МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

МЕЖГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ»

Кафедра «Финансы и бухгалтерский учёт»

Отчёт по лабораторной работе № 7

РЕГРЕССИОННЫЙ АНАЛИЗ

Выполнили

Студенты гр. АСОИ-181

Остапенко Александр Константинович

Лапин Никита Андреевич

Проверила

Ливинская Виктория Александровна

Могилёв, 2021

1. Вводим исходные данные

life\_expectancy<-WDI(indicator = "SP.DYN.LE00.IN",country ="all",start = 2015,end = 2015)

electricity\_access<-WDI(indicator = "EG.ELC.ACCS.ZS",country ="all",start = 2015,end = 2015)

age\_dep\_coeff<-WDI(indicator = "SP.POP.DPND",country ="all",start = 2015,end = 2015)

literacy\_rate<-WDI(indicator = "SE.ADT.LITR.ZS",country ="all",start = 2015,end = 2015)

1. Дискриптивно-статистический анализ данных. Для того, чтобы оценить степень однородности наших данных, воспользуемся коэффициентом вариации. Коэффициент вариации считается таким образом: стандартное отклонение нужно поделить на среднее значение. Обычно значение коэффициента вариации, взятое по модулю, лежит в пределах от 0 до 1, но иногда, если данные очень разнородны (стандартное отклонение большое), оно может быть больше 1.

Мы ожидаем коэффициенты вариации меньше 0,33.

sd(na.omit(life\_expectancy$SP.DYN.LE00.IN))/mean(na.omit(life\_expectancy$SP.DYN.LE00.IN))

sd(na.omit(electricity\_access$EG.ELC.ACCS.ZS))/mean(na.omit(electricity\_access$EG.ELC.ACCS.ZS))

sd(na.omit(age\_dep\_coeff$SP.POP.DPND))/mean(na.omit(age\_dep\_coeff$SP.POP.DPND))

sd(na.omit(literacy\_rate$SE.ADT.LITR.ZS))/mean(na.omit(literacy\_rate$SE.ADT.LITR.ZS))

Text

Description automatically generated

Так как коэффициенты вариации оказались меньше 0,33 (выборки однородны) мы имеем право использовать выборки в дальнейшем анализе.

1. Анализ линейных коэффициентов парной корреляции

data <- cbind(life\_expectancy, electricity\_access$EG.ELC.ACCS.ZS, age\_dep\_coeff$SP.POP.DPND, literacy\_rate$SE.ADT.LITR.ZS)

names(data)[3]<- "life\_expectancy"

names(data)[5]<- "electricity\_access"

names(data)[6]<- "age\_dep\_coeff"

names(data)[7]<- "literacy\_rate"

mean\_life\_expectancy <- mean(data$life\_expectancy[which(!is.na(data$life\_expectancy))])

mean\_electricity\_access <- mean(data$electricity\_access[which(!is.na(data$electricity\_access))])

mean\_age\_dep\_coeff <- mean(data$age\_dep\_coeff[which(!is.na(data$age\_dep\_coeff))])

mean\_literacy\_rate <- mean(data$literacy\_rate[which(!is.na(data$literacy\_rate))])

data$life\_expectancy[which(is.na(data$life\_expectancy))] <- mean\_life\_expectancy

data$electricity\_access[which(is.na(data$electricity\_access))] <- mean\_electricity\_access

data$age\_dep\_coeff[which(is.na(data$age\_dep\_coeff))] <- mean\_age\_dep\_coeff

data$literacy\_rate[which(is.na(data$literacy\_rate))] <- mean\_literacy\_rate

data %>%

select(life\_expectancy, age\_dep\_coeff, electricity\_access, literacy\_rate) %>%

psych::corr.test()

Table

Description automatically generated

Чем ближе коэффициент корреляции к +1, тем теснее прямая связь между признаками, чем ближе к -1, тем теснее обратная связь.

1. Построение уравнения множественной регрессии

linearRegression = lm(data$life\_expectancy ~ data$electricity\_access +data$age\_dep\_coeff+data$literacy\_rate)

coefficients(linearRegression)



1. Оценка коэффициента множественной корреляции (детерминации)

summary(linearRegression)

A screenshot of a computer

Description automatically generated with medium confidence

Коэффициент множественной корреляции (R) = 0.84, что свидетельствует о наличии связи между факторами

Коэффициент детерминации (D=R2) = 0.71, что свидетельствует о доли вариации результативного показателя, которая связана с вариацией факторных показателей.

1. Оценка статистической надёжности полученного уравнения регрессии

Чтобы провести оценку статистической надёжности полученного уравнения регрессии, необходимо проанализировать результат теста Фишера (F-статистика)



Считается, что чем больше значение F-statistic, тем лучше. Модель считается значимой, т.к. любой из её коэффициентов отличен от нуля.