УДК 621.019

# КОНТРОЛЬ ЗНАЧИМОСТИ ПРИЗНАКОВ ПРИ ОЦЕНКЕ ИНДИВИДУАЛЬНОЙ НАДЕЖНОСТИ ЭНЕРГОБЛОКОВ

**ФАРХАДЗАДЕ Э.М., МУРАДАЛИЕВ А.З., ФАРЗАЛИЕВ Ю.З.**

*АзНИПИИ Энергетики, г. Баку*

Значимость признаков при классификации статистических данных определяется численными значениями ошибок первого и второго рода, риском ошибочного решения, оптимальной величиной критического значения статистики, характеризующей представительность выборки данных.

Проблема достоверной оценки технического состояния (ТС) объектов энергетики и, в частности, энергоблоков (ЭБ) ГРЭС, приобретает все большую актуальность. На ГРЭС, где срок службы оборудования превышает срок службы установленный стандартами, наблюдается увеличение числа отказов, обусловленных износом, увеличение затрат на плановые ремонты, снижение предельно допустимых нагрузок и ухудшение основных производственных показателей. Существенно возрастает значимость периодического контроля непревышения износа предельно допустимого значения, испытаний , позволяющих установить предельно допустимые нагрузки, метрологического контроля, точности измерения диагностических параметров.

Количественно ТС характеризуется показателями индивидуальной надежности (безотказности, ремонтопригодности, сохраняемости, долговечности).

Само понятие индивидуальность, по сути, означает, что конкретный ЭБ по своим показателям существенно отличается от остальных, в частности, аналогичных ЭБ. Чтобы оценить показатели надежности конкретного ЭБ, очевидно, из общей совокупности статистических данных нужно выбрать данные, относящиеся к этому ЭБ. Эти данные будем называть выборкой. Исходное предположение об отличии ТС ЭБ должно быть проверено, поскольку различие оценок показателей надежности ЭБ может быть случайным и приводить к неверным решениям. Иначе говоря, выборка должна быть проверена на представительность (репрезентативность).

В [1] предложен критерий непосредственной проверки выборки на представительность $\Delta_m$. Если расчетное значение расхождения между функциями распределения совокупности данных $F_M^*(\tau)$ и выборки $F_m^*(\tau)$, равное $\Delta_m^* > \Delta_m(\alpha_k)$, где $\alpha_k$ - уровень значимости, то с коэффициентом доверия $R = 1-\alpha_k$ можно утверждать, что выборка непредставительна, т.е. «индивидуальность» по заданному признаку или признакам проявляется достаточно часто. Известно, что, приняв или отвергнув предположение о представительности выборки, полученной в результате очередного этапа классификации данных, и мы можем совершить ошибки первого (α) и второго (β) рода.

Традиционно ошибки принятия исходной гипотезы задаются и, как правило, принимается, что α≤0,1 и β≤0,1. Например, для α и β рекомендуется значения 0,1 при применении метода последовательного анализа [2].

Расчетное значение $\Delta_m^*$ является одной из возможных реализаций расхождения распределения $F_M^*(\tau)$ и $F_m^*(\tau)$. В этом нетрудно убедиться, если многократно моделировать реализации $F_m^{**}(\tau)$, случайно различающихся от экспериментальной $F_m^{*}(\tau)$. В реальных условиях истина не известна, а принятие решения связано с обеими ошибками. Если, например, мы для некоторого критического значения $\Delta_{m,k}$, принимаем решение о непредставительности выборки, то при этом совершаем ошибку, вероятность которой равна $\alpha(\Delta_{m,k})$. Одновременно мы совершаем ошибку $\beta(\Delta_{m,k})$, которая также характеризует ошибочность решения. Риск ошибочного решения в общем случае рассчитывается по формуле

$$\gamma(\Delta_{m,k}) = A \cdot \alpha(\Delta_{m,k}) + B \cdot \beta(\Delta_{m,k}) \qquad (1)$$

где А и В – постоянные, характеризующие относительную значимость, соответственно, $\alpha(\Delta_{m,k})$ и $\beta(\Delta_{m,k})$, т.е. А+В=1, а $0 \le \gamma(\Delta_{m,k}) \le 1$

Если последствия ошибочного решения равнозначны, то А=В=0,5. Коэффициенты А и В могут быть вычислены, исходя из затрат, обусловленных ошибочным решением. Однако, маловероятно, что в момент принятия решения известны экономические последствия ошибочных решений. Мы воспользуемся наиболее простым и физическим объяснимым подходом, в соответствии с которым значимость ошибок первого и второго ряда пропорциональна числу данных. Чем больше данных, тем достоверность решения будет больше. Пусть выборка состоит из m случайных величин, а совокупность – из М. Тогда, А=m/М, а В=(М - m)/М.

На рис.1 приведены характерные кривые распределения $\alpha^*(\Delta_m)$ и $\beta^*(\Delta_b)$ при числе случайных величин выборки m=10, числе случайных величин совокупности данных М=50.

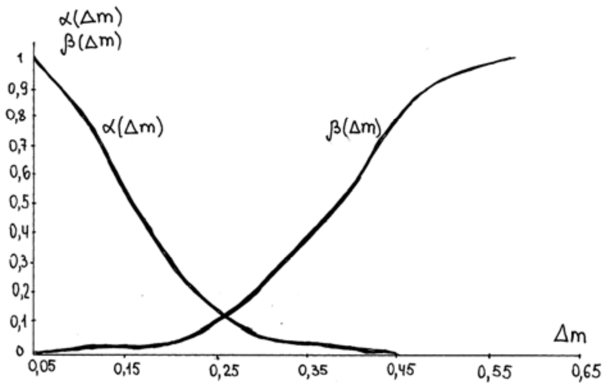


Рис.1 Графическая иллюстрация распределений ошибок первого $[\alpha(\Delta_m)]$ и второго рода $[\beta(\Delta_m)]$

Если обратиться к кривым $\alpha(\Delta_m)$ и $\beta(\Delta_m)$ (рис.1), то нельзя не заметить, что их сумма по мере роста $\Delta_m$ изменяется. Сначала она уменьшается, затем увеличивается. Очевидно, что некоторому дискретному значению $\Delta_m$ будет соответствовать минимальная ошибка принятия решения $\gamma(\Delta_m)$. Обозначим эту величину $\gamma(\Delta_{m,опт})$. Определив $\Delta_{m,опт}$, мы получаем принципиально новый эффект - $\alpha(\Delta_m)$ и $\beta(\Delta_m)$ не задаются, а вычисляются, исходя из требования $\min\left[A \cdot \alpha(\Delta_m) + B \cdot \beta(\Delta_m)\right]$. Казалось бы, определение $\Delta_{m,опт}$, тривиально. Если знать зависимости $\alpha(\Delta_m)$ и $\beta(\Delta_m)$, то, умножив их, соответственно, на коэффициенты А и В и сложив их для одних и тех же $\Delta_m$, определим $\gamma(\Delta_m)$ и далее $\Delta_{m,onm}$, соответствующего минимальному значению $\gamma(\Delta_m)$. На самом деле, оценка $\Delta_{m,onm}$ требует учета следующих особенностей:

1. Распределения $\alpha(\Delta_m)$ и $\beta(\Delta_m)$ дискретны, Обозначим число дискретизаций $\alpha(\Delta_m)$ и $\beta(\Delta_m)$, соответственно, через $r_\alpha$ и $r_\beta$, причем $r_\alpha \neq r_\beta$. Число одинаковых дискретизаций ($r_\Sigma$) определяется:

- величинами m и M. С ростом m и M число дискретизаций функций распределения $\alpha(\Delta_m)$ и $\beta(\Delta_m)$ хотя и возрастает, но величина $r_\Sigma$ по-прежнему может исчисляться единицами. Повышение $r_\Sigma$ достигается путем снижения числа значащих цифр до трех, что практически не вносит погрешности в оценку $\gamma(\Delta_m)$;
- характером расхождения $F_M^*(\tau)$ и $F_m^*(\tau)$. Чем расхождение больше, тем больше $r_\Sigma$
- соотношение m и M. При кратных m и M, величина $r_\Sigma$ резко уменьшается

2. Критерий проверки гипотезы о представительности выборки для конкретного $\Delta_m^*$ формируется исходя из сущности интегральных функций распределения $\alpha(\Delta_m)$, $\beta(\Delta_m)$

$$\alpha(\Delta_m) = 1 - F(\Delta_m / H_o) = 1 - P(\Delta_m \le \Delta_{m,k}) = \mathrm{P}\left(\Delta_\mathrm{m} > \Delta_{m,k}\right) = \sum_{i=k+1}^{r_1} P(\Delta_{m,i}) \qquad (2)$$

$$\beta(\Delta_m) = F(\Delta_m / H_1) = P(\Delta_m \le \Delta_{m,k}) = \sum_{i=1}^{k} P(\Delta_{m,i}) \qquad (3)$$

где $\Delta_{m,k}$ - некоторое заданное значение статистики $\Delta_m$, соответствующей к- ой дискретизации.

Если $\Delta_{m,\alpha} = \Delta_m(\alpha \le \alpha_k)$, где $\alpha_k$ - уровень значимости, а $\Delta_m(\alpha \le \alpha_k)_{\min}$ - минимальное из возможных дискретных значений $\Delta_m$, то критерий проверки гипотезы о представительности выборки ($H_0$) имеет вид:

- если $\Delta_\mathrm{m}^* > \Delta_{m,(\alpha_k)}$, то гипотеза $H_o$ отвергается. Здесь $\Delta_m^*$ - значение статистики $\Delta_\mathrm{m}$, рассчитанное по экспериментальным данным. Физически это означает, что $\Delta_m^*$ с вероятностью $R = 1 - \alpha_k$ не принадлежит множеству возможных дискретных значений квантилей распределения $\alpha^*(\Delta_m)$. Разделение данных оправдано;
- если $\Delta_m^* \le \Delta_m$, то гипотеза $H_o$ не противоречит данным выборки и, одновременно, это не означает, что выборка представительна;

Условие проверки гипотезы $H_1$ имеет вид:

- если $\Delta_m^* < \Delta_m(\beta \le \beta_k)_{\max} = \Delta_{m,\beta}$, то гипотеза $H_1$ отвергается, где $\beta_k$ - ошибка второго рода; $\Delta_m(\beta \le \beta_k)_{\max}$ - максимальное значение квантиля распределения $\beta(\Delta_m)$ при условии, что $\beta \le \beta_k$. Физически это означает, что рассматриваемая (экспериментальная) выборка и ей подобные, несмотря на наблюдаемое различие параметров распределения, не могут рассматриваться как непредставительные, т.е. разделение совокупности данных по заданной разновидности признаков на этом этапе нецелесообразно. Несмотря на то, что $\alpha_k$ и $\beta_k$ задаются, действительные значения ошибок первого и второго рода оказываются меньше, чем $\alpha_k$ и $\beta_k$, что объясняется дискретным характером распределения $\alpha(\Delta_m)$ и $\beta(\Delta_m)$

Сопоставив соотношения (1), (2) и условия критериев проверки гипотез $H_0$ и $H_1$, нетрудно заметить, что имеет место несоответствие условия, при котором гипотеза $H_o$ отвергается, и соответствующим ему значением $\alpha(\Delta_m)$. Формально имеет место несоответствие и условия, при котором отвергается гипотеза $H_1$ и соответствующим ему значением $\beta(\Delta_m)$.

Соответствие имеет место, если:

- при проверке гипотезы $H_o$ $\alpha(\Delta_m)$ принимается равной:

$$\alpha(\Delta_m) = P(\Delta_m \geq \Delta_{m,k}) = 1 - P(\Delta_m \leq \Delta_{m,k-1}) = 1 - F(\Delta_{m,k-1}) = \sum_{i=k}^{r_\alpha} P(\Delta_{m_i k}) \qquad (4)$$

-при проверке гипотезы $H_1$ $\beta(\Delta_m)$ принимается равной:

$$\beta(\Delta_m) = \sum_{i=1}^{k} P(\Delta_{m_i i})$$

поскольку $\beta(\Delta_m) < \beta(\Delta_{m,k})$

Эти равенства свидетельствует о том, что при проверке гипотезы $H_0$ величину $\Delta_m^*$ необходимо сопоставлять не с $\Delta_{m,r}(\alpha_k)$, а с $\Delta_{m,(r+1)}$ значением квантиля распределения $\alpha(\Delta_m)$, где r=1,$r_k$, а при проверке гипотезы $H_1$ величину $\Delta_m^*$ необходимо сравнивать с $\Delta_{m,r}(\beta_k)$, т.к. действительное значение $\beta_k$ меньше заданного.

3. С учетом практики инженерных расчетов, когда вероятность ошибочного решения не должна превышать величину 0,1 (или 0,05), целесообразно сохранить эти цифры и для $\gamma(\Delta_{m,onm})$. При этом предельно допустимые значения $\alpha(\Delta_m)$ и $\beta(\Delta_m)$ будут равны 0,2 (или 0,1).

В качестве примера в таблице 1 приведены результаты расчета распределений $\alpha(\Delta_m)$, $\beta(\Delta_m)$ ошибки принятия решения $\gamma(\Delta_m)$ для случая, когда m=10, M=50, A=0.2, B=0.8

Условимся, что $\alpha_k = \beta_k = \gamma_k = 0.1$

Анализ данных этой таблицы позволяет заключить:

- Принятые значения $\alpha_k$ и $\beta_k$, равные 0,1, среди оценок $\alpha^*(\Delta)$ и $\beta^*(\Delta)$отсутствуют. Условию $\Delta_m = 0{,}28$ при $\alpha = 0{,}092$. Для $\beta_k = 0{,}1$ величина $\Delta_m(\beta \leq \beta_k)_{\max} = 0{,}20$ при $\beta = 0{,}073$, что меньше заданного - $\{\gamma(\Delta_m)\}_{\min} = 0{,}121 > \gamma_k = 0{,}1$. Иначе говоря, риск ошибочного решения превышает допустимое значение. Разделение данных не целесообразно. Величина $\Delta_{m,onm} = 0{,}16$. Этому значению соответствует: ошибка первого рода - $\alpha(\Delta_{m,onm}) = 0{,}56$; ошибка второго рода - $\beta(\Delta_{m,onm}) = 0{,}011$. Таким образом, $\gamma(\Delta_{m,onm}) = 0{,}121$ обуславливается преимущественно значительной ошибкой первого рода.
- Без учета значимости ошибок первого и второго рода целесообразность разделения данных может быть проверена в соответствии с п.2. Этот подход используется также при $\gamma(\Delta_{m,onm}) = 0$, что физически соответствует существенному расхождению распределений $F_M^*(\tau)$ и $F_m^*(\tau)$.
- В условиях, когда $\gamma(\Delta_{m,onm}) > \gamma_k$, $\Delta_m(\alpha_k) \geq \Delta_m^*$ и $\Delta_{m(\beta_k)} \leq \Delta_m^*$ следует заключить, что данных недостаточно для заключения по заданной разновидности признака. Если же выбор должен быть сделан, например, один из двух энергоблоков должен быть отключен на плановый ремонт, то выбор может быть сделан на основе минимума

ошибок первого и второго рода для $\Delta_m^*$ с учетом их значимости. Если, в частности $\Delta_m^* = 0{,}24$, то поскольку $\gamma(\Delta_m) > 0{,}1$, $\Delta_{m(\alpha_k)} = 0{,}28 > 0{,}24$, $\Delta_{m(\beta_k)} = 0{,}20 < 0{,}24$, а $A \cdot \alpha(\Delta_m^*) = 0{,}045$ и $B \cdot \beta(\Delta_m^*) = 0{,}149$, то предпочтение гипотезы $H_1$ очевидно.

Необозримое множество значений $M$ и $n$, возможных различий распределений $F_M^*(\tau)$ и $F_m^*(\tau)$, отличительные выше особенности проверки гипотезы, громоздкость и трудоемкость вычислений обуславливают целесообразность перехода к автоматизированным вычислениям на ПК.

Таблица 1.

Экспериментальные данные распределений статистики $\Delta_m$ (m =10, M=50, δ= 0,7)

| № | $\Delta_m$ | $\alpha(\Delta_m)$ | $A \cdot \alpha(\Delta_m)$ | $\beta(\Delta_m)$ | $B \cdot \beta(\Delta_m)$ | $\gamma(\Delta_m)$ |
|---|---|---|---|---|---|---|
| 1 | 0.06 | 0.993 | 0,197 | 0,000 | 0 | |
| 2 | 0.08 | 0.963 | 0,193 | 0,000 | 0 | |
| 3 | 0.10 | 0.899 | 0,180 | 0,001 | 0,001 | 0,182 |
| 4 | 0.12 | 0.799 | 0,160 | 0.002 | 0,002 | 0,162 |
| 5 | 0.14 | 0.687 | 0,132 | 0.003 | 0,002 | 0,139 |
| 6 | 0.16 | 0.560 | 0,112 | 0.011 | 0,009 | 0,121 |
| 7 | 0.18 | 0.453 | 0,091 | 0.039 | 0,032 | 0,123 |
| 8 | 0.20 | 0.373 | 0,075 | 0.073 | 0,058 | 0,133 |
| 9 | 0.22 | 0.290 | 0,058 | 0.126 | 0,101 | 0,159 |
| 10 | 0.24 | 0.226 | 0,045 | 0.186 | 0,149 | 0,193 |
| 11 | *0.26* | *0.164* | 0,033 | *0.265* | 0,212 | 0,245 |
| 12 | 0.28 | 0.118 | 0,024 | 0.381 | 0,305 | 0,329 |
| 13 | 0.30 | 0.092 | 0,018 | 0.489 | 0,391 | 0,409 |
| 14 | 0.32 | 0.071 | 0,014 | 0.595 | 0,476 | 0,490 |
| 15 | 0.34 | 0.047 | 0,010 | 0.699 | 0,560 | 0,510 |
| 16 | 0.36 | 0.036 | 0,007 | 0.768 | 0,614 | 0,621 |
| 17 | 0.38 | 0.022 | 0,004 | 0.836 | 0,669 | 0,673 |
| 18 | 0.40 | 0.010 | 0,004 | 0.888 | 0,710 | 0,712 |
| 19 | 0.42 | 0.006 | 0,001 | 0.921 | 0,737 | 0,738 |
| 20 | 0.44 | 0.004 | 0,001 | 0.948 | 0,758 | 0,759 |
| 21 | 0.48 | 0.003 | 0,000 | 0.986 | 0,789 | 0,789 |
| 22 | 0.52 | 0.00 | 0,000 | 0.995 | 0,796 | 0,796 |
| 23 | 0.54 | | | 0.998 | 0,798 | 0,798 |
| 24 | 0.58 | | | 0.999 | | |
| 25 | 0.60 | | | 1.000 | | |

Графическая иллюстрация алгоритма принятия решения приведена на рис.2.

Пример. Требуется сопоставить длительность рабочего состояния (РС) $\tau_{раб,i}$ каждого ЭБ с i=1,8 за 2005 г. с усредненной длительностью РС всех ЭБ. Число реализаций длительности РС всех ЭБ M=73, а среднее значение $M_\Sigma^*[\tau_{раб}] = 700$ ч.

Результаты расчетов приведены в таблице 2. Дополнительно и вышеизложенному введены следующие обозначения: $m_i$ – число РС i – го ЭБ; $M_i^*[\tau_{раб}]$ - среднее значение длительности РС; Н и D – разделение статистических данных, соответственно, нецелесообразно или целесообразно; НИ– информация недостаточна.

Как следует из таблицы 2, средняя длительность РС $M_i^*[\tau_{раб}]$ первого и третьего ЭБ (соответственно равных 400 ч. и 455 ч.) существенно отличается (меньше), чем усредненная длительность РС всех ЭБ т.к. $\gamma_1(\Delta_{m,опт}) = 0$, а $\gamma_{\Sigma 3}(\Delta_{m,опт}) = 5{,}5\%$, что характеризует индивидуальные особенности. К этой же группе относится и восьмой ЭБ $\left[\Delta^*_{m,8} < \Delta_{m,опт}(0{,}1)\right]$, хотя и несколько в меньшей степени, т.к. $\gamma_8\left(\Delta_{m,опт}\right) = 16{,}5\%$ при $\beta_8(\Delta_m) \le 10\%$ и $\alpha_8(\Delta_m) \le 22\%$. С $\alpha_8(\Delta_m) \le 0{,}1\%$ расхождение величин $M^*_5\left[\tau_{раб}\right]$ и $M^*_6\left[\tau_{раб}\right]$ и $M^*_\Sigma\left[\tau_{раб}\right]$ могут быть приняты случайными.

Статические данные о $\tau_{раб}$ второго, шестого и седьмого ЭБ недостаточны для однозначного решения (в пределах допустимых вероятностей ошибок). Если же необходимо принять решение, то для второго ЭБ с ошибкой в 29% может быть принято заключение о неслучайном различии $M^*_\Sigma\left[\tau_{раб}\right] = 482$ ч. и $M^*_\Sigma\left[\tau_{раб}\right] = 701$ ч. Аналогичное заключение, но с большей ошибкой (40%) может быть принято и для седьмого ЭБ. Если же принято, что $M^*_\Sigma\left[\tau_{раб}\right]$ и $M^*_6\left[\tau_{раб}\right]$, различаются случайно то ошибка этого предположения будет равна 39%

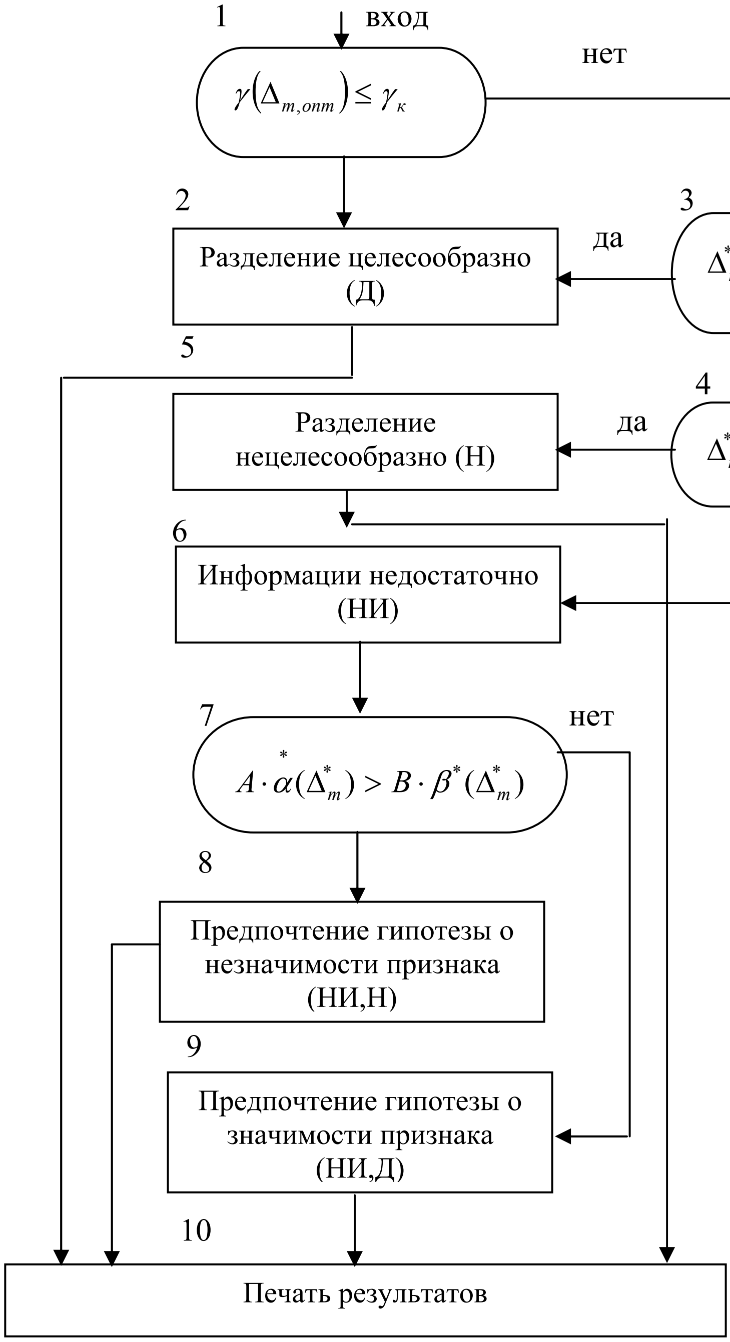


Рис.2. Алгоритм принятия решения при классификации статистических данных

Результаты сравнения индивидуальных и усредненных длительностей рабочих состояний ЭБ

Таблица 2

| i | $m_i$ | $M^*[\tau_i]$ | $\Delta_m^*$ | $\gamma^*(\Delta_{опт})$ | $\Delta_{m,\alpha}$ | $\Delta_{m,\beta}$ | $A \cdot \alpha(\Delta_m^*)$ | $B \cdot \beta(\Delta_m^*)$ | Решение | Оценка |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 16 | 401 | 0.27 | 0 | | | | | Д | 401 |
| 2 | 10 | 782 | 0.17 | 0.14 | 0,23 | 0,14 | 0,35 | 0,26 | НИ. Н | 699 (782) |
| 3 | 16 | 455 | 0.22 | 0 | | | | | Д | 455 |
| 4 | 8 | 747 | 0.04 | 0.25 | 0,24 | 0,10 | | | Н | 699 |
| 5 | 2 | 1242 | 0.05 | 0.32 | 0,28 | 0,06 | | | Н | 699 |
| 6 | 6 | 1093 | 0.24 | 0.17 | 0,26 | 0,14 | 0,23 | 0,50 | НИ. Д | 699 (1093) |
| 7 | 8 | 811 | 0.13 | 0.20 | 0,24 | 0,13 | | | Н | 699 |
| 8 | 7 | 1140 | 0.39 | 0.08 | | | | | Д | 1140 |


1. *Фархадзаде Э.М., Мурадалиев А.З., Фарзалиев Ю.З.* Оценка точности показателей индивидуальной надежности энергоблоков . Экоэнергетика, №1, 2006г.
2. *Шор Я.Б., Кузьмин Ф.И.* Таблицы для анализа и контроля надежности. М:, «Советское радио», 1968


## ENERJİBLOKLARININ FƏRDİ ETİBARLIĞININ QİYMƏTLƏNDİRİLMƏSİ ZAMANI ƏLAMƏTLƏRİN ƏHƏMİYYƏTLİLİYİNƏ NƏZARƏT


**FƏRHADZADƏ E.M., MURADƏLİYEV A.Z., FƏRZƏLİYEV Y.Z.**



Statistik verilənlərin təsnifatı zamanı, əlamətlərin əhəmiyyətliliyi verilənlərinin mötəbər seçimi xarakterizə edən, birinci və ikinci dərəcəli səhvlərin qiymətləmə, səhv qərar ehtimalı, statistikanın optimal kritik qiyməti ilə təyin olunur.


## THE SUPERVISION OF THE SIGNIFICANCE OF INDICATIONS OF THE ASSESSMENT INDIVIDUAL RELIABILITY OF POWER-GENERATING UNITS


**FARHADZADEH E.M., MURADALIYEV A.Z., FARZALIYEV J.Z.**



The importance of attributes at classification of the statistical data is defined by numerical meanings of mistakes of the first and second sort, risk of the erroneous decision and optimum size of critical meaning of statistics describing представительность samples of the data.