

# Różnorodność Przestrzeni Społecznej

*Szymon Talaga*

*17.12.2014*

Raport przedstawia wyniki analiz regresji badających związek pomiędzy różnorodnością przestrzeni społecznej (ocenianej na podstawie średniej entropii rozkładu typów (skupień) gości odwiedzających te same miejsca co respondent - zmienna `ent_avg`) a kapitałem społecznym operacjonalizowanym jako potencjał do mobilizacji zasobów oraz kontakty społeczne (`soccont`). Dodatkowo zostaną sprawdzone ewentualne interakcje ze skorelowanymi z kapitałem społecznym wymiarami przywiązania do miejsca.

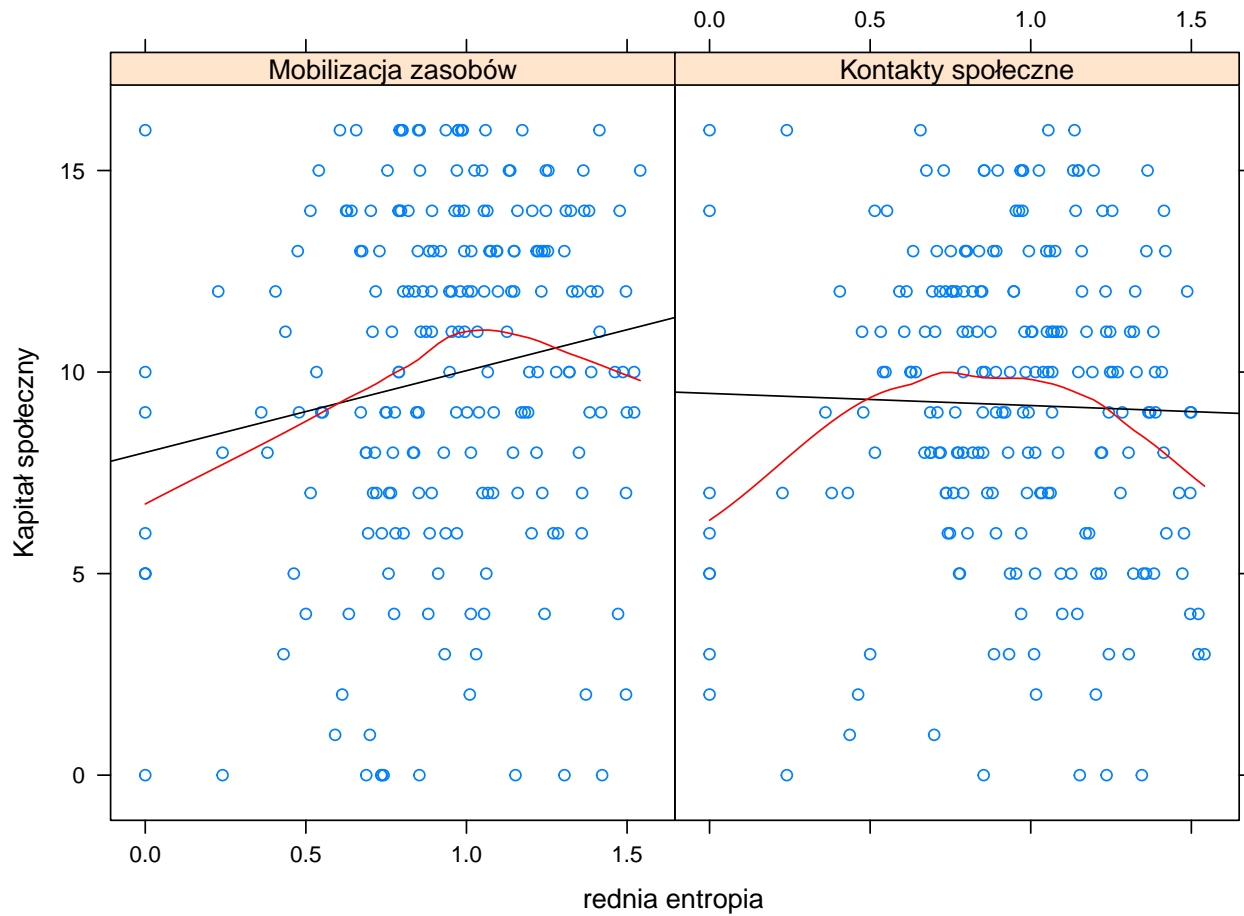
Wczytanie pakietów, funkcji oraz zbioru danych:

```
library(lattice)
library(latticeExtra)
library(car)
library(MASS)
library(psych)
library(reshape2)
library(lmtest)
library(knitr)
library(RColorBrewer)
source("../HelperFunctionsMisc/ComputingMisc.R")
load("../MainData/MainData11.RData")
```

Przygotowanie danych:

```
D.back = D
D = D.back[, c(1, 38:43, 68:72, 76)]
E = D[, c(1:3, 12)]
E = melt(E, id=c("id", "ent_avg"))
levels(E$variable) = c("Mobilizacja zasobów", "Kontakty społeczne")
```

Korelogramy dla entropii i kapitału społecznego z dodanymi trendami liniowymi oraz lokalnie ważonym wygładzeniem typu loess.



Widać, że zwłaszcza w przypadku skali soccont trend wygląda na kwadratowy.

Macierz korelacji pomiędzy kapiałem społecznym, entropią i przywiązaniem do miejsca:

```
##          resmb soccnt ent_v attgn attgv attds attnt
## resmob      1.00
## soccont     0.42  1.00
## ent_avg     0.16 -0.03  1.00
## attgn       0.20  0.22  0.17  1.00
## attgiven    0.02  0.05  0.07  0.57  1.00
## attdiscovered 0.24  0.42  0.16  0.61  0.30  1.00
## attnoatt    0.07 -0.06 -0.01 -0.39 -0.57 -0.09  1.00
```

Ewidentnie jedynie skala ogólna przywiązania oraz miejsce odkryte charakteryzują się wyraźnymi korelacjami z kapitałem społecznym. Dlatego tylko te dwa wymiary zostaną włączone do modeli w celu testowania interakcji.

Ograniczenie zbioru danych tylko do tych respondentów, dla których są obliczone wskaźniki entropijne.

```
D = D[!is.na(D$ent_avg), ]
```

## Mobilizacja zasobów

Model dwuzmiennowy:

```
resmob.lm1 = lm(resmob ~ ent_avg , data=D)
summary(resmob.lm1)
```

```
##
## Call:
## lm(formula = resmob ~ ent_avg, data = D)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.8925  -2.6899   0.8026   3.2157   7.9957
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.0043     0.8469   9.452  <2e-16 ***
## ent_avg        2.0307     0.8500   2.389   0.0178 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.228 on 216 degrees of freedom
## Multiple R-squared:  0.02574,    Adjusted R-squared:  0.02123
## F-statistic: 5.707 on 1 and 216 DF,  p-value: 0.01776
```

Model wyjaśnia 2,6% wariancji, zaś efekt entropii jest istotny na poziomie:  $t_{216} = 2,389$ ;  $p = 0,0178$ . Jednakże lokalnie ważona linia trendu sugeruje, że model uwzględniający związek kwadratowy może być bardziej adekwatny w tym przypadku.

Model kwadratowy:

```
resmob.lm2 = lm(resmob~ent_avg+I(ent_avg^2), data=D)
summary(resmob.lm2)
```

```
##
## Call:
## lm(formula = resmob ~ ent_avg + I(ent_avg^2), data = D)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.4136  -2.7039   0.6428   3.3427   9.8201
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.180     1.287   4.802 2.94e-06 ***
## ent_avg        7.395     2.984   2.478   0.0140 *
## I(ent_avg^2)  -3.229     1.723  -1.874   0.0622 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.204 on 215 degrees of freedom
## Multiple R-squared:  0.04141,    Adjusted R-squared:  0.03249
## F-statistic: 4.643 on 2 and 215 DF,  p-value: 0.01061
```

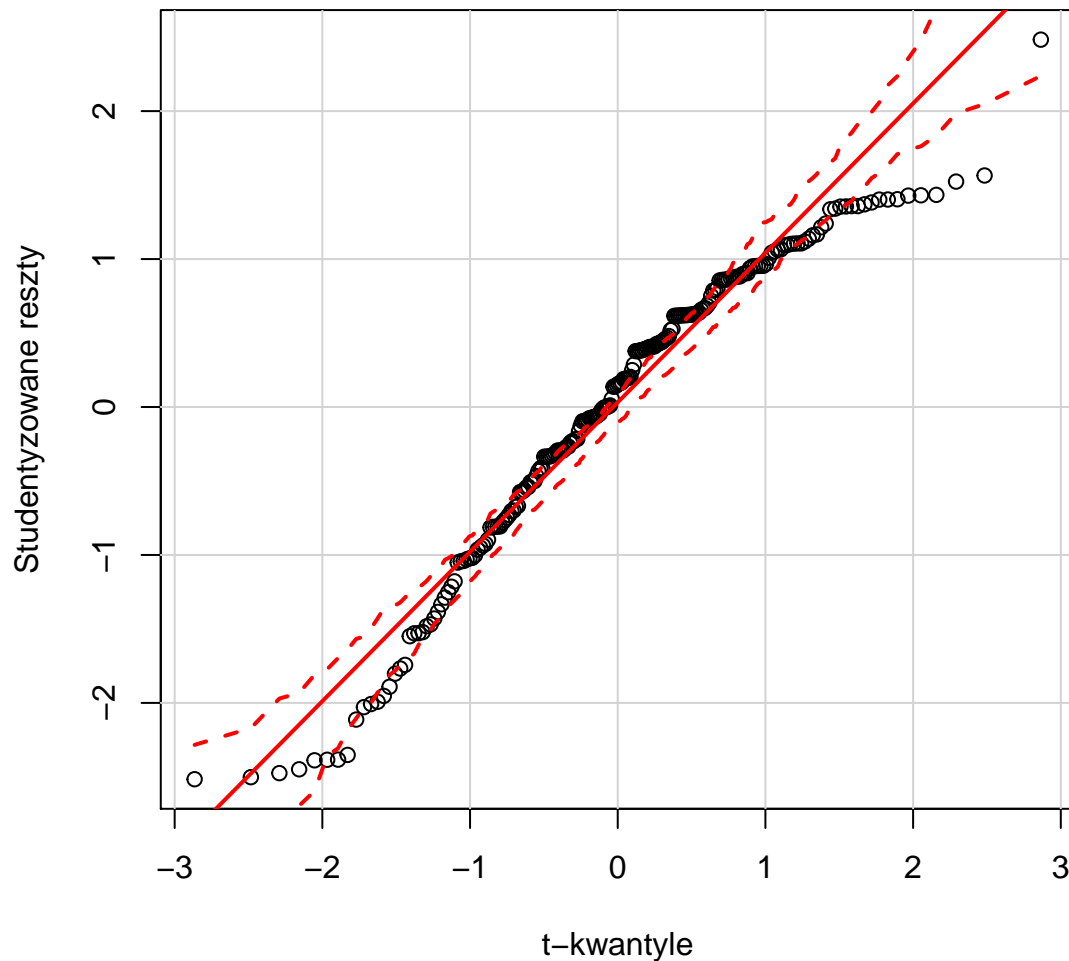
Model kwadratowy wyjaśnia 4,1% wariancji, co daje wzrost o 1,6%. Czynniki liniowy i kwadratowy są istotne na poziomie:  $t_{215} = 2,478$ ;  $p = 0,014$ . Czynniki kwadratowy okazał się być nieistotny:  $t_{215} = -1,874$ ;  $p = 0,062$ . Warto jednak zwrócić uwagę na fakt, że był on bliski istotności. Niemniej jednak na podstawie przeprowadzonej analizy

nie da się określić, czy faktycznie zachodzi zależność o charakterze kwadratowym, czy jest to jedynie osobliwość tego zbioru danych wynikająca z nielosowego doboru próby i niedoskonałości zastosowanych narzędzi pomiarowych. Oczywiście wzrost  $R^2$  również okazała się być nieistotny (tak samo nieznacznie):  $F_{1,215} = 3,513$ ;  $p = 0,062$ .

### Diagnostyka modelu

Normalność rozkładu reszt:

```
qqPlot(resmob.lm2, ylab="Studentyzowane reszty", xlab="t-kwantyle")
```



```
shapiro.test(resmob.lm2$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resmob.lm2$residuals
## W = 0.961, p-value = 1.109e-05
```

Jak pokazują wykres i wynik testu Shapiro-Wilka rozkład reszt modelu nie jest w pełni normalny. Wynika stąd, że estymowane współczynniki mogą nie być nieobciążone.

Test heteroskedastyczności Breuscha-Pagana:

```
bptest(resmob.lm2)
```

```
##  
## studentized Breusch-Pagan test  
##  
## data: resmob.lm2  
## BP = 1.3032, df = 2, p-value = 0.5212
```

Test dał nieistotny wynik, co oznacza, że model charakteryzuje się stałą wariancją reszt.

Test wartości odstających:

```
out.test = outlierTest(resmob.lm2)  
p = out.test$p  
p = p.adjust(p[p<=.05], n = length(p[p<=.05]))  
p <=.05
```

```
## 100  
## TRUE
```

Obserwacja numer 100 jest odstająca. Estymacja modelu bez jej udziału:

```
resmob.lm3 = lm(resmob ~ ent_avg + I(ent_avg^2), data=D, subset=-which(D$id==100))  
summary(resmob.lm3)
```

```
##  
## Call:  
## lm(formula = resmob ~ ent_avg + I(ent_avg^2), data = D, subset = -which(D$id ==  
## 100))  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -10.4285  -2.4291   0.6387   3.3373   9.8368   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)    6.163      1.288   4.784 3.21e-06 ***  
## ent_avg        7.499      2.990   2.508  0.0129 *    
## I(ent_avg^2)  -3.295      1.727  -1.908  0.0577 .    
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 4.208 on 214 degrees of freedom  
## Multiple R-squared:  0.04197,    Adjusted R-squared:  0.03301   
## F-statistic: 4.687 on 2 and 214 DF,  p-value: 0.01018
```

```
anova(resmob.lm3)
```

```
## Analysis of Variance Table  
##  
## Response: resmob
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## ent_avg      1  101.5  101.511   5.7322 0.01752 *
## I(ent_avg^2)  1   64.5   64.490   3.6417 0.05769 .
## Residuals    214 3789.7   17.709
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Brak wyraźnych zmian.

### Interakcja z przywiązanie do miejsca

[zmienne niezależne zostają we wszystkich analizach wycelowane, w celu ułatwienia interpretacji i ograniczenia błędów standardowych współczynników (Sosnowski, 2010)]

Ogólna skala przywiązania:

```
resmob.gen = lm(resmob ~ (I(center(ent_avg))+I(center(ent_avg^2)))*I(center(attgen)), data=D)
summary(resmob.gen)
```

```
##
## Call:
## lm(formula = resmob ~ (I(center(ent_avg)) + I(center(ent_avg^2))) *
##     I(center(attgen)), data = D)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5134  -2.4399   0.6485   3.2635   9.0329
##
## Coefficients:
##                                Estimate Std. Error t value
## (Intercept)                   9.94686    0.29620  33.581
## I(center(ent_avg))             4.10897    4.49968   0.913
## I(center(ent_avg^2))          -1.60781    2.37053  -0.678
## I(center(attgen))              0.08228    0.03966   2.075
## I(center(ent_avg)):I(center(attgen)) -0.19846    0.36325  -0.546
## I(center(ent_avg^2)):I(center(attgen)) 0.12304    0.20581   0.598
##                                Pr(>|t|)
## (Intercept)                   <2e-16 ***
## I(center(ent_avg))             0.3622
## I(center(ent_avg^2))           0.4984
## I(center(attgen))              0.0392 *
## I(center(ent_avg)):I(center(attgen)) 0.5854
## I(center(ent_avg^2)):I(center(attgen)) 0.5506
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.186 on 212 degrees of freedom
## Multiple R-squared:  0.06296,    Adjusted R-squared:  0.04086
## F-statistic: 2.849 on 5 and 212 DF,  p-value: 0.01637
```

```
anova(resmob.gen)
```

```
## Analysis of Variance Table
##
```

```
## Response: resmob
##
##           Df Sum Sq Mean Sq F value Pr(>F)
## I(center(ent_avg))          1  102.0  102.046   5.8240 0.01666
## I(center(ent_avg^2))        1   62.1   62.096   3.5440 0.06113
## I(center(attgen))           1   79.1   79.078   4.5132 0.03479
## I(center(ent_avg)):I(center(attgen))  1    0.1    0.110   0.0063 0.93697
## I(center(ent_avg^2)):I(center(attgen))  1    6.3    6.262   0.3574 0.55061
## Residuals                212 3714.6   17.522
##
## I(center(ent_avg))          *
## I(center(ent_avg^2))        .
## I(center(attgen))           *
## I(center(ent_avg)):I(center(attgen))
## I(center(ent_avg^2)):I(center(attgen))
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test brzegowy pokazał, że żaden z efektów nie jest istotnym dodatkiem do modelu bez jego udziału (Bieciak, 2012). Zaś test sekwencyjny (tabela z analizą wariancji) wykazał, że wprowadzie obie zmienne są istotne (poza czynnikiem kwadratowym), to ich interakcje nie są. Wobec tego związki z ogólną skalą przywiązania do miejsca nie będą rozpatrywane.

Miejsce odkryte:

```
resmob.disc = lm(resmob ~ (I(center(ent_avg))+I(center(ent_avg^2)))*I(center(attdiscovered)), data=D)
summary(resmob.disc)
```

```
##
## Call:
## lm(formula = resmob ~ (I(center(ent_avg)) + I(center(ent_avg^2))) *
##     I(center(attdiscovered)), data = D)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.7205  -2.6379   0.3912   3.0689   8.2885
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept)      9.7409    0.2922  33.332
## I(center(ent_avg))      9.6551    4.2477   2.273
## I(center(ent_avg^2))   -4.2165    2.2510  -1.873
## I(center(attdiscovered))  1.1300    0.3796   2.977
## I(center(ent_avg)):I(center(attdiscovered))  6.9112    3.6494   1.894
## I(center(ent_avg^2)):I(center(attdiscovered)) -3.7681    2.1480  -1.754
##              Pr(>|t|)
## (Intercept)      < 2e-16 ***
## I(center(ent_avg))  0.02403 *
## I(center(ent_avg^2)) 0.06242 .
## I(center(attdiscovered)) 0.00325 **
## I(center(ent_avg)):I(center(attdiscovered)) 0.05962 .
## I(center(ent_avg^2)):I(center(attdiscovered)) 0.08083 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 4.11 on 212 degrees of freedom
## Multiple R-squared:  0.09652,    Adjusted R-squared:  0.07521
## F-statistic: 4.529 on 5 and 212 DF,  p-value: 0.0006055
```

```
anova(resmob.disc)
```

```
## Analysis of Variance Table
##
## Response: resmob
##
##              Df Sum Sq Mean Sq F value
## I(center(ent_avg))      1  102.0  102.046   6.0403
## I(center(ent_avg^2))     1   62.1   62.096   3.6756
## I(center(attdiscovered))  1  157.4  157.363   9.3146
## I(center(ent_avg)):I(center(attdiscovered))  1    9.1    9.108   0.5391
## I(center(ent_avg^2)):I(center(attdiscovered))  1   52.0   51.991   3.0774
## Residuals              212 3581.6   16.894
##              Pr(>F)
## I(center(ent_avg))      0.014785 *
## I(center(ent_avg^2))     0.056560 .
## I(center(attdiscovered)) 0.002564 **
## I(center(ent_avg)):I(center(attdiscovered)) 0.463607
## I(center(ent_avg^2)):I(center(attdiscovered)) 0.080830 .
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test brzegowy wskazuje, że liniowy czynnik entropijny jest istotny,  $t_{212} = 2,273$ ;  $p = 0,024$ . Ponadto czynnik kwadratowy jest bliski istotności,  $t_{212} = -1,873$ ;  $p = 0,062$ . Efekt główny przywiązania również jest istotny,  $t_{212} = 2,977$ ;  $p = 0,003$ . Jednocześnie oba efekty interakcji otarły się o istotność - interakcja z czynnikiem liniowym,  $t_{212} = 1,894$ ;  $p = 0,060$ ; interakcja z czynnikiem kwadratowym,  $t_{212} = -1,754$ ;  $p = 0,081$ . Model jako całość wyjaśnia 9,6% wariancji.

Test sekwencyjny wykazał, że przy kolejnym dodawaniu istotny wzrost  $R^2$  obserwowany jest w przypadku liniowego czynnika entropijnego,  $F_{1,212} = 6,040$ ;  $p = 0,015$ , oraz przywiązani,  $F_{1,212} = 9,314$ ;  $p = 0,002$ . Ponadto czynnik kwadratowy okazał się być bardzo bliski istotności,  $F_{1,212} = 3,676$ ;  $p = 0,056$ . To samo tyczy się również interakcji przywiązania z czynnikiem kwadratowym,  $F_{1,212} = 3,077$ ;  $p = 0,081$ . Jako że wiele efektów, i w tym efekty interakcji, są bliskie istotności warto spróbować poddać interpretacji skonstruowane modele oraz dokonać ich wizualizacji.

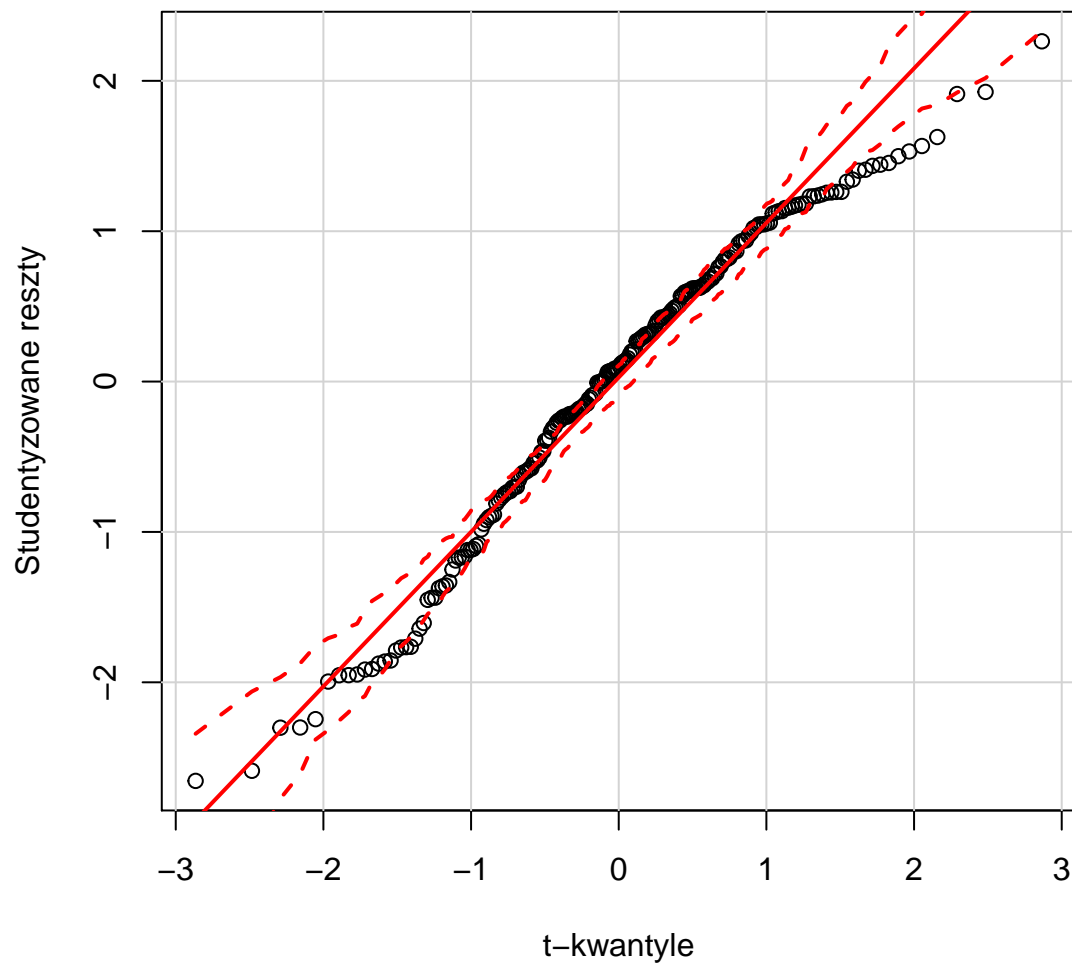
Najpierw jednak należy dokonać diagnostyki modelu z interakcjami.

## Diagnostyka

Normalność rozkładu reszt:

```
qqPlot(resmob.disc, ylab="Studentyzowane reszty", xlab="t-kwantyle")
```





```
shapiro.test(resmob.disc$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resmob.disc$residuals
## W = 0.9737, p-value = 0.0004244
```

Podobnie jak w modelu prostym reszty nie są idealnie normalne.

Heteroskedastyczność:

```
bptest(resmob.disc)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  resmob.disc
## BP = 10.4422, df = 5, p-value = 0.06363
```

Wartość testu Breuscha-Pagana jest bliska istotności, więc warto zastosować inny test w celu otrzymania pewniejszych wyników. Użyty zostanie test Goldfeldta-Quandta.

```
gqtest(resmob.disc)
```

```
##
## Goldfeld-Quandt test
##
## data: resmob.disc
## GQ = 0.9547, df1 = 103, df2 = 103, p-value = 0.5928
```

Test Goldfelda-Quandta jednoznacznie wskazuje na to, że problem heteroskedastyczności nie występuje.

Test wartości odstających:

```
out.test = outlierTest(resmob.disc)
p = out.test$p
p = p.adjust(p[p<=.05], n = length(p[p<=.05]))
p <= .05
```

```
## 100
## TRUE
```

Jedna odstająca obserwacja przy przyjęciu korekty Bonferroniego (id 100). Estymacja modelu bez jej udziału.

```
resmob.disc2 = lm(resmob ~ (I(center(ent_avg)) + I(center(ent_avg^2))) *
  I(center(attdiscovered)), data = D, subset=-which(D$id==100))
summary(resmob.disc2)
```

```
##
## Call:
## lm(formula = resmob ~ (I(center(ent_avg)) + I(center(ent_avg^2))) *
##     I(center(attdiscovered)), data = D, subset = -which(D$id ==
##     100))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.7490  -2.5860   0.4385   3.0835   8.2595
##
## Coefficients:
##                                Estimate Std. Error t value
## (Intercept)                   9.7564     0.2924  33.366
## I(center(ent_avg))             9.8912     4.2506   2.327
## I(center(ent_avg^2))          -4.3514     2.2530  -1.931
## I(center(attdiscovered))       1.1578     0.3802   3.045
## I(center(ent_avg):I(center(attdiscovered))  7.1583     3.6541   1.959
## I(center(ent_avg^2):I(center(attdiscovered)) -3.9158     2.1508  -1.821
##
##                                Pr(>|t|)
## (Intercept)                   < 2e-16 ***
## I(center(ent_avg))             0.02091 *
## I(center(ent_avg^2))           0.05478 .
## I(center(attdiscovered))       0.00262 **
## I(center(ent_avg):I(center(attdiscovered))  0.05144 .
## I(center(ent_avg^2):I(center(attdiscovered)) 0.07009 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.108 on 211 degrees of freedom
## Multiple R-squared:  0.09986,    Adjusted R-squared:  0.07853
## F-statistic: 4.682 on 5 and 211 DF,  p-value: 0.0004482

anova(resmob.disc2)

## Analysis of Variance Table
##
## Response: resmob
##
##              Df Sum Sq Mean Sq F value
## I(center(ent_avg))          1  101.5  101.511   6.0154
## I(center(ent_avg^2))         1   64.5   64.490   3.8216
## I(center(attdiscovered))      1  163.8  163.824   9.7080
## I(center(ent_avg)):I(center(attdiscovered))  1    9.3    9.265   0.5491
## I(center(ent_avg^2)):I(center(attdiscovered)) 1   55.9   55.933   3.3145
## Residuals                211 3560.6   16.875
##
##              Pr(>F)
## I(center(ent_avg))      0.01499 *
## I(center(ent_avg^2))    0.05192 .
## I(center(attdiscovered)) 0.00209 **
## I(center(ent_avg)):I(center(attdiscovered)) 0.45953
## I(center(ent_avg^2)):I(center(attdiscovered)) 0.07009 .
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

W tym przypadku usunięcie obserwacji doprowadziło do wzrostu wartości  $R^2$  do 10%. Jest to jednak nieznamna zmiana.

Jako, że model jest wielozmiennowy należy również sprawdzić, czy nie występuje problem silnych współliniowości pomiędzy zmiennymi. Posłuży do tego statystyka VIF. Jako, że czynnik kwadratowy jest funkcją tej samej zmiennej, co czynnik liniowy VIF zostanie obliczony w oparciu o model z wyłączeniem czynnika kwadratowego (oraz jego interakcji z przywiązaniem).

```
trunc.mod = lm(resmob ~ ent_avg * I(center(attdiscovered)), data=D)
vif(trunc.mod)
```

```
##              ent_avg              I(center(attdiscovered))
##              1.225086              6.571671
## ent_avg:I(center(attdiscovered))
##              6.413247
```

Jak widać żadna ze zmiennych nie ma VIF wyższego niż 10, co oznacza, że nie ma problemu silnej współliniowości (Biecek, 2012).

## Interpretacja

Model prosty (liniowy i kwadratowy czynnik entropijny) pokazuje, że mobilizacja zasobów jest krzywoliniowo zależna od różnorodności przestrzeni społecznej. Zbyt mała i zbyt duża różnorodność wiąże się z obniżonym potencjałem mobilizacji sieci społecznej. Wskazuje na to ujemny współczynnik dla czynnika kwadratowego. Model opisany jest równaniem:

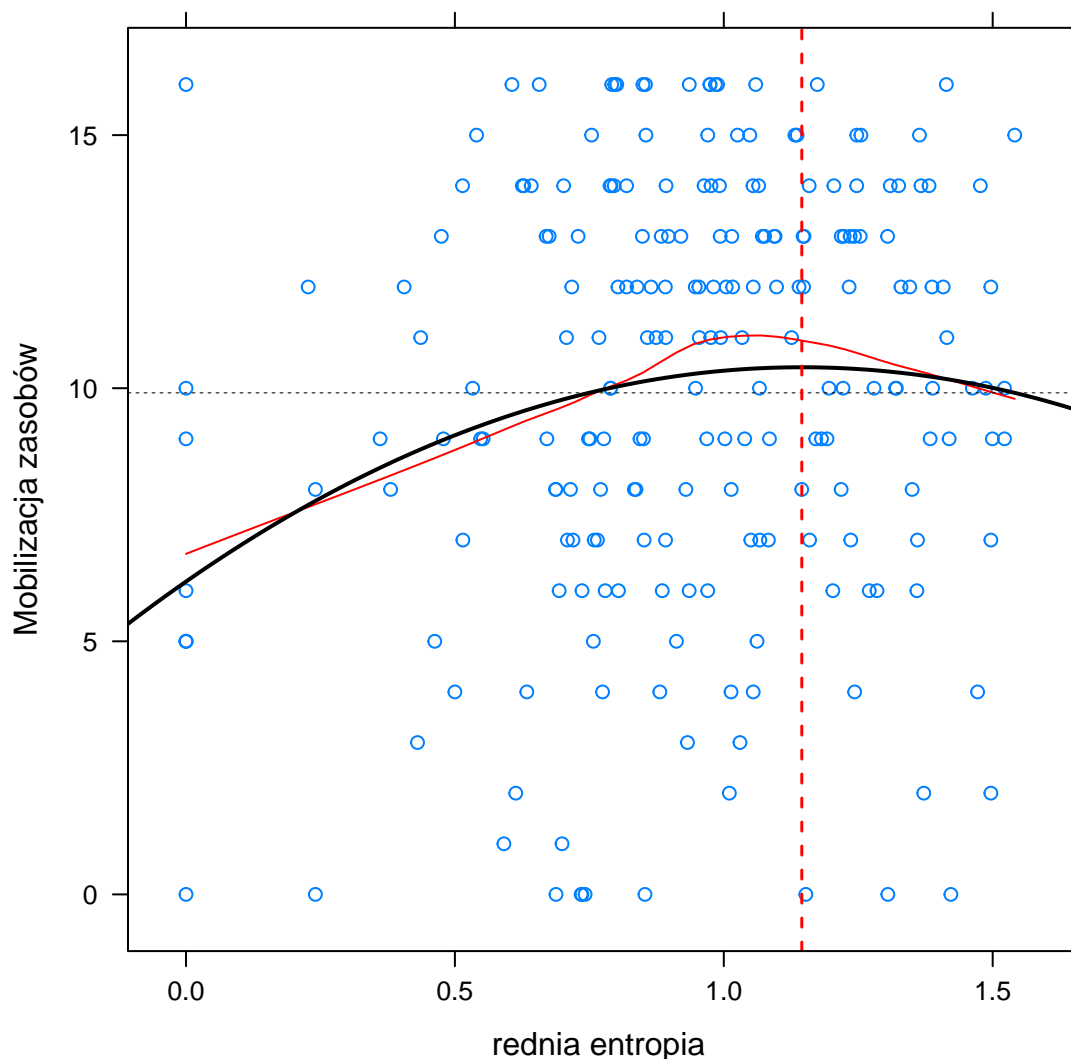
$$Y = 6,180 + 7,395H - 3.229H^2 + \epsilon$$

Jako, że jest to model kwadratowy warto obliczyć jego pochodną. Pozwoli to zobaczyć jaka jest funkcja wpływu entropii na mobilizację zasobów w zależności od jej poziomu. Pochodna opisana jest równaniem:

$$\frac{dY}{dH} = 7,395 - 6,458H$$

Przyrównując pochodną do zera można sprawdzić, gdzie jaka jest optymalny poziom różnorodności, który maksymalizuje potencjał mobilizacyjny sieci społecznej. Łatwo sprawdzić, że optymalną wartości jest:  $H = 1.145091$ . Oznacza to zatem, że średnia entropia przekraczająca trochę jeden bit jest optymalna. Ma to intuicyjną interpretację, wedle której najbardziej wspierające są środowiska, w których stosunkowo łatwo opisać ich typowych przedstawicieli.

Opisane zależności można przedstawić graficznie:

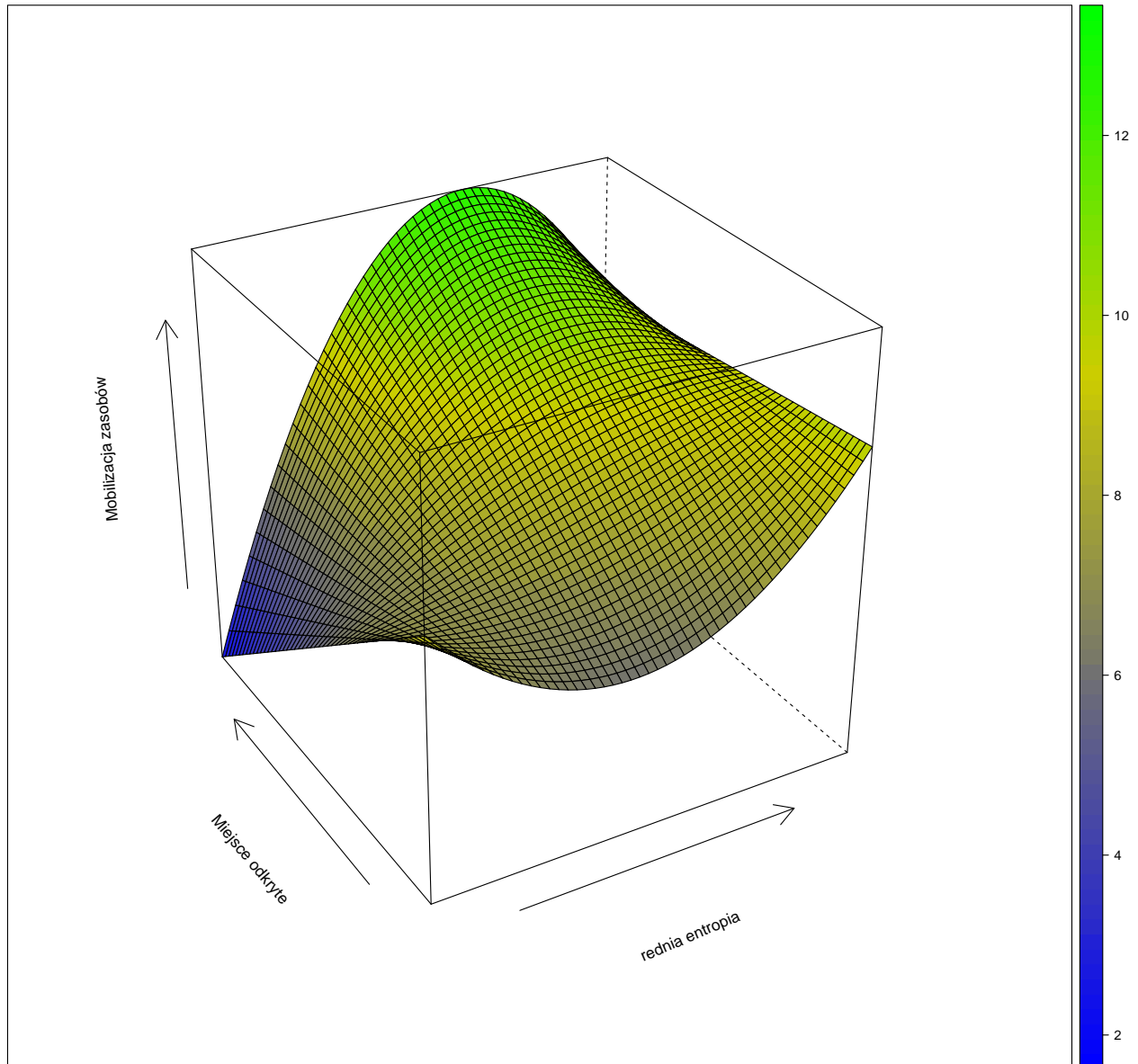


Na wykresie zaznaczono poziomą linię średnią dla mobilizacji zasobów oraz czerwoną pionową linią optymalną wartość średniej entropii. Jak widać krzywa równania modelu pokrywa się w znacznym stopniu z krzywą loess, co świadczy o przyzwoitym dopasowaniu modelu do empirycznego trendu danych. Warto jednak zwrócić uwagę, że ewentualna korzyść płynąca z odpowiedniego poziomu różnorodności jest relatywnie mała - grzbiet krzywej równania modelu wznosi się niewiele powyżej średniej dla mobilizacji zasobów. Oznacza to, że ważniejszym efektem różnorodności jest to, że jej zbyt wysoki bądź niski poziom może mieć zdecydowanie negatywny wpływ.

Model z interakcją z aktywnym przywiązaniem do miejsca ('Miejsce Odkryte') pokazuje, że przywiązanie nie ma istotnego wpływu na potencjał mobilizacyjny sieci społecznej, lecz odgrywa za to ważną rolę moderującą, ponieważ wzmacnia efekt różnorodności. Oznacza to, że w przypadku optymalnej różnorodności jej pozytywny wpływ na mobilizację jest zwiększony, zaś w przypadku za wysokiej bądź za niskiej różnorodności to jej negatywny efekt zostaje wzmożony. Równanie modelu (z pominięciem nieistotnego addytywnego efektu przywiązania) jest postaci:

$$Y = 4,874 + 9,655H - 4,217H^2 + 6,911HP - 3,768H^2P + \epsilon$$

Bardziej intuicyjny wgląd w sens modelu zapewni jego reprezentacja graficzna:



Wykres płaszczyzny regresji pozwala na pełniejsze zrozumienie szczególnego charakteru interakcji między różnorodnością otoczenia społecznego (entropią) a przywiązaniem do miejsca w rozumienia skali 'Miejsce Odkryte'. Widać wyraźnie, że najwyższy potencjał mobilizacji zasobów charakteryzuje osoby o umiarkowanej entropii otoczenia i wysokim przywiązaniu. Jednocześnie im większe odchylenie entropii od jej optymalnego poziomu, tym bardziej negatywny, zamiast pozytywnego, efekt wywiera przywiązanie do miejsca. Sugeruje to,

że osoby aktywnie przywiązane do miejsca potrzebują odpowiednio zróżnicowanego środowiska do czerpania korzyści ze swojego otoczenia społecznego.

## Kontakty społeczne

Model dwuzmiennowy:

```
soccont.lm1 = lm(soccont ~ ent_avg, data=D)
summary(soccont.lm1)

##
## Call:
## lm(formula = soccont ~ ent_avg, data = D)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.3972 -2.3173  0.6938  2.7663  6.8715
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.4694     0.7645  12.387  <2e-16 ***
## ent_avg      -0.3001     0.7673  -0.391   0.696
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.817 on 216 degrees of freedom
## Multiple R-squared:  0.0007077, Adjusted R-squared:  -0.003919
## F-statistic: 0.153 on 1 and 216 DF, p-value: 0.6961
```

Nieistotny wynik nie powinien dziwić. Przedstawiony na początku korelogram kazał przypuszczać, że związek liniowy pomiędzy zmiennymi nie występuje jako taki. Dlatego należy dopasować model kwadratowy.

Model kwadratowy:

```
soccont.lm2 = lm(soccont ~ ent_avg + I(ent_avg^2), data=D)
summary(soccont.lm2)

##
## Call:
## lm(formula = soccont ~ ent_avg + I(ent_avg^2), data = D)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.8354 -2.5647  0.3119  2.3093  9.3546
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.645     1.143   5.813  2.2e-08 ***
## ent_avg        8.004     2.651   3.019  0.00284 **
## I(ent_avg^2)  -4.999     1.530  -3.266  0.00127 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 3.734 on 215 degrees of freedom
## Multiple R-squared:  0.04795,    Adjusted R-squared:  0.0391
## F-statistic: 5.414 on 2 and 215 DF,  p-value: 0.00508
```

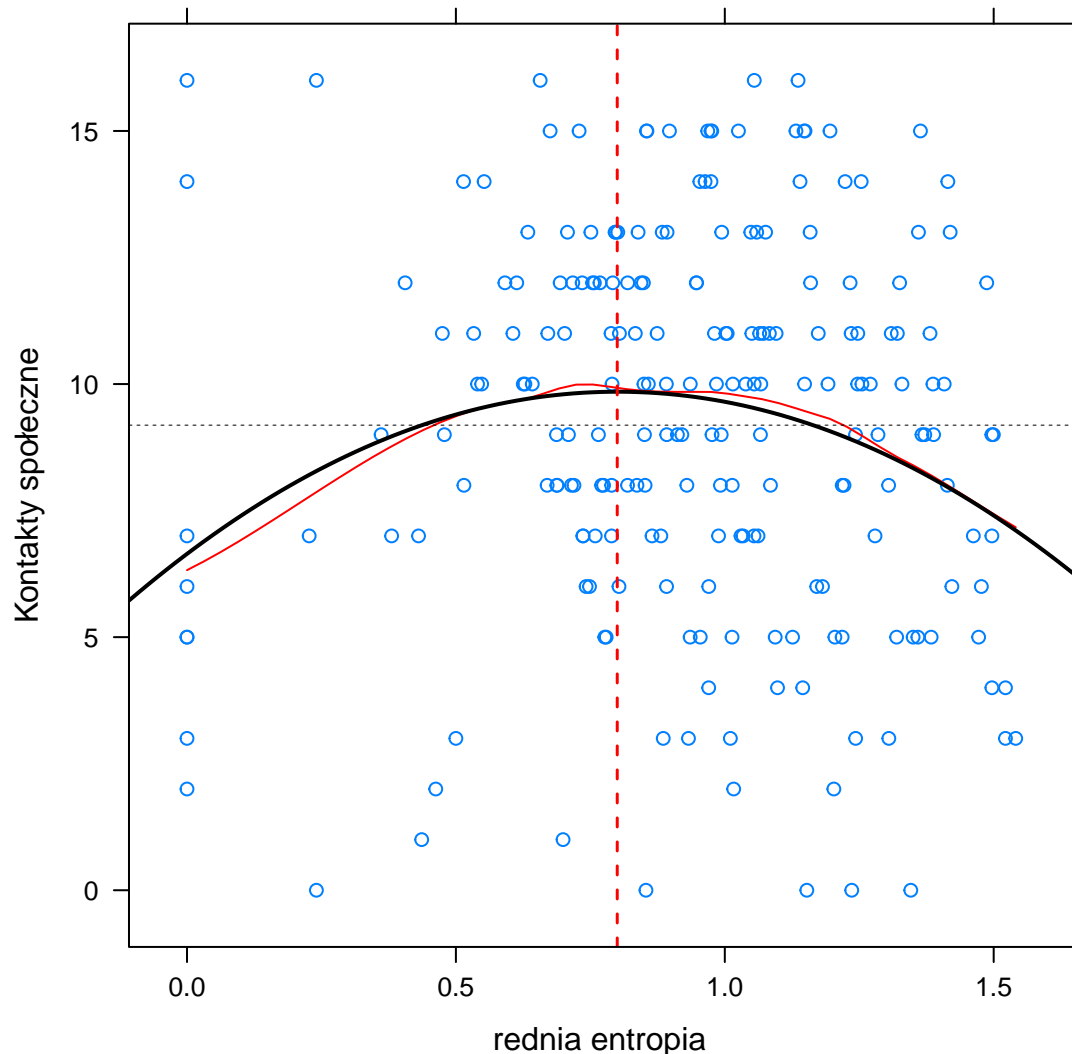
Po uwzględnieniu czynnika kwadratowego okazuje się, że zarówno on jak i czynnik liniowy jest istotnie związany z kontaktami społecznymi. Model całościowo zachowuje 4,8% wariancji. Czynnik liniowy jest istotny na poziomie:  $t_{215} = 3,019$ ;  $p = 0,003$ , a czynnik kwadratowy na poziomie:  $t_{215} = -3,266$ ;  $p = 0,001$ . Zatem model jest opisany równaniem:

$$Y = 6,645 + 8,004H - 4,999H^2 + \epsilon$$

Jego pochodna to:

$$\frac{dY}{dH} = 8,004 - 9,998H$$

Jako, że czynnik kwadratowy ma ujemny współczynnik, to ramiona paraboli wykresu równania opadają ku dołowi. Oznacza to, że funkcja ma globalne maksimum. Przekształcając pochodną w taki sam sposób jak poprzednio, można sprawdzić, że maksimum występuje dla punktu  $H = 0,800$ . Poniższy wykres przedstawia graficznie otrzymane wyniki:



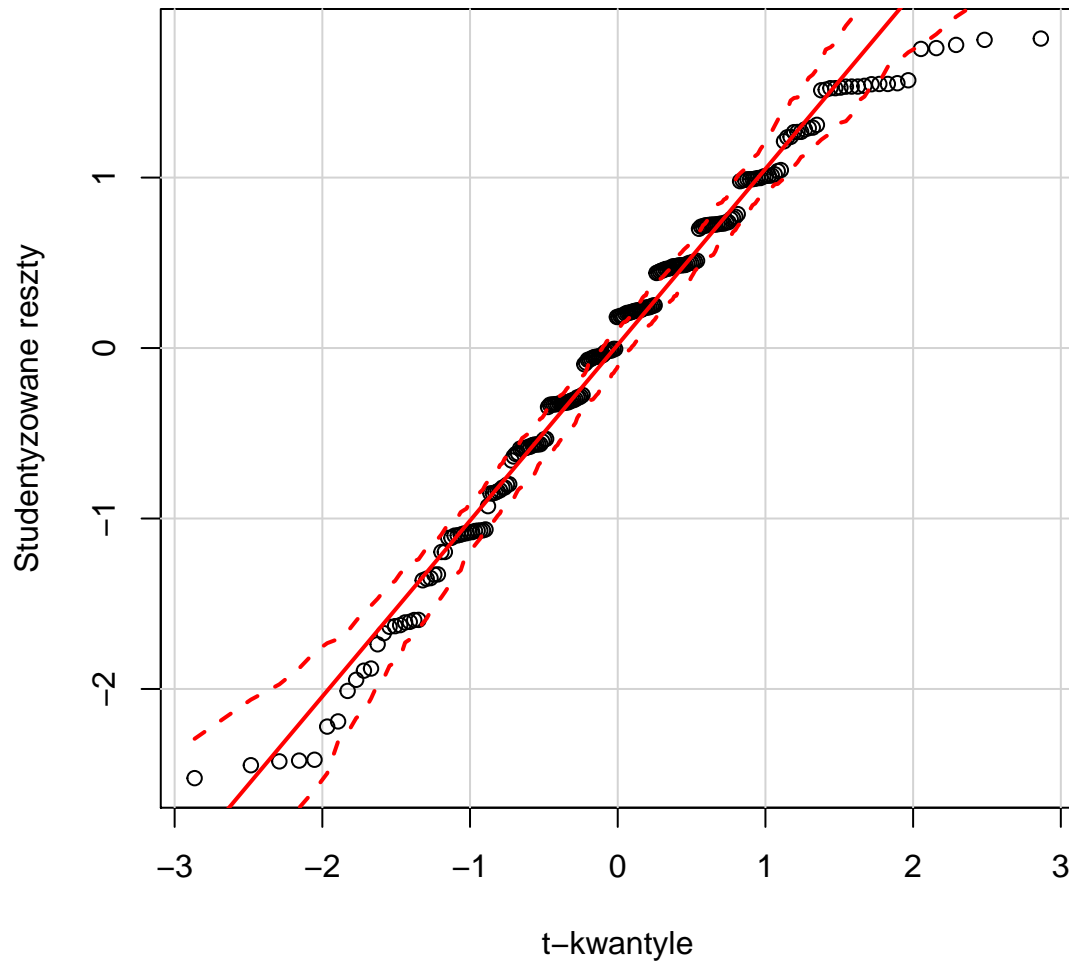
Podobnie jak poprzednio wykres funkcji równania regresji pokrywa się niemalże idealnie z krzywą loess. Co więcej w tym przypadku zakres danych pozwala się ujawnić kwadratowej zależności w większym odrobinę

większym stopniu. Wynik jest zatem podobny - istnieje optymalny poziom zróżnicowania środowiska społecznego, który jest optymalny pod względem kapitału społecznego rozumianego jako zbiór kontaktów i znajomości.

### Diagnostyka

Normalność reszt:

```
qqPlot(soccont.lm1, ylab="Studentyzowane reszty", xlab="t-kwantyle")
```



```
shapiro.test(soccont.lm1$residuals)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  soccont.lm1$residuals  
## W = 0.977, p-value = 0.001268
```

Stwierdzono odchylenie rozkładu od normalności, ale wykres kwantylowy sugeruje, że nie jest to odchylenie znaczne.

Heteroskedastyczność:



```
bptest(soccont.lm1)
```

```
##  
## studentized Breusch-Pagan test  
##  
## data: soccont.lm1  
## BP = 1.2808, df = 1, p-value = 0.2578
```

Test Breuscha-Pagana wskazuje, że wariancja reszt jest stała.

Test obserwacji odstających:

```
out.test = outlierTest(soccont.lm1)  
p = out.test$p  
p = p.adjust(p[p<=.05], n = length(p[p<=.05]))  
p <= .05
```

```
## 58  
## TRUE
```

Jedna silnie odstająca obserwacja (id 58). Estymacja modelu bez jej udziału.

```
soccont.lm2 = lm(soccont ~ ent_avg + I(ent_avg^2), data=D, subset=-which(D$id==58))  
summary(soccont.lm2)
```

```
##  
## Call:  
## lm(formula = soccont ~ ent_avg + I(ent_avg^2), data = D, subset = -which(D$id ==  
## 58))  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -9.8384 -2.5823  0.3144  2.3257  9.3582   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)      6.642      1.146   5.797  2.4e-08 ***  
## ent_avg          8.008      2.657   3.014  0.00289 **   
## I(ent_avg^2)    -4.995      1.534  -3.256  0.00131 **   
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 3.742 on 214 degrees of freedom  
## Multiple R-squared:  0.04783,    Adjusted R-squared:  0.03893   
## F-statistic: 5.375 on 2 and 214 DF,  p-value: 0.005279
```

```
anova(soccont.lm2)
```

```
## Analysis of Variance Table  
##  
## Response: soccont
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## ent_avg      1    2.04    2.039   0.1456 0.703199
## I(ent_avg^2)  1  148.52 148.518  10.6038 0.001312 **
## Residuals    214 2997.31   14.006
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Usunięcie obserwacji nie wpłynęło w żaden znaczący sposób na model. Zatem można się nią nie przejmować.

## Modele interakcyjne

Teraz zostaną przedstawione testy modeli zakładających interakcję entropii z przywiązaniem do miejsca - ogólnym bądź aktywnym ('Miejsce Odkryte'). Tak samo jak ostatnim razem i teraz przywiązanie do miejsca zostanie wycentrowanie w celu ułatwienia interpretacji i ustabilizowania błędów standardowych.

Interakcja z przywiązaniem ogólnym:

```
soccont.gen = lm(soccont ~ (ent_avg + I(ent_avg^2)) * I(center(attgen)), data=D)
summary(soccont.gen)
```

```
##
## Call:
## lm(formula = soccont ~ (ent_avg + I(ent_avg^2)) * I(center(attgen)),
##     data = D)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.4854 -2.5099  0.3409  2.3912  8.6175
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.4092     1.8334   2.950  0.00353 **
## ent_avg          10.3670     3.9835   2.602  0.00991 **
## I(ent_avg^2)     -6.0953     2.0986  -2.904  0.00407 **
## I(center(attgen)) -0.1255     0.1419  -0.884  0.37743
## ent_avg:I(center(attgen))  0.4326     0.3216   1.345  0.18000
## I(ent_avg^2):I(center(attgen)) -0.2057     0.1822  -1.129  0.26021
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.706 on 212 degrees of freedom
## Multiple R-squared:  0.07558,    Adjusted R-squared:  0.05378
## F-statistic: 3.467 on 5 and 212 DF,  p-value: 0.004942
```

```
anova(soccont.gen)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: soccont
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## ent_avg      1    2.23    2.229   0.1623 0.687457
## I(ent_avg^2)  1  148.78 148.784  10.8345 0.001167 **
## I(center(attgen))  1   58.55  58.548   4.2635 0.040157 *
## ent_avg:I(center(attgen))  1   10.96  10.956   0.7978 0.372758
```

```
## I(ent_avg^2):I(center(attgen))    1    17.50    17.501    1.2744    0.260210
## Residuals                        212    2911.27    13.732
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testy brzegowe współczynników modelu wykazały, że jedynie entropia (oba czynniki) jest istotna. Testy sekwencyjne kolejnych modeli wskazały na to samo. Należy zatem przyjąć, że między różnorodnością otoczenia społecznego i ogólnym wymiarem przywiązania do miejsca nie zachodzi interakcja.

Interakcja z przywiązaniem aktywnym ('Miejsce Odkryte')

```
soccont.disc = lm(soccont ~ (I(center(ent_avg)) + I(center(ent_avg^2))) * I(center(attdiscovered)), data = D)
summary(soccont.disc)
```

```
##
## Call:
## lm(formula = soccont ~ (I(center(ent_avg)) + I(center(ent_avg^2))) *
##     I(center(attdiscovered)), data = D)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.4199 -2.1714  0.1944  2.3255  9.1092
##
## Coefficients:
##                                Estimate Std. Error t value
## (Intercept)                   9.20626    0.24788   37.140
## I(center(ent_avg))             2.29690    3.60299    0.637
## I(center(ent_avg^2))          -2.08262    1.90937   -1.091
## I(center(attdiscovered))       1.86490    0.32202    5.791
## I(center(ent_avg)):I(center(attdiscovered)) -0.42823    3.09553   -0.138
## I(center(ent_avg^2)):I(center(attdiscovered)) -0.01188    1.82194   -0.007
##                                Pr(>|t|)
## (Intercept)                   < 2e-16 ***
## I(center(ent_avg))             0.524
## I(center(ent_avg^2))           0.277
## I(center(attdiscovered))       2.5e-08 ***
## I(center(ent_avg)):I(center(attdiscovered))  0.890
## I(center(ent_avg^2)):I(center(attdiscovered)) 0.995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.486 on 212 degrees of freedom
## Multiple R-squared:  0.1818, Adjusted R-squared:  0.1625
## F-statistic: 9.419 on 5 and 212 DF,  p-value: 4.011e-08
```

```
anova(soccont.disc)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: soccont
```

```
##                                Df    Sum Sq Mean Sq F value
## I(center(ent_avg))             1      2.23    2.23  0.1834
## I(center(ent_avg^2))           1    148.78   148.78 12.2406
```

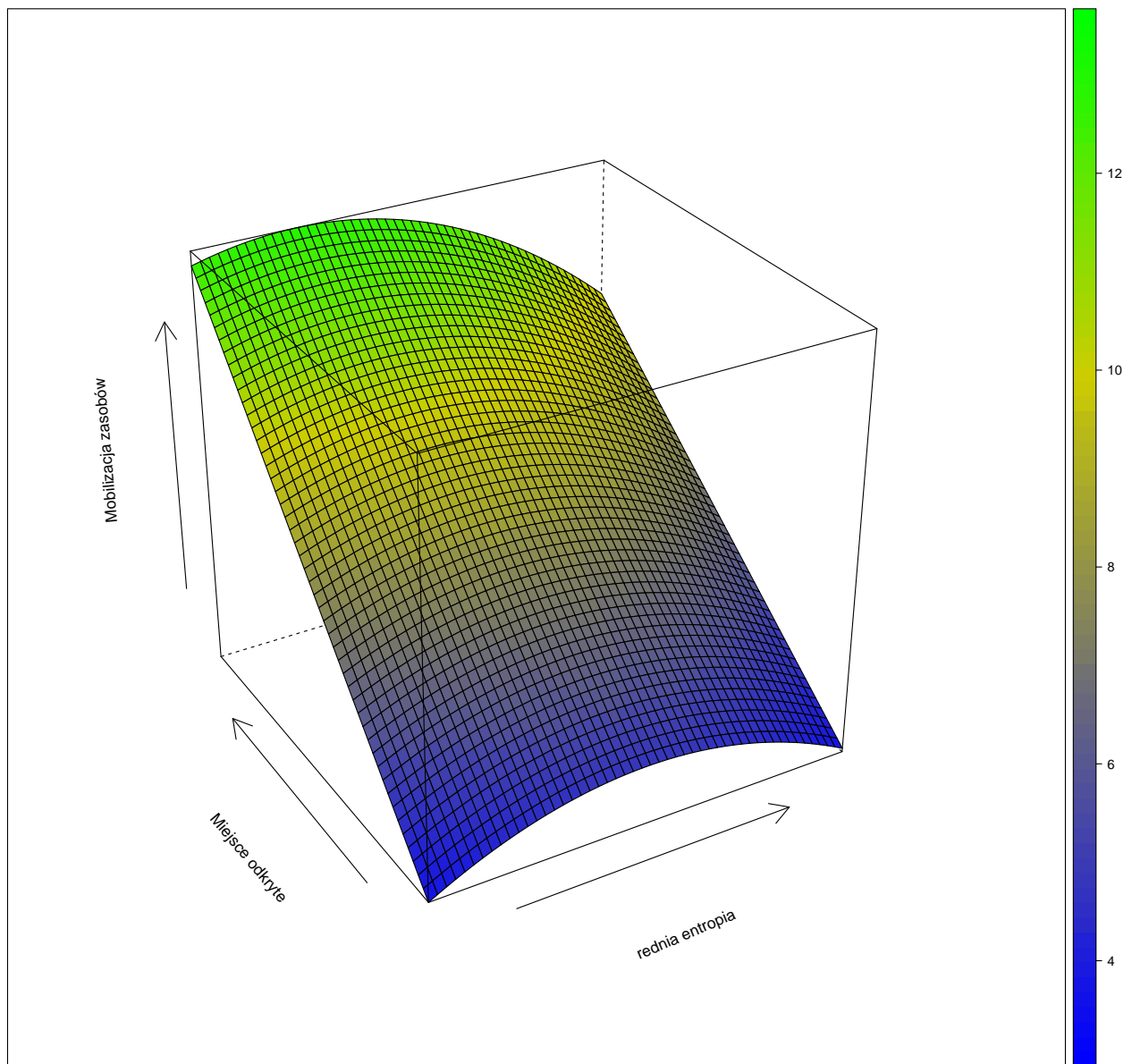
```

## I(center(attdiscovered))          1  418.61  418.61 34.4394
## I(center(ent_avg)):I(center(attdiscovered))  1    2.82    2.82 0.2319
## I(center(ent_avg^2)):I(center(attdiscovered))  1    0.00    0.00 0.0000
## Residuals                212 2576.85   12.15
##                               Pr(>F)
## I(center(ent_avg))          0.6689370
## I(center(ent_avg^2))        0.0005697 ***
## I(center(attdiscovered))    1.676e-08 ***
## I(center(ent_avg)):I(center(attdiscovered))  0.6305899
## I(center(ent_avg^2)):I(center(attdiscovered)) 0.9948020
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Stwierdzono, że model wyjaśnia aż 18,2% wariancji. Jest to oczywiście istotny wzrost względem 4,8% dla pierwszego modelu. Niemniej jednak żadne efekty interakcyjne nie są istotne. Testy brzegowe dla współczynników wykazały, że jedynie przywiązanie jest istotnym dodatkiem względem modelu bez swojego udziału,  $t_{212} = 5,791$ ;  $p < 0,001$ . Testy sekwencyjne kolejnych modeli pokazały, że model kwadratowy wyjaśnienie istotnie więcej niż sama średnia (co już wiadomo), oraz że dodanie addytywnego efektu przywiązania również istotnie (i to bardzo) wzmacnia moc predykcyjną modelu,  $F_{1,212} = 34,439$ ;  $p < 0,001$ . Efekty interakcji okazały się nieistotne.

Warto jednak zwizualizować również ten model, żeby zobaczyć, jak różni się zależność między przywiązaniem a entropią w przypadku braku interakcji:



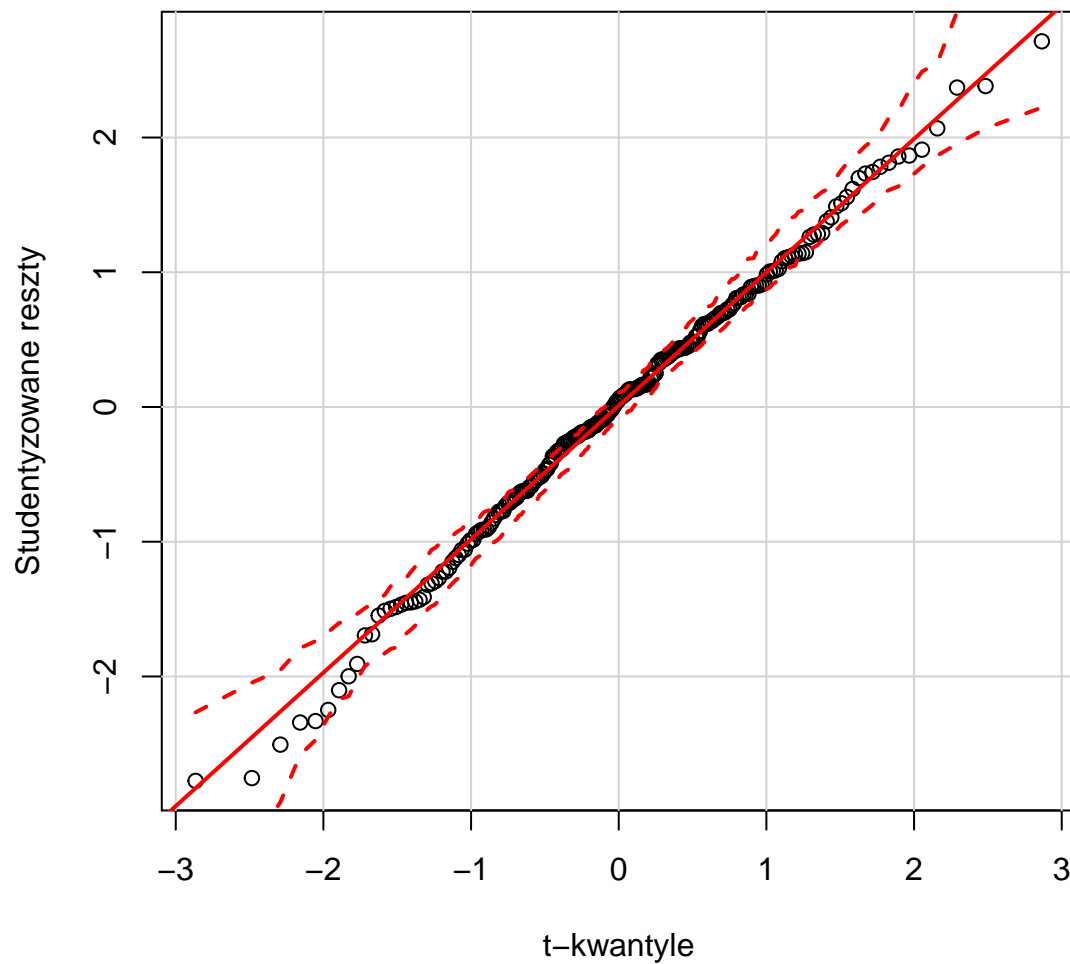
Jak widać w tym przypadku płaszczyzna jest prosta i wyższe przywiązanie za każdym razem wiąże się z większą ilością kontaktów społecznych. Lekka łukowata krzywizna płaszczyzny wzdłuż osi entropii jest oczywiście związana z kwadratowym charakterem jej związku z kontaktami społecznymi.

Na koniec należy przeprowadzić diagnostykę modelu z interakcją.

### Diagnostyka

Normalność reszt:

```
qqPlot(soccont.disc, ylab="Studentyzowane reszty", xlab="t-kwantyle")
```



```
shapiro.test(soccont.disc$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  soccont.disc$residuals
## W = 0.9953, p-value = 0.7483
```

Jak widać w tym przypadku wprowadzenie do modelu przywiązania do miejsca oraz jego efektów interakcji doprowadziło do pełnej zgodności rozkładu reszt z krzywą normalną.

Heteroskedastyczność:

```
bptest(soccont.disc)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  soccont.disc
## BP = 7.3352, df = 5, p-value = 0.1969
```

Zachowana jest również homoskedastyczność wariancji reszt.

Test obserwacji odstających:

```
out.test = outlierTest(soccont.disc, cutoff=Inf, n.max=Inf)
p = out.test$p
# Skorygowane wartości p dla obserwacji, które są istotnie odstające przy testach nieskorygowanych (pop
p = p.adjust(p[p<=.05], n = length(p[p<=.05]))
p <= .05
```

```
##      57      100      185      204      214      68      159      58      41      178      121      83
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

Brak silnie odstających obserwacji.

Statystyka VIF. Tak jak poprzednio VIF zostanie obliczony dla modelu z usunięciem czynnika kwadratowego, ponieważ jest on funkcją czynnika liniowego i w oczywisty sposób będzie zawyżał wyniki.

```
trunc.disc = lm(soccont ~ I(center(ent_avg))*I(center(attdiscovered)), data=D)
vif(trunc.disc)
```

```
##                                I(center(ent_avg))
##                                1.225086
##                                I(center(attdiscovered))
##                                1.074217
## I(center(ent_avg)):I(center(attdiscovered))
##                                1.279234
```

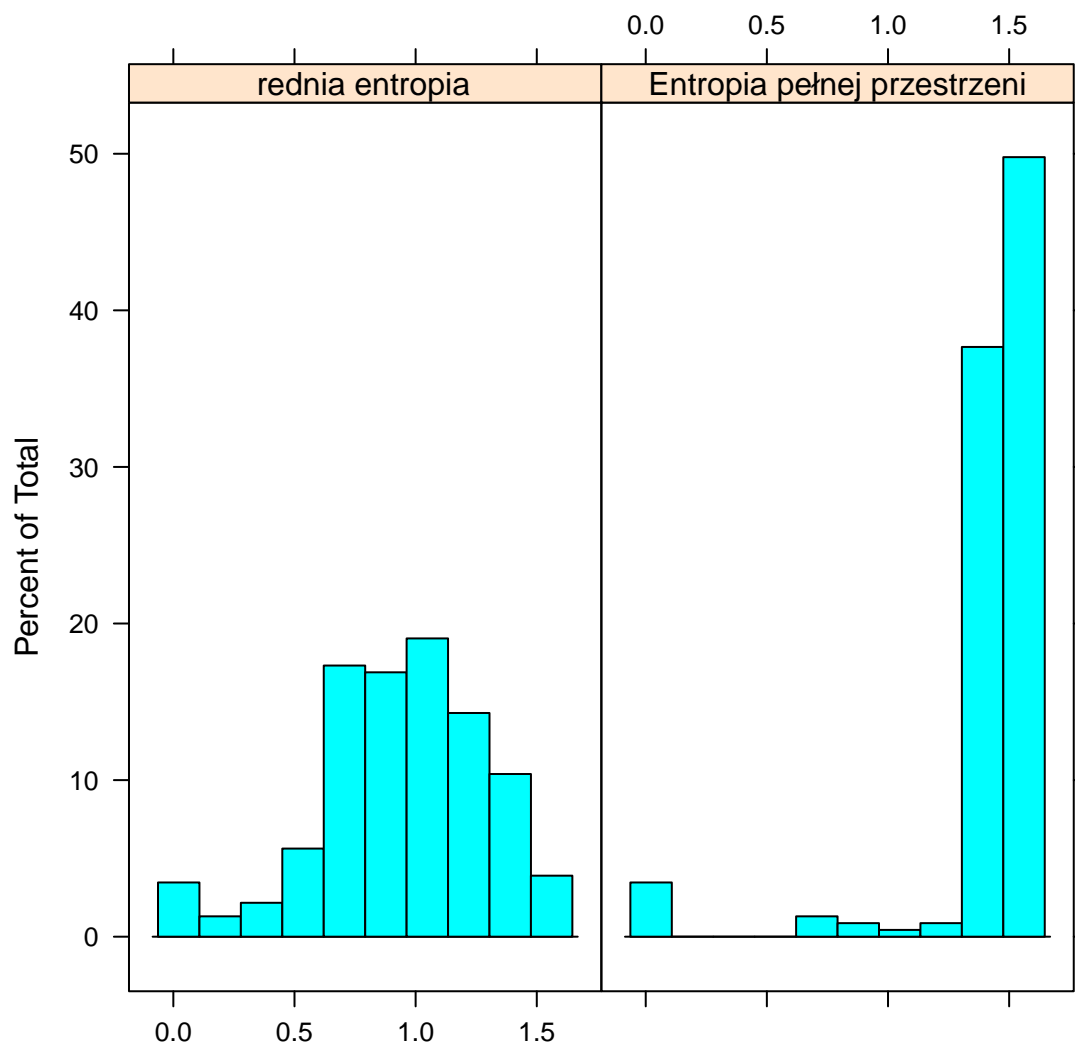
Wszystkie wartości VIF są bliskie 1, co oznacza, że problem silnych współliniowości praktycznie nie występuje.

### Entropia pełnej przestrzeni społecznej

Na koniec należy rozważyć jeszcze jedną kwestię. Średnia entropia odnosi się do tego jak przeciętnie zróżnicowane są miejsca odwiedzane przez daną osobę. W związku z tym odnosi się ona do średniego zróżnicowanie różnych podprzestrzeni (pól?) społecznych, z którymi osoba jest związana. W związku z tym nie jest ona bezpośrednio i jednoznacznie związana z różnorodnością całej przestrzeni społecznej jako takiej - którą można w tym kontekście rozumieć jako entropie rozkładu wszystkich osób, które chodzą do tych samych miejsc, co dany respondent. W związku z tym przeprowadzone analizy mówią o tym, jak przeciętne zróżnicowanie różnych podprzestrzeni przekłada się na kapitał społeczny. Dlatego warto przyjrzeć się temu, czy ogólne zróżnicowanie przestrzeni również może mieć na to jakiś wpływ.

Na początek warto przyjrzeć się rozkładowy entropii pełnej przestrzeni:

```
## No id variables; using all as measure variables
```

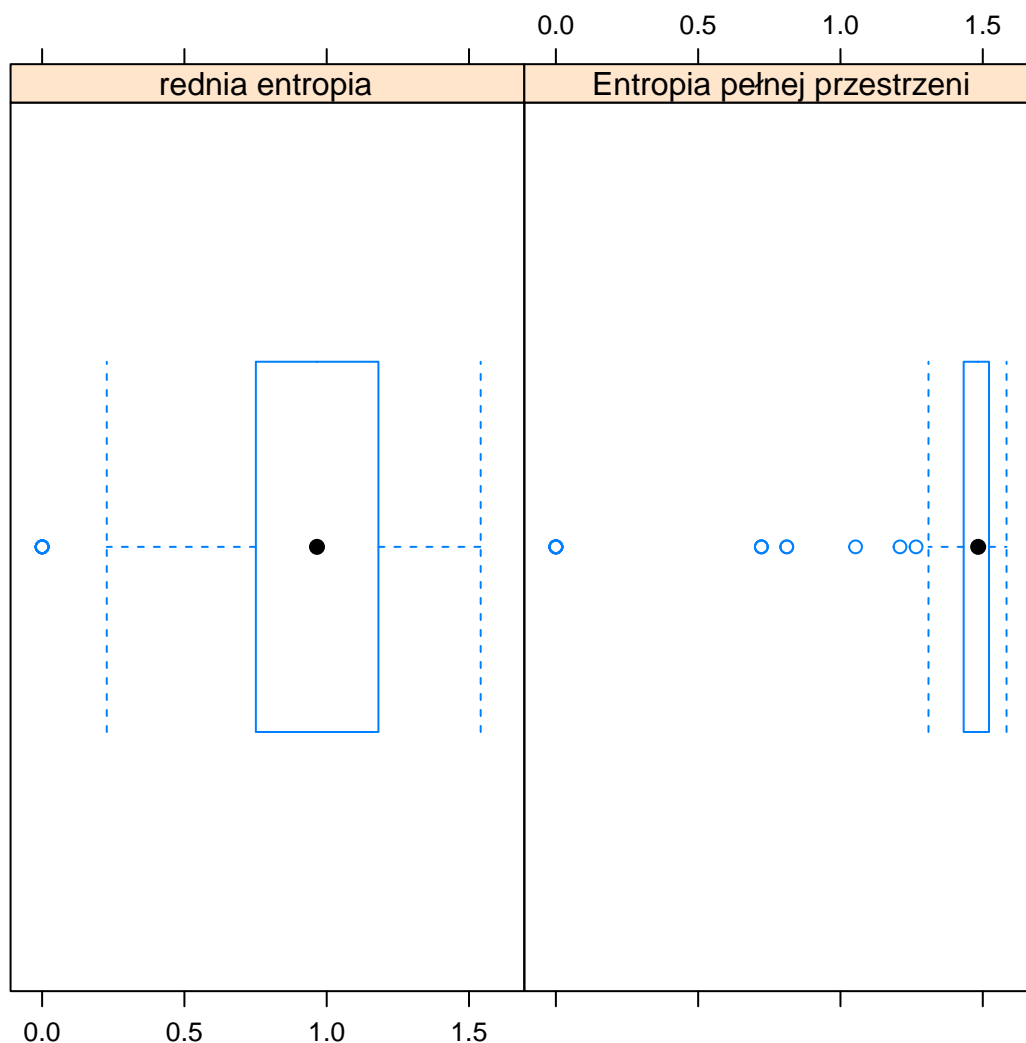


Widać, że właściwie wszystkie osoby charakteryzują się zbliżonym rozkładem entropii pełnej przestrzeni (i jest ona dosyć duża). Obok nich występują nieliczne obserwacje odstające o bardzo niskim zróżnicowaniu. Średnia entropia ma za to dosyć symetryczny i ładny rozkład. Sugeruje to, że entropia pełnej przestrzeni jest kiepską zmienną, która właściwie nie różnicuje pomiędzy osobami badanymi, a jedyne niskie wartości to obserwacje odstające, które w przeważającej większości są zapewne rezultatem małej motywacji przy wypełnianiu kwestionariusza badawczego i wskazania bardzo małej ilości miejsc.

Aby lepiej przedstawić tę zależność między dwoma rodzajami entropii warto przedstawić oba rozkłady przy użyciu wykresów pudełkowych:

```
bwplot(~ value | variable, data=data, horizontal=TRUE, xlab="")
```





Jak widać entropia pełnej przestrzeni jest praktycznie pozbawiona zmienności. Jest to o tyle ciekawe, że pokazuje, iż albo estymacja całkowitej różnorodności przestrzeni społecznej na podstawie miejsc jest nieefektywna, albo że pełna entropia jest ogólnie rzecz biorąc taka sama w społeczeństwie, a to, co się liczy to różnorodność poszczególnych podprzestrzeni czy pól.

**Koniec raportu**