

# R Notebook

## Survey selection

```
survey_3 = df_clean[df_clean$survey == "3",]
```

## Variable Aggregation

```
survey_3 <- survey_3 %>%  
  mutate(operazione_aggregata = case_when(  
    operazione %in% c(2, 3, 5, 6, 9, 10) ~ as.factor(operazione),  
    TRUE ~ "Altro"  
  ))  
survey_3$operazione_aggregata = as.factor((survey_3$operazione_aggregata))
```

```
survey_3 <- survey_3 %>%  
  mutate(regione_aggregata = case_when(  
    regione_des_ana %in% c(1, 12, 3, 17, 19) ~ as.factor(regione_des_ana),  
    TRUE ~ "Altro"  
  ))  
survey_3$regione_aggregata = as.factor((survey_3$regione_aggregata))
```

```
survey_3 <- survey_3 %>%  
  mutate(cs_aggregata = case_when(  
    cs_abi_num_comm %in% c(1,2,3,4,5,6,7,8,9) ~ as.factor(cs_abi_num_comm),  
    TRUE ~ "Altro"  
  ))
```

## NAs Imputation

```
survey_3 <- kNN(survey_3, variable = "risk_rating_comm")
```

## Dataset cleaning

```
survey_3_clean_3 <- survey_3[complete.cases(survey_3[, c("nps", "operazione_aggregata", "fascia_eta_cod  
c("nps", "operazione_aggregata", "fascia_eta_code_ana", "regione_aggregata", "
```

## Linear model

```
modello <- lm(nps ~ segmento_des_comm + operazione_aggregata + regione_aggregata + direzione + sesso_c
```

```
# 3. Visualizza i risultati del modello
```

```
summary(modello)
```

```
##
## Call:
## lm(formula = nps ~ segmento_des_comm + operazione_aggregata +
##     regione_aggregata + direzione + sesso_code_ana + fascia_eta_code_ana +
##     cliente_con_mutuo_flg_ana + cliente_investitore_flg_ana +
##     risk_rating_comm + cliente_solo_cc_flg_ana + cs_aggregata +
##     risk_rating_comm, data = survey_3_clean_3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.4478 -0.6748  0.3720  1.2692  1.9455
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      8.081151   0.194511  41.546 < 2e-16 ***
## segmento_des_comm3 -0.109354   0.045909  -2.382  0.017239 *
## operazione_aggregata2 -0.213274   0.076138  -2.801  0.005102 **
## operazione_aggregata3 -0.018814   0.066470  -0.283  0.777151
## operazione_aggregata5 -0.224421   0.073173  -3.067  0.002168 **
## operazione_aggregata6  0.275015   0.102224   2.690  0.007150 **
## operazione_aggregata9 -0.211981   0.085794  -2.471  0.013497 *
## operazione_aggregataAltro -0.200312   0.069114  -2.898  0.003760 **
## regione_aggregata12  0.055128   0.348411   0.158  0.874282
## regione_aggregata17  0.341129   0.549604   0.621  0.534823
## regione_aggregata19  1.106305   0.940728   1.176  0.239618
## regione_aggregata3  0.139687   0.261764   0.534  0.593603
## regione_aggregataAltro -0.036089   0.081274  -0.444  0.657023
## direzione3          0.018821   0.110069   0.171  0.864236
## direzione4         -0.029071   0.060659  -0.479  0.631765
## direzione5          0.071009   0.079414   0.894  0.371255
## direzione6         -0.058443   0.262953  -0.222  0.824120
## direzione7          0.225555   0.107945   2.090  0.036684 *
## direzione8          0.172616   0.101926   1.694  0.090381 .
## direzione9          0.147716   0.111085   1.330  0.183626
## sesso_code_ana2      0.041209   0.031944   1.290  0.197064
## fascia_eta_code_ana2  0.054455   0.047249   1.153  0.249138
## fascia_eta_code_ana3  0.010986   0.044107   0.249  0.803314
## fascia_eta_code_ana4 -0.057308   0.091132  -0.629  0.529459
## fascia_eta_code_ana5 -0.049861   0.058591  -0.851  0.394790
## fascia_eta_code_ana6  0.007931   0.106545   0.074  0.940666
## fascia_eta_code_ana7  0.133251   0.131840   1.011  0.312183
## cliente_con_mutuo_flg_ana2  0.198556   0.053842   3.688  0.000227 ***
## cliente_investitore_flg_ana2  0.127552   0.042737   2.985  0.002846 **
## risk_rating_comm2      0.095079   0.089280   1.065  0.286923
## risk_rating_comm3      0.150258   0.090394   1.662  0.096491 .
```

```
## risk_rating_comm4          0.055285    0.091317    0.605 0.544909
## risk_rating_comm5          0.206341    0.085138    2.424 0.015384 *
## risk_rating_comm6          0.257245    0.087125    2.953 0.003158 **
## risk_rating_comm7          0.181930    0.084976    2.141 0.032300 *
## cliente_solo_cc_flg_ana2    0.331652    0.173783    1.908 0.056363 .
## cs_aggregata2              -0.157759    0.146497   -1.077 0.281560
## cs_aggregata3              -0.242237    0.141636   -1.710 0.087244 .
## cs_aggregata4              -0.297019    0.137237   -2.164 0.030465 *
## cs_aggregata5              -0.231401    0.137382   -1.684 0.092142 .
## cs_aggregata6              -0.038888    0.139677   -0.278 0.780702
## cs_aggregata7              -0.021536    0.142544   -0.151 0.879913
## cs_aggregata8              0.118302    0.148163    0.798 0.424625
## cs_aggregata9              0.308312    0.162538    1.897 0.057873 .
## cs_aggregataAltro          0.205657    0.154177    1.334 0.182265
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.619 on 10441 degrees of freedom
## Multiple R-squared:  0.01936,    Adjusted R-squared:  0.01522
## F-statistic: 4.684 on 44 and 10441 DF,  p-value: < 2.2e-16
```

```
AIC(modello)
```

```
## [1] 39909.44
```

## Lasso Model

```
x <- model.matrix(nps ~ segmento_des_comm + operazione_aggregata + regione_aggregata + direzione +
y <- survey_3_clean_3$nps

lasso_model <- cv.glmnet(x, y, alpha = 1, intercept = TRUE)

lasso_coefs <- as.matrix(coef(lasso_model, s = "lambda.min"))

print(lasso_model$lambda.min) # Lambda ottimale
```

```
## [1] 0.00357845
```

```
coef(lasso_model, s = "lambda.min") # Coeff

## 66 x 1 sparse Matrix of class "dgCMatrix"
##                                     s1
## (Intercept)                       8.2731398566
## segmento_des_comm10                .
## segmento_des_comm11                .
## segmento_des_comm12                .
## segmento_des_comm13                .
## segmento_des_comm14                .
```

## segmento_des_comm2	.
## segmento_des_comm3	-0.1136348620
## segmento_des_comm4	.
## segmento_des_comm5	.
## segmento_des_comm6	.
## segmento_des_comm7	.
## segmento_des_comm8	.
## segmento_des_comm9	.
## operazione_aggregata2	-0.1662665576
## operazione_aggregata3	0.0105706447
## operazione_aggregata5	-0.1796328849
## operazione_aggregata6	0.2962047984
## operazione_aggregata9	-0.1621114007
## operazione_aggregataAltro	-0.1566277917
## regione_aggregata12	-0.0005644009
## regione_aggregata17	0.2307049811
## regione_aggregata19	0.9427874055
## regione_aggregata3	0.0187412837
## regione_aggregataAltro	.
## direzione10	.
## direzione11	.
## direzione2	-0.0413733772
## direzione3	-0.0566503430
## direzione4	-0.0744203664
## direzione5	.
## direzione6	.
## direzione7	0.1253514166
## direzione8	0.0778858249
## direzione9	0.0474057394
## sesso_code_ana2	0.0310481913
## fascia_eta_code_ana2	0.0407549651
## fascia_eta_code_ana3	.
## fascia_eta_code_ana4	-0.0432982186
## fascia_eta_code_ana5	-0.0441684060
## fascia_eta_code_ana6	.
## fascia_eta_code_ana7	0.0980005442
## fascia_eta_code_ana8	.
## cliente_con_mutuo_flg_ana2	0.1789043006
## cliente_investitore_flg_ana2	0.1083929713
## cliente_solo_cc_flg_ana2	0.2481834459
## cs_aggregata2	-0.0856948317
## cs_aggregata3	-0.1722565180
## cs_aggregata4	-0.2320573487
## cs_aggregata5	-0.1751494792
## cs_aggregata6	.
## cs_aggregata7	0.0026629249
## cs_aggregata8	0.1343991336
## cs_aggregata9	0.3074603970
## cs_aggregataAltro	0.2337013956
## risk_rating_comm10	.
## risk_rating_comm11	.
## risk_rating_comm12	.
## risk_rating_comm2	.
## risk_rating_comm3	0.0380694740

```
## risk_rating_comm4          -0.0293488393
## risk_rating_comm5          0.1023027990
## risk_rating_comm6          0.1522551824
## risk_rating_comm7          0.0782308326
## risk_rating_comm8          .
## risk_rating_comm9          .

# Identifica le variabili con coefficienti diversi da 0
selected_variables <- rownames(lasso_coefs)[lasso_coefs != 0]
selected_variables <- selected_variables[selected_variables != "(Intercept)"] # Escludi l'intercetta

# Estrai le colonne corrispondenti alle variabili selezionate
x_selected <- x[, selected_variables]

# Costruisci un nuovo modello lineare con solo le variabili selezionate
modello_finale <- lm(survey_3_clean_3$nps ~ x_selected)

# Visualizza il nuovo modello
summary(modello_finale)
```

```
##
## Call:
## lm(formula = survey_3_clean_3$nps ~ x_selected)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.4526 -0.6728  0.3693  1.2678  1.9986
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      8.19082    0.17750  46.145 < 2e-16 ***
## x_selectedsegmento_des_comm3 -0.11519    0.04528  -2.544 0.010971 *
## x_selectedoperazione_aggregata2 -0.21402    0.07609  -2.813 0.004921 **
## x_selectedoperazione_aggregata3 -0.01989    0.06640  -0.300 0.764559
## x_selectedoperazione_aggregata5 -0.22469    0.07309  -3.074 0.002116 **
## x_selectedoperazione_aggregata6  0.27559    0.10215   2.698 0.006989 **
## x_selectedoperazione_aggregata9 -0.21252    0.08570  -2.480 0.013155 *
## x_selectedoperazione_aggregataAltro -0.20029    0.06904  -2.901 0.003729 **
## x_selectedregione_aggregata12 -0.05353    0.23843  -0.224 0.822382
## x_selectedregione_aggregata17  0.37448    0.54333   0.689 0.490690
## x_selectedregione_aggregata19  1.13540    0.93657   1.212 0.225429
## x_selectedregione_aggregata3  0.03085    0.06406   0.482 0.630147
## x_selecteddirezione2 -0.04955    0.06864  -0.722 0.470330
## x_selecteddirezione3 -0.06744    0.07007  -0.962 0.335881
## x_selecteddirezione4 -0.07974    0.05527  -1.443 0.149135
## x_selecteddirezione7  0.13955    0.06665   2.094 0.036321 *
## x_selecteddirezione8  0.09289    0.06455   1.439 0.150154
## x_selecteddirezione9  0.06067    0.07171   0.846 0.397553
## x_selectedsesso_code_ana2  0.04090    0.03189   1.283 0.199691
## x_selectedfascia_eta_code_ana2  0.04784    0.03846   1.244 0.213524
## x_selectedfascia_eta_code_ana4 -0.06268    0.08768  -0.715 0.474730
## x_selectedfascia_eta_code_ana5 -0.05429    0.05312  -1.022 0.306785
## x_selectedfascia_eta_code_ana7  0.12982    0.12899   1.006 0.314250
## x_selectedcliente_con_mutuo_flg_ana2 0.20024    0.05342   3.748 0.000179 ***
```

```
## x_selectedcliente_investitore_flg_ana2  0.12942    0.04230    3.059 0.002223 **
## x_selectedcliente_solo_cc_flg_ana2      0.30603    0.14356    2.132 0.033055 *
## x_selectedcs_aggregata2                 -0.12547    0.08618   -1.456 0.145451
## x_selectedcs_aggregata3                 -0.20750    0.06571   -3.158 0.001594 **
## x_selectedcs_aggregata4                 -0.26136    0.05520   -4.735 2.22e-06 ***
## x_selectedcs_aggregata5                 -0.19646    0.05225   -3.760 0.000171 ***
## x_selectedcs_aggregata7                  0.01372    0.05745    0.239 0.811286
## x_selectedcs_aggregata8                  0.15737    0.06815    2.309 0.020963 *
## x_selectedcs_aggregata9                  0.33951    0.09439    3.597 0.000323 ***
## x_selectedcs_aggregataAltro              0.23687    0.08785    2.696 0.007021 **
## x_selectedrisk_rating_comm3              0.08687    0.06793    1.279 0.200958
## x_selectedrisk_rating_comm4             -0.00872    0.06829   -0.128 0.898397
## x_selectedrisk_rating_comm5              0.14269    0.05934    2.404 0.016213 *
## x_selectedrisk_rating_comm6              0.19334    0.06285    3.076 0.002102 **
## x_selectedrisk_rating_comm7              0.11671    0.05887    1.983 0.047433 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.619 on 10447 degrees of freedom
## Multiple R-squared:  0.01918,    Adjusted R-squared:  0.01561
## F-statistic: 5.376 on 38 and 10447 DF,  p-value: < 2.2e-16
```

```
AIC(modello_finale)
```

```
## [1] 39899.32
```

## Lasso interactions

```
x_selected <- as.data.frame(x_selected)
selected_variables <- colnames(x_selected) # Le variabili selezionate
# Assumendo che 'operazione_aggregata' sia la variabile categorica principale
formula_interazioni <- as.formula(paste("y ~ ( +", paste(selected_variables, collapse = " + "), ")^2"))
modello_interazioni <- lm(formula_interazioni, data = x_selected)
```

## Lasso with interactions vs lasso without interactions

```
anova(modello_finale, modello_interazioni)
```

```
## Analysis of Variance Table
##
## Response: survey_3_clean_3$nps
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x_selected  38   535.3  14.0872    5.3765 < 2.2e-16 ***
## Residuals 10447 27372.6   2.6201
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```