ОГЛАВЛЕНИЕ

[ВВЕДЕНИЕ 3](#_Toc516528907)

[1 Обзор предметной области 4](#_Toc516528908)

[2 Частотный анализ 5](#_Toc516528909)

[3 Определение наилучшей комбинации составляющих информационного блока социальной сети 6](#_Toc516528910)

[ЗАКЛЮЧЕНИЕ 24](#_Toc516528911)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 25](#_Toc516528912)

# ВВЕДЕНИЕ

В наши дни является обыденным делом использование веб-технологий в организации общения. Актуальность данного исследования определяется тем, что сегодня социальные сети играют в жизни современного человека огромную роль. Ежедневно миллионы людей могут не только общаться, но еще и продвигать свои услуги, заключать сделки, делиться профессиональными навыками, вести научную деятельность и т.д. именно в социальных сетях. Поэтому будет актуальным выяснить, как различные составляющие информационного блока влияют на его восприятие пользователями одной из социальных сетей по показателям, характеризующим вовлеченность аудитории. Эту задачу будем решать при помощи процедур частотного анализа, реализованных в среде пакета STATISTICA.

# Обзор предметной области

В данном исследовании рассмотрим социальную сеть ВКонтакте и ее информационные блоки с такими составляющими, как изображение, видео, ссылка и т.д. по следующим показателям: просмотры, лайки, комментарии, репосты.

Под информационным блоком будем подразумевать структуру данных социальной сети, называемую «пост».

Пост может содержать текстовую информацию, изображения, видеоматериалы, аудиофайлы, ссылки на сторонние ресурсы и т. д.

Посты в социальных сетях разделяют по характеру информации, содержащейся в них, на следующие виды:

1. информационные;
2. развлекательные;
3. инфографика;
4. рекламные (продуктовые) и др.

Далее будем рассматривать рекламные информационные блоки.

Также каждый пост характеризуется определенным набором характеристик: просмотры, лайки, комментарии, репосты.

Под количеством просмотром будем понимать количество пользователей, увидевших рекламный пост в своей новостной ленте.

Количество «лайков» – это число пользователей, которые посчитали данный информационный пост интересным.

Комментарием является сообщение от пользователя, прикрепляющееся к информационному блоку.

Количество «репостов» – это число пользователей, поделившихся данной записью с кругом своих «подписчиков».

# Частотный анализ

Таблицы частот, или одновходовые таблицы, представляют собой простейший метод анализа категориальных (номинальных) переменных. Часто их используют как одну из процедур разведочного анализа, чтобы просмотреть, каким образом различные группы данных распределены в выборке.

Кросстабуляция (сопряжение) — процесс объединения двух (или нескольких) таблиц частот так, что каждая ячейка (клетка) в построенной таблице представляется единственной комбинацией значений или уровней табулированных переменных. Таким образом, кросстабуляция позволяет совместить частоты появления наблюдений на разных уровнях рассматриваемых факторов. Исследуя эти частоты, можно определить связи между табулированными переменными. Обычно табулируются категориальные переменные или переменные с относительно небольшим числом значений. Если надо табулировать непрерывную переменную, то вначале ее следует перекодировать, разбив диапазон изменения на небольшое число интервалов.

# Определение наилучшей комбинации составляющих информационного блока социальной сети

В таблице 1 представлены результаты исследования 25 групп с аудиторией 84 – 173 тысяч человек, где строки соответствуют наблюдениям (группы), а столбцы переменным (составляющие информационного блока и показатели), где для переменных *Изображение*, *Видео*, *Текст*, *Ссылка*, *Репост* приняты следующие обозначения: 1 – присутствует, 0 – отсутствует.

Таблица 1

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Изображение | Видео | Текст | Ссылка | Репост | Просмотров | Лайков | Комментариев | Репостов |
| 1 | 1 | 0 | 0 | 1 | 1 | 2000 | 153 | 0 | 1 |
| 2 | 1 | 0 | 1 | 0 | 1 | 3000 | 51 | 0 | 4 |
| 3 | 0 | 0 | 1 | 1 | 1 | 3000 | 3 | 1 | 0 |
| 4 | 0 | 1 | 1 | 0 | 1 | 7000 | 2 | 0 | 0 |
| 5 | 1 | 0 | 0 | 0 | 1 | 4000 | 83 | 0 | 0 |
| 6 | 1 | 0 | 1 | 1 | 1 | 4000 | 29 | 2 | 0 |
| 7 | 1 | 0 | 0 | 1 | 0 | 4000 | 91 | 2 | 0 |
| 8 | 1 | 0 | 1 | 1 | 0 | 12000 | 13 | 5 | 0 |
| 9 | 1 | 0 | 1 | 1 | 0 | 3000 | 3 | 0 | 0 |
| 10 | 1 | 0 | 1 | 1 | 0 | 3000 | 176 | 0 | 1 |
| 11 | 1 | 0 | 1 | 1 | 0 | 2000 | 111 | 0 | 0 |
| 12 | 1 | 0 | 1 | 0 | 1 | 3000 | 10 | 0 | 0 |
| 13 | 1 | 0 | 0 | 0 | 1 | 10000 | 63 | 1 | 7 |
| 14 | 0 | 0 | 1 | 1 | 0 | 14000 | 40 | 0 | 0 |
| 15 | 1 | 0 | 1 | 1 | 1 | 9000 | 45 | 0 | 1 |
| 16 | 1 | 0 | 0 | 1 | 1 | 11000 | 76 | 0 | 0 |

Продолжение таблицы 1

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 17 | 1 | 0 | 1 | 1 | 0 | 14000 | 42 | 3 | 0 |
| 18 | 1 | 0 | 1 | 1 | 1 | 1000 | 11 | 0 | 0 |
| 19 | 1 | 0 | 1 | 1 | 1 | 2000 | 8 | 0 | 0 |
| 20 | 1 | 0 | 1 | 1 | 1 | 2000 | 4 | 0 | 0 |
| 21 | 1 | 0 | 1 | 1 | 0 | 9000 | 8 | 0 | 0 |
| 22 | 1 | 0 | 1 | 0 | 1 | 3000 | 123 | 0 | 15 |
| 23 | 1 | 0 | 0 | 0 | 1 | 4000 | 94 | 0 | 1 |
| 24 | 1 | 0 | 1 | 1 | 1 | 5000 | 26 | 0 | 0 |
| 25 | 1 | 0 | 1 | 1 | 0 | 5000 | 14 | 5 | 0 |

Для построения таблиц кросстабуляции необходимо перекодировать переменные *Просмотров*, *Лайков*, *Комментариев*, *Репостов*, разбив диапазон изменения на небольшое число интервалов.

Разобьем значения переменных на равные диапазоны и введем для них следующие обозначения: *Очень мало*, *Мало*, *Много*, *Очень много* (таблица 2).

Таблица 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Просмотров | Лайков | Комментариев | Репостов |
| Очень мало | 0-3 тыс. | 0-40 | 0-1 | 0-3 |
| Мало | 3-6 тыс. | 41-80 | 2-3 | 4-7 |
| Много | 6-9 тыс. | 81-120 | 4-5 | 7-10 |
| Очень много | 9 тыс. и более | 121 и более | 6 и более | 11 и более |

В результате получим следующую таблицу (таблица 2):

Таблица 3

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Изображение | Видео | Текст | Ссылка | Репост | Просмотров | Лайков | Комментариев | Репостов |
| 1 | 1 | 0 | 0 | 1 | 1 | очень мало | очень много | очень мало | очень мало |

Продолжение таблицы 3

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | 1 | 0 | 1 | 0 | 1 | мало | мало | очень мало | мало |
| 3 | 0 | 0 | 1 | 1 | 1 | мало | очень мало | очень мало | очень мало |
| 4 | 0 | 1 | 1 | 0 | 1 | много | очень мало | очень мало | очень мало |
| 5 | 1 | 0 | 0 | 0 | 1 | мало | много | очень мало | очень мало |
| 6 | 1 | 0 | 1 | 1 | 1 | мало | очень мало | мало | очень мало |
| 7 | 1 | 0 | 0 | 1 | 0 | мало | мало | мало | очень мало |
| 8 | 1 | 0 | 1 | 1 | 0 | очень много | очень мало | много | очень мало |
| 9 | 1 | 0 | 1 | 1 | 0 | мало | очень мало | очень мало | очень мало |
| 10 | 1 | 0 | 1 | 1 | 0 | мало | очень много | очень мало | очень мало |
| 11 | 1 | 0 | 1 | 1 | 0 | очень мало | много | очень мало | очень мало |
| 12 | 1 | 0 | 1 | 0 | 1 | мало | очень мало | очень мало | очень мало |
| 13 | 1 | 0 | 0 | 0 | 1 | очень много | мало | очень мало | много |
| 14 | 0 | 0 | 1 | 1 | 0 | очень много | очень мало | очень мало | очень мало |

Продолжение таблицы 3

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 15 | 1 | 0 | 1 | 1 | 1 | очень много | мало | очень мало | очень мало |
| 16 | 1 | 0 | 0 | 1 | 1 | очень много | мало | очень мало | очень мало |
| 17 | 1 | 0 | 1 | 1 | 0 | очень много | мало | мало | очень мало |
| 18 | 1 | 0 | 1 | 1 | 1 | очень мало | очень мало | очень мало | очень мало |
| 19 | 1 | 0 | 1 | 1 | 1 | очень мало | очень мало | очень мало | очень мало |
| 20 | 1 | 0 | 1 | 1 | 1 | очень мало | очень мало | очень мало | очень мало |
| 21 | 1 | 0 | 1 | 1 | 0 | очень много | очень мало | очень мало | очень мало |
| 22 | 1 | 0 | 1 | 0 | 1 | мало | очень много | очень мало | очень много |
| 23 | 1 | 0 | 0 | 0 | 1 | мало | много | очень мало | очень мало |
| 24 | 1 | 0 | 1 | 1 | 1 | мало | очень мало | очень мало | очень мало |
| 25 | 1 | 0 | 1 | 1 | 0 | мало | очень мало | много | очень мало |

Для построения таблиц кросстабуляции в меню *Анализ* выберем пункт *Основные статистики и таблицы*.



Рисунок 1

В диалоговом окне выберем *Таблицы сопряженности, флагов и заголовков* (рисунок 2).



Рисунок 2

Появится следующее окно:



Рисунок 3

Зададим таблицы сопряженности, как показано на рисунке 4.



Рисунок 4

Перейдем во вкладку *Опции* и выберем *Выделить частоты* $>$ *10*, *Проценты по строке* и *Проценты по столбцу* для пункта *Таблицы* и *Корреляция Спирмена* для пункта *Статистики для двухвходовых таблиц*.



Рисунок 5

Затем перейдем во вкладку *Дополнительно* и нажмем кнопку *Подробные двухвходовые таблицы* (рисунок 6).



Рисунок 6



Рисунок 7

На рисунке 7 представлена итоговая 2-входовая таблица с информацией о просмотрах информационных блоков, содержащих изображение. По результатам видно, что 27% рассматриваемых информационных блоков, содержащих изображение, имеют очень много просмотров.



Рисунок 8

По результатам рисунка 8 можем увидеть, что информационные блоки, не содержащие видео, набирают мало просмотров, однако, есть и блоки, попадающие в категорию с очень большим количеством просмотров и не содержащие видео. Информационный блок, содержащий видео, всего один и он набирает большое количество просмотров.



Рисунок 9

Влияние на просмотры наличия текста в информационных блоках показано на рисунке 9. Блок с текстом набирает много просмотров в 1 случае из 19, очень много в 5 случаях.



Рисунок 10

6 из 18 блоков, содержащих ссылку, набирают очень много просмотров (рисунок 10).



Рисунок 11

Из 16 информационных блоков, содержащих репост много просмотров набирает 1, очень много – 3 (рисунок 11). Из 9 блоков, не содержащих ссылку очень много просмотров набирают 4.

Таким образом, можем сделать вывод, что 27% рассматриваемых информационных блоков, содержащих изображение, имеют более 9000 просмотров. Информационные блоки, не содержащие видео, набирают до 3000 просмотров, однако, есть и блоки, попадающие в категорию с количеством просмотров более 9000 и не содержащие видео. Информационный блок, содержащий видео, всего один и он набирает до 9000 количество просмотров. Блок с текстом в основном набирает более 6000 просмотров в 1 случае из 19, более 9000 просмотров в 5 случаях. 6 из 18 блоков, содержащих ссылку, набирают более 9000 просмотров. Из 16 информационных блоков, содержащих репост более 6000 просмотров набирает 1, более 9000 – 3. Из 9 блоков, не содержащих ссылку более 9000 просмотров набирают 4.

Рассмотрим влияние наличия изображения на количество лайков на рисунке 12.



Рисунок 12

Очень мало лайков набирают все информационные блоки без изображения. Очень много и много лайков набирают по 3 информационных блока с изображениями из 22.



Рисунок 13

Очень мало лайков набирает половина информационных блоков без видео (рисунок 13).



Рисунок 14

По результатам рисунка 14 видно, что очень много лайков у 2 блоков с текстом из 19 и у 1 блока без текста из 6.



Рисунок 15

У информационных блоков со ссылкой очень много лайков в 1 из 18 случаев, в 2 – очень много. У блоков без ссылки в 2 из 7 случаев много лайков, в 1 из 7 очень много (рисунок 15).



Рисунок 16

Информационные блоки, содержащие репост набирают очень много или много лайков в 2 из 16 случаев.

Можно сделать вывод, что до 40 лайков набирают все информационные блоки без изображения. От 81 лайка набирают 6 информационных блоков с изображениями из 22. До 40 лайков набирает половина информационных блоков без видео. Более 121 лайка у 2 блоков с текстом из 19 и у 1 блока без текста из 6. У информационных блоков со ссылкой более 81 лайка в 1 из 18 случаев, в 2 – более 121. У блоков без ссылки в 2 из 7 случаев до 120 лайков, в 1 из 7 более 121. Информационные блоки, содержащие репост набирают более 121 лайка в 2 из 16 случаев.

Рассмотрим влияние наличия в информационном блоке тех или иных составляющих на количество комментариев.



Рисунок 17

По результатам рисунка 17 можем сделать следующий вывод: информационные блоки с изображением получают много комментариев в 2 из 22 случаев. Все блоки без изображения получают очень мало комментариев.



Рисунок 18

При отсутствии видео информационный блок получает очень мало комментариев в 19 из 24 случаев (рисунок 18).



Рисунок 19

Много комментариев получают 2 информационных блока с текстом из 19 (рисунок 19).



Рисунок 20

Из результатов рисунка 20 следует, что все информационные блоки без ссылки получают очень мало комментариев.



Рисунок 21

Информационные блоки, не содержащие репост, получают больше комментариев, чем блоки с репостом (рисунок 21).

Таким образом, можно сделать вывод, что информационные блоки с изображением получают до 5 комментариев в 2 из 22 случаев. Все блоки без изображения получают до 1 комментария. При отсутствии видео, информационный блок получает 0-1 комментариев в 19 из 24 случаев. Более 4 комментариев получают 2 информационных блока с текстом из 19. Все информационные блоки без ссылки получают до 1 комментария.

Рассмотрим влияние наличия в информационном блоке тех или иных составляющих на количество репостов.



Рисунок 22

Много и очень много репостов получают по 1 информационному блоку из 22, содержащих изображение (рисунок 22).



Рисунок 23

Из результатов рисунка 23 следует, что при отсутствии видео информационный блок получает очень мало репостов в 21 из 24 случаев.



Рисунок 24

Очень много репостов получает 1 информационный блок, содержащий текст, из 19 (рисунок 24).



Рисунок 25

Из информационных блоков, не содержащих ссылку, много и очень много репостов набирают по 1 блоку из 7.



Рисунок 26

По 1 информационному блоку из 16, содержащих репост, набирают много и очень много репостов (рисунок 26).

Следовательно, можем говорить о том, что более 7 репостов получают 2 информационных блока из 22, содержащих изображение. При отсутствии видео информационный блок получает до 3 репостов в 21 из 24 случаев. 11 и более репостов получает 1 информационный блок, содержащий текст, из 19. Из информационных блоков, не содержащих ссылку, 7 и более репостов набирают по 1 блоку из 7. По 1 информационному блоку из 16, содержащих репост, набирают более 7 репостов.

Исходя из вышеизложенных данных, приходим к следующему выводу:

1. появление изображения в рекламном «посте» приводит к увеличению просмотров и репостов. При отсутствии изображения количество лайков и комментариев уменьшается;
2. информационные блоки, не содержащие видео, могут набирать большое количество просмотров, так же, как и блоки, не содержащие видео. Отсутствие видео приводит к уменьшению числа комментариев, лайков и репостов;
3. «пост» с текстом набирает меньшее число просмотров, лайков, репостов и комментариев, нежели без;
4. влияние репоста в рекламном «посте» не отражается на числе просмотров, лайков и комментрариев;
5. наличие ссылки не влияет на число лайков и комментариев, но влияет на число просмотров. Так, у «поста» без ссылки просмотров меньше, чем у «поста» со ссылкой.

# ЗАКЛЮЧЕНИЕ

В ходе данного исследования была рассмотрена социальная сеть ВКонтакте и ее информационные блоки с такими составляющими, как изображение, видео, ссылка и т.д. по следующим показателям: просмотры, лайки, комментарии, репосты.

При помощи процедур частотного анализа, реализованных в среде пакета STATISTICA, было проанализировано как различные составляющие информационного блока влияют на его восприятие пользователями одной из социальных сетей по показателям, характеризующим вовлеченность аудитории.

Таким образом, предложенный метод исследования информационных блоков социальной сети путем совмещения частот появления наблюдений на разных уровнях рассматриваемых факторов позволяет определить связи между переменными и таким образом выяснить, какой набор составляющих является наиболее выигрышным для рекламного «поста». Метод прост в реализации, так как основан на использовании стандартных процедур статистических пакетов.

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Халафян А.А. STATISTICA 6. Статистический анализ данных. М.: Бином, 2010. 528 с.