

DOI: 10.35401/2500-0268-2019-15-3-74-76

**В.П. Леонов**

## ТРЕБОВАНИЯ И РЕКОМЕНДАЦИИ ПО ОПИСАНИЮ И ИСПОЛЬЗОВАНИЮ СТАТИСТИЧЕСКОГО АНАЛИЗА В ИССЛЕДОВАНИЯХ

✉ В.П. Леонов, к.т.н., доцент, НЦ БИОСТАТИСТИКА, редактор сайта БИОМЕТРИКА (<http://www.biometrika.tomsk.ru/>), e-mail: leo.biostat@gmail.com

**Цитировать:** Леонов В.П. Требования и рекомендации по описанию и использованию статистического анализа в исследованиях. Инновационная медицина Кубани. 2019;15(3):74-76. DOI: 10.35401/2500-0268-2019-15-3-74-76

**V.P. Leonov**

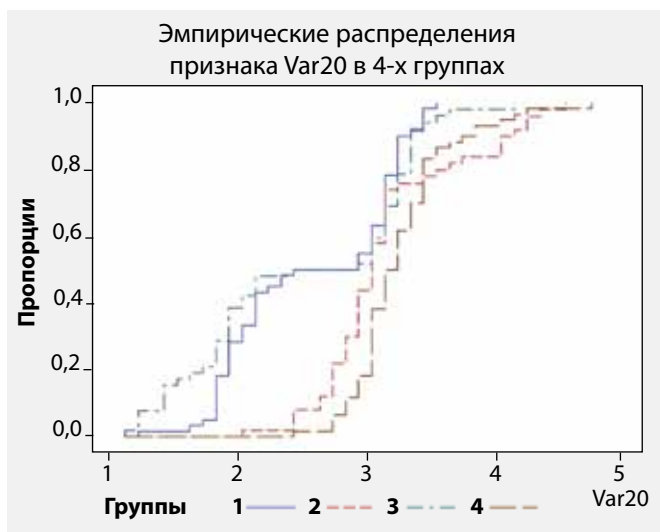
## REQUIREMENTS AND RECOMMENDATIONS FOR DESCRIPTION AND USE OF STATISTICAL ANALYSIS IN RESEARCH

✉ V.P. Leonov, CTS, Assistant professor, NC Biostatistics, editor of the site "Biometrika", e-mail: leo.biostat@gmail.com

**Cite this article as:** Leonov V.P. Requirements and recommendations for description and use of statistical analysis in research. Innovative Medicine of Kuban. 2019;15(3):74-76. DOI: 10.35401/2500-0268-2019-15-3-74-76

*Продолжение. Начало статьи в №1(2019)*

13. Для лучшего понимания специфики идентичности или различия сравниваемых групп, рекомендуется также производить проверку гипотез о равенстве групповых медиан. А для наглядного отражения наличия или отсутствия различия распределений сравниваемых групп по наиболее важному количественному признаку, рекомендуем приводить график с эмпирическим распределением всех сравниваемых групп. Ниже приведён пример такого графика, на котором наглядно видно, что при значениях признака Var20 до величины порядка 3, эти 4 подгруппы фактически распределены на 2 подгруппы. А при значениях признака Var20 более 3, эти 2 подгруппы сближаются, а степень различия между 4 группами становится существенно меньше.



Отмечаем, что подобный графический анализ, по сути, является исследованием интенсивности взаи-

мосвязи между одним дискретным (группирующим) признаком и признаком количественным. А также изучается наличие/отсутствие близости значений анализируемого количественного признака в сравниваемых группах. Решение о публикации подобных графиков в рукописи статьи принимают сами авторы публикации.

14. При описании вычисляемых парных или множественных взаимосвязей в различных методах статистического анализа нельзя приводить оценки интенсивности этих связей как «слабые», «средние» или «сильные». Следует приводить имеющиеся в этих методах числовые значения интенсивности этих взаимосвязей. Например, «Коэффициент корреляции Пирсона равен 0,62», «Коэффициент связи V-Крамера = 0,52», «Коэффициент контингенции равен 0,46», «Коэффициент канонической корреляции равен 0,678» и т.п. Также, приводя числовое значение конкретного статистического критерия типа « $F = 3,25$ », необходимо при этом также приводить степень свободы типа « $df = 3$ », и значение достигнутого уровня статистической значимости для данного критерия вида « $p = 0,0227$ ».

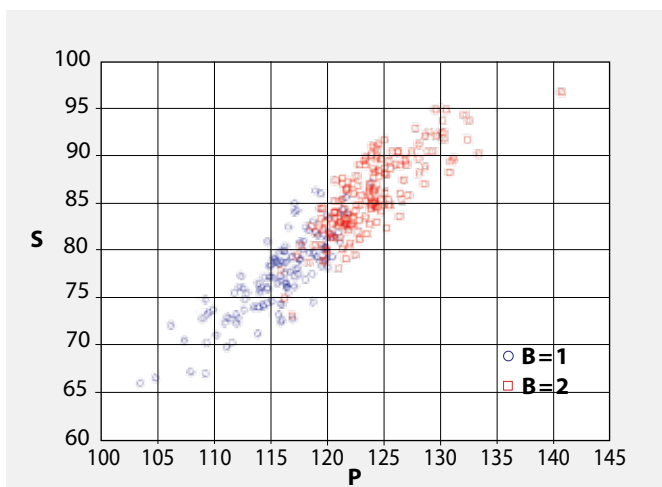
15. При анализе взаимосвязи между двумя количественными признаками часто используются методы корреляционного анализа. В таком случае следует производить аргументацию выбора вида корреляционного анализа, каждый из которого имеет свои ограничения и приоритеты. После чего сообщать название выбранного типа корреляционной связи. Например, корреляция Пирсона, корреляция Спирмена, ранговая корреляция  $\tau$ -Кендалла, частная корреляция, множественная корреляция, корреляционное отношение, автокорреляция и т.д. При описании результатов корреляционного анализа обязательно нужно вначале сформулировать цель данного анализа, при этом со-

Pearson Correlation Coefficients / Prob >  R  under Ho: Rho=0 / Number of Observations							
VAR2							
	VAR2	VAR11	VAR14	VAR8	VAR16	VAR26	VAR17
	1,0000	-0,74835	-0,4357	-0,41895	0,23766	-0,12345	0,13456
	0,0	0,0001	0,035	0,023	0,0693	0,345	0,146
	93	93	83	85	91	92	83

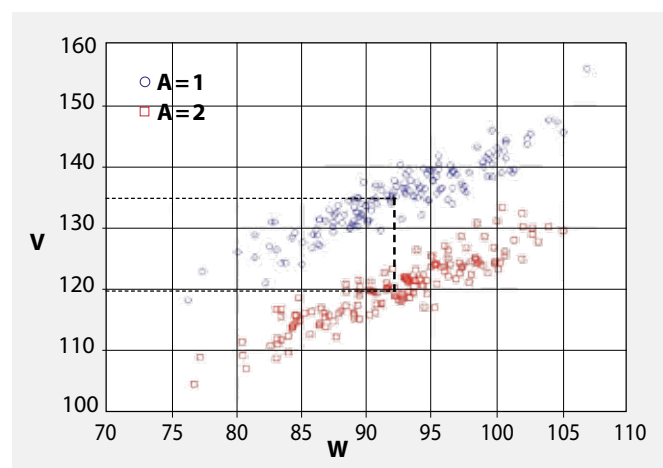
Spearman Correlation Coefficients / Prob >  R  under Ho: Rho=0 / Number of Observations							
VAR2							
	VAR2	VAR14	VAR11	VAR26	VAR16	VAR17	VAR15
	1,0000	-0,8746	-0,69726	-0,34345	0,24486	0,14234	0,14678
	0,0	0,0002	0,0006	0,0054	0,094	0,179	0,194
	93	83	93	92	91	83	73

обшив используемые количественные признаки и название варианта корреляционного анализа. Для каждого приводимого коэффициента корреляции обязательно приводятся объёмы анализируемых наблюдений, достигнутый уровень статистической значимости. Также следует учитывать тот факт, что в разных парных взаимосвязях количественных признаков имеются разные виды зависимостей. Например, для одной пары признаков линейная связь, а для другой – нелинейная. Поэтому для оценки и сравнения этих видов взаимосвязей следует применять как минимум 2 вида корреляционных зависимостей. Далее представлен такой пример, в котором использованы корреляции Пирсона и корреляция Спирмена. В приводимых ниже двух таблицах, результаты парных корреляций отсортированы по уменьшению модуля коэффициента корреляции.

По результатам, представленным в двух этих таблицах, видно, что для пары признаков VAR2 и VAR11 коэффициенты корреляции по Пирсону и Спирмену очень близки, что подтверждает наличие линейной взаимосвязи. Тогда как для пары признаков VAR2 и VAR14 коэффициент корреляции Пирсона в 2 раза меньше коэффициента Спирмена, что свидетельствует о доминировании нелинейного вида взаимосвязи данной пары признаков.

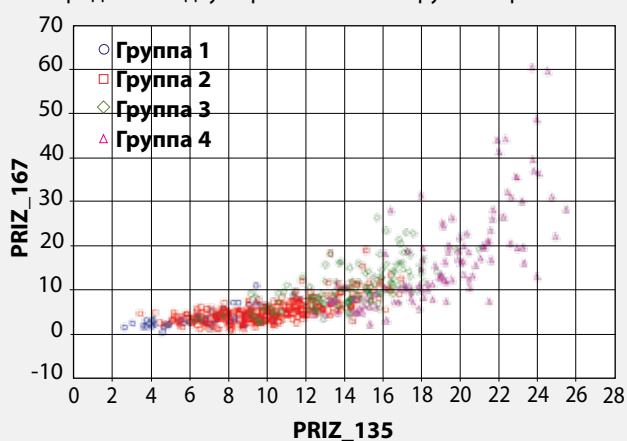


16. В подобных исследованиях между двумя количественными признаками рекомендуется использовать графическое изображение распределения наблюдений в осях двух данных количественных признаков, подтверждающих выбор данного вида корреляции. Если по этим двум признакам также исследуются взаимосвязи с некоторыми группирующими



признаками, то рекомендуем при этом в графике вводить отображения наблюдений принадлежностью к подгруппе конкретного группирующего признака. Ниже приведён график распределения признаков P и

Распределение двух признаков в 4-х группах признака GR



S, относящимся к двум группам по признаку B.

Как видно из этого рисунка, эти подгруппы соединяются по обоим количественным признакам, и имеют практически одинаковые уравнения регрессии.

На следующем графике подобные две анализируемые группы не соприкасаются полностью, но при этом их уравнения регрессии отличаются практически лишь свободным членом.

Тогда как на приведённом ниже графике распределения двух признаков PRIZ\_135 и PRIZ\_167 в 4-х группах признака GR имеют совершенно иные параметры и уравнения взаимосвязи.

По всем наблюдениям 4-х групп зависимость между парой количественных признаков является нелинейной. Причём в зависимости от принадлежности этих двух признаков к конкретным группам, меняются величины их средних значений и дисперсий, а также направления связей. При этом можно

произвести оценку уравнений регрессии отдельно для каждой из таких групп, что позволит определить характер измерения этих зависимостей при переходе от одной группы к другой. Исходя из вида распределения наблюдений в осях двух количественных признаков, делается и выбор вида парного регрессионного анализа. Например, если наблюдения базы данных имеют линейный характер распределения, то выбирается линейная парная регрессия. Если же вид распределения сугубо нелинейный, то производится выбор соответствующей нелинейной парной регрессии. Это может быть степенная регрессия, например, с наличием квадрата конкретного признака, либо логарифмическая или экспоненциальная парная регрессия и т.д. Решение о публикации подобных графиков в рукописи статьи принимают сами авторы публикации.

*Продолжение следует.*