг в j-й выборке.

Далее подсчитывается статистика критерия

$$\chi^{2**} = \frac{\sum_{j=1}^{k} n_j\,[\bar{x}_j - \frac{1}{2}(N+1)]^2}{\frac{1}{12}(N^2 - 1)}\,. \qquad (27)$$

Затем определяется табличное значение критерия Пирсона $\chi^2_{k-1;\,p}$ и проверяется неравенство

$$\chi^{2**} \le \frac{N}{N-1}\,\chi^2_{k-1;\,0{,}95}\,. \qquad (28)$$

Если оно выполняется, то гипотеза о равенстве средних принимается, в противном случае отвергается в пользу альтернативной.

Если средние неравны, то отбрасывается среднее с максимальным отклонением от общего среднего и все расчеты повторяются. Эта операция продолжается до тех пор, пока неравенство не будет выполняться.

Заключение. Показатели производственного травматизма при простом сравнении их средних значений времени чаще всего не дают однозначного ответа на вопрос о значимости их различия. Такая задача с высокой степенью надежности может быть решена методами математико-статистического анализа, что позволяет более рационально использовать средства на предупреждение несчастных случаев.

Литература

1. Сорокин, Н. А. Математико-статисти- ческий анализ травматизма – постановка задачи / Н. А. Сорокин // Труды БГТУ. Сер. II, Лесная и деревообраб. пром-сть. – 2008. – Вып. XVI. – С. 345–347.

2. Джонсон, Н. Статистика и планирование эксперимента в технике и науке / Н. Джонсон, Ф. Лион. – М.: Мир, 1980. – 610 с.

3. Государственная система обеспечения единства измерений. Прямые измерения с многократными наблюдениями. Методы обработки результатов наблюдений. Основные положения: ГОСТ 8.207–76. – Введ. 01.01.1977. – М.: Стандартинформ, 2006. – 7 с.