Анализ факторов загрязнения атмосферы токсичным газом CO методом двухфакторного дисперсионного анализа

Анна Третьякова, группа: М610М, предмет: ППС в ЗПД

06 января 2018 г

## Введение

В 2015 году, по данным Росстата и Росприроднадзора, в государственном докладе «О состоянии и об охране окружающей среды Российской Федерации в 2015 году», опубликованный на сайте Министерства природных ресурсов и экологии, общий объем выбросов загрязняющих веществ в атмосферу в 2015 году составил 31,3 млн т, из которых на стационарные источники пришлось 17,3 млн.т, а на автотранспорт – 13,8 млн т.[2]

После обнародования данного доклада начались разговоры о том, как снизить ущерб для атмосферы, была разработана программа “НДТ”(Наилучшие доступные технологии) призванная заменить на предприятиях старое и оказывающее наиболее опасное воздействие на атмосферу оборудование на современное и экологически чистое, уже в 2016 были зафиксированы первые положительные результаты, которые обуславливались снижением количества выбросов в атмосферу, а 2017 год должен был лишь закрепить этот успех.

с автотранспортом дела обстоят намного хуже, весомой альтернативы бензиновым и дизельным двигателям нет, электротраспорт и гибридные модели реального применения не нашли в первую очередь из за дороговизны и недоказанности реального снижения вреда для атмосферы, а количество машин только растёт[3]

В своей работе я применю двухфакторный дисперсионный анализ для анализа и сравнения объёма загрязнения атмосферы в 2015 и 2017 году от стационарного и нестационарного источника выборосов одним из наиболее токсичных газов - CO, который оказывает негативное влияние заболеваемость органов дыхания преимущественно у детей живущих в наиболее загазованных регионах РФ, например Красноярске и Челябинске.[3]

## Описание используемых методов

Дисперсионный анализ — это статистический метод оценки связи между факторными и результативным признаками в различных группах, отобранный случайным образом, основанный на определении различий значений признаков.

Для проведении дисперсионного анализа требуется:

1.Нормальность распределения анализируемых групп или соответствие выборочных групп генеральным совокупностям с нормальным распределением.[1]

2.Независимость (не связанность) распределения наблюдений в группах.[1]

Данные для исследования объёма выброса CO в атмосферу за 2015 и 2017 год по всем субъектам РФ были взяты с портала открытых данных РФ - <http://data.gov.ru/>.

Все вычисления и визуализация результата будут произведены с помощью программного обеспечения R-Studio.

Перед применением дисперсионного анализа, исходные данные были обработаны, а именно:

1. Была выделена количественная (зависимая) переменная - уровень выброса CO по каждому субъекту РФ в килотоннах, а в качестве номинативных (независимых) переменных были взяты: Источник выбросов и год.

В результате получилась выборка из 32 наблюдений поделённая на 4 группы: По источнику выброса (N-нестационарный,S-стационарный) и году (2015,2017), в котором проводились замеры, в каждой группе по 8 значений, где каждое значение - это совокупная информация по загрязнению атмосферы газом CO в тысячах килотонн по всем городам входящим в тот или иной субъект РФ.

1. Данные были проверены на соотвествие гипотезы о нормальном распределении, с помощью теста “Шапиро-Уилка” в R.

Для этого была использована функция: shapiro.test(). В первую очередь загрузим данные из файла с расширением .csv в R исследуемые данные, сохраним в переменной readfile и зафиксируем функцией attach() для удобства работы с ними.

readfile<-read.csv("C:\\data.csv",header=TRUE,stringsAsFactors=FALSE,sep=';')  
attach(readfile)

Выведем имена заголовков таблицы для дальнейшего использования их в работе и отразим загруженные данные в виде таблицы, которая выведет данные по группам и количеству значений в каждой группе.

names(readfile)

## [1] "X.соотнош." "Ист.выбр." "Год"

table(Ист.выбр., Год)

## Год  
## Ист.выбр. 2015 2017  
## N 8 8  
## S 8 8

Затем применим тест “Шапиро-Уилка”, который в качестве гипотезы предполагает, что данные распределены в соответсвии с законом о нормальном распределении, если P-уровень значимости превышает 0.05, то гипотеза принимается, если меньше, то отклоняется.

shapiro.test(X.соотнош.)

##   
## Shapiro-Wilk normality test  
##   
## data: X.соотнош.  
## W = 0.91771, p-value = 0.01803

Результат P>0.05, гипотеза принимается, значения подчиняются нормальному распределению.

Рассчитаем для каждой нашей группы значения среднего, дисперсии и стандартного отклонения,в R - это можно сделать одной функцией aggregate () и сохраним результат в переменных для последующего удобства воспроизведения результа.

mean<-aggregate(X.соотнош., by=list(Ист.выбр.,Год), FUN=mean)  
var<-aggregate(X.соотнош., by=list(Ист.выбр.,Год), FUN=var)  
sd<-aggregate(X.соотнош., by=list(Ист.выбр.,Год), FUN=sd)  
data.frame(Группа\_1=mean$Group.1,Группа\_2=mean$Group.2,Среднее=mean$x,Дисперсия=var$x,Ст.откл=sd$x)

## Группа\_1 Группа\_2 Среднее Дисперсия Ст.откл  
## 1 N 2015 0.67750 0.08876429 0.2979334  
## 2 S 2015 0.65375 0.03908393 0.1976966  
## 3 N 2017 0.88375 0.10574107 0.3251785  
## 4 S 2017 0.48125 0.03515536 0.1874976

Дисперсионный анализ включает в себя:

Формулирование и гипотез, которые в результате эксперимета необходимо принять или отвергнуть, вычисление влияния внутригрупповой изменчивости фактора и межгрупповой изменчивости фактора с учётом значения их степеней свободы, вычисление дисперсии ошибки,которая учитывает влияние всех факторов, в том числе и не учтенных, рассчёт значения критерия Фишера - F для каждой гипотезы, а затем, по значению F, определение P-уровеня значимости, который позволяет принять или отвергнуть сформулированные ранее гипотезы,визуализацию и интерпретацию результатов.[1]

Сформулируем гипотезы:

: Влияние факторов, а также их взаимодействие не оказывает никакого влияния на зависимую переменную.

: Влияние факторов, а также их взаимодействие оказывает существенное влияния на зависимую переменную.

Для расчёта значений внутригруповой и межгрупповой изменчивости, а также значений критерий Фишера и соответствующего ему P-уровня значимости, в языке R есть функция aov(), воспользуемся ей и сохраним результат в переменной ANOVA.

ANOVA<-aov(X.соотнош.~ Ист.выбр.\*Год)

Затем с помощью функции summary() получим итоговый результат.

## Результаты

В результате проведённого двухфакторного анализа, мы получили следующее:

summary(ANOVA)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Ист.выбр. 1 0.3634 0.3634 5.409 0.0275 \*  
## Год 1 0.0023 0.0023 0.034 0.8552   
## Ист.выбр.:Год 1 0.2869 0.2869 4.270 0.0481 \*  
## Residuals 28 1.8812 0.0672   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Значение P-уровня значимости для первого фактора ниже 0.05, следовательно нулевая гипотеза отвергается и принимается альтернативная - источник выбросов влияет на объём выброса CO в атмосферу.

Значение P-уровня значимости для второго фактора выше 0.05, следовательно нулевая гипотеза принимается - внутригрупповая изменчивость фактора B никак не влияет на зависимую переменную.

Для взаимодействия факторов уровень P также меньше 0.05, что даёт нам право отвергнуть нулевую гипотезу и принять альтернативную, которая гласит, что влияние 2-ух факторов повлияли на объём выбросов CO в атмосферу.

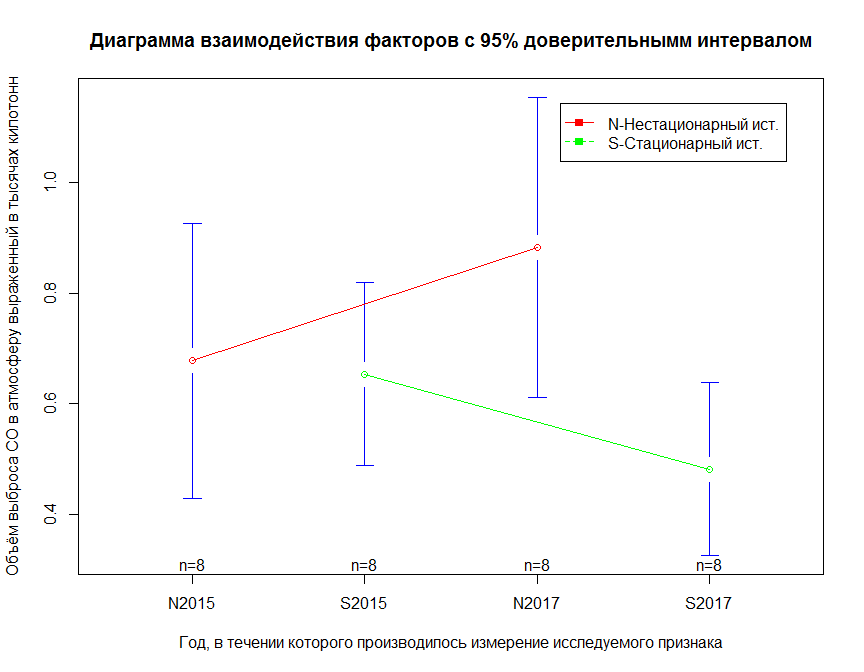
Результат проведённого двухфакторного анализа изображён на графике (рис. 1). 

Рисунок 1 - Диаграмма взаимодействия факторов на исследуемую переменную с 95%-ми доверительными интервалами

## Заключение

На графике (см.рис. 1), хорошо видно, что эксперимент с переходом предприятий на реализуемую программу НДТ полностью себя оправдывает, уровень выбросов СО в 2017 году от стационарного источника уменьшился, и хочется верить, что подобное падение будет только расти. Полный переход на прграмму НДТ начнётся в 2019 году и продолжится в течении 4 лет, будет очень интересно через 2 года вновь провести анализ для сравнения результатов.

К большому сожалению, из графика (см.рис. 1) также видно, что урон от нестационарного транспорта только возрос и для снижения ущерба от транспорта необходимо вести активную работу, исследуя возможность замены топлива на более экологически чистое и безвредное для окружающей среды и здоровья населения.

## Список используемой литературы

[1]Шеффе Г. Дисперсионный анализ. - М.:Наука,1980.

[2]Государственном доклад «О состоянии и об охране окружающей среды Российской Федерации в 2015 году»[Электронный ресурс]. – Режим доступа: <http://www.mnr.gov.ru/gosdoklad-eco-2015/index.html>.(Дата обращения: 28.12.2017).

[3]В погоне за воздухом [Электронный ресурс]. – Режим доступа: <http://bellona.ru/2017/05/31/industrial-pollution-russia/>.(Дата обращения: 30.12.2017).