Regresijska analiza i modeliranje sile rezanja

Table of Contents

[Opis Eksperimentalnog istraživanja 1](#_Toc482709927)

[Regresijska Analiza 2](#_Toc482709928)

[Analiza eksperimentalnih podataka 2](#_Toc482709929)

[Korelacijska analiza parametara 5](#_Toc482709930)

[Šta p-vrijednosti (p-value) govori o statističkim podacima 8](#_Toc482709931)

[Linearni regresijski model 9](#_Toc482709932)

[Određivanje najpouzdanijeg regresijskog modela 11](#_Toc482709933)

[Predviđanje vrijednosti sile rezanja 13](#_Toc482709934)

# Opis Eksperimentalnog istraživanja

Izvršeno je eksperimentalno istraživanje sile rezanja, pri čemu se mjerila sila rezanja u funkciji hrapavosti obrađene površine i parametara obrade. Ulazni parametri koji su bili predmet mjerenja predstavljali su:

* Hrapavost obradne površine [MPA],
* Brzina rezanja, [m/mm],
* Posmak obrade, [mm/o],
* Dubina rezanja [mm].

Sila rezanja mjerila se u sva tri pravca i to:

* Aksijalna Sila rezanja, [N]
* Radijalna Sila rezanja, [N] i
* Tangencijalna Sila rezanja [N].

Ukupno je izvršeno 38 mjerenja, pri čemu su se varirali ulazni parametri u sljedećim intervalima:

* Brzina rezanja, , u intervalu od 50 - 300,
* Posmak obrade, , u intervalu od 0,05 do 0,2,
* Dubina rezanja, u intervalu od 0.15 do 0.4.

Zadatak ovog rada predstavlja regresijsku analizu i modeliranje ukupne sile rezanja u funkciji ulaznih parametara, da bi se dobili pouzdani modeli za predviđanje parametre obrade.

Eksperimentalni podaci učitani su iz tekstualne datoteke, te formatirani i primpremljeni za analizu. naime sljedeća R skripta prikazuje primjer učitavanja podataka i formatiranja kolona u tipove koji su spremni za nalizu. Eksperimentalni podaci prikazani su u narednoj tabeli:

library(readr)  
cf\_data <- read\_delim("C:/sc/git/github/MSMO1102/Vjezbe/Sila rezanja/sila\_rezanja\_data.txt",   
 "\t", escape\_double = FALSE, col\_types = cols(HRc = col\_double(),   
 RB = col\_skip(), f = col\_double(),   
 vc = col\_double()), locale = locale(decimal\_mark = ",",   
 grouping\_mark = "."), trim\_ws = TRUE,   
 skip = 10)  
cf\_data

## # A tibble: 38 × 7  
## HRc vc f ap Fa Fr Ft  
## <dbl> <dbl> <dbl> <dbl> <int> <int> <int>  
## 1 45 100 0.10 0.20 56 104 128  
## 2 45 150 0.05 0.20 20 51 50  
## 3 45 150 0.08 0.20 28 69 75  
## 4 45 150 0.10 0.20 28 71 83  
## 5 45 150 0.10 0.30 60 118 135  
## 6 45 150 0.10 0.40 82 129 174  
## 7 45 150 0.20 0.38 40 120 151  
## 8 45 200 0.10 0.20 33 79 91  
## 9 50 100 0.10 0.20 41 111 106  
## 10 50 150 0.05 0.20 35 103 69  
## # ... with 28 more rows

# Regresijska Analiza

Na osnovu eksperimentalnog istraživanja koje je sprovedeno, te identificiranih ulaznih i izlaznih parametara regresijska analiza sprovodi se na:

* 4 ulazna parametra i
* 1 izlazna varijabla koju čini totalnu silu rezanja .

Rezultat regresijske analize predstavljat će izračunavanje regresijskih modela totalne sile rezanja u funkciji ulaznih parametara, odnosno: .

## Analiza eksperimentalnih podataka

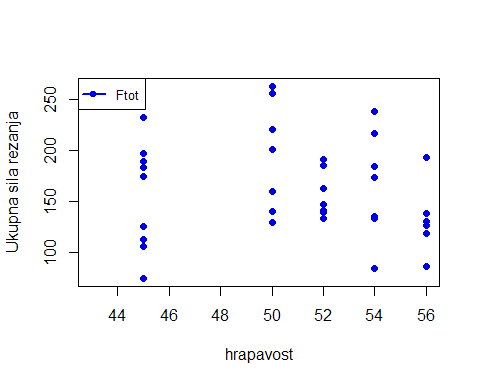
Na samom početku potrebno je izvršiti analizu dobijenih eksperimentalnih podataka. Njbrži način dobijanja globalne slike o podacima jeste predstaviti podatke u obliku grafikona.

Kako smo ranije naglasili izlazna varijable predstavlja totalna sila rezanja $F\_{tot}, na samom početku potrebno je iz postojećih vrijednosti za sve tri komponente odrediti za savko mjerenje totalnu silu. To znači da ćemo dodati novu kolonu u eksperimentu koja će predstavljati ukupnu silu rezanja . To ćemo lahko uraditi ukoliko izvršimo sljedeću R skriptu

#F\_tot nova kolona u eksperimentalnim podacima koja predstavlja totalnu silu rezanja,  
# a dobijena je jednostavnim izračunavanjem internziteta sve tri kkomponente  
cf\_data$Ftot=with(cf\_data, sqrt(Fa^2+Ft^2+Fr^2));

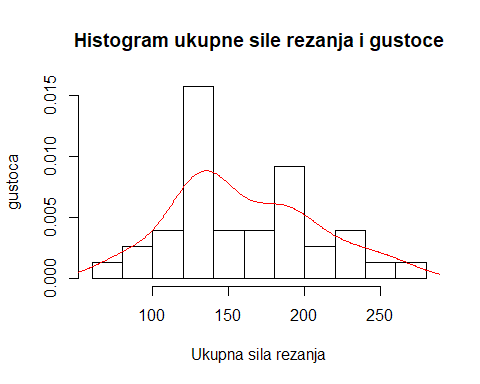
Izvršićemo grafički prikaz podataka u obliku dijagrama raspršenosti i vidjeti kako su podaci organizirani iz grafičke perspektive. Prikažimo dijagrame raspršenosti preko funkcije "plot":

# Define colors to be used for cars, trucks, suvs  
plot\_colors <- c(rgb(r=0.0,g=0.0,b=0.9), "red", "forestgreen")  
  
plot(cf\_data$HRc,cf\_data$Ftot, col=plot\_colors[1], type = "p", pch = 19, xlim = range(43,56), ylim = range(min(cf\_data$Ftot),max(cf\_data$Ftot)), ann = F)  
  
# nazivi x i y osa  
title(xlab="hrapavost")  
title(ylab="Ukupna sila rezanja")  
  
# Legenda na gornjem lijevom uglu dijagrama  
legend("topleft", names(cf\_data[c(8)]), cex=0.8, col=plot\_colors,   
 lwd=2, pch = 19);

 Prethodni dijagram pokazuje silu rezanja u funkciji hrapavosti obratka. Kako se može vidjeti iz prezentiranog dijagrama ne može se odredti linearna ovisnost sile rezanja i hrapavosti. Kvadratna ovisnost može se nazirati iz prikazane vrijednosti.

Donjim dijagramom moguće je pokazati raspodjelu sile rezanja.

hist(cf\_data$Ftot, prob = TRUE,   
 main = "Histogram ukupne sile rezanja i gustoće",   
 xlab ="Ukupna sila rezanja" ,  
 ylab = "gustoća" );  
lines(density( cf\_data$Ftot), col="red" )

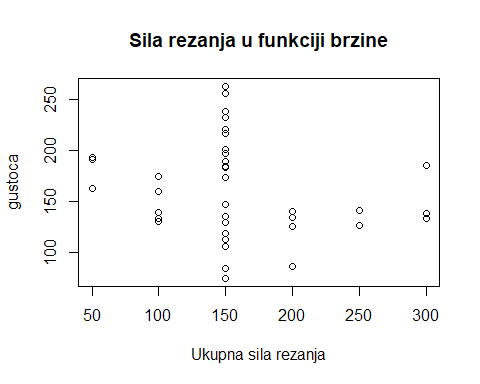


Iz raspodjele sile rezanja uočavamo normalnu distribuciju, kojoj ukupna sile rezanja teži.

## Korelacijska analiza parametara

U ovom dijelu analiziraćemo međusobnu zavisnost odnosno korelaciju među varijablama.Najjednostavnii pristup određivanja uzajamne korelacije među varijablama predstavlja jednostavan dijagrama raspršenosti između dvije varijable. Da bi grafički prikazali uzajamnu korrelaciju između brzine rezanja i totalne sile rezanja koristićemo funkciju plot:

plot(cf\_data$vc,cf\_data$Ftot, main = "Sila rezanja u funkciji brzine",   
 xlab ="Ukupna sila rezanja" ,  
 ylab = "gustoća" );

 Iz prethodnog dijagrama možemo uočiti da nema značajnije korelacije između ove dvije varijable. Ukoliko bi htjeli da se uvjerimo u to pozvaćemo funkciju za izračunavanje korelacije među vdije varijable:

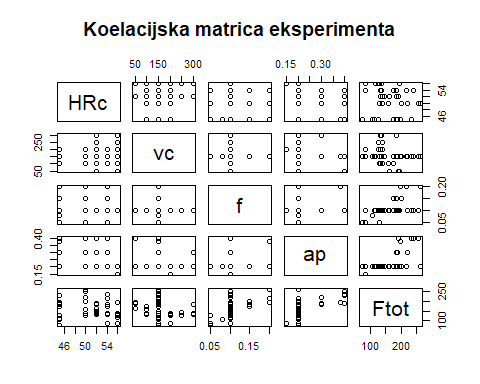
cor(cf\_data$vc,cf\_data$Ftot)

## [1] -0.1988477

Vidimo da je što predstavlja vrlo malu korelaciju među varijablama.

Korištenjem funkcije možemo dobiti grafičku reprezentacije korelacije među svakom od variabli u eksperimentu.

pairs(cf\_data[c(1,2,3,4,8)],main = "Koelacijska matrica eksperimenta")

 Iz gornjih dijagrama vidmo da posmak i dubina rezanja imaju određenu vrstu korelacije sa silom rezanja.

Odredimo sve vrijednosti korelacije u našim eksperimentalnim podacima tako što ćemo funkciji corel proslijediti varijablu cf\_data..

cor(cf\_data[c(1,2,3,4,8)])

## HRc vc f ap Ftot  
## HRc 1.00000000 0.10643454 -0.05828263 -0.22035056 -0.1198057  
## vc 0.10643454 1.00000000 -0.03319519 -0.12193985 -0.1988477  
## f -0.05828263 -0.03319519 1.00000000 0.09076166 0.5787220  
## ap -0.22035056 -0.12193985 0.09076166 1.00000000 0.6425137  
## Ftot -0.11980572 -0.19884771 0.57872202 0.64251374 1.0000000

Iz korelacijske tabele u kojoj su prikazane sve vrijednosti koeficijenta korelacije uočavamo da postoji donekle veza srednje jačine (vrijednosti oko ) između totalne sile rezanja i posmaka koja iznosi , te totalne sile rezanja i dubine rezanja koja iznosi . Da bi korelacijska analiza bila i potvrđena potrebno je izvršiti test korelacijske analize te odrediti jačinu korelacije među parametrima. Testiranje korelacijske analize vršimo preko:

cor.test(cf\_data$vc,cf\_data$Ftot)

##   
## Pearson's product-moment correlation  
##   
## data: cf\_data$vc and cf\_data$Ftot  
## t = -1.2174, df = 36, p-value = 0.2314  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.4875390 0.1290384  
## sample estimates:  
## cor   
## -0.1988477

Rezultat korelacijskog testa za brzinu i silu rezanja pokazuju da je a zbog čega zaključujemo da je korelacija ove dvije varijable vrlo slaba, odnosno test uticajnosti slabu korelaciju.

Testirajmo značajnost korelacije između dubine i sile rezanja .

cor.test(cf\_data$ap,cf\_data$Ftot)

##   
## Pearson's product-moment correlation  
##   
## data: cf\_data$ap and cf\_data$Ftot  
## t = 5.0309, df = 36, p-value = 1.367e-05  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.4062809 0.7982382  
## sample estimates:  
## cor   
## 0.6425137

Iz prikazanog rezultata korelacijskog testa vidimo da dubina i sila rezanja imaju snažnu korelaciju koju potvrđuje test značajnosti korelacije i što predstavlja vrlo malu vrijednost u odnosu na graničnu vrijednost 0.05. Također, iz testa možemo uočiti interval pouzdanosti koji se kreće od 0.4062809 0.7982382.

# Šta p-vrijednosti (p-value) govori o statističkim podacima

Kada se izvršava testiranje hipoteze u statistici, p-vrijednost (p-value) pomaže da se odredi značajnost rezultata testa. Test hipoteze se koristi za testiranje validacije pretpostavke koja se formira nad populacijom. Pretpostavka da nema nikakve zavisnosti između dvije varijable zovemo nulta hipoteza. Alternativne hipoteze predstavljaju hipoteze koje predstavljaju kontradikciju nulte hipoteze, odnosno opovrgavaju nultu hipotezu. Sve hipoteze koriste p-vrijednost da bi izmjerile jačinu dokaza. P-vrijednost predstavlja broj između 0 i 1 koji se interpretira na sljedeći način: \* Mala vrijednosti p-value (oko ≤ 0.05) pokazuje jak dokaz protiv nulte hipoteze. \* Velika vrijednost p-value (> 0.05) pokazuje slabe dokaze protiv nulte hipoteze, tako da nemamo argumenata da odbacimo nultu hipotezu. \* Vrijdnost p-value koja je blizu granične vrijednosti (0.05) predstavlja marginalnu ali ipak dovoljno da se u nekim slučajevima nulta hipoteza odbaci.

## Linearni regresijski model

Nakon analize podataka moguće je definisati prvi model koji će predstavljati linearni regresijski model.Iako već iz korelacijske analize možemo uvidjeti da linearni model neće biti pouzdan, ipak sprovedimo ovu proceduru. Linearni regresijski model definišemo preko funkcije.

l\_model = lm(Ftot ~ HRc + vc + f + ap, data = cf\_data);  
summary(l\_model)

##   
## Call:  
## lm(formula = Ftot ~ HRc + vc + f + ap, data = cf\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -82.339 -14.373 2.177 15.281 49.590   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -26.61275 67.11153 -0.397 0.694   
## HRc 0.64188 1.17448 0.547 0.588   
## vc -0.08733 0.07321 -1.193 0.241   
## f 714.48284 130.47126 5.476 4.52e-06 \*\*\*  
## ap 387.49965 64.36058 6.021 9.08e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 27.02 on 33 degrees of freedom  
## Multiple R-squared: 0.7006, Adjusted R-squared: 0.6644   
## F-statistic: 19.31 on 4 and 33 DF, p-value: 2.858e-08

Iz prikazanog primjera vidimo da smo funkciji definisali formulu modela , pri čemu varijabla koja se nalazi na lijevoj strani tilda simbola predstavlja izlazu varijablu, a s desne strane se nalaze prediktori koji su povezani + operacijom. + operacija označava linearnu zavisnost parametara od izlazne varijable.

Rezultat modeliranja prikazan je gornjom tabelom pri kojoj:

* - odgovara (Intercept),
* - odgovara ,
* - odgovara ,
* - odgovara i
* - odgovara .

Na osnovu rezultata naš linearni regresijski model:

.

Iz prikazanog F - testa uticajnosti parametara vidimo da su samo i parametri uticajni, dok je hrapavost i brzina rezanja imaju visoku vrijednost a što daje malo dokaza za odbijanje nulte hipoteze. S druge strane uočavamo da je , iz kojeg možemo zaključiti da model ne opisuje podatke sa visokim koeficijentom korelacije. Iz tog razloga preporučuje se da se model proširi sa interakcijama odnosno parametrima višeg stepena.

U tom pogledu ispitajmo model pri kojem ćemo uključiti sve interakcije među ulaznim parametrima

l\_model1 = lm(Ftot ~ HRc + vc + f + ap + HRc \* vc + HRc\*f + HRc\*ap + vc\*f + vc\*ap + f\*ap, data = cf\_data);  
summary(l\_model1)

##   
## Call:  
## lm(formula = Ftot ~ HRc + vc + f + ap + HRc \* vc + HRc \* f +   
## HRc \* ap + vc \* f + vc \* ap + f \* ap, data = cf\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -29.045 -11.765 -2.727 8.913 39.980   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.166e+02 3.163e+02 -0.685 0.49900   
## HRc 2.008e+00 5.658e+00 0.355 0.72535   
## vc -1.377e+00 1.177e+00 -1.170 0.25168   
## f 4.769e+03 1.892e+03 2.521 0.01768 \*   
## ap 9.009e+02 7.764e+02 1.160 0.25571   
## HRc:vc 1.307e-02 2.094e-02 0.624 0.53756   
## HRc:f -4.164e+01 3.244e+01 -1.284 0.20973   
## HRc:ap 7.125e-01 1.339e+01 0.053 0.95793   
## vc:f NA NA NA NA   
## vc:ap 2.652e+00 8.882e-01 2.986 0.00581 \*\*   
## f:ap -8.463e+03 1.778e+03 -4.759 5.37e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 18.92 on 28 degrees of freedom  
## Multiple R-squared: 0.8755, Adjusted R-squared: 0.8354   
## F-statistic: 21.87 on 9 and 28 DF, p-value: 2.07e-10

Iz prikazanog test vidimo da su uticajni parametri i te . Isto tako možemo uočiti da je u pitanju vrlo jaka stepen uticajnosti jer je p-vrijednost vlo mala. Isto tako ima veću vrijednost u odnosu na čisti linearni model. Međutim, regresijsku analizu potrebno je vršiti dalje te provjeriti modele višeg stepena.

U ovoj fazi potrebno je testirati regresijske modele višeg reda. U tom pogledu posljednji model ćemo proširiti sa kvadratnim članovima ulaznih parametara. Pa imamo:

l\_model2 = lm(Ftot ~ HRc + vc + f + ap + HRc \* vc + HRc\*f + HRc\*ap + vc\*f + vc\*ap + f\*ap + I(HRc^2) + I(vc^2) + I(f^2) + I(ap^2), data = cf\_data);  
summary(l\_model2)

##   
## Call:  
## lm(formula = Ftot ~ HRc + vc + f + ap + HRc \* vc + HRc \* f +   
## HRc \* ap + vc \* f + vc \* ap + f \* ap + I(HRc^2) + I(vc^2) +   
## I(f^2) + I(ap^2), data = cf\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -29.380 -8.515 -0.750 6.112 43.832   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.559e+03 7.122e+02 -2.189 0.03853 \*   
## HRc 5.473e+01 2.665e+01 2.053 0.05108 .   
## vc -1.810e+00 1.153e+00 -1.569 0.12963   
## f 4.539e+03 1.847e+03 2.457 0.02161 \*   
## ap 1.614e+03 1.095e+03 1.474 0.15352   
## I(HRc^2) -5.119e-01 2.608e-01 -1.962 0.06143 .   
## I(vc^2) 6.223e-04 6.083e-04 1.023 0.31646   
## I(f^2) 1.183e+03 1.994e+03 0.593 0.55840   
## I(ap^2) -3.704e+02 1.015e+03 -0.365 0.71834   
## HRc:vc 1.764e-02 2.043e-02 0.863 0.39646   
## HRc:f -4.286e+01 3.321e+01 -1.290 0.20921   
## HRc:ap -8.845e+00 1.360e+01 -0.651 0.52154   
## vc:f NA NA NA NA   
## vc:ap 2.505e+00 9.529e-01 2.629 0.01472 \*   
## f:ap -8.584e+03 1.967e+03 -4.363 0.00021 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 18.23 on 24 degrees of freedom  
## Multiple R-squared: 0.9009, Adjusted R-squared: 0.8473   
## F-statistic: 16.79 on 13 and 24 DF, p-value: 5.728e-09

Iz posljednjeg testa regresijskog modela vidimo da iznosi , a podešeni (adjusted) R^2 što predstavlja visoku vrijednost s kojim možemo smatrati model pouzdanim. Isto tako iz priloženog testa možemo vidjeti da su uticajni regresijski koeficijenti , , , , i .

# Određivanje najpouzdanijeg regresijskog modela

Iz prethodne analize moguće je izabrati samo one prediktore i njihove interakcije odnosno kvadratne vrijednosti koji su se pokazali značajni kroz ove tri analize, te na osnovu toga odrediti model samo u odnosu na signifikantne parametre. Iz posljednjeg primjera vidimo da svi ulazni parametri imaju određenu signifikantnost jer se vrijednosti kreću oko granične vrijednosti od . Isto tako iz prethodnih modela vidimo da interakcije koje su signifikantne predstavljaju te . Na kraju kvadratni članovi koji su se pokazali signifikantni predstvalja hrapavost . Ako spomenute parametre uključimo u regresijski model, a ostale izostavimo dobijamo sljedeći model:

l\_model3 = lm(Ftot ~ HRc + vc + f + ap + vc\*ap + f\*ap + I(HRc^2), data = cf\_data);  
summary(l\_model3)

##   
## Call:  
## lm(formula = Ftot ~ HRc + vc + f + ap + vc \* ap + f \* ap + I(HRc^2),   
## data = cf\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -24.357 -11.903 -0.519 6.739 45.854   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1336.6155 583.3110 -2.291 0.029133 \*   
## HRc 50.4831 23.6622 2.133 0.041178 \*   
## vc -0.6383 0.1906 -3.348 0.002204 \*\*   
## f 2226.1039 305.8814 7.278 4.20e-08 \*\*\*  
## ap 750.6930 179.6725 4.178 0.000234 \*\*\*  
## I(HRc^2) -0.5036 0.2358 -2.135 0.041005 \*   
## vc:ap 2.4545 0.8042 3.052 0.004725 \*\*   
## f:ap -6487.1567 1244.0646 -5.214 1.27e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 17.64 on 30 degrees of freedom  
## Multiple R-squared: 0.884, Adjusted R-squared: 0.8569   
## F-statistic: 32.66 on 7 and 30 DF, p-value: 2.397e-12

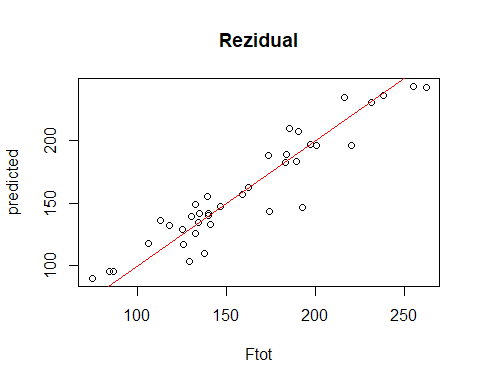
Iz testa vidimo da su svi odabrani parametri signifikatni (značajni), da je i podeseni visoki 0.88 i 0.85 i ujednačnei, koji u konačnici predstavljaju pouzdanije vrijednosti u odnosu na prethodni model. Na osnovu izloženog možemo konačno odrediti najpouzdaniji model koji predstavlja:

$F\_{tot} = -1336.6155 + HRc 50.4831 - 0.6383 v\_c + 2226.1039 f + 750.6930 a\_p -0.5036 HRc^2 + 2.4545 v\_c a\_p + -6487.1567 f a\_p.

d = cf\_data  
d$predicted <- predict(l\_model3) # Save the predicted values  
d$residuals <- residuals(l\_model3) # Save the residual values  
# Quick look at the actual, predicted, and residual values  
d[c(8,9,10)]

## # A tibble: 38 × 3  
## Ftot predicted residuals  
## <dbl> <dbl> <dbl>  
## 1 174.17233 143.62362 3.054870e+01  
## 2 74.16873 89.82102 -1.565229e+01  
## 3 105.68822 117.68120 -1.199297e+01  
## 4 112.75637 136.25465 -2.349827e+01  
## 5 189.07406 183.26943 5.804629e+00  
## 6 231.60527 230.28422 1.321052e+00  
## 7 196.97969 196.97969 5.556917e-14  
## 8 124.94399 128.88567 -3.941682e+00  
## 9 158.86472 156.83426 2.030461e+00  
## 10 128.82158 103.03166 2.578993e+01  
## # ... with 28 more rows

plot(d[c(8,9)], main="Rezidual")  
lines(x=c(0,300), y=c(0,300), col="red")



# Predviđanje vrijednosti sile rezanja

Nakon što smo odredili najpouzdaniji model, u zadnjoj fazi modeliranja regresijskom analizi model je potrebno testirati kako predviđa vrijednost izlazne varijable za ulazne parametre koje nisu iz skupa podataka nad kojim je model određen. U tom pogledu formirajmo novi skup podataka. Novi skup podakata formirajmo od dvije nove vrste u kojim ulazni parametri imaju vrijednosti van prethodnog skupa nad kojim je model izračunat. Vrijednosti za silu rezanja u ovom stadiju ne trebamo jer vrijednosti će biti izračunate preko regresijskog modela. Naime u ovoj fazi potrebno je koristiti novi skup podataka pri kojem testiramo kako model koji je dobijen predviđa vrijednost dsile rezanja. Novi sku podataka sastoji se od dvije vrste.

testData = data.frame(  
 HRc = c(51, 49),  
 vc = c(135, 110),  
 f = c(0.12,0.13),  
 ap =c(0.25,0.26)  
);  
  
testData

## HRc vc f ap  
## 1 51 135 0.12 0.25  
## 2 49 110 0.13 0.26

Na osnovu ulaznih podataka koje smo definisali, izračunajmo vrijednosti ukupne sile rezanja.

# za nove vrijednosti ulaznih parametara izračunavamo silu rezanja pomoću dobijenog modela  
predvidjanje = predict(l\_model3,testData)  
predvidjanje

## 1 2   
## 185.0494 193.2341

Viz gornjeg rezultata vidimo da je ukupna sile rezanja za ulazne parametre a za .