УДК 658.382

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

МАТЕМАТИКО-СТАТИСТИЧЕСКИЙ АНАЛИЗ ТРАВМАТИЗМА – СРАВНЕНИЕ ДВУХ СРЕДНИХ ПОКАЗАТЕЛЕЙ

Простое сравнение двух средних показателей травматизма обычно не дает ответа на вопрос о значимости их различия. Предлагается решение этой задачи анализа травматизма (сравнение двух средних показателей травматизма у двух организаций, в различные смены, сезоны, по признаку пола, по видам работ, оборудования и т.д.) методами математической статистики. Такой анализ позволяет с высокой степенью надежности получать достоверные результаты о значимости различий двух средних показателей производственного травматизма по различным признакам. С его помощью можно определить наиболее существенные факторы, влияющие на уровень травматизма, и наиболее эффективно планировать расходование средств на его профилактику.

Simple comparison of two average indexes of a traumatism usually does not give the answer to a question on the importance of their distinction. The decision of this problem of the analysis of a traumatism (comparison of two average indexes of a traumatism at two organisations, in various changes, seasons, on the basis of a floor, by kinds of works, the equipment etc.) methods of mathematical statistics is offered. Such analysis allows to receive with high degree of reliability authentic results about the importance of distinctions of two average indexes of an industrial traumatism to various signs. With its help it is possible to define the most essential factors influencing level of a traumatism and most effectively to plan an expenditure of means for its preventive maintenance.

Введение. При анализе травматизма возникают задачи сравнения двух средних показателей для оценки их различия: среднего по республике и по отрасли, среднего по отрасли и конкретного предприятия, средних по структурным подразделениям, видам работ, оборудования, в различные смены, по признаку пола и т. д. Пути решения таких задач методами математической статистики сформулированы в публикации [1].

Основная часть. Выбор критерия при сравнении средних двух совокупностей определяется видом распределения исходных данных, вследствие чего необходима предварительная проверка каждой совокупности на нормальность распределения [2].

При сравнении двух средних при нормальном распределении выборочных данных следует различать два случая. Первый, когда обе выборочные совокупности друг от друга не зависят, и второй, когда значения варианта одной из выборок зависят от значений варианта другой, или когда варианты выборок связаны попарно.

При сравнении двух средних независимых выборочных совокупностей задача решается следующим образом (рисунок).

Вычисляется число степеней свободы меньшей из выборок:

$$\nu_{\min} = n_{\min} - 1, \tag{1}$$

где $n_{\min}$ – объем меньшей из выборок.

Определяется минимально необходимое число наблюдений для сравнения дисперсий двух совокупностей по статистическим таблицам, например [3]. Если это число больше $n_{\min}$, то рассматривается вопрос либо об увеличении выборки, либо о применении непараметрического критерия.

В противном случае проверяется гипотеза о равенстве дисперсий. Вычисляются обе дисперсии и находится отношение

$$\frac{S_1^2}{S_2^2} \geq 1. \tag{2}$$

Затем подсчитывается число степеней свободы первой выборки:

$$\nu_1 = n_1 - 1, \tag{3}$$

где n_1 – объем первой выборки; и второй:

$$\nu_2 = n_2 - 1, \tag{4}$$

где n_2 – объем второй выборки.

Далее по статистическим таблицам, например [3], определяется табличное значение критерия Фишера и проверяется неравенство вида

$$\frac{S_1^2}{S_2^2} \leq F_{\nu_1;\,\nu_2;\,0,95}. \tag{5}$$

Если это выражение выполняется, то гипотеза о равенстве дисперсий двух совокупностей принимается.

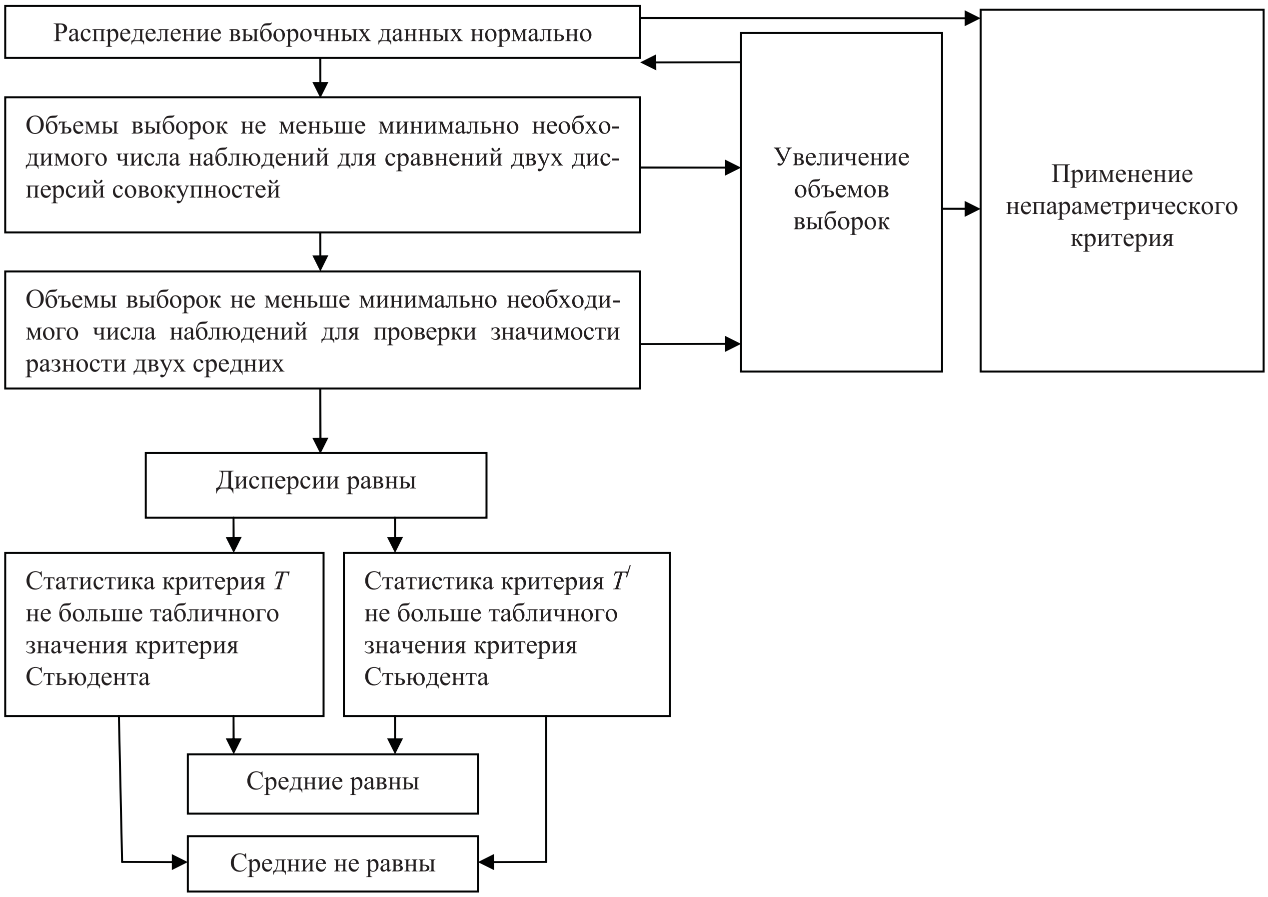
В дальнейшем вычисляется средневзвешенная дисперсия

$$S_p^2 = \frac{\nu_1 S_1^2 + \nu_2 S_2^2}{\nu_1 + \nu_2}, \tag{6}$$

затем величина

$$D = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{S_p}, \tag{7}$$

где $\bar{X}_1$ и $\bar{X}_2$ – средние соответственно первой и второй выборок.



Принципиальная схема сравнения двух средних при независимых выборках

Далее находится минимально необходимое число наблюдений для проверки значимости разности двух средних, например [3], и сравнивается с объемом выборки $n_{\min}$. Если это число больше значения $n_{\min}$, то рассматривается вопрос об увеличении объема выборки или применяется непараметрический критерий.

Если же минимально необходимое число наблюдений меньше или равно значению $n_{\min}$, то в случае равенства дисперсий вычисляется статистика критерия

$$T = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{S_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \tag{8}$$

и определяется число степеней свободы

$$\nu = n_1 + n_2 - 2 . \tag{9}$$

Если дисперсии неравны, то статистика критерия определяется по следующей формуле:

$$T^{/} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} , \tag{10}$$

а число степеней свободы

$$\nu = \frac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{S_2^2}{n_2}\right)^2}{n_2 - 1}} . \tag{11}$$

В заключение в обоих случаях определяется по статистическим таблицам, например [3], значение критерия Стьюдента $t_{\nu;\,0,95}$ и проверяется неравенство вида

$$T \le t_{\nu;\,0,95} \tag{12}$$

или

$$T^{/} \le t_{\nu;\,0,95} . \tag{13}$$

Если оно выполняется, то средние значения показателей равны между собой, а их отличие обусловлено случайными колебаниями. Если неравенство не выполняется, то средние различаются значимо.

Задача сравнения двух средних при парных или коррелированных выборках возникает в тех случаях, когда значения варианта в выборках связаны попарно или значения варианта одной из выборок зависят от значений варианта другой. Объемы выборок должны быть равны, в противном случае следует использовать непараметрический критерий.

Сначала определяются разности между параллельными наблюдениями

$$\Delta_i = x_{1i} - x_{2i}\ , \tag{14}$$

где x_{1i} – i-тый член первой выборки; x_{2i} – i-тый член второй выборки.

Затем находится средняя разность между параллельными наблюдениями

$$\overline{\Delta} = \frac{1}{n}\sum_{i=1}^{n}\Delta_i\ . \tag{15}$$

Вычисляется дисперсия

$$S_{\overline{\Delta}}^2 = \frac{1}{n\,(n-1)}\sum_{i=1}^{n}(\Delta_i - \overline{\Delta})^2\ . \tag{16}$$

Определяется статистика критерия

$$T^{//} = \frac{\overline{\Delta}}{S_{\overline{\Delta}}}\ . \tag{17}$$

Находится табличное значение критерия Стьюдента $t_{n-1;\,0,95}$.

Средние равны, если выполняется неравенство вида

$$\left|T^{//}\right| \le t_{n-1;\,0,95}\ . \tag{18}$$

Задача сравнения двух средних при распределении выборочных данных, отличном от нормального, предполагает применение критерия, свободного от распределения. Одним из них является критерий, основанный на так называемых нормальных метках. Он обладает достаточно хорошей чувствительностью, т. е. его относительная мощность оказывается выше, чем можно было бы ожидать, с учетом того, что при его использовании приходится пренебрегать частью информации.

Методика применения критерия следующая. По статистическим таблицам, например [3], определяется по значению $N = n_1 + n_2$ столбец нормальных меток и все члены обоих выборок заменяются метками. Причем член выборок с максимальным значением заменяется первой меткой, ближайший к нему – второй и т. д. в порядке убывания. Метки берутся сначала сверху вниз, а затем снизу вверх, но уже с отрицательным знаком. Если при замене выясняется, что несколько членов совокупности имеют одинаковые значения, то они заменяются метками, определенными как среднеарифметическое табличных меток. Сумма всех меток равна нулю.

Далее подсчитывается сумма меток, относящихся к первой выборке $\sum x_{1i}$.

Затем все члены совокупности меток возводятся в квадрат и определяется их сумма $\sum\sum x_{ji}^2$.

Вычисляется статистика

$$U = \frac{N}{\sqrt{n_1\,n_2}}\,\frac{\left|\sum_{i=1}^{n1} x_{1i}\right|}{\sqrt{\sum_{j=1}^{2}\sum_{i=1}^{nj} x_{ji}^2}}\ . \tag{19}$$

По формуле (9) определяется число степеней свободы и находится табличное значение Стьюдента $t_{\nu;\,p}$.

В дальнейшем проверяется неравенство вида

$$U \le t_{\nu;\,0,95}\ . \tag{20}$$

Если оно выполняется, то средние равны, а их различие объясняется случайными колебаниями. В противном случае средние показатели отличаются значимо.

Заключение. Показатели производственного травматизма при простом сравнении их двух средних значений 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.

При этом необходимо стремиться к применению параметрических критериев, которые являются более сильными, т. е. дают более достоверные результаты. Но они рассчитаны на нормальное распределение исходных данных.

Если распределение исходных данных отлично от нормального, следует применять непараметрические критерии.

Предложенный анализ показателей травматизма позволяет достоверно определить различие двух средних показателей и более рационально использовать средства на предупреждение несчастных случаев.

Литература

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

2. Сорокин, Н. А. Математико-статистический анализ травматизма – решение задач / Н. А. Сорокин // Труды БГТУ. Сер. II, Лесная и деревообраб. пром-сть. – 2009. – Вып. XVII. – С. 285–287.

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

Поступила 01.04.2010