

Неволин И.В

Москва, ЦЭМИ РАН

О РОЛИ ХАРАКТЕРИСТИК РЕСПОНДЕНТОВ ПРИ ИССЛЕДОВАНИИ СПРОСА НА КИНОФИЛЬМЫ.

***Аннотация:** Статья показывает различия в публикуемых данных о демографических характеристиках кинозрителей. При этом данные, представляющие взрослое население страны, оказываются неадекватными, поскольку не учитывают значительный сегмент – зрители до 18 лет, которые, по оценкам из разных источников, составляют не менее 10% кинозрителей. Анализ методов прогнозирования кассовых сборов показывает, что информация из социальных сетей, хотя и генерируется нерепрезентативной выборкой, являются хорошим индикатором коммерческого успеха кинофильма. Объяснение того, что нерепрезентативная выборка позволяет предсказывать кассовые сборы, может состоять в том, что активность в социальных сетях, связанная с конкретным кинофильмом, коррелирует с рекламным бюджетом, и от того, насколько прицельной окажется реклама, зависит воздействие на аудиторию. Именно при подготовке рекламной кампании учитываются демографические и социальные характеристики зрителей, и поэтому лишь представители целевой аудитории напишут отзыв об ожидаемом фильме. Таким образом, учёт портрета зрителя лишь косвенно используется в современных методах прогнозирования кассовых сборов кинофильмов.*

Социально-экономическое прогнозирование активно привлекает массивы индивидуальных данных, формируя выборки для анализа. При этом для обобщения выводов актуален вопрос о репрезентативности выборки, т.е. о том, насколько точно доступные данные характеризуют генеральную совокупность. Когда известно распределение генеральной совокупности по наблюдаемым характеристикам, возможно построить репрезентативную выборку, и для этого существуют методики. Традиционно, знание о распределении внутри генеральной совокупности получалось исследователями в результате переписи того или иного вида. Интернет сделал возможным непосредственное сканирование сетевых сообществ и предоставил, тем самым, принципиально новую возможность работы с данными – работу с полным массивом данных о сообществе вместо репрезентативной выборки. При этом, однако, возникает проблема распространения выводы за пределы сетевого сообщества, поскольку остаётся неизвестной его место в генеральной совокупности «реальных», внесетевых сообществ. Под определением «реальный» здесь понимаются привычные на обывательском уровне сообщества людей как противоположность Интернет-сообществам. Даже тематические сообщества, в которых интернет-пользователи объединены вокруг одной предметной области, проявляют себя по-разному, и, видимо, имеют различные распределения пользователей по социальным характеристикам. Так, в одном из исследований [6] обнаружено, что пользователи веб-сайтов Internet Movie Database и Rotten Tomatoes – интернет-площадок, посвящённых кинофильмам, - по-разному оценивают одни и

те же фильмы. На сайте IMDb рейтинг фильма систематически ниже, чем на сайте Rotten Tomatoes. То есть аудитория IMDb в целом настроена более критически. При этом оба сайта имеют одинаковую направленность: информирование аудитории о произведениях киноиндустрии.

Многочисленные исследования указывают на то, что интернет-пользователи в целом отличаются от общей массы населения. Например, Фонд общественного мнения (ФОМ) регулярно проводит исследования, которые позволяют выявить демографические и поведенческие отличия интернет-пользователей (см., например [9]). В практической социологии можно встретить рекомендации ограничить онлайн-исследования населения пилотажными опросами, которые позволяют собрать набор мнений по обсуждаемой проблематике и являются качественными по своей сути [12]. При этом онлайн исследования предоставляют следующие преимущества перед непосредственными опросами населения: возможность недорогими средствами достичь аудитории в различных регионах страны, быстро обрабатывать поступающие данные и использовать в исследованиях мультимедийные возможности. Последнее оказалось привлекательным для исследования массового спроса на кинофильмы, и позволило уточнить методику выявления воздействия трейлеров (рекламных роликов кинофильмов) на эмоциональную сферу зрителей [11]. На этапе адаптации известной методики использовалась очень узкая аудитория, и при этом основным фокусом было удобство работы с методикой. Когда вопрос с адаптацией был решён, и удалось сократить количество вопросов, исследование перешло в Интернет, где можно достичь более широкой аудитории, и встраивание видео-роликов в опрос оказалось критически важным. Вместе с тем, в качестве практического применения методика предполагает возможность содержательных выводов о формировании отношения зрителей к кинофильму по трейлеру и о потенциальном объёме кассовых сборов. В этом контексте знание демографических и социальных характеристик, в перспективе, имеет два важных значения. Во-первых, они позволят оценить, насколько исследуемая выборка отличается от общей массы кинозрителей и, во-вторых, позволят сформировать рекомендации относительно сегмента зрительской аудитории, на которой стоит сфокусироваться кинопродюсерам и прокатчикам. Знание социальных и демографических характеристик зрителей чрезвычайно важно для киноиндустрии, поскольку является некоторым предиктором потенциального размера зрительской аудитории конкретного фильма и позволяет спланировать кинопрокат: сколько копий, в каких регионах, в какое время выводить на экраны. С другой стороны, информация о том, какая аудитория интересуется тем или иным фильмом позволяет прокатчикам точнее выбирать рекламу, которая показывается зрителям перед сеансом – имеются данные, что реклама и продажи в баре составляют, соответственно, 20-25% и 20-30% выручки кинотеатров [7].

Итак, характеристики аудитории представляют коммерческую ценность для отрасли и научный интерес для исследователей, но возникает вопрос о том, как их получить. На рынках потребительских товаров и услуг, и в киноиндустрии в частности, они чрезвычайно полезны, но не настолько, чтобы заинтересовать государство в проведении масштабной переписи. Бизнесу остаётся лишь периодически проводить опросы своих клиентов, чтобы узнать потребителя. Полезными оказываются данные клубных карт, для получения которых необходимо пройти анкетирование. Ещё одним источником информации служат банковские транзакции – по истории платежей видно, кто у кого совершал покупки – в частности, приобретал билеты на киносеанс. Но эта информация совершенно недоступна. В настоящей статье предпринята попытка сравнить характеристики аудитории, принявшей участие в онлайн-исследовании, с посетителями кинотеатров и населением в целом.

В рамках исследования предпочтений кинозрителей разработана методика оценки восприятия трейлера (рекламного ролика) кинофильма [11]. Когда методика прошла пилотную апробацию, решено перенести её в Интернет, где можно привлечь более широкую аудиторию. Для этого создан специальный веб-сайт, а само исследование рекламировалось через сообщества социальной сети ВКонтакте. Рекламные кампании исследования проводились в январе и в феврале 2014 года, и для отслеживания посещаемости веб-сайта исследования использовался сервис Яндекс.Метрика¹. Данный сервис позволяет собирать статистику доступа к сайту (количество подключений, просмотренных страниц, число отказов и т.п.), а также извлекать характеристики пользователей: пол, возраст, коммерческие интересы, регион. Характеристики пользователей определяются с помощью специально разработанного алгоритма классификации Матрикснет [10], который постоянно обновляется, чтобы поспевать за динамичными изменениями в сети.

В период анкетирования Яндекс.Метрика зафиксировала 1367 и 286 посещений, соответственно, в январе и феврале 2014г. Причём интерес аудитории спадает чрезвычайно быстро: самая большая активность в первый час, затем, по-видимому, реклама вытесняется из новостной ленты более новыми сообщениями. На Рисунке 2 пик в районе 18 часов соответствует старту рекламной кампании 17 января 2014г.

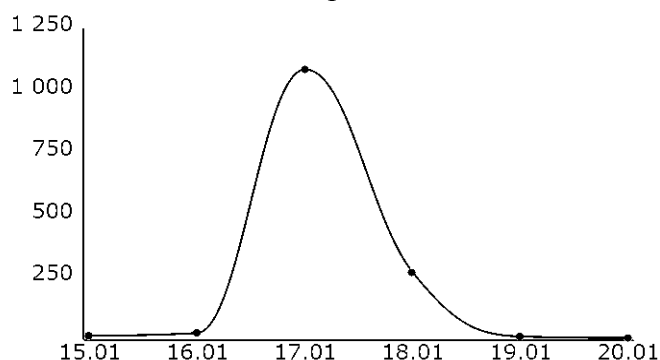


Рисунок 1. Динамика числа визитов во времени. По оси абсцисс – дни 15-20 января 2014г., по оси ординат – количество посетителей страницы. Источник: Яндекс.Метрика для сайта исследования.

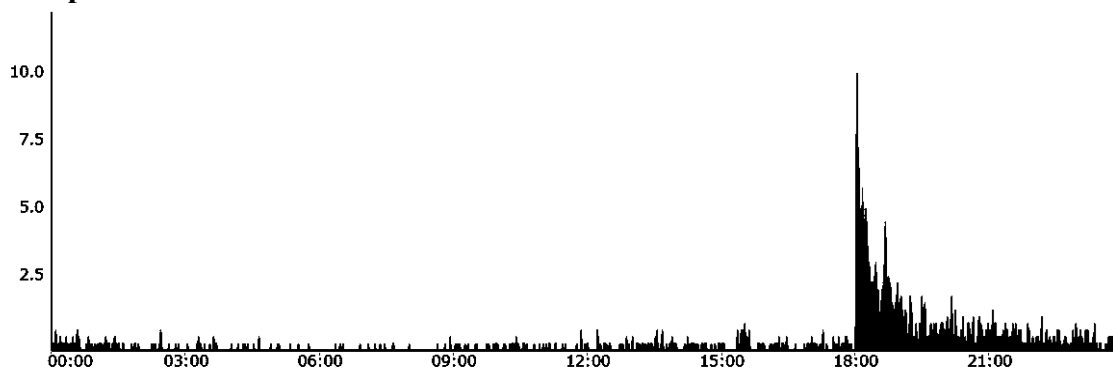


Рисунок 2. Средняя нагрузка на главную страницу сайта в течение суток. По оси абсцисс – время, по оси ординат – среднее количество запросов к веб серверу, на котором хранится интернет-страница. Источник: Яндекс.Метрика для сайта исследования.

Яндекс.Метрика классифицирует пользователей по демографическим показателям, регионам и коммерческим интересам. Таблица 1 иллюстрирует демографические и региональные характеристики посетителей сайта во время каждого анкетирования. Именно

¹ <http://metrika.yandex.ru>

эти характеристики и необходимо сравнить с данными из других источников. Для краткости данные опроса января 2014г. названы янв.14, а февраля 2014г. – фев.14.

	янв.14	фев.14
Пол		
Мужской	46,8	67,7
Женский	53,2	32,3
Возраст		
До 18	16,20	5,0
18-24	45,7	36,8
25-34	28,8	47,5
35-44	5,6	4,6
45 и старше	3,7	6,1
Регион		
Москва и Санкт-Петербург	28,6	55,0
Регионы	50,2	33,2
Другие страны	21,2	11,8

Таблица 1. Распределение посетителей (в процентах) веб-сайта исследования кинозрителей. Источник: Яндекс.Метрика для сайта исследования.

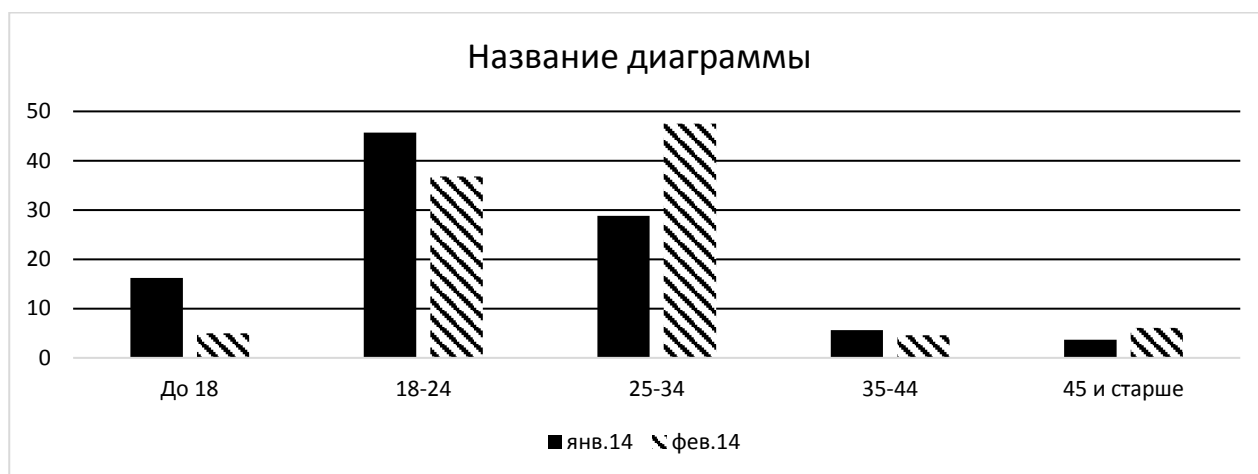


Рисунок 3. Сравнение распределения посетителей сайта исследования по возрастным группам. Источник: Яндекс.Метрика для сайта исследования.

Исследования демографических характеристик посетителей кинотеатров, а также их сопоставление с таковыми для интернет-аудитории, - достаточно редкое явление. Известны обследования Невафильм [13] и Фонда общественное мнение [9]. Также доступны результаты наблюдений кинотеатральной сети [7].

Согласно данным Невафильм (см. Таблицу 2), профили аудиторий – посетителей кинотеатров и интернет-пользователей, - заметно отличаются. Так, например, в случае интернет-аудитории сокращается, хотя не исчезает, перевес женщин, Москва и Санкт-Петербург играют значительную роль в общей картине. При этом заметно, что классификация респондентов Невафильм по возрастным группам отличается от той, которая используется Яндекс.Метрика. Для краткости выборки Невафильм посетителей кинотеатров зимой 2012г. и весной 2013г. обозначены, соответственно, теарт-12, театр-13, а пользователей Рунета - рунет-12, рунет-13.

	Посетители кинотеатров		Пользователи Рунета	
	Зима 2012	Весна 2013	Зима 2012	Весна 2013
Пол				
Мужской	37,1	35,5	43,1	43,6
Женский	62,9	64,5	56,9	56,4
Возраст				
12-17	15,0	13,9	5,8	11,6
18-24	40,0	44,7	50,2	52,0
25-29	22,4	21,6	24,3	19,5
30-39	15,8	14,9	16,9	12,3
40-49	5,0	3,7	2,4	2,6
50 и более	2,0	1,3	0,5	2,0
Место жительства				
Москва и Санкт-Петербург	15,9	8,0	41,5	31,0
Регионы	84,1	92,0	37,9	55,1
Другие страны	-	-	20,6	13,9

Таблица 2. Российские кинозрители. Источник: Невафильм

Ценным является исследование Фонда общественного мнения [9]: в июне 2013г. проводился опрос, в котором присутствовала и такая формулировка: «Скажите, пожалуйста, как часто вы смотрите кинофильмы в кинотеатре, киноклубе?» Установлено, что 66% респондентов в возрасте 18-30 лет ходят в кино, по крайней мере, один раз в месяц, в возрасте 31-45 лет – 39%, 46-60 лет – 18% и, наконец, старше 60 – 4%. Каковы же пропорции возрастных групп посетителей кинотеатров? Основываясь на данных того же опроса, можно рассчитать, какая доля опрошенных приходится на те или иные возрастные группы респондентов: 18-30 лет – 28%, 31-45 лет – 29%, 46-60 лет - 11%, старше 60 – 32%. Тогда можно рассчитать возрастной профиль посетителей кинотеатров. Аналогично, пол посетителей можно получить по данным официальной статистики, на основе которых ФОМ формирует выборки для своих исследований. Характеристики посетителей кинотеатров, рассчитанные по данным ФОМ и Росстата², приведены в Таблице 3. Для краткости выборка Фонда общественного мнения обозначена ФОМ-13.

Пол	
Мужской	44,9
Женский	55,1
Возраст	
18-30	55,9
31-45	34,2
46-60	6,0
старше 60	3,9

Таблица 3. Группировка посетителей кинотеатров на основе данных ФОМ. Источник: расчеты автора по данным ФОМ и Росстата.

Из представленных таблиц 1-3 видны различия группировки: Яндекс, ФОМ и Невафильм объединяют пользователей/ респондентов в различные возрастные группы. Данные можно было бы сравнить с наблюдениями Росстата по возрастным группам, как с некоторой объективной шкалой. Однако бессмысленность такого подхода состоит в том, что аудитория кинотеатров сильно отличается от профиля населения в целом, как показывает исследование ФОМ. Сравнению поддаются лишь распределения по половой структуре, и можно заметить, что выборки фев.13, театр-12 и театр-13 сильно отличаются от ФОМ-13.

² Данные о половозрастной структуре населения за 2013г. взяты с сайта gks.ru.

Для количественной оценки различий используется методика на основе кривых Лоренца, которая позволяет сравнивать разнообразные числовые распределения [1]. Эта методика позволяет оценить устойчивость распределений по числовым показателем (в данном случае по возрастным группам) и позволяет выявлять дисбаланс в системах, которые характеризуются числовыми рядами.

Суть метода, основанного на кривых Лоренца, заключается в рассмотрении накопленных сумм упорядоченного числового ряда. Рассматривая числовой ряд как характеристику статистической системы, ему можно поставить в соответствие единственную числовую характеристику. Если S_n/S_N – относительная накопленная сумма n элементов упорядоченного ряда N чисел, то кривая Лоренца может быть аппроксимирована уравнением вида:

$$\left(1 - \frac{S_n}{S_N}\right)^\alpha + \left(\frac{n}{N}\right)^\alpha = 1,$$

где α – параметр характеризующий пропорции распределения. При этом показано, что система наиболее сбалансирована при $\alpha=1,8$. Для рассматриваемых числовых рядов рассчитаны значения параметра α , и результаты представлены в таблице.

	театр-12	театр-13	рунет-12	рунет-13	янв.14	фев.14	ФОМ-13
α	1,6	1,6	2	1,9	1,7	1,8	0,6

Таблица 4. Значения параметра для различных выборок. Источник: расчёты автора.

По параметру α близки распределения театр-12, театр-13 и янв.14, и лишь данные ФОМ-13 сильно выделяются из общей картины. Объяснение состоит в том, что в опросах ФОМ принимают участие граждане старше 18 лет, однако, наблюдения показывают высокую долю несовершеннолетних посетителей кинотеатров: более 10%, согласно данным Невафильм. Таким образом, можно сделать вывод, что к результатам ФОМ необходимо относиться очень осторожно при изучении предпочтений кинозрителей. Уже упомянутое обследование ФОМ [9] показывает сильные предпочтения по отношению к телевизору при выборе источника кино. Таким образом, данные ФОМ, вероятно, применимы для анализа телевизионных потребностей аудитории, но ни в коем случае не применимы для анализа посетителей кинотеатров. Сравнение распределений по возрастным группам наглядно иллюстрирует Рисунок 4.

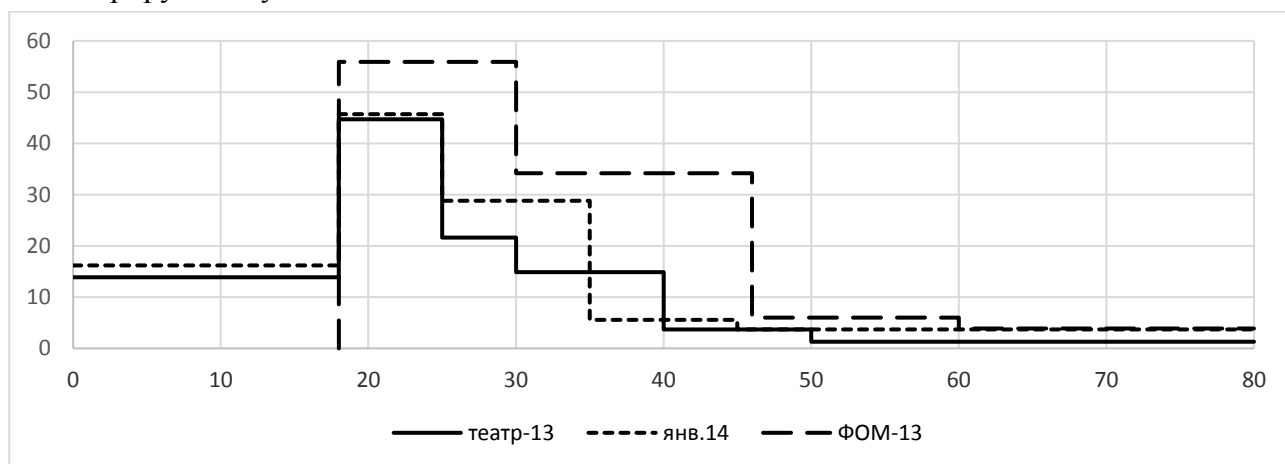


Рисунок 4. Распределения выборок по возрастным группам. По оси абсцисс отложен возраст, по оси ординат – доля респондентов выборки в возрастной группе. Границы возрастных групп указаны в соответствии с Яндекс.Метрика, Невафильм Research и ФОМ.

Анализ кривых Лоренца показывает стабильность данных о посетителях кинотеатров, а также, что данные янв.14г. близки к аудитории кинотеатров. Однако следует проявить осторожность в выводе о том, что интернет-аудитория подходит для изучения посетителей кинотеатров. Хотя демографические показатели оказались сопоставимыми для январской выборки, сетевое сообщество имеет сильные социальные различия: исследования показывают, что интернет-пользователи моложе и однороднее, чем общество в целом – средний возраст и дисперсия ниже, проявляются отличия первых по занятости и уровню дохода, региональные различия стираются – важнее становится, есть доступ к сети или нет [8]. В целом, для опросов через Интернет характерно следующее: они отражают мнение активной части общества, и сами по себе не являются репрезентативными.

Практика показывает различия слоёв общества в предпочтениях относительно каналов доступа к кино, а также самих фильмов – демографические и социальные характеристики являются значимыми предикторами выбора. Между тем, интернет-аудитория, хотя и не является репрезентативным представителем всего общества, широко используется для прогнозирования кассового успеха кинофильмов (см., например, [2,5]). Развитие Интернета, распространение социальных сетей открыли перед исследователями возможности к доступу и обработке огромного массива данных. Так, отзывы, рецензии и комментарии в интернете весьма успешно используются для прогнозирования кассовых сборов: на основе классификации сообщений пользователей по настроениям удаётся построить математические модели прогнозирования высокой точности. С другой стороны, типичным является подход к прогнозированию кассовых сборов на основе характеристик кинофильмов - актёрского состава, режиссёра, жанра и т.п., - который также показывает хорошую точность прогнозирования (см, например, работу [4] о воздействии жанра, актёрского состава и отзывов на кассовые сборы). Между тем, ни первые методы, ни вторые не используют характеристики пользователей/ зрителей ни в качестве объясняющих переменных, ни при построении выборки сообщений для анализа. Однако производители и прокатчики отмечают ценность знаний о том, кто является их зрителем. Решение указанного противоречия – важность демографического и социального портрета зрителя, но его отсутствие в моделях прогнозирования, - кроется в том, что режиссёры и продюсеры учитывают портрет в производстве и рекламе кинофильма. Сценарий, актёрский состав и прочие атрибуты подбираются под конкретную аудиторию. Далее, рекламные ролики, плакаты и площадки их размещения выбираются, исходя из предпочтений целевой аудитории. Данные социальных сетей, как демонстрирует работа [5], могут использоваться для построения модели сарафанного радио (word of mouth) и, таким образом, являются индикатором воздействия на поклонников кино – для анализа выбираются сообщения всех пользователей в которых содержится название того или иного фильма. С другой стороны, в той же работе отмечается связь между дневным рекламным бюджетом и активностью в социальных сетях. Также в [3] продемонстрирована связь между продажами билетов в кино, рекламным бюджетом, отзывами прессы и зрителей. Таким образом, можно предположить, что характеристики кинозрителей учитываются при подготовке рекламной кампании, и большой рекламный бюджет позволяет, с одной стороны, подготовить качественные материалы, с другой – для доступа к целевой аудитории направить их по наиболее эффективным каналам. И тогда воздействие на зрителя вполне объяснимо коррелирует с его отзывами в социальных сетях. В этой связи стоит надеяться, что продолжающиеся исследования воздействия трейлеров кинофильмов на эмоциональную сферу зрителя прольют свет на механизмы привлечения и удержания внимания, а также на принятие решения о просмотре фильма.

Список использованной литературы:

1. Antoniou I., Ivanov V.V., Korolev Yu.L., Kryanev A.V., Matokhin V.V., Suchanecki Z. Analysis of resources distribution in economics based on entropy // *Physica A.* – 2002. – Vol. 304 – P. 525-534
2. Asur S., Huberman B. A. Predicting the future with social media / *Proceedings IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology.* - 2010. – pp.492-499.
3. Beuscart J. S., Mellet K. The impact of advertising on box office. An empirical analysis of the French motion picture industry / *Working Paper, ACEI, 2012.*
4. Desai K. K., Basuroy S. Interactive influence of genre familiarity, star power, and critics' reviews in the cultural goods industry: The case of motion pictures / *Psychology & Marketing.* – 2005. – Vol. 22. – Issue. 3. – pp. 203-223.
5. Ishii A. et al. The «hit» phenomenon: a mathematical model of human dynamics interactions as a stochastic process // *New journal of physics.* – 2012. – Vol. 14. – Issue. 6. – 063018.
6. Wong F. M. F., Sen S., Chiang M. Why watching movie tweets won't tell the whole story? / *Proceedings of the 2012 ACM workshop on Workshop on online social networks.* – ACM, 2012. – p. 61-66.
7. Гладких И.В., Михайлина А.П. «Кронверк Синема»: сколько стоит билет в кино? (учебный кейс) / *Вестник Санкт-Петербургского университета. Серия 8: Менеджмент.* - 2011. - №3. - с.145-159.
8. Звоновский В.Б. Исследование возможности исследования онлайн опросов для репрезентации больших социальных групп // *Онлайн исследования в России 2.0 / Под редакцией Шашкина А.В., Девятко И.Ф., Давыдова С.Г.* – М.: РИЦ «Северо-восток», 2010. с.249-260.
9. Кино: что и как смотрят россияне. Телефонный опрос «ТелеФОМ» 22-23 июня 2013г. URL <http://fom.ru/posts/download/11192> (дата обращения 09.10.2014г.)
10. Кураленок И., Щекалев А. GPU в задачах машинного обучения / *Открытые системы. СУБД.* – 2013. - №8 – с.44-46.
11. Ноакк Н.В., Неволин И.В., Татарников А.С. Методика прогнозирования выручки от проката кинофильмов // *Финансовая аналитика: проблемы и решения.* – 2012 – №48(138). – с.17-24
12. Петер Д. Качественные онлайн исследования // *Онлайн исследования в России 2.0 / Под редакцией Шашкина А.В., Девятко И.Ф., Давыдова С.Г.* – М.: РИЦ «Северо-восток», 2010. с.151-168.
13. Пользователи Рунета предпочитают торренты, а жители малых городов недовольны качеством кинотеатров. Исследование Невафильм Research, 2013. URL http://research.nevafilm.ru/reports/cinemagoers/2stage/russian_spectators_2.pdf (дата обращения 09.10.2014г.)
14. Россияне о кино: между чувством и долгом. Еженедельный опрос «ФОМнибус» 29-30 сентября 2012г. URL <http://fom.ru/posts/download/10788> (дата обращения 09.10.2014г.)