Министерство образования и науки Республики Казахстан

Северо-Казахстанский государственный университет им. М. Козыбаева

Факультет информационных технологий

Кафедра "Математика и информатика"

КУРСОВАЯ РАБОТА

На тему: "Методы математической статистики и теории вероятностей в выборе фильма""

по специальности 5В060200 - "Информатика"

Выполнил Е.К. Касымов

студент группы Ин (е) - 13

Научный руководитель В.П. Куликова

к. т. н., доцент

Петропавловск 2015

***Содержание***

Введение

1. Частота встречаемости зарубежных/отечественных фильмов на сайте Megogo.net

1.1 Теоретическое описание фильмов сайта Megogo.net

1.2 Популярность "отечественных" фильмов

2. Сравнение размаха количества просмотренных фильмов отечественного и зарубежного производства

2.1 Жанры киноиндустрии

2.2 Сравнения размаха количества просмотренных фильмов отечественных и зарубежных кинокомпаний

Заключение

Список использованной литературы

Приложение

# ***Введение***

Кинематограф - отрасль человеческой деятельности, заключающаяся в создании движущихся изображений. Иногда также упоминается как синематограф и кинематография. Кинематограф был изобретен в конце XIX века и стал крайне популярен в XX веке.

В понятие кинематографа входят киноискусство - вид современного изобразительного искусства, произведения которого создаются при помощи движущихся изображений, и киноиндустрия (кинопромышленность) - отрасль экономики, производящая кинофильмы, спецэффекты для кинофильмов, мультипликацию, и демонстрирующая эти произведения для зрителей. Произведения киноискусства создаются при помощи кинотехники. Изучением кинематографа занимается наука киноведение. Сами кинофильмы могут сниматься в различных жанрах.

Все больше телезрителей сегодня остаются недовольны политикой телевизионных каналов и постепенно уходят от экранов телевизоров и переходят к мониторам компьютеров. И это неудивительно, ведь в сети Интернет можно найти все что угодно. База интернет киноиндустрии пополняется каждый день большим количеством фильмов. На сегодняшний день можно смело говорить о том, что интернет - это огромная всеобъемлющая реальность, возможностям которой нет предела. К одной из таких возможностей можно смело причислить онлайн кинотеатры.

Все больше пользователей каждый день убеждаются что просмотр фильмов онлайн это очень просто и удобно. Сайты онлайн кинотеатров предоставляют широкий выбор художественных фильмов, а также самые различные передачи, мультфильмы и сериалы. У такого подхода есть огромное количество преимуществ. Во-первых, для того чтобы посмотреть интересующий вас фильм вам не понадобиться даже выходить из дома. Вы можете смотреть все новинки, устроившись на любимом диване. Во-вторых, смотреть фильмы вы можете в любое удобное для вас время. Ну и в-третьих, для того что бы посмотреть фильм вам не понадобиться тратить деньги на покупку или прокат дисков, так как онлайн кинотеатр - это абсолютно бесплатное удовольствие.

Еще одним неоспоримым достоинством онлайн кинотеатров можно считать то что на таких ресурсах вы можете найти абсолютно все, начиная от признанных произведений метров киноискусства, заканчивая никому не известными режиссерскими работами. Обычно сайты онлайн кинотеатров имеют очень простой интерфейс, ориентироваться в котором очень просто, так вы сможете отыскать любой фильм, как по названию, так и по году выпуска или по фамилии режиссера, снявшего данный фильм. Кроме всего прочего вы сможете найти здесь анонсы к фильмам и отзывы уже посмотревших его людей. Таким образом, получается, что вы можете не только смотреть фильмы на сайтах онлайн кинотеатров, но и обсуждать фильм с теми, кто его видел.

Кинематограф занимает значительную часть современной культуры многих стран. Во многих странах киноиндустрия является значимой отраслью экономики. Производство кинофильмов сосредоточено на киностудиях. Фильмы демонстрируются в кинотеатрах, по телевидению, распространяются "на видео" в форме видеокассет и видеодисков, а с появлением скоростного интернета стало доступным скачивание кинофильмов в форме видеофайлов на специализированных сайтах или посредством пиринговых сетей, а также просмотр онлайн.

*Актуальность* выбранной темы обусловлена тем, что в нынешнее время в просторах интернета содержится огромное количество доступных для просмотра кинокартин. Вместе с этим усложняется и выбор просмотра нужного фильма, который будет действительно обладать всеми качествами, которыми и должен обладать фильм способный впечатлить в полной мере, вызвать бурю эмоций или же наоборот просто расслабиться.

выбор фильм статистика математическая

*Цель курсовой работы:* Определить популярность отечественных и зарубежных фильмов сайта megogo.net, и проверить размах количества просмотренных фильмов отечественных и зарубежных кинокомпаний.

Для достижения цели необходимо решить следующие *задачи*:

· Собрать необходимые статистические данные

· Определить, насколько популярны фильмы отечественных кинокомпаний.

· Сравнить размах количества просмотренных фильмов отечественных и зарубежных кинокомпаний.

· Сделать соответствующие выводы.

*Объект исследования -* отечественные и зарубежные фильмы.

*Предмет исследования* - частоты встречаемости и размах количества фильмов отечественных и зарубежных кинокомпаний.

*Методы исследования*, применяемые в данной работе:

1) тест для единственной выборки

) параметрический метод Стьюдента

3) тест экстремальных реакций Мозеса, позволяющий сравнить размах цен двух независимых выборок

*Основные положения, выносимые автором на защиту:* частоты встречаемости зарубежных, отечественных фильмов, количества фильмов отечественных и зарубежных кинокомпаний

*Практическая значимость.* Исследования, проведенные в курсовой работе, непременно имеют практическую значимость, а именно разработка инструмента, который необходим помощи выбора фильма, подходящего для зрителя.

*Объем работы* - 25с.

# ***1. Частота встречаемости зарубежных/отечественных фильмов на сайте Megogo.net***

# ***1.1 Теоретическое описание фильмов сайта Megogo.net***

После тяжелого дня некоторые люди хотят расслабиться за чашечкой теплого чая и просмотром хорошего интересного фильма, на который можно потратить основную часть своего отдыха. Выбор фильма - важен, для того что бы получить волну впечатлений. Наша выборка будет сделана из сайта megogo.net. MEGOGO - Украинская меди компания, предоставляющая услуги онлайн-кинотеатра на просторах сайта megogo.net можно найти огромное количество фильмов разных жанров. Как и на любом сайте интернета посвящённому онлайн просмотру фильмов, здесь можно встретить самые различные жанры и категории фильмов. Но среди всего многообразия фильмов, все фильмы можно условно разделить на две группы, в зависимости от страны в которой был снят фильм. - "отечественные" и "зарубежные". Под "отечественными" понимаются фильмы, снятые в Казахстане в таких киностудиях как "Телекомпания Калиот <http://frc.kz/kultura-iskusstvo/kultura-iskusstvo-shou-biznes/kinostudii-kinokompanii/47-249-743-2013043-0-0.html>", "Кинокомпания Евразия фильм <http://frc.kz/kultura-iskusstvo/kultura-iskusstvo-shou-biznes/kinostudii-kinokompanii/47-249-743-2026140-0-0.html>", "Life production <http://frc.kz/kultura-iskusstvo/kultura-iskusstvo-shou-biznes/kinostudii-kinokompanii/47-249-743-2032657-0-0.html>" и т.д. "зарубежные же" же - фильмы, производимые в других странах, таких как, Япония, Германия, Америка и т.п. Если просмотреть сайт megogo.net, то нетрудно заметить, что одни фильмы встречаются чаще других, то есть одни фильмы более популярны, чем другие. И такой фактор, как популярность фильма занимает не последнее место в выборе просматриваемого фильма зрителем. При выборе фильма важно учитывать популярность фильма, так как просмотр популярного фильма дает уверенность будет впечатляющим и запомнится своим сюжетом. Зайдя на сайт можно выбрать нужную вам категорию жанра, согласно которому и будет совершаться выбор в просмотре фильма.

# ***1.2 Популярность "отечественных" фильмов***

Для определения популярности фильмов необходимо определить, какие фильмы встречаются чаще. Зайдя на просторы интернета был выбран первый сайт, предоставляющий услуги онлайн просмотра фильмов. Выбор пал на украинскую медиа-компанию Megogo, база зрителей которой постоянно растет.

На сайте megogo.net, случайным образом было проверено 50 фильмов, которые распределились, как показано в таблице 1.1, в зависимости от страны производителя следующим образом:

Таблица 1.1

Распределение фильмов по странам

|  |  |  |  |
| --- | --- | --- | --- |
| Группа | Отечественные | Зарубежные | Итого |
| Количество | 15 | 35 | 50 |

Можно ли на основании полученных данных сказать, что все фильмы

пользуются одинаковой популярностью?

В данном случае имеется одна выборка и необходимость проверить, соответствует ли распределение результатов в выборке предполагаемому распределению.

Так как в данном случае у нас имеется небольшая выборка, используется непараметрический тест. Для решения подобных задач применяется один из наиболее популярных непараметрических тестов - тест .

В качестве теоретического распределения в нашем случае используется равномерное распределение. Смысл его в том, что все результаты считаются равновероятными. При наличии двух групп фильмов вероятность встретить фильм, принадлежащую какой-либо из них, должна быть одна и та же и равна 1/2=0,5.

С учетом данного обстоятельства окончательный вариант

расчетной таблицы для данного примера представлен в таблице 1.2.

Таблица 1.2

Теоретическое и эмпирическое распределение фильмов

|  |  |  |  |
| --- | --- | --- | --- |
| Группа | Отечественные | Зарубежные | Итого |
| Теоретическое количество | 25 | 25 | 50 |
| Эмпирическое количество | 15 | 35 | 50 |

Формулируем нулевую и альтернативную гипотезы и задаем уровень значимости .

Н0: Вероятность встретить на сайте megogo.net фильм отечественного производства, равна вероятности встретить фильм, произведенный в зарубежных странах (Если  меньше ).

Н1: Вероятность встретить на сайте megogo.net фильм отечественного производства, не равна вероятности встретить фильм, произведенный в зарубежных странах (Если  больше или равно ).

Вычисляем сумму отклонений между наблюдаемыми и теоретическими значениями по формуле:

,

где  - наблюдаемые, или эмпирические, значения (частоты) для

каждой из категорий,  - ожидаемые, или теоретические, значения (частоты) для каждой из категорий таблицы

Подставим соответствующие значения  и  в расчетную формулу:



Таблица 1.3

Распределение теоретических и эмпирических частот

|  |  |  |  |
| --- | --- | --- | --- |
|  | Категории | | |
|  | 1 | 2 | итого |
| Ожидаемые и эмпирические частоты | , , 50 |  |  |

Полученное значениесравнивается со значением , которое берется из таблицы критических значений для теста  (см. Приложение Таблица 1) в зависимости от выбранного уровня значимости  и числа степеней свободы. В свою очередь, число степеней свободы для теста зависит от размера расчетной таблицы и равно. В нашем случае .

Если  больше или равно , то нулевая гипотеза отклоняется и принимается альтернативная (эмпирическое распределение значимо отличается от теоретического: фильмы одних стран-производителей встречаются на просторах сайта megogo.net чаще или реже, чем фильмы других).

Из таблицы 1 находим, что для и , .

Поскольку  гораздо больше, чем , то нулевая гипотеза отклоняется и принимается альтернативная. То есть, Вероятность встретить на сайте megogo.net фильм отечественного производства, не равна вероятности встретить фильм, произведенный в зарубежных странах.

Для определения популярности "отечественных" фильмов можно воспользоваться и параметрическими методами. Для этого необходимо собрать несколько больше статистических данных.

Для исследования были собраны данные из 100 выборок, в каждой из которых рассматривалось по 20 фильмов. Статистические данные представлены в таблице 1.4.

Таблица 1.4

Статистические данные фильмов из 100 выборок

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Номер выборки | Число отечественных фильмов | Номер выборки | Число отечественных фильмов | Номер выборки | Число отечественных фильмов | Номер выборки | Число отечественных фильмов |
| 1 | 3 | 26 | 3 | 51 | 3 | 76 | 5 |
| 2 | 2 | 27 | 2 | 52 | 2 | 77 | 4 |
| 3 | 2 | 28 | 1 | 53 | 4 | 78 | 3 |
| 4 | 4 | 29 | 3 | 54 | 4 | 79 | 3 |
| 5 | 1 | 30 | 3 | 55 | 2 | 80 | 3 |
| 6 | 3 | 31 | 3 | 56 | 5 | 81 | 3 |
| 7 | 2 | 32 | 2 | 57 | 4 | 82 | 2 |
| 8 | 1 | 33 | 4 | 58 | 3 | 83 | 2 |
| 9 | 3 | 34 | 4 | 59 | 3 | 84 | 4 |
| 10 | 3 | 35 | 2 | 60 | 3 | 85 | 1 |
| 11 | 3 | 36 | 5 | 61 | 3 | 86 | 3 |
| 12 | 2 | 37 | 4 | 62 | 2 | 87 | 2 |
| 13 | 4 | 38 | 3 | 63 | 2 | 88 | 1 |
| 14 | 4 | 39 | 3 | 64 | 4 | 89 | 3 |
| 15 | 2 | 40 | 3 | 65 | 1 | 90 | 3 |
| 16 | 5 | 41 | 3 | 66 | 3 | 91 | 3 |
| 17 | 4 | 42 | 2 | 67 | 2 | 92 | 2 |
| 18 | 3 | 43 | 2 | 68 | 1 | 93 | 4 |
| 19 | 3 | 44 | 4 | 69 | 3 | 94 | 4 |
| 20 | 3 | 45 | 1 | 70 | 3 | 95 | 2 |
| 21 | 3 | 46 | 3 | 71 | 3 | 96 | 5 |
| 22 | 2 | 47 | 2 | 72 | 2 | 97 | 4 |
| 23 | 2 | 48 | 1 | 73 | 4 | 98 | 3 |
| 24 | 4 | 49 | 3 | 74 | 4 | 99 | 3 |
| 25 | 1 | 50 | 3 | 75 | 2 | 100 | 3 |

Так как параметрические методы применяются при нормальном распределении, проверим, наши данные на нормальность.

Для того чтобы распределение было нормальным, нужно чтобы среднее, мода и медиана были равны.

Модой () называется наиболее часто встречающееся значение среди имеющихся. Для того чтобы разобраться с модой, построим дополнительную таблицу. Поместим в нее значения встречающихся "отечественных" фильмов от минимального количества (1) до максимального (5), и укажем, сколько раз встречается то или иное количество фильмов (табл. 1.5).

Таблица 1.5

Частота встречаемости "отечественных" фильмов

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Число отечественных фильмов | 1 | 2 | 3 | 4 | 5 |
| Частота | 10 | 25 | 40 | 20 | 5 |

Как видно из таблицы, чаще всего на каждые 20 фильмов встречается 3 "отечественных" (это значение встречается чаще всего - 40 раз). Поэтому мода появления "отечественных" фильмов равна-3.

Новая таблица содержит ту же самую информацию, что и предыдущая, но она заметно короче, данные в ней упорядочены по возрастанию, и с ней удобней работать.

Медиана () представляет собой значение, которое делит упорядоченные данные пополам таким образом, что одна половина данных оказывается меньше медианы, а другая - больше.

Нахождение медианы не носит столь наглядного характера, как нахождение моды. Для определения медианы приходится прибегать к дополнительным преобразованиям и вычислениям. Во-первых, дополним таблицу еще двумя строками и получим таблицу 1.6.

Таблица 1.6

Частота, накопленная частота и процент встречаемости фильмов

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Число отечественных фильмов | 1 | 2 | 3 | 4 | 5 |
| Частота | 10 | 25 | 40 | 20 | 5 |
| Накопленная частота | 10 | 35 | 75 | 95 | 100 |
| % | 10 | 35 | 75 | 95 | 100 |

В первой из дополнительных строк запишем значения так называемых "накопленных" (или кумулятивных) частот, которое будет равно 100.

Во-вторых, запишем в следующую графу, какой процент от 100

составляет каждое значение накопленных частот. В нашем случае значения третей и четвертой строк совпадают, так как частота равняется 100. Вообще могут получаться не одинаковые значения.

Попытаемся понять смысл полученного в последней графе результата.

При переходе от столбца со значением числа фильмов "2" к столбцу со значением "3" за плечами остается 35% всех результатов. А при переходе от столбца со значением "3" к столбцу "4" за плечами уже 75%. Медиана - это та точка, которая делит все данные в отношении 50: 50. Очевидно, требуемая точка где-то внутри столбца со значением "3". То есть 

Для нахождения среднего используется простая формула, смысл которой в том, чтобы сложить все значения (в нашем случае значения количества встреченных "отечественных" фильмов) и разделить полученный результат на число значений (в нашем случае 100).

Дальше можно идти двумя путями.

Во-первых, начать непосредственно складывать все 100 значений из первой таблицы.

Во-вторых, догадаться, что если некоторые значения количества просмотренных фильмов встречаются несколько раз, то можно воспользоваться данными из таблицы 1.2 и перейти от сложения повторяющихся значений к умножению этих значений на число повторов (например, число машин 1 встречается в первой таблице 10 раз, значит вместо 1 + 1 + 1 + 1+1+1+1+1+1 можно записать 1x10). Тогда:



Среднее значение получилось близким к трем.

Исходя их найденных значений моды, медианы и среднего, можно утверждать, что данное распределение можно считать нормальным, так как значения моды и медианы абсолютно равны, а значение среднего отличается на 0,15. Если взять большее количество испытаний, то среднее значение будет приближаться к трем.

На диаграмме 1, можно просмотреть, что закон распределения стремится к нормальному.



Диаграмма 1.

Теперь приступим к проверке гипотезы.

Н0: Вероятность встретить на сайте megogo.net фильм отечественного производства, равна вероятности встретить фильм, произведенный в зарубежных странах.

Н1: Вероятность встретить на сайте megogo.net фильм отечественного производства, не равна вероятности встретить фильм, произведенный в зарубежных странах.

Наш прогноз , выборка , средняя .

Меры центральной тенденции показывают, вокруг каких значений

группируется большинство экспериментальных данных. Обычно в качестве "центра" такого группирования рассматривается среднее (X).

Меры изменчивости говорят о том, в какой степени полученные результаты отклоняются от "центра группирования", что чаще всего приводит к определению меры отклонения экспериментальных данных от среднего. В принципе, в качестве меры изменчивости можно было бы использовать среднее значение отклонений текущих значений от среднего. Для этого необходимо определить, насколько каждое значение количества фильмов отклоняется в большую или меньшую сторону от X-2,895 затем сложить все результаты и разделить на число значений. К сожалению, этот путь невозможен, поскольку, как правило, отклонения от среднего в большую сторону (со знаком "+") и в меньшую сторону (со знаком "-") компенсируют друг друга и в сумме дают ноль. Для решения этой проблемы лучше использовать не отклонение от среднего, а квадрат этого отклонения, потому что такая процедура позволяет избавиться от влияния знака. Вначале делается та же операция - определяется, насколько каждое значение возраста отклоняется в большую или меньшую сторону от Х= 2,85. Затем каждый из полученных результатов возводится в квадрат, все складывается и делится на число значений. Получаемая таким образом мера изменчивости называется дисперсией. Данные приведены в таблице 1.7.

Таблица 1.7. Вычисление дисперсии

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Число фильмов | Частота | (Число фильмов-Среднее) | (Число фильмов - Среднее) 2 | (Частота) \* (фильмов - Среднее) 2 |
| 1 | 10 | -1,85 | 3,42 | 34,2 |
| 2 | 25 | -0,85 | 0,72 | 18 |
| 3 | 40 | 0,15 | 0,02 | 0,8 |
| 4 | 20 | 1,15 | 1,32 | 26,4 |
| 5 | 5 | 2,15 | 4,62 | 23,1 |

.

Подставим в эту формулу необходимые значения из таблицы



К сожалению, дисперсия оказывается не очень удобным показателем меры изменчивости.

Для решения возникшей проблемы вместо значения дисперсии используется квадратный корень из нее. Полученное таким образом новое значение называется стандартным отклонением и является наиболее популярной мерой изменчивости. Стандартное отклонение часто обозначается как  (сигма):



Для нашего случая:

.

Проверим гипотезу по формуле Стьюдента .

Выбирается уровень значимости a и критическая область А, так, чтобы условная вероятность попадания критерия в А при условии справедливости гипотезы равнялась a.

Выберем a=0.05 (вероятность не принять верный прогноз). По табл. распределения Стьюдента находим критическое значение (квантиль) (см. Приложение Таблица2)



Критерия

Ю Р{|t|>1,725}=0.05,Ю критическая область А:

|t|>1,725.





То есть прогноз должен быть отвергнут, так как отклонение от него в сторону повышения не является случайностью.

Гипотеза о том, что прогноз верен, проверена по критерию  на уровне значимости 5% и отвергнута.

# ***2. Сравнение размаха количества просмотренных фильмов отечественного и зарубежного производства***

# ***2.1 Жанры киноиндустрии***

В современном мире киноиндустрии существует множество жанров, согласно которым классифицируют все фильмы. Фильмы распределяют в зависимости, от эмоций, которые хотел вызвать у зрителей режиссёр данной картины.

Артхаус - фильм, нацеленный не на массовую аудиторию. Подобные фильмы прежде всего некоммерческие, самостоятельно сделанные фильмы, а также фильмы, снятые маленькими киностудиями или, так называемые, короткометражные фильмы.

Биография - жанр кинематографа, повествующий о судьбе знаменитой или известной личности. Фильм-биография пытается всесторонне передать жизнеописание личности или хотя бы годы, сыгравшие наиболее важную роль в жизни человека.

Боевик - жанр кинематографа, в котором основное внимание уделяется насилию: перестрелкам, дракам, погоням и т.п. Такие фильмы часто обладают высоким бюджетом, изобилуют каскадёрскими трюками и спецэффектами.

Военный - исторический художественный фильм, реконструирующий события реально происходившей войны или сражения, приёмы и организацию боя. В центре художественной композиции такого фильма обычно находится сцена главного сражения, съёмки которого сочетают широкие панорамные планы с крупными планами героев фильма.

Документальный - Документальным называется фильм, в основу которого легли съёмки подлинных событий и лиц. Зачастую фильмы опираются на документальные источники.

Драма - жанр кинематографа, который зачастую отображает преимущественно бытовую сюжетику и более приближенную к обыденной реальности стилистику. Драмы специфически изображают, как правило, частную жизнь человека и его конфликт с обществом.

Комедия - жанр кинематографа, характеризующийся юмористическим или сатирическим подходом, а также не стандартные пути решения проблемы которые помогают вызвать положительные эмоции о зрителя.

Криминал, гангстерский - категория фильмов об организованной преступности (гангстерских синдикатах), а также о профессиональных правонарушителях Приключение - жанр кино, в отличие от боевика, в приключенческих фильмах акцент смещён с грубого насилия на смекалку персонажей, умение перехитрить, обмануть злодея.

Фантастика - фильмы данного жанра преимущественно трёх тематических направлений: космические путешествия, появление инопланетных существ на Земле и положение человека в изменившихся условиях будущего

Фентези - произведения фэнтези чаще всего напоминают историко-приключенческий роман, действие которого происходит в вымышленном мире, близком к реальному Средневековью, герои которого сталкиваются со сверхъестественными явлениями и существами.

# ***2.2 Сравнения размаха количества просмотренных фильмов отечественных и зарубежных кинокомпаний***

За исходные данные было решено взять 130 фильмов, просмотренных за последнее время. Эти фильмы были разбиты по странам производителям и по жанрам. Было выбрано 7 жанров отечественных кинокомпаний и 7 зарубежных. Результаты выборки приведены в таблице 2.1

Таблица 2.1

|  |  |  |
| --- | --- | --- |
| Жанр | Страна выпуска | |
|  | Отечественные фильмы | Зарубежные фильмы |
| Артхаус | 3 | 7 |
| Биография | 12 | 4 |
| Военный | 15 | 5 |
| Документальный | 8 | 2 |
| Драма | 6 | 16 |
| Комедия | 9 | 19 |
| Сказки | 10 | 14 |

Используем тест экстремальных реакций Мозеса для сравнения размаха количества просмотренных фильмов отечественных и зарубежных кинокомпаний.

В тесте экстремальных реакций Мозеса одна выборка выступает в качестве контрольной группы, а другая - в качестве экспериментальной.

В качестве экспериментальной группы выбирается та, в которой, предполагается, что результаты будут носить более экстремальный характер. В нашем случае это группа зарубежных фильмов.

В качестве меры изменчивости, позволяющей сравнивать "ширину" обеих выборок, используется один из наиболее редко применяемых статистических показателей - размах. Размах S определяется как разность между максимальным и минимальным значением плюс единица: 

Размах для контрольной группы "отечественные фильмы": 

Размах для экспериментальной группы "зарубежные фильмы": 

Как видим, результаты в экспериментальной группе носят более экстремальный характер.

В тесте экстремальных реакций Мозеса размах (по отношению к рангам) вычисляется для контрольной группы. Затем, на основе значения размаха, сведений о размере выборок и других показателях, вычисляется вероятность получения такого результата.

Если полученное значение вероятности оказывается меньшим или равным выбранному значению уровня значимости , то нулевая гипотеза отвергается.

Если значение вероятности больше выбранного значения уровня значимости , то оснований отвергнуть нулевую гипотезу нет.

Применим тест экстремальных реакций Мозеса к нашим данным.

Выберем уровень значимости и сформулируем нулевую и альтернативную гипотезы.

: Страна производитель не влияет на количество просмотренных фильмов разных жанров.

: Количество просмотренных зарубежных фильмов носит более экстремальный характер по сравнению с количеством просмотренных отечественных фильмов.

Объединим обе выборки, расположив значения цен в порядке возрастания, присвоим каждому значению свой ранг и укажем кому - "отечественным фильмам" (L) или "зарубежным фильмам" (А) - принадлежит то или иное значение (табл. 2.2):

Таблица 2.2. Упорядоченные значения количества для объединенной выборки

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 14 | 15 | 16 | 19 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |

На основании результатов ранжирования определяем размах для контрольной группы зарубежных фильмов.

Как видно из таблицы, минимальное значение ранга в контрольной группе равно 4, максимальное значение ранга равно 13.

Размах .

Однако в статистике размах считается ненадежной мерой изменчивости, поскольку он слишком зависит от колебаний крайних значений. Для придания размаху большей стабильности в тесте экстремальных реакций Мозеса слева и справа отбрасывается несколько крайних значений, и размах вычисляется для оставшихся значений.

Обозначим как  то, сколько значений слева и справа будет отброшено. Для малых выборок  обычно равно 1. В общем случае рекомендуется слева и справа отсекать по 2,5% значений (то есть всего, слева и справа, отсекается не более 5%).

Отбросим от рангов для контрольной группы по одному крайнему значению слева и справа (то есть отбросим значения 4 и 13). Новое минимальное значение ранга для контрольной группы будет равно 5, а новое максимальное значение будет равно 9. Обозначим новое значение размаха как . Оно будет равно.

Можно показать, что при любых размерах экспериментальной и контрольной группы значение размаха  не может быть меньше, чем  и больше, чем .

Обозначим как  то, насколько наблюдаемое значение размаха больше минимально возможного.

В нашем случае минимально возможное значение размаха равно .

Поскольку -и , то .

После проделанной подготовки можно приступить к вычислению требуемой вероятности. Формула для ее вычисления выглядит следующим образом:

.

, если  и , если .

Подставляем в формулу наши данные: ; ; .



Полученное значение вероятности меньше выбранного значения уровня значимости а=0,05. Нулевая гипотеза отвергается, и принимается альтернативная. Количество просмотренных зарубежных фильмов носит более экстремальный характер по сравнению с количеством просмотренных отечественных фильмов.

# ***Заключение***

В результате исследования фильмов сайта megogo.net и количества просмотренных фильмов отечественных и зарубежных кинокомпаний, было установлено, что популярность зарубежных фильмов больше, чем отечественных. Это говорит о том, что фильмы зарубежных кинокомпаний набирают большое количество просмотров, по сравнению с фильмами отечественных производителей.

Оценка вероятности встретить "отечественный" фильм проходила двумя способами - с помощью параметрического и непараметрического методов исследования. В качестве параметрического метода использовался критерий Стьюдента, и разумеется данный метод более точен по сравнению с непараметрическим методом. Использовавшимся непараметрическим методом был тест  для единственной выборки. Так как в данном методе использовалась только одна выборка, данный тест дает менее точные данные. Так как выборка одна, то результат зависит от случая.

Сравнения размаха количества фильмов отечественных и зарубежных кинокомпаний показал, что размах зарубежных картин является более экстремальным, в сравнении с отечественными картинами. Это говорит о том, что нельзя утверждать о том, что фильмы одной страны лучше или хуже просматриваемы чем фильмы другой страны, так как каждый фильм несет в себе частичку эмоций, которую способен почувствовать зритель.

Окончательный выбор в любом случае останется за зрителем, но в результате исследования было установлено, что зарубежные фильмы более популярны, следовательно, их качество и бюджет выше, чем у фильмов отечественного производства.

# ***Список использованной литературы***

1. Резник А.Д. Книга для тех, кто не любит статистику, но вынужден ею пользоваться - СПб.: Речь, 2008. - 265 с.: илл.

2. Бочаров П.П., Печинкин А.В. Теория вероятностей. Математическая статистика. - М.: ФИЗМАТЛИТ, 2005. - 296 с.

. Кремер Н.Ш. Теория вероятностей и математическая статистика 2-е изд., перераб. И доп. - М.: Юнити-Дана, 2004. - 573 с.

. Гмурман В. Теория вероятностей и математическая статистика. - М.: "Высшая Школа", 2003. - 480 с.

. Тюрин Ю.Н. Непараметрические методы статистики. - М.: Знание, 1978. - 64 с.

. Правила ранжирования, http://psystat. at.ua/publ/1-1-0-34 (Актуальная дата: 26.11.2014 г.).

. Непараметрическая статистика и подгонка распределения, http://www.novsu.ru/file/109286 (Актуальная дата: 26.11.2014 г.).

. Вентцель Е.С. Теория вероятностей: - М.: Наука, 1969. - 576 с.

. Сидни Люмет. Как делается кино. - М.: "Высшая школа", 2003. - 300 с.:

. Вуколов Э. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STASTICA и EXCEL. - М.: "ИНФРА-М", 2004. - 464 с.

11. Справочная информация по сайту megogo https: // ru. wikipedia.org/wiki/Megogo.net/ <https://ru.wikipedia.org/wiki/Megogo.net/>

. Критерий Хи-квадрат, http://psystat. at.ua/publ/1-1-0-29 (Актуальная дата: 26.11.2014 г.).

. Тест Мозеса http://hr-portal.ru/spss/Glava14/Index4. php <http://hr-portal.ru/spss/Glava14/Index4.php>

. Александров Григорий <http://paraknig.ru/a/%D0%90%D0%BB%D0%B5%D0%BA%D1%81%D0%B0%D0%BD%D0%B4%D1%80%D0%BE%D0%B2%20%D0%93%D1%80%D0%B8%D0%B3%D0%BE%D1%80%D0%B8%D0%B9>, Жанры кино,-2010.

. Макаров Е. MathCAD: Учебный курс. - СПб.: Питер, 2009 г. - 384

# ***Приложение***

Таблица 1. Таблица критических значений для теста 

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| k /α | 0,01 | 0,025 | 0,05 | 0,95 | 0,975 | 0,99 |
| 1 | 6,63490 | 5,02389 | 3,84146 | 0,00393 | 0,00098 | 0,00016 |
| 2 | 9,21034 | 7,37776 | 5,99146 | 0,10259 | 0,05064 | 0,02010 |
| 3 | 11,34487 | 9,34840 | 7,81473 | 0,35185 | 0,21580 | 0,11483 |
| 4 | 13,2767 | 11,14329 | 9,48773 | 0,71072 | 0,48442 | 0,29711 |
| 5 | 15,08627 | 12,8325 | 11,0705 | 1,14548 | 0,83121 | 0,55430 |
| 6 | 16,81189 | 14,44938 | 12,59159 | 1,63538 | 1,23734 | 0,87209 |
| 7 | 18,47531 | 16,01276 | 14,06714 | 2,16735 | 1,68987 | 1,23904 |
| 8 | 20,09024 | 17,53455 | 15,50731 | 2,73264 | 2,17973 | 1,64650 |
| 9 | 21,66599 | 19,02277 | 16,91898 | 3,32511 | 2,70039 | 2,08790 |
| 10 | 23, 20925 | 20,48318 | 18,30704 | 3,94030 | 3,24697 | 2,55821 |
| 11 | 24,72497 | 21,92005 | 19,67514 | 4,57481 | 3,81575 | 3,05348 |
| 12 | 26,21697 | 23,33666 | 21,02607 | 5,22603 | 4,40379 | 3,57057 |
| 13 | 27,68825 | 24,7356 | 22,36203 | 5,89186 | 5,00875 | 4,10692 |
| 14 | 29,14124 | 26,11895 | 23,68479 | 6,57063 | 5,62873 | 4,66043 |
| 15 | 30,57791 | 27,48839 | 24,99579 | 7,26094 | 6,26214 | 5,22935 |
| 16 | 31,99993 | 28,84535 | 26,29623 | 7,96165 | 6,90766 | 5,81221 |
| 17 | 33,40866 | 30, 19101 | 27,58711 | 8,67176 | 7,56419 | 6,40776 |
| 18 | 34,80531 | 31,52638 | 28,86930 | 9,39046 | 8,23075 | 7,01491 |
| 19 | 36, 19087 | 32,85233 | 30,14353 | 10,11701 | 8,90652 | 7,63273 |
| 20 | 37,56623 | 34,16961 | 31,41043 | 10,85081 | 9,59078 | 8,26040 |
| 21 | 38,93217 | 35,47888 | 32,67057 | 11,59131 | 10,2829 | 8,89720 |
| 22 | 40,28936 | 36,78071 | 33,92444 | 12,33801 | 10,98232 | 9,54249 |
| 23 | 41,63840 | 38,07563 | 35,17246 | 13,09051 | 11,68855 | 10, 19572 |
| 24 | 42,97982 | 39,36408 | 36,41503 | 13,84843 | 12,40115 | 10,85636 |
| 25 | 44,31410 | 40,64647 | 37,65248 | 14,61141 | 13,11972 | 11,52398 |
| 26 | 45,64168 | 41,92317 | 38,88514 | 15,37916 | 13,84391 | 12, 19815 |
| 27 | 46,96294 | 43, 19451 | 40,11327 | 16,15140 | 14,57338 | 12,87850 |
| 28 | 48,27824 | 44,46079 | 41,33714 | 16,92788 | 15,30786 | 13,56471 |
| 29 | 49,58788 | 45,72229 | 42,55697 | 17,70837 | 16,04707 | 14,25645 |
| 30 | 50,89218 | 46,97924 | 43,77297 | 18,49266 | 16,79077 | 14,95346 |
| 31 | 52, 19139 | 48,23189 | 44,98534 | 19,28057 | 17,53874 | 15,65546 |
| 32 | 53,48577 | 49,48044 | 46, 19426 | 20,07191 | 18,29076 | 16,36222 |
| 33 | 54,77554 | 50,72508 | 47,39988 | 20,86653 | 19,04666 | 17,07351 |
| 34 | 56,06091 | 51,96600 | 48,60237 | 21,66428 | 19,80625 | 17,78915 |
| 35 | 57,34207 | 53, 20335 | 49,80185 | 22,46502 | 20,56938 | 18,50893 |
| 36 | 58,61921 | 54,43729 | 50,99846 | 23,26861 | 21,33588 | 19,23268 |
| 37 | 59,89250 | 55,66797 | 52, 19232 | 24,07494 | 22,10563 | 19,96023 |
| 38 | 61,16209 | 56,89552 | 53,38354 | 24,8839 | 22,87848 | 20,69144 |
| 39 | 62,42812 | 58,12006 | 54,57223 | 25,69539 | 23,65432 | 21,42616 |
| 40 | 63,69074 | 59,34171 | 55,75848 | 26,5093 | 24,43304 | 22,16426 |
| 41 | 64,95007 | 60,56057 | 56,94239 | 27,32555 | 25,21452 | 22,90561 |
| 42 | 66, 20624 | 61,77676 | 58,12404 | 28,14405 | 25,99866 | 23,65009 |
| 43 | 67,45935 | 62,99036 | 59,30351 | 28,96472 | 26,78537 | 24,39760 |
| 44 | 68,70951 | 64, 20146 | 60,48089 | 29,78748 | 27,57457 | 25,14803 |
| 45 | 69,95683 | 65,41016 | 61,65623 | 30,61226 | 28,36615 | 25,90127 |
| 46 | 71, 20140 | 66,61653 | 62,82962 | 31,43900 | 29,16005 | 26,65724 |
| 47 | 72,44331 | 67,82065 | 64,00111 | 32,26762 | 29,95620 | 27,41585 |
| 48 | 73,68264 | 69,02259 | 65,17077 | 33,09808 | 30,75451 | 28,17701 |
| 49 | 74,91947 | 70,22241 | 66,33865 | 33,93031 | 31,55492 | 28,94065 |
| 50 | 76,15389 | 71,42020 | 67,50481 | 34,76425 | 32,35736 | 29,70668 |

Таблица 2

Таблица критических значений Стьюдента

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| df\p | 0.40 | 0.25 | 0.10 | 0.05 | 0.025 | 0.01 | 0.005 | 0.0005 |
| 1 | 0.324920 | 1.000000 | 3.077684 | 6.313752 | 12.70620 | 31.82052 | 63.65674 | 636.6192 |
| 2 | 0.288675 | 0.816497 | 1.885618 | 2.919986 | 4.30265 | 6.96456 | 9.92484 | 31.5991 |
| 3 | 0.276671 | 0.764892 | 1.637744 | 2.353363 | 3.18245 | 4.54070 | 5.84091 | 12.9240 |
| 4 | 0.270722 | 0.740697 | 1.533206 | 2.131847 | 2.77645 | 3.74695 | 4.60409 | 8.6103 |
| 5 | 0.267181 | 0.726687 | 1.475884 | 2.015048 | 2.57058 | 3.36493 | 4.03214 | 6.8688 |
| 6 | 0.264835 | 0.717558 | 1.439756 | 1.943180 | 2.44691 | 3.14267 | 3.70743 | 5.9588 |
| 7 | 0.263167 | 0.711142 | 1.414924 | 1.894579 | 2.36462 | 2.99795 | 3.49948 | 5.4079 |
| 8 | 0.261921 | 0.706387 | 1.396815 | 1.859548 | 2.30600 | 2.89646 | 3.35539 | 5.0413 |
| 9 | 0.260955 | 0.702722 | 1.383029 | 1.833113 | 2.26216 | 2.82144 | 3.24984 | 4.7809 |
| 10 | 0.260185 | 0.699812 | 1.372184 | 1.812461 | 2.22814 | 2.76377 | 3.16927 | 4.5869 |
| 11 | 0.259556 | 0.697445 | 1.363430 | 1.795885 | 2.20099 | 2.71808 | 3.10581 | 4.4370 |
| 12 | 0.259033 | 0.695483 | 1.356217 | 1.782288 | 2.17881 | 2.68100 | 3.05454 | 4.3178 |
| 13 | 0.258591 | 0.693829 | 1.350171 | 1.770933 | 2.16037 | 2.65031 | 3.01228 | 4.2208 |
| 14 | 0.258213 | 0.692417 | 1.345030 | 1.761310 | 2.14479 | 2.62449 | 2.97684 | 4.1405 |
| 15 | 0.257885 | 0.691197 | 1.340606 | 1.753050 | 2.13145 | 2.60248 | 2.94671 | 4.0728 |
| 16 | 0.257599 | 0.690132 | 1.336757 | 1.745884 | 2.11991 | 2.58349 | 2.92078 | 4.0150 |
| 17 | 0.257347 | 0.689195 | 1.333379 | 1.739607 | 2.10982 | 2.56693 | 2.89823 | 3.9651 |
| 18 | 0.257123 | 0.688364 | 1.330391 | 1.734064 | 2.10092 | 2.55238 | 2.87844 | 3.9216 |
| 19 | 0.256923 | 0.687621 | 1.327728 | 1.729133 | 2.09302 | 2.53948 | 2.86093 | 3.8834 |
| 20 | 0.256743 | 0.686954 | 1.325341 | 1.724718 | 2.08596 | 2.52798 | 2.84534 | 3.8495 |