Семинар 6. Красотища

Июнь, 18, 2018

Шаманское заклинание для настройки глобальных опций отчёта:

Подключаем пакеты

library(tidyverse) # обработка данных, графики...  
library(skimr) # описательные статистики  
library(rio) # импорт фантастического количества форматов данных  
library(broom) # метла превращает результаты оценивания моделей в таблички  
library(GGally) # больше готовых графиков  
library(sjPlot) # ещё больше графиков  
library(lmtest) # диагностика линейных моделей  
library(sjstats) # удобные мелкие функции для работы с моделями  
library(sandwich) # оценка Var для гетероскедастичности  
library(AER) # работа с инструментальными переменными  
library(Ecdat) # много-много разных наборов данных  
library(huxtable) # красивые таблички в html, tex  
library(stargazer) # красивые таблички в html, tex  
library(texreg) # и снова красивые таблички в html, tex :)  
library(estimatr) # модели с робастными ошибками

Заметим, что в данном случае, важно, какой пакет подключать раньше, AER, или Ecdat. Оба пакета содержат встроенный набор данных с названием Mroz, и эти наборы данных чуть-чуть отличаются. По умолчанию, активен набор данных Mroz из последнего подключённого пакета. Если нужен другой набор данных, то до него можно дотянуться, указав имя пакета в начале, AER::Mroz.

# Регрессия при гетероскедастичных ошибках

Если не предполагать, что дисперсии ошибок одинаковы для всех наблюдений, то построенные нами доверительные интервалы и выполненная проверка гипотез — полный отстой :)

Есть два подхода работать с робастными стандартными ошибками. Можно переоценить модель с помощью функций из пакета estimatr, а можно использовать ранее оцененные модели, указывая каждый раз нужную оценку ковариационной матрицы.

Загружаем данные и объявляем переменные факторными:

pulse <- import('data/pulse.txt')  
pulse\_fct <- pulse %>%  
 mutate\_at(vars(-Weight, -Height, -Age, -Pulse1, -Pulse2), factor)

Оценим модели без учёта гетероскедастичности:

model\_r <- lm(data = pulse\_fct, Pulse2 ~ Pulse1)  
model\_ur <- lm(data = pulse\_fct, Pulse2 ~ Weight + Pulse1 + Ran + Smokes)  
summary(model\_ur)

##   
## Call:  
## lm(formula = Pulse2 ~ Weight + Pulse1 + Ran + Smokes, data = pulse\_fct)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -39.631 -3.717 0.170 4.360 42.354   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 56.12572 12.14842 4.620 1.10e-05 \*\*\*  
## Weight 0.02187 0.09113 0.240 0.811   
## Pulse1 0.89417 0.10421 8.580 9.77e-14 \*\*\*  
## Ran2 -52.20287 2.75215 -18.968 < 2e-16 \*\*\*  
## Smokes2 1.90805 4.52173 0.422 0.674   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 14.1 on 104 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.8078, Adjusted R-squared: 0.8004   
## F-statistic: 109.3 on 4 and 104 DF, p-value: < 2.2e-16

Способ с переоцениванием модели:

model\_r\_rob <- lm\_robust(data = pulse\_fct, Pulse2 ~ Pulse1)  
model\_ur\_rob <- lm\_robust(data = pulse\_fct, Pulse2 ~ Weight + Pulse1 + Ran + Smokes)  
summary(model\_ur\_rob)

##   
## Call:  
## lm\_robust(formula = Pulse2 ~ Weight + Pulse1 + Ran + Smokes,   
## data = pulse\_fct)  
##   
## Standard error type: HC2   
##   
## Coefficients:  
## Estimate Std. Error Pr(>|t|) CI Lower CI Upper DF  
## (Intercept) 56.12572 18.32138 2.787e-03 19.7937 92.4577 104  
## Weight 0.02187 0.08825 8.047e-01 -0.1531 0.1969 104  
## Pulse1 0.89417 0.18665 5.541e-06 0.5240 1.2643 104  
## Ran2 -52.20287 3.14263 5.027e-31 -58.4348 -45.9709 104  
## Smokes2 1.90805 2.34735 4.182e-01 -2.7468 6.5629 104  
##   
## Multiple R-squared: 0.8078 , Adjusted R-squared: 0.8004   
## F-statistic: 78.48 on 4 and 104 DF, p-value: < 2.2e-16

Проверка гипотез и доверительные интервалы с робастно оценёнными моделями:

coeftest(model\_ur\_rob)

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 56.125722 18.321380 3.0634 0.002787 \*\*   
## Weight 0.021875 0.088252 0.2479 0.804727   
## Pulse1 0.894167 0.186648 4.7907 5.541e-06 \*\*\*  
## Ran2 -52.202870 3.142630 -16.6112 < 2.2e-16 \*\*\*  
## Smokes2 1.908051 2.347350 0.8129 0.418159   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

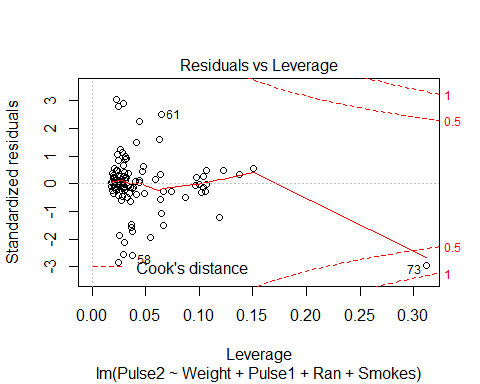
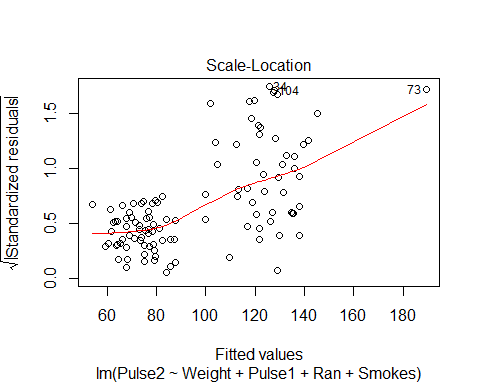
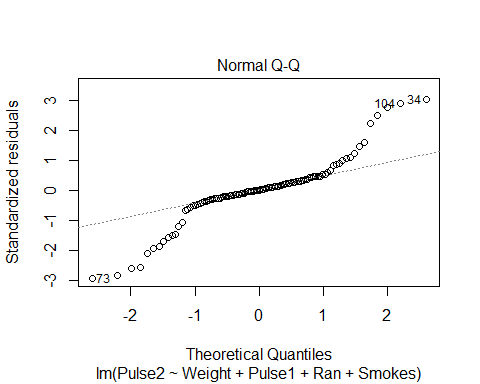
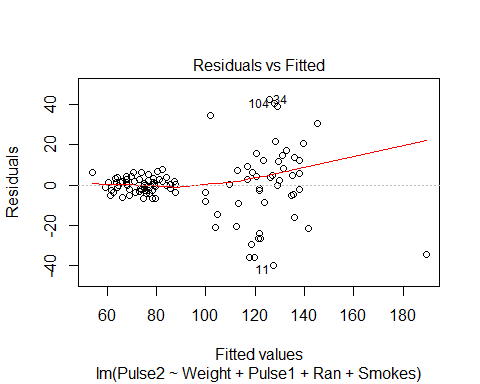
coefci(model\_ur\_rob)

## 2.5 % 97.5 %  
## (Intercept) 19.7937381 92.4577056  
## Weight -0.1531328 0.1968823  
## Pulse1 0.5240371 1.2642973  
## Ran2 -58.4348228 -45.9709175  
## Smokes2 -2.7468326 6.5629340

coefci(model\_ur)

## 2.5 % 97.5 %  
## (Intercept) 32.0349494 80.2164943  
## Weight -0.1588415 0.2025909  
## Pulse1 0.6875121 1.1008223  
## Ran2 -57.6604911 -46.7452493  
## Smokes2 -7.0587150 10.8748163

plot(model\_ur)



resid(model\_ur\_rob)

## NULL

Сравнение двух робастно оценённых моделей:

waldtest(model\_r\_rob, model\_ur\_rob)

┌────────────────────────────────────┐ │ Res.Df Df Chisq Pr(>Chisq) │ ├────────────────────────────────────┤ │ 107         │ │ 104 3 293 3.32e-63 │ └────────────────────────────────────┘

Column names: Res.Df, Df, Chisq, Pr(>Chisq)

Второй способ с использование опции и изначальными моделями, оценёнными без поправки на гетероскадастичность.

coeftest(model\_ur, vcov. = vcovHC)

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 56.125722 21.191573 2.6485 0.009345 \*\*   
## Weight 0.021875 0.095912 0.2281 0.820039   
## Pulse1 0.894167 0.217629 4.1087 7.954e-05 \*\*\*  
## Ran2 -52.202870 3.252667 -16.0492 < 2.2e-16 \*\*\*  
## Smokes2 1.908051 2.516461 0.7582 0.450029   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

coefci(model\_ur, vcov. = vcovHC)

## 2.5 % 97.5 %  
## (Intercept) 14.1020380 98.1494057  
## Weight -0.1683218 0.2120713  
## Pulse1 0.4626001 1.3257342  
## Ran2 -58.6530314 -45.7527089  
## Smokes2 -3.0821863 6.8982876

Тест на гетероскедастичность

Результаты в R чуть отличаются от результатов в stata. По-умолчанию, R использует корректировку HC3 для подсчёта стандартных ошибок коэффициентов, а stata — менее удачную, устаревшую HC1. Сравнение есть у [Achim Zeileis](https://cran.r-project.org/web/packages/sandwich/vignettes/sandwich.pdf). Если хочется воспроизвести именно корректировку HC1, то запросто :)

coeftest(model\_ur, vcov. = vcovHC(model\_ur, type = "HC1"))

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 56.125722 16.388352 3.4247 0.0008819 \*\*\*  
## Weight 0.021875 0.084090 0.2601 0.7952746   
## Pulse1 0.894167 0.165376 5.4069 4.111e-07 \*\*\*  
## Ran2 -52.202870 3.121163 -16.7255 < 2.2e-16 \*\*\*  
## Smokes2 1.908051 2.250136 0.8480 0.3984014   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Сравним две исходных модели с учётом поправки на гетероскедастичность:

waldtest(model\_r, model\_ur, vcov = vcovHC)

┌─────────────────────────────────┐ │ Res.Df Df F Pr(>F) │ ├─────────────────────────────────┤ │ 107            │ │ 104 3 91.8 4.24e-29 │ └─────────────────────────────────┘

Column names: Res.Df, Df, F, Pr(>F)

* Какая модель предпочтительнее?
* Упражнение 8.

Получите таблицу с тестами на значимость коэффициентов в модели house\_ur, используя корректировку HC3. А затем сравните две модели house\_r и house\_ur с учётом поправки на гетероскедастичность.

# coeftest(\_\_\_, vcov. = \_\_\_, type = \_\_\_\_)  
# waldtest(\_\_\_, \_\_\_, vcov = vcovHC)

# Инструментальные переменные

Исследователи всегда мечтают обнаружить не просто статистическую связь, а причинно-следственную. Проще всего было бы обнаружить связь с помощью рандомизированного эксперимента: поделить случайно индивидов случайно на две группы, одну попросить стирать обычным порошком, а вторую — новым чудопорошком. Важно упаковать порошки совершенно одинаково и говорить одни и те же слова :)

Мы часто имеем с данными наблюдений. Рандомизированного эксперимента при этом не было. В редких случаях можно попытаться обнаружить причинно-следственную связь на данных наблюдений.

Рассмотрим набор данных по предложению труда женщин Mroz из пакета Ecdat.

glimpse(Mroz)

## Observations: 753  
## Variables: 18  
## $ work <fct> no, no, no, no, no, no, no, no, no, no, no, no, no,...  
## $ hoursw <int> 1610, 1656, 1980, 456, 1568, 2032, 1440, 1020, 1458...  
## $ child6 <int> 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, ...  
## $ child618 <int> 0, 2, 3, 3, 2, 0, 2, 0, 2, 2, 1, 1, 2, 2, 1, 3, 2, ...  
## $ agew <int> 32, 30, 35, 34, 31, 54, 37, 54, 48, 39, 33, 42, 30,...  
## $ educw <int> 12, 12, 12, 12, 14, 12, 16, 12, 12, 12, 12, 11, 12,...  
## $ hearnw <dbl> 3.3540, 1.3889, 4.5455, 1.0965, 4.5918, 4.7421, 8.3...  
## $ wagew <dbl> 2.65, 2.65, 4.04, 3.25, 3.60, 4.70, 5.95, 9.98, 0.0...  
## $ hoursh <int> 2708, 2310, 3072, 1920, 2000, 1040, 2670, 4120, 199...  
## $ ageh <int> 34, 30, 40, 53, 32, 57, 37, 53, 52, 43, 34, 47, 33,...  
## $ educh <int> 12, 9, 12, 10, 12, 11, 12, 8, 4, 12, 12, 14, 16, 12...  
## $ wageh <dbl> 4.0288, 8.4416, 3.5807, 3.5417, 10.0000, 6.7106, 3....  
## $ income <int> 16310, 21800, 21040, 7300, 27300, 19495, 21152, 189...  
## $ educwm <int> 12, 7, 12, 7, 12, 14, 14, 3, 7, 7, 12, 14, 16, 10, ...  
## $ educwf <int> 7, 7, 7, 7, 14, 7, 7, 3, 7, 7, 3, 7, 16, 10, 7, 10,...  
## $ unemprate <dbl> 5.0, 11.0, 5.0, 5.0, 9.5, 7.5, 5.0, 5.0, 3.0, 5.0, ...  
## $ city <fct> no, yes, no, no, yes, yes, no, no, no, no, no, no, ...  
## $ experience <int> 14, 5, 15, 6, 7, 33, 11, 35, 24, 21, 15, 14, 0, 14,...

Отберём работавших женщин.

labor <- filter(Mroz, wagew > 0)

Мы хотим оценить причинно-следственный эффект от дополнительного года обучения на заработную плату женщины.

На первый взгляд разумно попробовать обычный МНК:

model\_lm <- lm(data = labor, log(wagew) ~ educw + experience + I(experience^2))  
summary(model\_lm)

##   
## Call:  
## lm(formula = log(wagew) ~ educw + experience + I(experience^2),   
## data = labor)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.22168 -0.22193 0.00529 0.24018 1.17754   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.0310084 0.1295490 0.239 0.81098   
## educw 0.0834844 0.0090775 9.197 < 2e-16 \*\*\*  
## experience 0.0238019 0.0087690 2.714 0.00699 \*\*   
## I(experience^2) -0.0003113 0.0002532 -1.229 0.21983   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3784 on 332 degrees of freedom  
## Multiple R-squared: 0.2524, Adjusted R-squared: 0.2457   
## F-statistic: 37.37 on 3 and 332 DF, p-value: < 2.2e-16

Эта регрессия прекрасно оценивает силу статистической взаимосвязи и показывает, насколько различается зарплата двух женщин, у которых разное образование, и одинаковый опыт работы.

Эта регрессия вовсе не говорит, насколько увеличится зарплата данной женщины, если она надумает ещё год поучиться.

Проблема состоит в ненаблюдаемых характеристиках, например, в способностях жещнины. Логично предположить, что ненаблюдаемые способности связаны положительно и с зарплатой и с продолжительностью обучения.

При всей спорности предположения, будем считать, что

где ошибка некоррелирована с остальными переменными в правой части.

Предположим, что образование матери и отца, educwm и educwf не коррелированы с ошибкой . Тогда их можно использовать как инструменты для переменной educw\_i.

В качестве инструментальных переменных для educw возьмём образование матери и отца, educwm и educwf:

model\_iv <- ivreg(data = labor,  
 log(wagew) ~ educw + experience + I(experience^2) |  
 experience + I(experience^2) + educwm + educwf)  
summary(model\_iv, diagnostics = TRUE)

##   
## Call:  
## ivreg(formula = log(wagew) ~ educw + experience + I(experience^2) |   
## experience + I(experience^2) + educwm + educwf, data = labor)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.2646 -0.2287 0.0176 0.2501 1.1828   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.0924839 0.2642086 -0.350 0.72653   
## educw 0.0934865 0.0207397 4.508 9.1e-06 \*\*\*  
## experience 0.0232873 0.0088372 2.635 0.00881 \*\*   
## I(experience^2) -0.0002929 0.0002560 -1.144 0.25335   
##   
## Diagnostic tests:  
## df1 df2 statistic p-value   
## Weak instruments 2 331 39.395 4.49e-16 \*\*\*  
## Wu-Hausman 1 331 0.288 0.592   
## Sargan 1 NA 0.136 0.713   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3791 on 332 degrees of freedom  
## Multiple R-Squared: 0.2497, Adjusted R-squared: 0.2429   
## Wald test: 15.91 on 3 and 332 DF, p-value: 1.085e-09

После вертикальной палочки | в команде ivreg указывают экзогенные переменные, используемые в качестве инструментов.

Если верить в то, что образование матери и отца некоррелировано с ненаблюдаемыми способностями женщины, тогда коэффициент при образовании женщины educw показывает, насколько изменится зарплата женщины при увеличении образования на один год, постоянном опыте и постоянных способностях.

Автоматически команда summary с опцией diagonostics = TRUE проводит три теста:

1. Тест на слабые инструменты

Нулевая гипотеза: : инструментальные переменные не коррелированы с потенциально эндогенным регрессором

1. Тест Хаусмана на равенство оценок IV и обычного МНК.

Нулевая гипотеза: : ковариация потенциально эндогенного регрессора с ошибкой равна нулю.

1. Тест Саргана

Нулевая гипотеза: : инструментальные переменные правильно подобраны, то есть они не коррелированы с ошибкой, а невключённые инструменты и не должны быть включены в модель.

* Какие гипотезы отвергаются, какие не отвергаются в нашем случае?

Легко использовать IV с робастными к гетероскедастичности ошибками

summary(model\_iv, diagnostics = TRUE, vcov. = vcovHC)

##   
## Call:  
## ivreg(formula = log(wagew) ~ educw + experience + I(experience^2) |   
## experience + I(experience^2) + educwm + educwf, data = labor)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.2646 -0.2287 0.0176 0.2501 1.1828   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.0924839 0.2568502 -0.360 0.71902   
## educw 0.0934865 0.0207925 4.496 9.57e-06 \*\*\*  
## experience 0.0232873 0.0086521 2.692 0.00747 \*\*   
## I(experience^2) -0.0002929 0.0002538 -1.154 0.24928   
##   
## Diagnostic tests:  
## df1 df2 statistic p-value   
## Weak instruments 2 331 34.686 2.11e-14 \*\*\*  
## Wu-Hausman 1 331 0.269 0.604   
## Sargan 1 NA 0.136 0.713   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3791 on 332 degrees of freedom  
## Multiple R-Squared: 0.2497, Adjusted R-squared: 0.2429   
## Wald test: 20.21 on 3 and 332 DF, p-value: 4.733e-12

Есть и отдельная функция в пакете estimatr:

model\_iv\_robust <- iv\_robust(data = labor,  
 log(wagew) ~ educw + experience + I(experience^2) |  
 experience + I(experience^2) + educwm + educwf)  
summary(model\_iv\_robust)

##   
## Call:  
## iv\_robust(formula = log(wagew) ~ educw + experience + I(experience^2) |   
## experience + I(experience^2) + educwm + educwf, data = labor)  
##   
## Standard error type: HC2   
##   
## Coefficients:  
## Estimate Std. Error Pr(>|t|) CI Lower CI Upper DF  
## (Intercept) -0.0924839 0.2549478 7.170e-01 -0.5940007 0.4090329 332  
## educw 0.0934865 0.0206390 8.252e-06 0.0528868 0.1340863 332  
## experience 0.0232873 0.0085373 6.717e-03 0.0064932 0.0400813 332  
## I(experience^2) -0.0002929 0.0002493 2.407e-01 -0.0007833 0.0001974 332  
##   
## Multiple R-squared: 0.2497 , Adjusted R-squared: 0.2429   
## F-statistic: 20.53 on 3 and 332 DF, p-value: 3.151e-12

* Упражнение 9.

Для упражнения возьмём данные о потреблении сигарет в США CigaretteeSW из пакета AER. Его описание можно прочесть в справке!

Возьмём наблюдения только за 1995 год и создадим несколько новых переменных: логарифм реальных цен lrprice, логарифм дохода на душу населения lrincome, логарифм количества пачек сигарет на человека lquant и реальный налог на сигареты tdiff.

data("CigarettesSW")  
cig <- subset(CigarettesSW, year == 1995)  
cig$lrprice <- log(cig$price / cig$cpi)  
cig$lrincome <- log(cig$income / cig$population / cig$cpi)  
cig$lquant <- log(cig$packs)  
cig$tdiff <- (cig$taxs - cig$tax) / cig$cpi  
  
# glimpse(\_\_\_)  
# skim(\_\_\_)

Постройте регрессию логарифма количества пачек сигарет на человека lquant на логарифм реальной цены одной пачки lrprice, используя в качестве инстурмента реальный налог tdiff.

# cig\_iv1 <- ivreg(data = \_\_\_, \_\_\_ ~ \_\_\_ | \_\_\_)  
# summary(\_\_\_, diagnostics = \_\_\_)

Добавьте к объясняющим переменным логарифм реального дохода на душу населения lrincome, а к инструментам — реальный налог на сигареты tax / cpi.

# cig\_iv2 <- ivreg(data = \_\_\_, \_\_\_ ~ \_\_\_ + \_\_\_ | \_\_\_\_ + \_\_\_ + I(tax/cpi))  
# summary(\_\_\_, \_\_\_ = \_\_\_)

Мы доносим информацию до читателя в графиках, таблицах и тексте.

R позволяет создавать документы в совершенно разных форматах, мы познакомимся с тремя: html, docx, tex.

# Формат html

Прелесть html состоит в богатстве оформления и динамических элементах.

Начнём с нескольких таблиц.

Составим по-быстрому табличку описательных статистик:

skim\_to\_wide(diamonds) %>%  
 filter(type != 'factor') %>%  
 select(variable, complete, mean, sd, p0, p50, hist) %>%  
 hux(add\_colnames = TRUE) %>% print\_html()

variable

complete

mean

sd

p0

p50

hist

price

53940

3932.8

3989.44

326

2401

carat

53940

0.8

0.47

0.2

0.7

depth

53940

61.75

1.43

43

61.8

table

53940

57.46

2.23

43

57

x

53940

5.73

1.12

0

5.7

y

53940

5.73

1.14

0

5.71

z

53940

3.54

0.71

0

3.53

Оформляем наши таблички симпатичнее:

htmlreg(list(model\_r, model\_ur, model\_r\_rob, model\_ur\_rob),  
 include.ci = FALSE,  
 star.symbol = '\\\*')

Statistical models

Model 1

Model 2

Model 3

Model 4

(Intercept)

31.85

56.13\*\*\*

31.85\*

56.13\*\*

(16.44)

(12.15)

(14.25)

(18.32)

Pulse1

0.86\*\*\*

0.89\*\*\*

0.86\*\*\*

0.89\*\*\*

(0.21)

(0.10)

(0.19)

(0.19)

Weight

0.02

0.02

(0.09)

(0.09)

Ran2

-52.20\*\*\*

-52.20\*\*\*

(2.75)

(3.14)

Smokes2

1.91

1.91

(4.52)

(2.35)

R2

0.13

0.81

0.13

0.81

Adj. R2

0.12

0.80

0.12

0.80

Num. obs.

109

109

109

109

RMSE

29.57

14.10

29.57

14.10

\*\*\*p < 0.001, \*\*p < 0.01, \*p < 0.05

Вариант с доверительными интервалами:

htmlreg(list(model\_r, model\_ur, model\_r\_rob, model\_ur\_rob), ci.force = TRUE)

Statistical models

Model 1

Model 2

Model 3

Model 4

(Intercept)

31.85

56.13\*

31.85\*

56.13\*

[-0.38; 64.07]

[32.32; 79.94]

[3.59; 60.10]

[19.79; 92.46]

Pulse1

0.86\*

0.89\*

0.86\*

0.89\*

[0.44; 1.28]

[0.69; 1.10]

[0.49; 1.23]

[0.52; 1.26]

Weight

0.02

0.02

[-0.16; 0.20]

[-0.15; 0.20]

Ran2

-52.20\*

-52.20\*

[-57.60; -46.81]

[-58.43; -45.97]

Smokes2

1.91

1.91

[-6.95; 10.77]

[-2.75; 6.56]

R2

0.13

0.81

0.13

0.81

Adj. R2

0.12

0.80

0.12

0.80

Num. obs.

109

109

109

109

RMSE

29.57

14.10

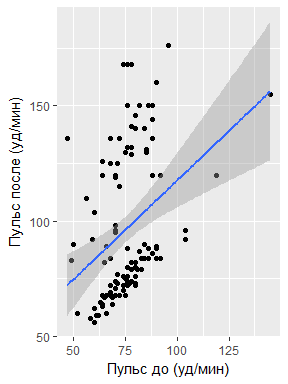
29.57

14.10

\* 0 outside the confidence interval

Просто картинка в html с подписями и ссылкой:

ggplot(data = pulse\_fct, aes(x = Pulse1, y = Pulse2)) +  
 geom\_point() +  
 geom\_smooth(method = 'lm') + labs(x = 'Пульс до (уд/мин)', y = 'Пульс после (уд/мин)')



Для примера процитируем несколько источников, Einstein et al. (1935), Афанасьев and Василевский (1992), Cobb (2011), Гультяев (2008), Doe and Pam (2011).

Другие стилевые заготовки для оформления цитат можно скачать в архиве [zotero](https://www.zotero.org/styles).

Примеры таблиц для html с [пакетом huxtable](https://hughjonesd.github.io/huxtable/), интерактивные элементы, например, карты, с [htmlwidgets](http://gallery.htmlwidgets.org).

И библиографию в студию:

Cobb, G.W., 2011. Teaching statistics: Some important tensions. Chilean Journal of Statistics 2, 31–62.

Doe, J., Pam, S., 2011. The website title.

Einstein, A., Podolsky, B., Rosen, N., 1935. Can quantum-mechanical description of physical reality be considered complete? Physical review 47, 777.

Афанасьев, В.В., Василевский, О.Н., 1992. Расчеты электрических цепей на программируемых микрокалькуляторах. Энергоиздат, Москва.

Гультяев, А., 2008. Microsoft office project 2007 professional. управление проектами: Практическое пособие. СПб:.КОРОНА-Век.