Министерство просвещения Российской Федерации

Полное наименование учебного заведения: Назарбаев Интеллектуальная школа физико-математического направления (НИШ)

Полное наименование конкурса: VI Международный конкурс исследовательских работ школьников "Research start" 2024/2025

Форма конкурсной работы

**Название конкурсной работы:**

**Статистическое исследование общественного мнения**

Выполнил: Баширов Даниал Самигуллаевич

ученик 12 класса

Руководитель: Аскарова Нургуль Мырзагуловна

учитель математики

2024/25 учебный год

**Введение.**

Мой любимый предмет в школе – математика. Одним из самых интересных разделов математики является статистика, разрабатывающая методы регистрации, описания и анализа данных наблюдений и экспериментов с целью построения вероятностных моделей массовых случайных явлений.

Математическая статистика используется для представления данных в форме графиков и таблиц, с которыми очень удобно работать для прогнозирования и проверки гипотез – предположений, которые можно подтвердить или опровергнуть.

На первый взгляд, кажется, что математическая статистика важна только для точных наук. Но это не так. Статистика используется во всех научных дисциплинах. Методы и модели статистики эффективны в социально-экономических, управленческих, технических исследованиях, в медицине, истории, в любой прикладной отрасли и области знания. Успешно используется статистика и в социологии.

Социология – одна из наиболее важных общественных наук. Стоит отметить, что математические методы наиболее активно стали использоваться социологами с 70-х годов прошлого столетия. Теория измерений и нечеткие множества, математические методы классификации и многомерное шкалирование, непараметричекая статистика и статистика нечисловых данных - эти способы стали применяться социологами уже тогда.

 Цели моего проекта:

 - доказать, что математическая статистика важна и эффективна для выявления общественного мнения;

 - использовать и пополнить свои знания по разделу «Статистика»;

 - разобрать и понять методы выявления общественного мнения;

 - изучить подробно один из методов получения общественного мнения (опрос);

 - узнать, каким должен быть размер выборки (количество людей, которое должно ответить на опрос), и как данный показатель вычисляется;

 - рассмотреть практику и порядок проведения исследований опытными социологами и социологическими центрами;

 - исследовать некоторые опросы, проводимые студентами НИШ;

 - провести на практике исследования с целью проверки гипотез.

 Свой проект я считаю актуальным, поскольку на современном этапе развития общества роль общественного мнения неуклонно возрастает. Чтобы найти правильное решение по той или иной проблеме, необходимо знать реальное состояние масс, интересы, настроения, чувства отдельных социальных групп и общества в целом. В то же время, невозможно получить необходимую и достоверную информацию без математически точного исследования, для чего нам необходима статистическая грамотность.

 В большинстве стран мира статистическое образование объявлено национальным приоритетом, так как в условиях глобализации и модернизации экономики граждане должны владеть не только иностранными языками и компьютерными технологиями, но и ориентироваться в информационном поле, понимать, какие социально-экономические и политические процессы происходят в стране и целом в мире.

 Работая над проектом, я изучил много литературы по темам общественного мнения и математической статистики, ознакомился с методами получения общественного мнения, узнал и проверил на практике необходимые формулы и вычисления.

1. **Понятие общественного мнения. Его виды и значение.**

На сегодняшний день мировое сообщество стремится стать более демократичным. У любого человека появляется возможность высказать своё мнение по какому-либо вопросу. На основе точек зрения каждого индивидуума и формируется общественное мнение, исследованием которого активно занимаются социологи.

 Согласно общепринятому определению, общественное мнение - форма массового сознания, в которой проявляется отношение различных групп людей к событиям и процессам действительной жизни, затрагивающим их интересы и потребности. Роль общественного мнения достаточно велика, ведь оно способно манипулировать человеческим сознанием и повлиять на человека с целью изменения его поведения или внесения в его поведение некоторых изменений. Эти изменения могут повлиять на человека таким образом, что он будет воспринимать социальную реальность так, как это выгодно тем, кто занимается манипуляциями.

Отражение общественного мнения как социального института демонстрирует возможность гласного, публичного высказывания населения, а также развитие общественно-политических отношений.

 Существуют несколько видов общественного мнения.

* Оценочное. Оно отражает отношение к тем или иным социальным факторам. Благодаря этому способу, опрашиваемый может дать оценку какому-либо явлению или процессу, а исследователь получает возможность узнать, как данный объект работает на деле.
* Конструктивное. Здесь после определения мнения населения возникает необходимость принимать какие-то меры по решению поднимаемых злободневных проблем общественной жизни.
* Аналитическое. В этом случае, на основе общественного мнения проводится анализ полученных результатов, исследователями выделяются ключевые аспекты, имеющие огромное влияние на данную проблему.
* Регулятивное. Данный вид требует внедрения определённых норм общественных отношений.

Ниже в Таблице представлены основные функции, которые выполняет общественное мнение.

|  |  |
| --- | --- |
| **Функция** | **Разъяснение** |
| Экспрессивная  | Выражает точку зрения различных групп населения |
| Познавательная | Дает информацию о позиции, которую выражает народ |
| Консультативная | Дает варианты разрешения социальных проблем |
| Директивная | За выполнением общественного мнения следуют реальные действия |

 Таким образом, общественное мнение выполняет довольно важные для развития общества и политической системы функции.

**2 .Опросы как метод выявления общественного мнения.**

Существует множество форм изучения общественного мнения. К ним можно отнести митинги и забастовки, брифинги и пресс-конференции, предвыборные кампании, встречи с представителями общественных движений и так далее. Кроме прямого наблюдения общественного мнения можно выделить анализ документов и опросы.

Самым распространённым и популярным методом, который я собираюсь исследовать в данном проекте, являются опросы. Опросы общественного мнения проводятся для количественного определения общественного мнения, а определять количественные показатели без математики невозможно. Обычно прохождение опросов не занимает большого времени у населения, ибо в сравнении с другими формами социологических исследований (интервью, подкаст и др.) они не требуют развёрнутого ответа на вопрос. Зачастую, при ответе на тонкий вопрос опрашиваемый просто выбирает правильный ответ из нескольких предложенных (таких как - доволен / не доволен, не против / против и т. д.). При опрашивании определенного количества человек, итоги опроса записываются в виде процентов, после чего на основе этих цифровых результатов исследователи выводят общественное мнение. Стоит отметить, что после получения результатов обычному человеку интересно бывает узнать, совпадает ли его мнение с мнением всего населения. На результаты опросов нередко ссылаются и депутаты, поддерживающие спорные законопроекты, а сами опросы иногда способны играть даже роль референдумов (решений, принятие которых отдаются народу).

При этом сами опросы не следует путать с выборами. Если опросы проводятся под руководством социологов, то выборы – под эгидой правительства. Социологи обычно предпочитают проводить опросы до и после выборов, чтобы сопоставить итоги с результатами, полученными в результате опросов. В дальнейшем, благодаря полученным результатам, социологи определяют, сфальсифицированы результаты выборов или нет. Соответственно, если итоги выборов существенно отличаются от результатов социологических опросов, у населения возникает много вопросов и сомнений касательно прозрачности и честности проведенных выборов.

В то же время необходимо учитывать, правильно ли проводился сам опрос. Так, в последнее время можно заметить, что мнение обычного человека не всегда совпадает с мнением всего населения, которое преподносится нам социологами. Это, в свою очередь, приводит к потере доверия к таким опросам и нежеланию человека к участию в них.

В нашей школе ученики также любят проводить опросы. Опросы являются важной частью проектной (или курсовой) работы, и во время защиты проекта ученики должны делать акцент на результаты своих исследований. Проектная работа является обязательным аспектом по таким предметам, как КСМ (Казахстан в современном мире) и ГППР (Глобальные перспективы и проектная работа). Проекты по данным учебным предметам направлены на исследование и решение какой-либо глобальной проблемы или проблемы страны, поэтому для учащихся важно изначально правильно определить детали поднимаемой проблемы, а затем грамотно и внимательно провести исследование (в том числе и опрос).

Таким образом, существует множество методов социологических опросов, в том числе заочных. Студенты НИШ часто предпочитают наиболее удобные и доступные для них онлайн - опросы.

На сегодняшний день разработано множество приложений по созданию опросов (например, Google Forms, Microsoft Forms и т. д.), которые способны легко создавать диаграммы, иллюстрирующие процентное соотношение людей, голосовавших за тот или иной вариант ответа. С помощью этих приложений достаточно просто получить результаты опроса.

В то же время, при проведении онлайн-опросов, студенты иногда допускают ошибки, из которых и вытекают недостатки этого метода исследования. Поэтому, в своём исследовании я собираюсь математическим путем выявить имеющиеся недоработки и понять, как можно улучшить качество проведения опросов так, чтобы их результаты действительно могли отражать реальность.

**3.** **Исследование: как улучшить процесс проведения опроса**

Своё исследование я начал с изучения статистических аспектов, необходимых при проведении опроса. Я узнал, как определяется размер выборки тех, кто будет принимать участие в опросе. Я определил главное - выборка должны быть репрезентативной, т. е. соответствовать характеристикам генеральной совокупности.

Очень часто при проведении опросов нам говорят, что мы не учитываем мнение всего населения, к которому относится какой-либо опрос. Но учесть мнение всего населения очень тяжело, да и невозможно. Так, при проведении всеобщих казахстанских опросов генеральной совокупностью является всё население страны, а это 20 млн. человек. Опросить каждого никак не получится – это займёт слишком много времени. Также обязательно найдутся люди, которые не захотят проходить опрос либо не способны участвовать в нем ввиду языковых или иных ограничений. Поэтому выборка должна составляться так, чтобы она была репрезентативна, и вероятность попадания ее в доверительный интервал была высокой. Такая выборка называется случайной, но она имеет цель максимально отразить мнение всего населения.

Рассмотрим несколько видов выборок:

*Районированная выборка*. Здесь исследователь разделяет генеральную совокупность на однородные группы (например, по возрасту, полу, физическим возможностям). Иногда данные группы должен определить сам исследователь, поскольку тема проекта может оказывать на эти факторы глубокое влияние. Затем формируются простые случайные выборки из каждой группы.

*Серийная выборка*. В данном случае генеральная совокупность делится на малые группы (серии), после чего формируется простая случайная выборка этих групп, и отобранные группы рассматриваются полностью. Применение подобной процедуры является выгодным, когда затраты на поездку между случайно выбранными элементами слишком высоки.

*Систематическая выборка*. Она начинается со случайно выбранного элемента, после которого в выборку будет попадать каждый k-й элемент. Например, при исследовании движения на автомагистрали можно будет проверить каждый сотый автомобиль на пункте оплаты. Это удобно реализовывать, если число машин плавно меняется с течением времени.

Если использовать вышеперечисленные виды выборок, то мы получим нормальное распределение и у нас появится высокая вероятность того, что полученный результат будет соответствовать ожидаемому.

При нормальном распределении вероятность часто берётся как 95%. Это означает, что если у нас есть уровень достоверности 95% и мы бы провели опрос 100 раз, то он дал бы точно такие же результаты в 95 случаях из 100.

Для каждого значения вероятности есть своё значение функции нормального распределения Гаусса (для 95% это значение будет равно 1,96). Данное значение называется z-баллом. Для z-преобразования существует следующая формула со средним значением μ и стандартным отклонением σ.

$ z=\frac{x-μ}{δ}$

Определить значение z-балла для любого нормального распределения можно по общей таблице для стандартного нормального распределения.



Чем выше вероятность, тем выше будет z-балл. Соответственно, необходимый размер выборки прямо пропорционален данному z-баллу. При таких условиях, для наибольшего доверия к результатам, нам надо опросить как можно больше респондентов.

Помимо уровня достоверности, важную роль играет доверительный интервал. Он также выражается в процентах. С помощью него мы можем узнать предел погрешности, а именно тот интервал, где может находиться полученный нами результат. Например, при результате 55% и погрешности 3% наш полученный результат X будет находиться в промежутке 52%<X<58%. Значение размера выборки обратно пропорционально значению доверительного интервала, поэтому для того, чтобы погрешность была маленькая, размер выборки должен быть больше. Но это не означает, что в целях исключения погрешности мы должны опрашивать тысячи людей. Наоборот, с увеличением размера выборки, наша погрешность будет очень медленно уменьшаться, а отклонений станет больше. В итоге это будет не выгодно, как экономически, так и математически.

В нашем случае достаточно взять в основу такие допустимые погрешности, как 2-5 %, которые на результат сильно не повлияют.

На графике ниже видно, как изменяется значение допустимой ошибки в зависимости от размера выборки.

Существует ещё один показатель, который важно учитывать – доля выборки. Точность результатов исследования здесь также варьируется в зависимости от процента выборки, выбравшей тот или иной ответ. Если 97% населения выбирают "да", а 3% - "нет", то вероятность ошибки невелика. Но если 70% населения выберут "да", а 30% - "нет", - вероятность ошибки будет выше, независимо от размера выборки.

При формировании размера выборки из n человек, необходимого для заданного уровня точности, исследователям рекомендуется использовать процент с наихудшим вариантом, т.е. 50%, так как предугадать заранее, сколько людей выберет тот или иной вариант ответа, невозможно.

Каждый голосующий – это результат испытания по схеме Бернулли с некоторой вероятностью голоса «за» (p) и вероятностью голоса «против» (1-р). Исходя из этого, стандартное отклонение будет равно:

$$δ=\frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

 Для больших значений n у нас получится нормальное распределение.

 Выведем общую формулу для размера выборки:

$$z=\frac{x-μ}{δ}=\frac{e}{δ}=\frac{e}{\sqrt{\frac{p(1-p)}{n}}}$$

$$z^{2}=\frac{e^{2}}{\frac{p(1-p)}{n}}$$

$$n=\frac{z^{2}×p(1-p)}{e^{2}}$$

 – итоговая формула для размера выборки, где n – размер выборки, z – z-балл (значение функции для нормального распределения), е – допустимая погрешность (в виде десятичной дроби).

Данная формула помогает определить необходимое количество людей, которых необходимо опросить при известном нам значении допустимой ошибки.

Из этой же формулы можно вывести формулу для допустимой погрешности, которая вычислит погрешность из количества тех людей, которых мы опросили:

$$e=z×\sqrt{\frac{p(1-p)}{n}}$$

Принимая во внимание, что социологи зачастую берут значение доли выборки, как p=0,5 (как самый затруднительный случай), мы можем вывести более упрощённые формулы:

$$n=\frac{z^{2}×0.5×(1-0.5)}{e^{2}}=\frac{z^{2}×0.25}{e^{2}}=\left(\frac{0.5z}{e}\right)^{2}$$

$$e=\frac{0.5z}{\sqrt{n}}$$

Стоит отметить, что кроме использования формул, найти вероятность того, что полученный результат попадёт в доверительный интервал, можно по схеме нормального распределения.

График представляет собой холм, у которого вершина будет совпадать со значением ожидаемого результата (р). Площадь под графиком равна 1. При этом, при значительной вероятности нормального распределения большая площадь графика будет сосредоточена у значений доверительного интервала. Чем ниже доверительный интервал, тем холм будет уже, и это, соответственно, не вызовет у нас сомнения в том, что результат опроса близок к ожидаемому результату.

Ниже представлена схема нормального распределения с уровнем достоверности 95% с ожидаемым результатом р и погрешностью е. 95% площади под графиком будет находиться в интервале р-е < p < p+e. Остальная площадь под графиком (5%) будет находиться за границами доверительного интервала. И чем дальше мы будем уходить от значения границ доверительного интервала, тем ближе и ближе график будет приближаться к оси Ох. Исходя из этого, мы поймем, что вероятность такого отклонения очень маленькая. Подобные схемы в применении очень удобны. Ко всему прочему, они помогают увидеть, насколько результаты опроса отличаются от итоговых результатов.

 

Далее я решил посмотреть, как проводят опросы различные опытные социологи или официальные социологические центры, ведь у них можно узнать о наиболее эффективных методах, используемых центрами при формировании вопросов опрашиваемому оппоненту.

В проекте я проанализирую два опроса.

1. Опрос, проводимый центром «Demoscope» по выявлению предпочтений казахстанцев по строительству АЭС.

Как известно, сейчас активно поднимается вопрос о строительстве в Казахстане атомной электростанции. Мнения казахстанцев по поводу строительства данного сооружения значительно разделились. Известно, что в ближайшем времени по данному поводу будет проводиться референдум, по результатам которого и будет выдвинуто окончательное решение. Учитывая актуальность ситуации, Демоскоп решил провести опрос, чтобы определить общественное мнение не только по отношению к строительству АЭС, но и к референдуму (включая доверие к его результатам).

Опрос проводился через телефонные интервью среди абонентов Казахстана по сотовой связи. Основа выборки – база номеров, полученных с помощью генератора случайных чисел..

Ознакомившись с методами и результатами опроса, можно отметить следующее:

* Все данные были сформированы по полу, возрасту, региону, а также по предпочитаемому языку (язык, на котором отвечали респонденты, кстати, выделил разногласия по некоторым вопросам).
* Учитывалась степень осведомлённости респондентов о строительстве (оказалось, что 36,8% людей не знали о строительстве АЭС вообще).
* Среди вариантов ответа на самый главный вопрос - «Вы за или против строительства АЭС?» были такие варианты ответа: «однозначно поддерживаю», «скорее всего, поддерживаю», «скорее всего, не поддерживаю», «точно не поддерживаю», а также «затрудняюсь ответить». Эти ответы достаточно хорошо обозначили позицию казахстанцев.
* Важно было узнать мотивы тех, кто поддерживает или не поддерживает строительство АЭС, поэтому опрос включал в себя дополнительный вопрос, который определял основные причины, по которым казахстанцы были за / против строительства АЭС.
* Через исследование также была поднята проблема дефицита электроэнергии в Казахстане. В итоге выяснилось, что те, у кого есть такая проблема, выступают за немедленное строительство.
* При определении мнения относительно референдума, Демоскоп выявлял примерный процент казахстанцев, который собирается идти на референдум - / скорее всего, пойдёт / однозначно не пойдёт / скорее, не пойдёт / пока не знают. При этом около 57% людей доверяют результатам референдума.
* Ещё одна важная деталь, которая была проанализирована – предпочтение казахстанцев той или иной стране, которая должна строить АЭС. Оказалось, что явное предпочтение здесь также отсутствует: Россия обгоняет Южную Корею, но незначительно. Интересно, что те кто отвечал на русском языке, предпочитают сотрудничество с Россией, а те, кто на казахском языке – вариант с Южной Кореей.

 Таким образом, можно сделать вывод, что тема была изучена социологами достаточно глубоко. По данным Демоскопа, максимальный размер погрешности с вероятностью 95% не превысил 3%. Отсюда следует, что размер выборки (по вышеприведённой формуле) равен примерно 1067 человек.

1. Опрос, проводимый «Левада-центр».

 «Левада-центр» - Российский аналитический центр, проводящий опросы и исследующий мнение жителей Российской Федерации. В 2022 году центр провел эксперимент, в ходе которого была проверена следующая гипотеза: «респонденты, которые не поддержали руководство страны, сегодня часто отказываются от участия в опросе».

 Основные результаты эксперимента показали, что в основном отказывается от опроса молодежь от 18-24 лет, охотно принимают участие в опросах респонденты в возрасте 40-54 лет. Доля отказов среди респондентов с разным уровнем образования одинакова, но люди с высшим образованием брали трубку чаще.

 В целом эксперимент не подтвердил гипотезы о том, что респонденты, которые не одобряют деятельность руководства страны, чаще отказываются от участи в опросах. Также не подтвердилось и предположение о том, что повторно участвующие в опросах люди, стали более позитивно относиться к происходящим событиям.

 Если обратиться к методологии проведенного эксперимента, то нужно отметить, что на начальном этапе были выделены социально-демографические характеристики: пол, возраст, образование, тип населенного пункта, образование. Параметры выборки были максимально приближены к параметрам генеральной совокупности (в данном случае – взрослого населения страны). Далее были рассчитаны и использованы коэффициенты достижимости (доля успешных интервью), отказов (доля отказов) и контактов (доля номеров, с которыми удалось установить контакт). Интересно, что общий уровень отказов в данном эксперименте оказался ниже, чем в опросах со случайной выборкой номеров. Такой результат получился вследствие того, что в выборку экспериментального опроса были включены те абоненты, которые уже участвовали ранее в опросах и поэтому рассматривались социологами как лояльно относящиеся к подобным исследованиям субъекты.

В целом исследования аналитического центра «Левада-центр» можно считать тщательно подготовленными и продуманными. Анализ эксперимента мне очень понравился.

 Изучение методов исследования опытных социологических центров оказались полезными для моего проекта.

 Наконец, я посмотрел, насколько правильно проводят опросы студенты нашей школы.

 Так, осенью 2023-2024 учебного года многие 12-классники отправляли свои опросы по предмету КСМ на электронную корпоративную почту “Outlook”. Каждый опрос затрагивал какую-то проблему Казахстана по разным направлениям (политика, экономика, социальная жизнь населения, география и т. п.). Поскольку проекты должны были обязательно касаться Казахстана, то естественно, что в генеральной совокупности должны быть только казахстанцы.

 Изучив опросы 12-классников, я, прежде всего, обратил внимание на то, кому по корпоративной почте был отправлен данный запрос. Здесь удалось разделить авторов опросов на 2 категории: первая группа отправляла опрос студентам только НИШ г. Актобе (иногда учителям), вторая группа - рассылала опрос по «Outlook» студентам НИШ разных городов. При этом инициаторы опросов из второй группы отдельно опрашивали более взрослых людей, учитывая в самой форме опроса такие характеристики, как пол, возраст (иногда – регион, в котором проживает опрашиваемый). Таким образом, вторая группа студентов более тщательно подходила к выбору респондентов.

 Помимо этого, было интересно наблюдать за содержанием вопросника и проанализировать то, как данный опрос может помочь определить общественное мнение респондентов.

 Участвуя с 7-го класса в проектной работе, сам я столкнулся с неактивным участием студентов НИШ в проводимых мною опросах. Постепенно я понял, что одной из причин этого является неправильное формирование группы респондентов. Кроме того, я увидел, что большинство опрашиваемых респондентов избегают давать ответы на вопросы, требующие собственного ответа, предпочитая выбирать готовые ответы. Вопросы, которые не имели отметки - «обязательные для ответа», часто оставались без внимания. В результате, я не смог получить полные ответы на интересуемые меня вопросы и, соответственно, обеспечить выявление по изучаемой проблеме максимально верного общественного мнения.

**4.** **Исследование с помощью проверки гипотез – что эффективнее: обычное анкетирование или онлайн-опрос**?

В проекте я намерен сравнить традиционное анкетирование с одним из популярных методов опроса на сегодняшний день - онлайн-опросом. С помощью проверки гипотез я собираюсь узнать, насколько эффективен тот или иной метод изучения общественного мнения, и действительно ли репрезентативная выборка помогает добиться точного результата в обоих случаях.

**4.1. Анкетирование**

Одним из подвидов опроса является анкетирование. В отличие от онлайн – опросов, анкетирование позволяет непосредственно выбрать личность опрашиваемого и контактировать с ним. Соответственно, можно быстрее сформировать группу людей, объединенных по одному признаку. В то же время анкетирование может быть и анонимным. Анкетирование используется в науке, культуре, образовании, медицине, кадровой политике и в бизнесе. С его помощью можно собрать информацию за короткий срок.

 В социологии анкетирование – это метод опроса, который используется для составления статистических или динамических статистических представлений о состоянии общества, общественного мнения, состояния различно рода напряженности с целью прогнозирования действий или событий.

Анкетирование может использоваться для проверки гипотез, поэтому я решил на деле проверить, как работают математические формулы.

Для проверки гипотез я решил провести среди студентов 11 классов НИШ Актобе с русским языком обучения небольшое анкетирование.

Всего в нашей школе таких классов 3 и в них обучаются 65 человек.

При проверке гипотез я действовал по нижеприведённому алгоритму.

Первым шагом в проверке гипотез являются определение целей исследования и правильная формулировка нулевой и альтернативной гипотез. Нулевая обычно предполагает отсутствие существенных различий между двумя переменными, а альтернативная – наличие взаимосвязи или различий.

Далее следовало разработать анкету, содержащую вопросы, соответствующие целям исследования. Анкета должна быть понятной и удобной для заполнения. Я составил анкету, в которой было всего 8 легких вопросов.

Затем мне нужно было сделать правильную выборку респондентов, чтобы результаты анкетирования были репрезентативными. В данном случае я выбрал критерий по полу. Из опрашиваемых 65 человек 41 – представители мужского и 24 – представители женского пола. Отсюда, отношение количества мальчиков к девочкам равняется примерно 1,7. Получив ответы от всех 65 человек, я смешал анкеты и выбрал из них рандомно 10 анкет (для сопоставления общественного мнения 10 человек с общественным мнением 65), из заполненных мальчиками было 6 анкет, девочками - 4. Отношение в данном случае равнялось 1,5 – ближе к 1,7, из чего можно сделать вывод, что характеристика выборки здесь практически совпадала с характеристикой всей генеральной совокупности. Поэтому выборку можно считать стратифицированной.

Следующим шагом в проверке гипотез является сбор данных. Здесь, как уже указывалось, важно было обеспечить репрезентативность выборки по отношению к интересующей нас совокупности, убедившись в заполнении анкеты респондентами в полном объеме. Это необходимо для получения достоверных результатов.

 На последующем этапе проводился анализ результатов, которые позволили узнать, может ли нулевая гипотеза быть отвергнута в пользу альтернативной. Анализ данных - это статистическая обработка данных и формулирование выводов.

 Последний этап анкетирования - интерпретация результатов статистического анализа. Здесь мы определяем, являются ли полученные результаты статистически значимыми, подтверждают ли они гипотезу либо отвергают ее.

Проверить я решил следующие гипотезы:

Гипотеза 1:

* Поскольку НИШ предоставляет огромные возможности для обучения в топовых университетах мира, то ученики нашей школы стремятся поступить в зарубежные вузы. (Респондентам задавался вопрос о том, собираются ли они поступать в зарубежный вуз после окончания школы).

Гипотеза 2:

* При стратифицированной выборке (если соотношение выбранных людей по одному демографическому признаку равно соотношению людей всей генеральной совокупности) мнение 10 человек сможет отразить мнение 65 человек (всей параллели).

 Ниже представлены результаты проведенного мною анкетирования:

*Результаты анкетирования при проверке всей генеральной совокупности (65 человек*)

*Результаты анкетирования при выборке из 10 человек*.

Получив результаты, я приступил к проверке двух гипотез.

1. Гипотеза №1

Нулевая гипотеза, где результат наблюдений является полностью случайным, выглядит так:

Н0: р=0,5

 Альтернативная гипотеза, где существует некое реальное воздействие, а наблюдаемые данные обусловлены этим воздействием (в данном случае, большим количеством возможностей в НИШ):

 H1: p>0,5

 Из второй таблицы среди 10 человек выявлено, что 8 человек предпочитают учиться заграницей (80%, р=0,8) и 2 человека хотят продолжать обучение в Казахстане (20%, р=0,2).

 Уровень значимости, по которому мы проверяем нулевую гипотезу, берётся как 95%. В таблице для нормального распределения значение 95% соответствует значению z-балла 1,645. Если полученный результат будет ниже – нулевая гипотеза работает, а если больше или равно – нулевая гипотеза может отклоняться.

 Чтобы проверить, отвергнется ли нулевая гипотеза или нет, мы используем следующую формулу для z-балла:

* z=$\frac{p-p0}{\sqrt{p0(1-p0)/n}}$
* Подставляем: z=$\frac{0,8-0,5}{\sqrt{0,5\*(1-0,5)/10}}=1,897$
* Как видно, 1,645<1,897. Итого, альтернативная гипотеза подтвердилась, и нулевая гипотеза отклоняется (фактор действительно оказывает влияние на выбор учеников).
* *Схема нормального распределения: полученное значение z-балла выходит за пределы доверительного интервала, что подтверждает альтернативную гипотезу.*
1. Гипотеза №2.

Проверять вторую гипотезу можно на основе любого вопроса. Я выбрал вопрос про то, сколько времени старшекласснику нужно для выполнения домашнего задания. Поскольку здесь предложенные варианты ответов были даны в виде временных интервалов, то данные значения являются непрерывными случайными величинами. Сначала для удобства создадим таблицу, где отображаются середины интервалов и количество учеников, выбравших тот или иной интервал, как с 65 старшеклассниками, так и с 10 (исходя из результатов моего анкетирования).



 Чтобы проверить гипотезу, среднее значение времени для выборки из 10 человек у нас должно попасть в интервал (X-z\*E; X+z\*E), где Х – среднее значение времени для всей генеральной совокупности, z – показатель значимости (берётся как 1,96 – 95%) и Е – стандартная ошибка выборки (погрешность).

 Вычислим среднее арифметическое для выборки (х) и генеральной совокупности (Х):

 $x= \frac{0,5\*2+2\*6+4\*1}{10}=1,7$

 X $= \frac{0,5\*18+2\*33+4\*3}{65}=1,34$

 То есть, значение 1,7 должно попасть в интервал. Теперь, чтобы определить границы интервала, найдём ошибку выборки. Сначала находим исправленную дисперсию, которая высчитывается по данной формуле:

 $S^{2}= \frac{n}{n-1}\*\left(\frac{\sum\_{i=1}^{n}ft^{2}}{n}-x^{2}\right)$

 Подставляем:

 $S^{2}= \frac{10}{10-1}\*\left(\frac{2\*0,5\*0,5+6\*2\*2+1\*4\*4}{10}-1,7^{2}\right)=1,29$

 Стандартное отклонение равно:

$S= \sqrt{1,29}$ = 1,136

Находим погрешность: $E=\frac{S}{\sqrt{n}}=\frac{1,136}{\sqrt{10}}=0,36$

Находим границы интервала:

(1,34-1,96\*0,36; 1,34+1,96\*0,36) = (0,6344; 2,0456)

0,6344 < 1,7 < 2,0456 – как видим, значение 1,7 входит в интервал.

Гипотеза подтвердилась: мнение 10 учеников отразило мнение 65.

**4.2. Онлайн-опрос**

Далее я решил отправить ученикам всех 12 классов на общую корпоративную почту Microsoft Outlook онлайн-опрос. Содержание опроса примерно такое же, как и в анкетировании, но в этот раз данные должны были быть взвешены по языку обучения ученика. Несмотря на то, что опрос был адресован 160 оппонентам, ответило на него всего 16. При этом отношение участников из казахских классов к участникам из русских классов значительно отличалось от фактического отношения учеников казахских классов к ученикам русских классов (6:10, в то время как в генеральной совокупности это отношение составляет 4:3). То есть выборка здесь не была стратифицирована, ввиду чего было маловероятно, что на основе 16 ответов мы можем достичь точного ответа. Поэтому из 16 ответов я решил убрать некоторые ответы учеников русских классов так, чтобы выборка реально соответствовала характеристикам генеральной совокупности. Поскольку мы доказали, что при стратифицированной выборке мнение даже 10 человек сможет отразить мнение всей генеральной совокупности, мы можем её сопоставить с нерепрезентативной выборкой из 16 человек, определяя все основные статистические показатели.

Здесь проверялись следующие гипотезы:

Гипотеза 1:

Наша школа является школой физико-математического направления, поэтому большинство учеников этой школы будут связывать свою будущую профессию с математикой;

Гипотеза 2:

Мнение людей, попавших в нестратифицированную выборку, не способно отразить мнение людей, попавших в стратифицированную выборку.

Ниже представлены результаты онлайн-опроса:

*****Результаты онлайн-опроса среди всех, принявших участие в онлайн-опросе*

*Результаты онлайн-опроса при репрезентативной выборке из 10 человек*

Проверка гипотезы №1:

* Нулевая гипотеза (Н0) р=0,5
* Альтернативная гипотеза (Н1) р>0,5: из таблицы следует, что р=9/16=0,5625 (56,25% 12-классников предпочитают выбрать профессию, связанную с математикой).
* Уровень значимости берём, как 95% (по z-значению 1,645)
* Используем формулу z=$\frac{p-p0}{\sqrt{p0(1-p0)/n}}$
* Подставляем: z=$\frac{0,5625-0,5}{\sqrt{0,5\*(1-0,5)/16}}=0,5$
* 0,5 < 1,645, т.е. альтернативная гипотеза отвергается.
* Итог: нулевая гипотеза подтверждается.



*Схема нормального распределения: полученное значение z-балла входит в доверительный интервал, что опровергает альтернативную гипотезу.*

Проверка гипотезы №2:

Для проверки гипотезы об эффективности нерепрезентативной выборки, я задал вопрос о том, что учащиеся ставят на первое место: академическую успеваемость, олимпиады или научные проекты.

В таблице из 16 человек показано, что для большинства (9 человек) больше всего в приоритете академическая успеваемость – 56,25% респондентов, в то время как стратифицированная выборка из 10 человек показала что таких учеников всего 40 процентов. Для того чтобы узнать, можно ли действительно доверять этому результату, необходимо проверить, входят ли эти 56,25% учеников в доверительный интервал (Х-1,96\*E; X+1,96\*E).

Для этого сначала находим стандартное отклонение: Е= $\sqrt{p0(1-p0)/n}$ = $\sqrt{0,5\*(1-0,5)/16}=0,125$

Находим границы доверительного интервала (0,4-1,96\*0,125; 0,4+1,96\*0,125) = (0,155; 0,645)

0,155<0,5625<0,645, и отсюда вытекает вывод, что нестратифицированная выборка тоже может отражать мнение всей генеральной совокупности.

* Итог: гипотеза отвергается.
* *Схема нормального распределения: полученное значение z-балла входит в доверительный интервал, что подтверждает гипотезу.*

Главной причиной отвергания гипотезы является то, что количество людей, участвующих в выборке, оказалось слишком маленьким. Как указывалось ранее, чем больше у нас количество опрошенных, тем меньше будет доверительный интервал. В данном случае, доверительный интервал оказался слишком большим – 49%. Наша гипотеза нашла бы подтверждение в том случае, если бы количество опрошенных было бы как минимум 37 человек (при проведении нижеперечисленных вычислений).

Е (стандартное отклонение) = $\frac{0,5625-0,4}{1,96}=0,0829$

n = $\frac{p\*(1-p)}{E^{2}}=\frac{0,5\*(1-0,5)}{0,0829^{2}}=36,4≈37$

**4.3. Вывод исследования.**

На основе всех математических вычислений, можно сделать вывод, что традиционное анкетирование более эффективно, чем онлайн-опрос.

* Анкетирование позволяет нам самим сформулировать стратифицированную выборку (так чтобы в равной мере учитывались мнения различных групп людей по какому-либо признаку), в то время как при отправке онлайн-опроса всем респондентам мы никак не можем повлиять на репрезентативность, из-за чего мнение одной группы людей, которая во всей генеральной совокупности находится в меньшинстве, может перевесить мнение большинства.
* Большинство не имеет желание проходить онлайн-опросы либо не находит времени. Мы получаем очень мало ответов, что приводит к большой погрешности в наших результатах. Причина – отсутствие личного взаимодействия с респондентами.

**Заключение.**

Таким образом, математическая статистика находит большое применение даже в такой социально-гуманитарной науке, как социология. На основе статистических вычислений можно работать с гипотезами, которые ставят и проверяют исследователи, находить количество людей, которых необходимо опросить, и достигать на основе полученных результатов успешности проведения опроса. Поскольку исследования общественного мнения сейчас актуальны во многих демократических странах, очень важно проводить вычисления корректно. К примеру, та же стратифицированная выборка позволяет существенно снизить отклонение от результата, который мог бы быть определён среди всей генеральной совокупности. В итоге, большинство обычных людей уже будут доверять результатам опроса, проводимых, например, при объявлении выборов и референдумов.

 Исследования, проводимые в рамках данного проекта, могут помочь в будущем и ученикам нашей школы. С помощью формул и методов, которые отражены в проекте, возможно не только правильно проводить исследования по таким предметам, как ГППР и КСМ в рамках школьного обучения, но и повысить свою статистическую грамотность в целом. Статистическая грамотность, в свою очередь, позволит всем нам ориентироваться в обществе; быть экономически и политически развитыми; разбираться в вопросах, связанных с общественным мнением и общественными отношениями и прогнозировать будущее общества.

 Статистика как наука не останавливается на месте. Ученые постоянно разрабатывают новые методы и формулы, с помощью которых государство и высшие органы смогут достоверно узнавать о предпочтениях граждан и решать, на какие аспекты следует обратить внимание. К примеру, довольно успешно социологи начинают уже использовать математические матрицы, которые, на самом деле, являются системой всей нашей жизни. Более того, последние научные исследования доказывают, что математика способна не только прогнозировать, но и изменять общественное мнение.

 В этом и заключается перспективность взаимодействия общественного мнения и математической статистики.

**Список использованной литературы**

Demoscope (октябрь 2023). ОПРОС: Мнение казахстанцев о строительстве АЭС разделилось. <https://demos.kz/opros-mnenie-kazahstancev-o-stroitelstve-ajes-razdelilos/>

Good calculators (дата неизвестна). Sample Size Calculator. <https://goodcalculators.com/sample-size-calculator/>

Агапеева К. & Волков Д. (июнь 2022). Готовность участвовать в опросах: результаты эксперимента. Источник: Левада-Центр (включён в реестр иноагентов). <https://www.levada.ru/2022/06/14/gotovnost-uchastvovat-v-oprosah-rezultaty-eksperimenta/>

Кунилова К. (ноябрь 2023 – последнее обновление). Общественное мнение с позиции социологии. Справочник от автор24. <https://spravochnick.ru/sociologiya/obschestvennoe_mnenie_s_pozicii_sociologii/?ysclid=lqzfibffzh254829480>

Юдин Г. Б. (включён в реестр иноагентов) (2020). Книга «Общественное мнение, или Власть цифр». Издательство Европейского университета в Санкт-Петербурге.