МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение

высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КубГУ»)**

**Факультет математики и компьютерных наук**

**Кафедра математических и компьютерных методов**

**Отчёт  
по производственной практике**

**Б2.О.02.01(П)**

**технологической**

**(проектно-технологической) практике**

студента Пасько Дмитрия Анатольевича группа 102/1

Направление подготовки 02.04.01 Математика и компьютерные науки

Направленность (профиль) «Математическое и компьютерное   
 моделирование»

Форма обучения очная

Квалификация выпускника магистр

Краснодар

2020

**ЛИСТ ПРОВЕДЕНИЯ ИНСТРУКТАЖЕЙ**

**КубГУ, кафедра математических и компьютерных методов**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Вид  инструктажа | Дата  проведения  инструктажа | Подпись  инструктирующего  Качанова И.В. | Подпись  инструктируемого |
| 1 | Инструктаж по охране труда | 15.06.2020 |  |  |
| 2 | Инструктаж по технике  безопасности | 15.06.2020 |  |  |
| 3 | Инструктаж по пожарной безопасности | 15.06.2020 |  |  |
| 4 | Инструктаж по ознакомлению с правилами внутреннего трудового распорядка | 15.06.2020 |  |  |

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение

высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КубГУ»)**

**Факультет математики и компьютерных наук**

**Кафедра математических и компьютерных методов**

**ИНДИВИДУАЛЬНОЕ ЗАДАНИЕ  
по производственной практике  
(технологической**

**(проектно-технологической) практике)  
Б2.О.02.01(П)**

студента Пасько Дмитрия Анатольевича группа 102/1

**Цели**: систематизация, обобщение и углубление теоретических знаний, формирование практических умений, общепрофессиональных и профессиональных компетенций на основе изучения работы организаций, в которых студенты проходят практику, проверка готовности студентов к самостоятельной трудовой деятельности, а также к продолжению обучения в аспирантуре.

**Задачи технологической (проектно-технологической) практики:** состоят в исследовании конкретной предметной области: построение или изучение существующей математической либо компьютерной модели, анализ математической и вычислительной корректности поставленной задачи, разработка алгоритма решения задачи, программирование на языке высокого уровня, отладка программы и тестирование ее, анализ полученных результатов на их соответствие реальному объекту исследования, внедрение разработок в производственный процесс. В ходе практики студентам предоставляется возможность проведения самостоятельной работы и экспериментальных исследований по заранее разработанной совместно с научным руководителем программе. Предпочтительным является выполнение разработок и исследований по теме выпускной квалификационной работы.

**Место прохождения практики:**ФГБОУ ВО «КубГУ», кафедра МКМ.

**Индивидуальное задание**: разработать или изучить используемую существующую математическую модель, применяемую для решения поставленной научным руководителем задачи по теме выпускной квалификационной работы. Исследовать методы, применяемые в рассматриваемой математической модели. Разработать программную реализацию одного из методов. Провести численные эксперименты и проанализировать полученные результаты.

**Срок прохождения практики:** 15.06.2020-12.07.2020

**Перечень заданий студенту-практиканту**

|  |  |
| --- | --- |
| Содержание  программы практики | Задание  студенту-практиканту |
| **1. Подготовительный этап.**  Беседа руководителя проектно-технологической практики со студентами об организации работы, ведении документации и критериях оценивания работы студентов на практике.  Инструктаж по технике безопасности. | 15.06.2020  Ознакомиться с выданным заданием на практику.  Составить план работы на практике.  Расписаться в журнале регистрации инструктажа по технике безопасности. |
| **2.Практический этап.**  Поиск дополнительной информации (книги, статьи, программы) по теме научной работы. Решение поставленной научной задачи. | 16.06.2020-08.07.2020  Исследовать предметную область, изучить литературу по аналогичным задачам, построить математическую модель, разработать алгоритм решения задачи, создать компьютерную модель, ее протестировать и апробировать на реальных данных. |
| **3.Заключительный этап.**  Подведение итогов проделанной работы. Представление материалов по проектно-технологической практике ее руководителю. Выставление оценок. | 09.07.2020-12.07.2020  Обработать и проанализировать полученные результаты, подготовить отчет по практике. Подготовится к защите отчета на кафедре и пройти саму процедуру защиты. |

**Задание получил:** Пасько Д. А

(подпись)

**Задание выдано:** 15.06.2020

**Задание выдал:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Качанова И. А.

(подпись)

**СОДЕРЖАНИЕ**

ВВЕДЕНИЕ

Вспомогательные функции

Простейшая линейная модель

Уточнение коэффициентов

Уточнение коэффициентов: зависимость от диапазона

Лучшие модели

Сравнение линейных моделей

ЗАКЛЮЧЕНИЕ

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

## ВВЕДЕНИЕ

Вычислительные возможности позволяют создать и проверить огромное количество моделей. Цель исследователя – предложить несколько вариантов и выбрать среди них лучший. Кроме этого, нужно посмотреть, как работает модель на разных группах данных, чтобы обнаружить выбросы или обнаружить, что на такой-то группе модель вообще не может работать, из-за чего такую группу придётся исключить.

Итак, в модели представляют интерес следующие показатели:

* ошибки в разных группах данных;
* выбросы и влиятельные наблюдения;
* статистическая значимость модели;
* оценки кросс-валидации.

Требуется найти оптимальную линейную модель для исходного набора данных. Для этого создадим несколько возможных моделей, сравним их и выберем лучшую.

## 1 Вспомогательные функции

Напишем несколько вспомогательных функций (на языке R) для оценивания моделей:

**Error**=function(target,weight) (target-weight)^2 %>% mean() %>% sqrt()

**Show**=function(vals,df=data){

cat('Ошибки модели (выведены 15 элементов с наибольшими ошибками): \n')

err=df$RM-vals

cbind(Fact=round(vals),Target=df$RM,

ERROR=abs(df$RM-vals),

ErrorPercent=abs(err)/df$RM\*100,

df[,c(2:15)]) %>% tbl\_df() %>% select(-Age,-Mail,-Experience,-AgeGroup,-IndexGroup,-BodyType,-Height)%>% arrange(-ERROR,-ErrorPercent,Weight) %>%

filter(ERROR>1)%>%head(15) %>% kable() %>% print()

cat("\n")

rg=range(err)#;print(err);print(rg)

if(rg[1]<0)cat("------------> Наибольшая ошибка в большую сторону:",-rg[1],"\n")

if(rg[2]>0)cat("------------> Наибольшая ошибка в меньшую сторону:",rg[2],"\n")

s=sum(abs(err)/df$RM\*100>maxerror)

len=length(err)

cat("Модель ошиблась более чем на",maxerror,"% в",s,"случаях из",len,"(",s/len\*100,"%)\n")

s=sum(abs(err)>maxerror)

cat("Модель ошиблась более чем на",maxerror,"кг в",s,"случаях из",len,"(",s/len\*100,"%)\n")

cat("----------------> Статистика по ошибкам в процентах:\n")

(abs(df$RM-vals)/df$RM\*100) %>% summary() %>% print()

cat("-------------------> Среднеквадратичная ошибка:", Error(vals,df$RM),"\n")

}

**ShowErrors**=function(model,power.coef=1,sum.coef=0){

Show(predict(model,data)\*power.coef+sum.coef)

cat("Оценка кросс-валидации для всего набора данных",

boot::cv.glm(data,glm(formula = model$call$formula,data=data),K=10)$delta[1],"\n")

cat("Оценка кросс-валидации для не более чем 10 повторений",

boot::cv.glm(data %>% filter(Count<11),glm(formula = model$call$formula,data=data %>% filter(Count<11)),K=10)$delta[1],"\n")

cat("Оценка кросс-валидации для не более чем 6 повторений",

boot::cv.glm(data %>% filter(Count<7),glm(formula = model$call$formula,data=data %>% filter(Count<7)),K=10)$delta[1],"\n")

}

**ResAn**=function(res){

p=ggplot(data %>% mutate(res=res),aes(x=CountGroup,y=res))+

geom\_boxplot()+labs(x="Группа повторений",y="Остатки (цель - предсказание)",title="Распределения остатков в зависимости от группы повторений")+theme\_bw()

print(p)

(p+facet\_grid(vars(Action))) %>% print()

(p+facet\_grid(vars(BodyType),vars(Action)))%>% print()

return(0)

}

**ResVal**=function(vals)ResAn(data$RM-vals)

**mysummary**=function(mdl){

cat("-----> ОБЩАЯ ИНФОРМАЦИЯ О МОДЕЛИ:\n");cat("\n")

gvlma::gvlma(mdl) %>% summary() %>% print();cat("\n")

cat("-----> БАЗОВЫЕ ГРАФИКИ:\n");cat("\n")

par(mfrow=c(2,2))

plot(mdl)

par(mfrow=c(1,1)) ;cat("\n")

cat("-----> ТЕСТ НА НОРМАЛЬНОСТЬ РАСПРЕДЕЛЕНИЯ ОСТАТКОВ\n");cat("\n")

shapiro.test(mdl$residuals) %>% print();cat("\n")

qqPlot(mdl,main="Q-Q plot")

cat("-----> ФАКТОР ИНФЛЯЦИИ ДИСПЕРСИЙ:\n");cat("\n")

vif(mdl)%>% print();cat("\n")

cat("-----> ТЕСТ НА АВТОКОРРЕЛЯЦИЮ:\n");cat("\n")

durbinWatsonTest(mdl) %>% print();cat("\n") #тест на автокорреляцию

cat("-----> ТЕСТ НА ВЫБРОСЫ И ВЛИЯТЕЛЬНЫЕ НАБЛЮДЕНИЯ:\n");cat("\n")

outs=outlierTest(mdl)

outs%>% print()

influ=influencePlot(mdl,main="Диаграмма влияния",sub="Размеры кругов пропорциональны расстояниям Кука")

gb=capture.output(influ %>% print())

cat("-----> ВЫБРОСЫ И ВЛИЯТЕЛЬНЫЕ НАБЛЮДЕНИЯ:\n")

data[c(outs$p %>% names(),influ %>% rownames()) %>% as.numeric(),] %>% unique() %>%

select(-Mail,-Action2,-Body2,-AgeGroup,-CountGroup,-Index) %>% arrange(-RM,-Count) %>% kable() %>% print();cat("\n")

}

**all**=function(modelka){

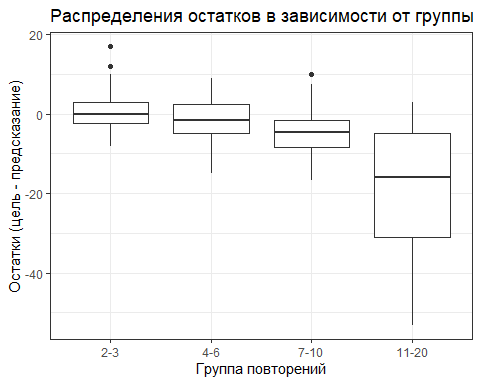
modelka %>% ShowErrors()

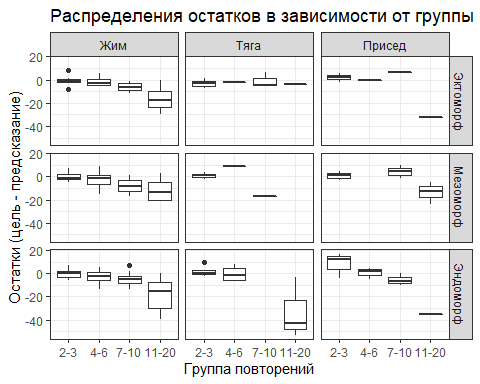
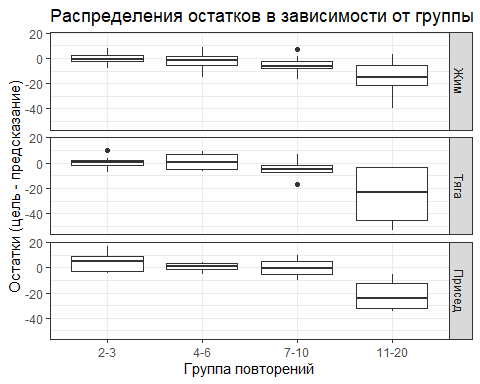
modelka %>% predict(data) %>% ResVal()

modelka %>% mysummary()

}

**2 Простейшая линейная модель**

Применив формулу (модель Вендлера) ко всем наблюдениям, получим оценки (Fact), которые будут отличаться от истинного (Target). Разница между этими величинами – остатки – будут следующие:



Здесь такая логика: чем лучше модель, тем “ящики” тоньше и тем их середины ближе к 0.

В данном случае видно, что **модель Вендлера хорошо работает почти на всех данных из диапазона 2-3 и на немалой части данных из диапазона 4-6**. Для большего числа повторений модель даёт завышенные оценки.

Если посмотреть на численные ошибки

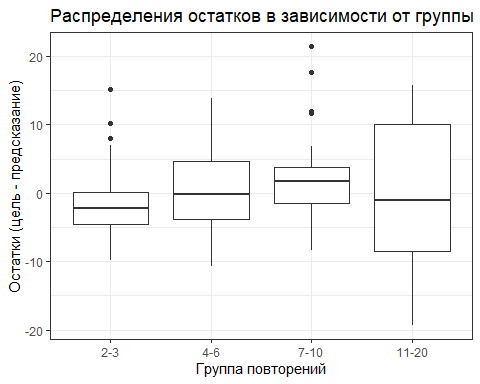
## ------------> Наибольшая ошибка в большую сторону: 53.22   
## ------------> Наибольшая ошибка в меньшую сторону: 17.018   
## Модель ошиблась более чем на 5 % в 48 случаях из 162 ( 29.62963 %)  
## Модель ошиблась более чем на 5 кг в 69 случаях из 162 ( 42.59259 %)  
## ----------------> Статистика по ошибкам в процентах:  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.009091 1.466058 2.895000 4.466610 6.231771 24.950000   
## -------------------> Среднеквадратичная ошибка: 10.79902

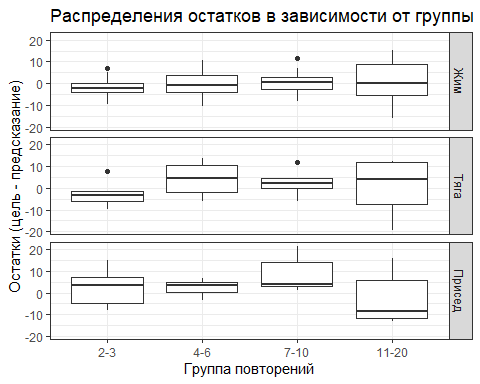
увидим, что модель заметно ошибается в 30% случаев и в среднем на почти 5%.

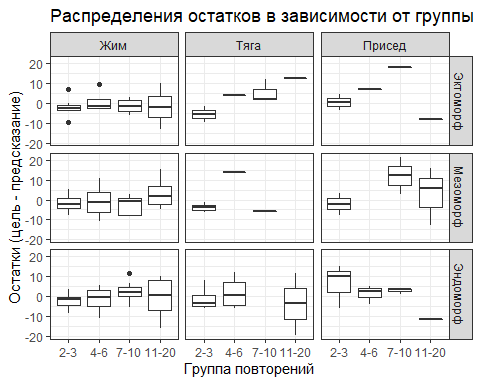
**3 Уточнение коэффициентов**

Теперь возьмём модель вида и подберём коэффициенты лучшим образом. В итоге:

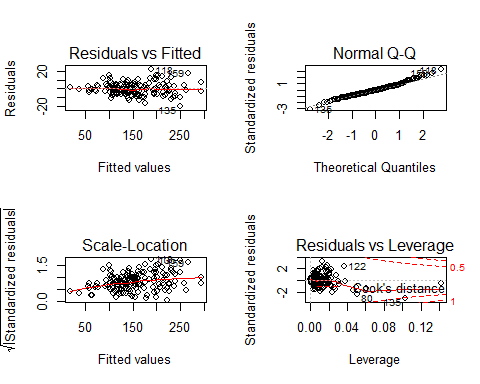
## ------------> Наибольшая ошибка в большую сторону: 19.33226   
## ------------> Наибольшая ошибка в меньшую сторону: 21.46997   
## Модель ошиблась более чем на 5 % в 39 случаях из 162 ( 24.07407 %)  
## Модель ошиблась более чем на 5 кг в 66 случаях из 162 ( 40.74074 %)  
## ----------------> Статистика по ошибкам в процентах:  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.001034 1.344302 3.072494 3.463881 4.816583 13.279030   
## -------------------> Среднеквадратичная ошибка: 6.656834   
## Оценка кросс-валидации для всего набора данных 46.67453   
## Оценка кросс-валидации для не более чем 10 повторений 31.99091   
## Оценка кросс-валидации для не более чем 6 повторений 30.00755



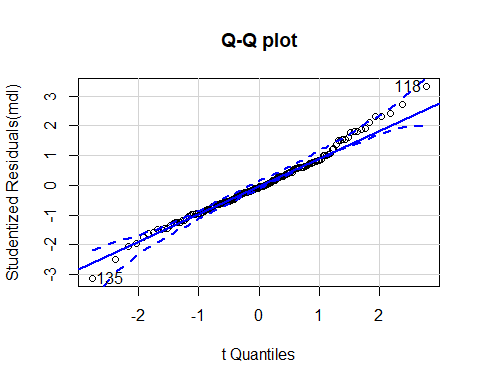




## -----> ОБЩАЯ ИНФОРМАЦИЯ О МОДЕЛИ:  
##   
##   
## Call:  
## lm(formula = RM ~ MRM + MRM:Count - 1, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.3323 -4.3200 -0.4872 3.9362 21.4700   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## MRM 1.0470734 0.0067757 154.53 <2e-16 \*\*\*  
## MRM:Count 0.0209793 0.0008774 23.91 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.698 on 160 degrees of freedom  
## Multiple R-squared: 0.9983, Adjusted R-squared: 0.9983   
## F-statistic: 4.664e+04 on 2 and 160 DF, p-value: < 2.2e-16  
##   
##   
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS  
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:  
## Level of Significance = 0.05   
##   
## Call:  
## gvlma::gvlma(x = mdl)   
##   
## Value p-value Decision  
## Global Stat 15.1022 0.004494 Assumptions NOT satisfied!  
## Skewness 3.1467 0.076078 Assumptions acceptable.  
## Kurtosis 1.9628 0.161218 Assumptions acceptable.  
## Link Function 0.2725 0.601627 Assumptions acceptable.  
## Heteroscedasticity 9.7202 0.001823 Assumptions NOT satisfied!  
## Value p-value Decision  
## Global Stat 15.1022329 0.004493809 Assumptions NOT satisfied!  
## Skewness 3.1467476 0.076078463 Assumptions acceptable.  
## Kurtosis 1.9627685 0.161217544 Assumptions acceptable.  
## Link Function 0.2725495 0.601626657 Assumptions acceptable.  
## Heteroscedasticity 9.7201673 0.001822569 Assumptions NOT satisfied!  
##   
## -----> БАЗОВЫЕ ГРАФИКИ:



##   
## -----> ТЕСТ НА НОРМАЛЬНОСТЬ РАСПРЕДЕЛЕНИЯ ОСТАТКОВ  
##   
##   
## Shapiro-Wilk normality test  
##   
## data: mdl$residuals  
## W = 0.98782, p-value = 0.1726



## -----> ФАКТОР ИНФЛЯЦИИ ДИСПЕРСИЙ:  
##   
## MRM MRM:Count   
## 3.057194 3.057194

Выше приведена вся необходимая информация о качестве модели (назовём эту модель ). Из неё важнее всего следующее:

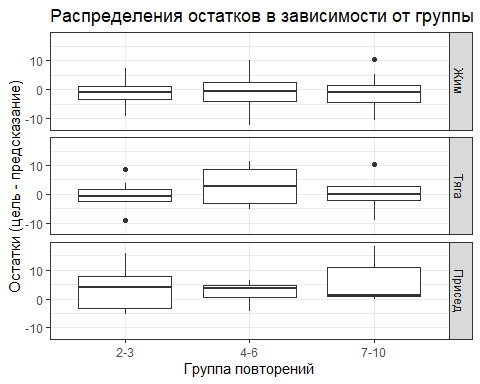
* На выборочных данных модель ошибается максимум на 21кг и 13%;
* В среднем модель ошибается на 3%;
* В 75% случаев ошибка не превышала 5%;
* Наибольший разброс ошибок приходится на диапазон 11-20;
* Модель статистически значима и удовлетворяет всем нужным требованиям, кроме гетероскедастичности;
* Немало больших ошибок приходится на диапазон 11-20 повторений, содержащий меньше всего наблюдений.

**4 Уточнение коэффициентов: зависимость от диапазона**

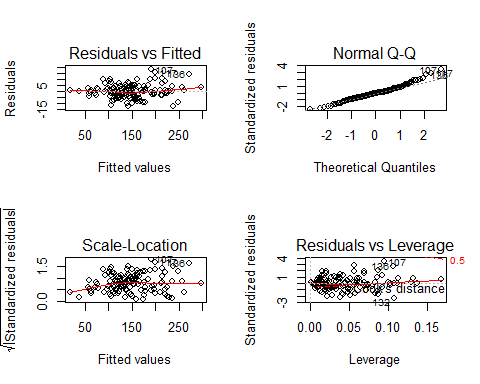
Теперь сделаем поправку для коэффициентов в зависимости от факторных переменных. Путём подбора удалось найти две схожие модели (назовём их и ), немного отличающиеся оценками кросс-валидации на разных диапазонах.

Далее приведён анализ для модели , в которой и и имеют поправку на диапазон повторений.

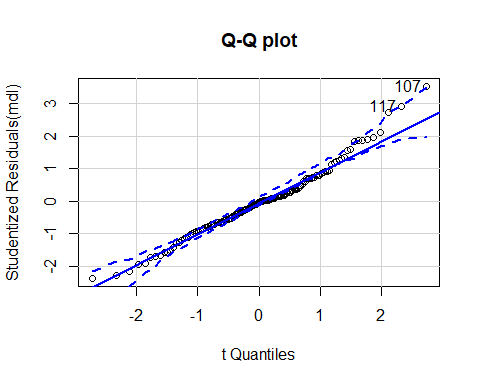
## ------------> Наибольшая ошибка в большую сторону: 12.46227   
## ------------> Наибольшая ошибка в меньшую сторону: 18.24883   
## Модель ошиблась более чем на 5 % в 28 случаях из 139 ( 20.14388 %)  
## Модель ошиблась более чем на 5 кг в 44 случаях из 139 ( 31.65468 %)  
## ----------------> Статистика по ошибкам в процентах:  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00977 0.91712 2.53843 2.98739 4.29788 12.47021   
## -------------------> Среднеквадратичная ошибка: 5.529521   
## Оценка кросс-валидации для всего набора данных 34.03878   
## Оценка кросс-валидации для не более чем 10 повторений 34.20106   
## Оценка кросс-валидации для не более чем 6 повторений 29.84518



## -----> ОБЩАЯ ИНФОРМАЦИЯ О МОДЕЛИ:  
##   
##   
## Call:  
## lm(formula = RM ~ MRM:Count:CountGroup + MRM:CountGroup - 1,   
## data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.4623 -3.6210 -0.1014 2.5825 18.2488   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## MRM:CountGroup2-3 0.988699 0.029669 33.324 < 2e-16 \*\*\*  
## MRM:CountGroup4-6 0.997533 0.035651 27.981 < 2e-16 \*\*\*  
## MRM:CountGroup7-10 1.039799 0.055896 18.603 < 2e-16 \*\*\*  
## MRM:Count:CountGroup2-3 0.039572 0.011331 3.492 0.000650 \*\*\*  
## MRM:Count:CountGroup4-6 0.031609 0.007065 4.474 1.63e-05 \*\*\*  
## MRM:Count:CountGroup7-10 0.023854 0.006434 3.708 0.000306 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.653 on 133 degrees of freedom  
## Multiple R-squared: 0.9987, Adjusted R-squared: 0.9987   
## F-statistic: 1.768e+04 on 6 and 133 DF, p-value: < 2.2e-16  
##   
##   
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS  
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:  
## Level of Significance = 0.05   
##   
## Call:  
## gvlma::gvlma(x = mdl)   
##   
## Value p-value Decision  
## Global Stat 8.359 0.07929 Assumptions acceptable.  
## Skewness 2.795 0.09455 Assumptions acceptable.  
## Kurtosis 2.394 0.12180 Assumptions acceptable.  
## Link Function -2.689 1.00000 Assumptions acceptable.  
## Heteroscedasticity 5.858 0.01550 Assumptions NOT satisfied!  
## Value p-value Decision  
## Global Stat 8.358530 0.07929330 Assumptions acceptable.  
## Skewness 2.795168 0.09454886 Assumptions acceptable.  
## Kurtosis 2.394055 0.12179736 Assumptions acceptable.  
## Link Function -2.689003 1.00000000 Assumptions acceptable.  
## Heteroscedasticity 5.858310 0.01550367 Assumptions NOT satisfied!  
##   
## -----> БАЗОВЫЕ ГРАФИКИ:



##   
## -----> ТЕСТ НА НОРМАЛЬНОСТЬ РАСПРЕДЕЛЕНИЯ ОСТАТКОВ  
##   
##   
## Shapiro-Wilk normality test  
##   
## data: mdl$residuals  
## W = 0.98086, p-value = 0.0484



## -----> ФАКТОР ИНФЛЯЦИИ ДИСПЕРСИЙ:  
##   
## GVIF Df GVIF^(1/(2\*Df))  
## MRM:CountGroup 60763.96 3 6.270098  
## MRM:Count:CountGroup 60763.96 3 6.270098

Аналогичные выводы:

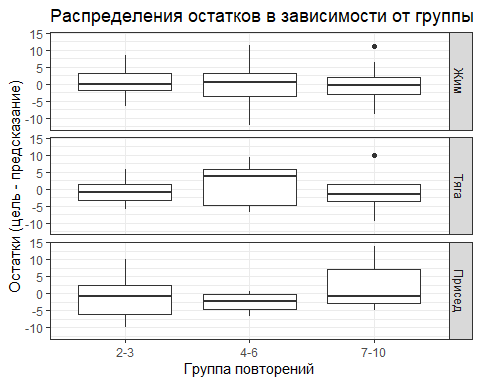
* На выборочных данных модель ошибается максимум на 18кг и 13%;
* В среднем модель ошибается на 3%;
* В 80% случаев ошибка не превышала 5%;
* Наибольший разброс ошибок приходится на диапазон 4-6;
* Модель статистически значима и удовлетворяет всем нужным требованиям, кроме гетероскедастичности.

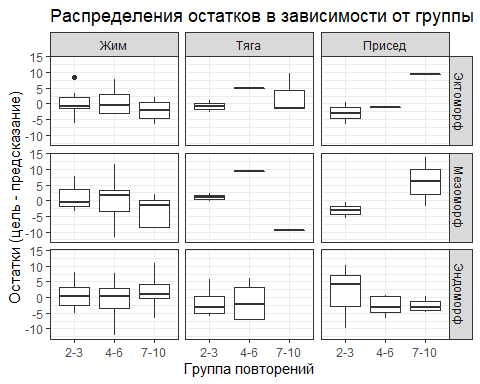
**5 Лучшие модели**

Путём пошагового отбора переменных были найдены ещё две равнозначные модели ( и ), которые, несмотря на б**о**льшую сложность в сравнении с предыдущими, имеют лучшие оценки при кросс-валидации.

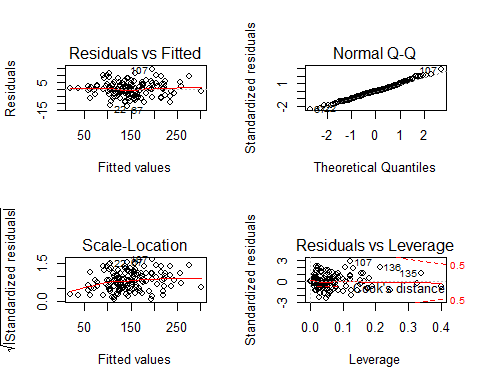
Статистика для модели :

## ------------> Наибольшая ошибка в большую сторону: 12.09735   
## ------------> Наибольшая ошибка в меньшую сторону: 13.90043   
## Модель ошиблась более чем на 5 % в 21 случаях из 139 ( 15.10791 %)  
## Модель ошиблась более чем на 5 кг в 44 случаях из 139 ( 31.65468 %)  
## ----------------> Статистика по ошибкам в процентах:  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.01759 0.94961 2.08660 2.83833 3.94832 12.31595   
## -------------------> Среднеквадратичная ошибка: 5.026615   
## Оценка кросс-валидации для всего набора данных 31.61373   
## Оценка кросс-валидации для не более чем 10 повторений 29.42087   
## Оценка кросс-валидации для не более чем 6 повторений 29.97883

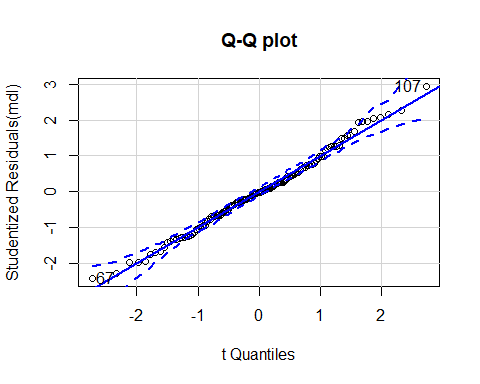




## -----> ОБЩАЯ ИНФОРМАЦИЯ О МОДЕЛИ:  
##   
##   
## Call:  
## lm(formula = RM ~ I((MRM/Index)^6) + MRM:CountGroup + MRM:Action +   
## MRM:CountGroup:Count - 1, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.0974 -3.2898 -0.1568 3.0781 13.9004   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## I((MRM/Index)^6) -2.989e-05 1.125e-05 -2.657 0.00887 \*\*   
## MRM:CountGroup2-3 9.963e-01 2.848e-02 34.984 < 2e-16 \*\*\*  
## MRM:CountGroup4-6 9.711e-01 3.401e-02 28.558 < 2e-16 \*\*\*  
## MRM:CountGroup7-10 1.077e+00 5.195e-02 20.738 < 2e-16 \*\*\*  
## MRM:ActionТяга 2.893e-02 1.014e-02 2.854 0.00502 \*\*   
## MRM:ActionПрисед 4.984e-02 9.648e-03 5.166 8.78e-07 \*\*\*  
## MRM:CountGroup2-3:Count 3.325e-02 1.067e-02 3.116 0.00226 \*\*   
## MRM:CountGroup4-6:Count 3.578e-02 6.842e-03 5.229 6.62e-07 \*\*\*  
## MRM:CountGroup7-10:Count 1.833e-02 6.011e-03 3.050 0.00278 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.198 on 130 degrees of freedom  
## Multiple R-squared: 0.999, Adjusted R-squared: 0.9989   
## F-statistic: 1.395e+04 on 9 and 130 DF, p-value: < 2.2e-16  
##   
##   
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS  
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:  
## Level of Significance = 0.05   
##   
## Call:  
## gvlma::gvlma(x = mdl)   
##   
## Value p-value Decision  
## Global Stat 2.19263 0.7004 Assumptions acceptable.  
## Skewness 0.36972 0.5432 Assumptions acceptable.  
## Kurtosis 0.02639 0.8710 Assumptions acceptable.  
## Link Function 1.38357 0.2395 Assumptions acceptable.  
## Heteroscedasticity 0.41295 0.5205 Assumptions acceptable.  
## Value p-value Decision  
## Global Stat 2.19262765 0.7003792 Assumptions acceptable.  
## Skewness 0.36971780 0.5431581 Assumptions acceptable.  
## Kurtosis 0.02639011 0.8709512 Assumptions acceptable.  
## Link Function 1.38356953 0.2394941 Assumptions acceptable.  
## Heteroscedasticity 0.41295021 0.5204759 Assumptions acceptable.  
##   
## -----> БАЗОВЫЕ ГРАФИКИ:



##   
## -----> ТЕСТ НА НОРМАЛЬНОСТЬ РАСПРЕДЕЛЕНИЯ ОСТАТКОВ  
##   
##   
## Shapiro-Wilk normality test  
##   
## data: mdl$residuals  
## W = 0.99366, p-value = 0.7982



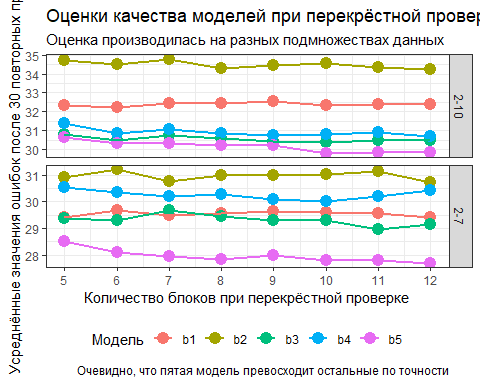
## -----> ФАКТОР ИНФЛЯЦИИ ДИСПЕРСИЙ:  
##   
## GVIF Df GVIF^(1/(2\*Df))  
## I((MRM/Index)^6) 2.561915 1 1.600598  
## MRM:CountGroup 72443.114733 3 6.456535  
## MRM:Action 2.898866 2 1.304839  
## MRM:CountGroup:Count 72727.913065 3 6.460759

Аналогичные выводы:

* На выборочных данных модель ошибается максимум на 13кг и 12%;
* В среднем модель ошибается на менее 3%;
* В 85% случаев ошибка не превышала 5%;
* В целом, на каждом диапазоне имеется почти одинаковый разброс;
* Модель статистически значима и удовлетворяет всем нужным требованиям.

**6 Сравнение линейных моделей**

В результате перекрёстной проверки для всех моделей при числе блоков 5, 6, 7, 8, 9, 10, 11, 12 было обнаружено, что модель значительно превосходит остальные модели, причём это верно, если делать проверку как на всём диапазоне повторений (2-10), так и на более близком к силовому диапазоне 2-7:



На втором месте после модели идёт более простая модель .

**Что это за модели?** Посмотрим на модель :

##   
## Call:  
## lm(formula = RM ~ MRM:Count:CountGroup + MRM:Action - 1, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11.1658 -3.3898 0.0727 3.0208 13.5361   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## MRM:ActionЖим 0.998895 0.019934 50.111 < 2e-16 \*\*\*  
## MRM:ActionТяга 1.012918 0.020963 48.320 < 2e-16 \*\*\*  
## MRM:ActionПрисед 1.036228 0.021859 47.406 < 2e-16 \*\*\*  
## MRM:Count:CountGroup2-3 0.031222 0.007776 4.015 9.86e-05 \*\*\*  
## MRM:Count:CountGroup4-6 0.029581 0.004048 7.308 2.26e-11 \*\*\*  
## MRM:Count:CountGroup7-10 0.027353 0.002420 11.301 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.31 on 133 degrees of freedom  
## Multiple R-squared: 0.9989, Adjusted R-squared: 0.9988   
## F-statistic: 2.004e+04 on 6 and 133 DF, p-value: < 2.2e-16

Она имеет вид

где равен 0.9988949 для жима, 1.0129181 для тяги и 1.0362281 для приседа, а есть поправка на диапазон повторений, равная 0.0312223 для диапазона 2-3, 0.0295811 – для 4-6 и 0.0273531 – для 7-10. Как видно, первые три коэффициента близки к единице, однако различия между ними имеют значение; заметим также, что поправка на диапазон повторений 2-3 довольно близка к коэффициенту Вендлера (0.0333).

Модель имеет два дополнительных слагаемых:

##   
## Call:  
## lm(formula = RM ~ I((MRM/Index)^6) + MRM:CountGroup + MRM:Action +   
## MRM:CountGroup:Count - 1, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.0974 -3.2898 -0.1568 3.0781 13.9004   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## I((MRM/Index)^6) -2.989e-05 1.125e-05 -2.657 0.00887 \*\*   
## MRM:CountGroup2-3 9.963e-01 2.848e-02 34.984 < 2e-16 \*\*\*  
## MRM:CountGroup4-6 9.711e-01 3.401e-02 28.558 < 2e-16 \*\*\*  
## MRM:CountGroup7-10 1.077e+00 5.195e-02 20.738 < 2e-16 \*\*\*  
## MRM:ActionТяга 2.893e-02 1.014e-02 2.854 0.00502 \*\*   
## MRM:ActionПрисед 4.984e-02 9.648e-03 5.166 8.78e-07 \*\*\*  
## MRM:CountGroup2-3:Count 3.325e-02 1.067e-02 3.116 0.00226 \*\*   
## MRM:CountGroup4-6:Count 3.578e-02 6.842e-03 5.229 6.62e-07 \*\*\*  
## MRM:CountGroup7-10:Count 1.833e-02 6.011e-03 3.050 0.00278 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.198 on 130 degrees of freedom  
## Multiple R-squared: 0.999, Adjusted R-squared: 0.9989   
## F-statistic: 1.395e+04 on 9 and 130 DF, p-value: < 2.2e-16

Её вид, аналогично:

Здесь при последнем слагаемом равен -0.0000299, – индекс массы тела.

## ЗАКЛЮЧЕНИЕ

В ходе проделанной работы были рассмотрены несколько линейных моделей для определения повторного максимума в силовых упражнениях. Лучшая модель ошибается максимум на 3%.

Также из самого вида лучшей линейной модели следует, что **для определения ПМ нужны лишь следующие данные**:

* МПМ;
* Число повторений;
* Движение;
* Рост и вес спортсмена.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

[1] Хэтфилд Ф. К., Всестороннее руководство по развитию силы // Красноярск: Союзспорт, 1992. - 284 с.

[2] Вадим Протасенко. Думай! Или «Супертренинг» без заблуждений.

[3] Уикем Х., Гроулмунд Г. Язык R в задачах науки о данных: импорт, обработка, визуализация и моделирование данных. : Пер. с англ. — СПб. : ООО “Диалектика”, 2018. — 592 с.

[4] Мастицкий С. Э., Шитиков В. К. Статистический анализ и визуализация данных с помощью R. – М.: ДМК Пресс, 2015. – 496 с.

[5] Роберт И. Кабаков. R в действии. Анализ и визуализация данных в программе R. / пер. с англ. Полины А. Волковой. – М: ДМК Пресс, 2016. – 588 с.

[6] Джеймс Г., Уиттон Дм Хасти Т., Тибширани Р. Введение в статистическое обучение с примерами на языке R. Пер. с англ. С. Э. Мастицкого - М.: ДМК Пресс, 2016. - 450 с:

[7] Крупкина, Т. В. Математическая статистика [Электронный ресурс] : курс лекций / Т. В. Крупкина, А. К. Гречкосеев. – Электрон. дан. (3 Мб). – Красноярск : ИПК СФУ, 2009.

**ОЦЕНОЧНЫЙ ЛИСТ**

**результатов выполнения заданий   
проектно-технологической практики**

**по направлению подготовки**

**02.04.01 Математика и компьютерные науки**

Студент: Пасько Дмитрий Анатольевич Курс 1 Группа 102/1

Тип практики: производственная практика

|  |  |  |
| --- | --- | --- |
| № | ОБЩАЯ ОЦЕНКА  (отмечается руководителем практики) | Оценка |
|  | УК-6 – Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки |  |
|  | ОПК-2 – Способен создавать и исследовать новые математические модели в естественных науках, совершенствовать и разрабатывать концепции, теории и методы |  |
|  | ОПК-3 – Способен самостоятельно создавать прикладные программные средства на основе современных информационных технологий и сетевых ресурсов, в том числе отечественного производства |  |
|  | ПК-5 – Способен использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и  пакетов прикладных программ моделирования |  |

Итоговая оценка по производственной учебной практике

(технологической

(проектно-технологической) практике) (дифференцированный зачет) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Руководитель практики \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Качанова И.А.

(подпись)