

**Uniwersytet Warszawski**  
Wydział Matematyki, Informatyki i Mechaniki

**Paweł Ambor**

Nr albumu: 311509

**Modelowanie ścieżek przy  
zastosowaniu metody częściowych  
najmniejszych kwadratów**

Praca licencjacka  
na kierunku MATEMATYKA  
w zakresie JEDNOCZESNYCH STUDIÓW  
EKONOMICZNO-MATEMATYCZNYCH

Praca wykonana pod kierunkiem  
**dra inż. Przemysława Biecka**  
Instytut Matematyki Stosowanej i Mechaniki

Wrzesień 2013

## **Oświadczenie kierującego pracą**

Potwierdzam, że niniejsza praca została przygotowana pod moim kierunkiem i kwalifikuje się do przedstawienia jej w postępowaniu o nadanie tytułu zawodowego.

Data

Podpis kierującego pracą

## **Oświadczenie autora (autorów) pracy**

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że przedstawiona praca nie była wcześniej przedmiotem procedur związanych z uzyskaniem tytułu zawodowego w wyższej uczelni.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Data

Podpis autora (autorów) pracy

## Streszczenie

W niniejszej pracy przedstawiono najważniejsze informacje na temat modelowania ścieżek przy pomocy metody częściowych najmniejszych kwadratów. Pierwszy rozdział zawiera podstawowe informacje teoretyczne na temat tego typu modelowania wraz z omówieniem metod porównywania grup. Następnie przedstawiony został opis pakietu `plspm` w środowisku  $\mathcal{R}$  służący do estymacji współczynników tego typu modeli. Znaczna część pracy została poświęcona analizie danych rzeczywistych polegającej na estymacji poziomu szczęścia Polaków w 2011 roku oraz porównaniu jego poziomu u kobiet i u mężczyzn.

## Słowa kluczowe

modelowanie ścieżek, metoda częściowych najmniejszych kwadratów,  $\mathcal{R}$ , funkcja `plspm()`, funkcja `plspm.groups()`, szczęście

## Dziedzina pracy (kody wg programu Socrates-Erasmus)

11.2 Statystyka

## Klasyfikacja tematyczna

62J99, 62P25

## Tytuł pracy w języku angielskim

Partial least squares path modelling



# Spis treści

0.1. Oznaczenia . . . . .	9
<b>1. Wstęp teoretyczny . . . . .</b>	<b>11</b>
1.1. Informacje wstępne . . . . .	11
1.1.1. Rodzaje zmiennych . . . . .	11
1.1.2. Oznaczenia . . . . .	12
1.1.3. Model wewnętrzny . . . . .	13
1.1.4. Model zewnętrzny . . . . .	14
1.1.5. Algorytm estymacji współczynników . . . . .	14
1.2. Analiza porównawcza grup . . . . .	16
1.2.1. Bootstrap test $t$ . . . . .	17
1.2.2. Procedura permutacji . . . . .	18
<b>2. Opis funkcji w pakiecie <math>\mathcal{R}</math> . . . . .</b>	<b>19</b>
2.1. Funkcja <code>plspm()</code> . . . . .	19
2.2. Funkcja <code>plspm.groups()</code> . . . . .	20
2.3. Funkcje graficzne . . . . .	21
<b>3. Analiza danych rzeczywistych . . . . .</b>	<b>23</b>
3.1. Wstęp . . . . .	23
3.2. Struktura modelu . . . . .	23
3.2.1. Model wewnętrzny . . . . .	24
3.2.2. Opis bloku pierwszego . . . . .	26
3.2.3. Opis bloku drugiego . . . . .	26
3.2.4. Opis bloku trzeciego . . . . .	26
3.2.5. Opis bloku czwartego . . . . .	26
3.2.6. Opis bloku piątego . . . . .	27
3.2.7. Opis bloku szóstego . . . . .	27
3.3. Wstępna analiza danych . . . . .	29
3.4. Estymacja parametrów modelu . . . . .	30
3.5. Diagnostyka modelu . . . . .	31
3.6. Prezentacja wyników . . . . .	34
3.7. Analiza grup . . . . .	36
3.8. Podsumowanie . . . . .	40
<b>4. Zakończenie . . . . .</b>	<b>41</b>
<b>A. Pytania oraz wyniki ankiety . . . . .</b>	<b>43</b>
<b>Bibliografia . . . . .</b>	<b>49</b>



# Spis rysunków

1.1. Przykładowy pełny model PLS-PM . . . . .	12
1.2. Przykładowy model wewnętrzny PLS-PM . . . . .	13
1.3. Przykładowe modele wewnętrzne PLS-PM . . . . .	15
3.1. Model wewnętrzny . . . . .	24
3.2. Piramida Maslow'a . . . . .	28
3.3. Modele zewnętrzne . . . . .	28
3.4. Model wewnętrzny wraz z wyestymowanymi współczynnikami . . . . .	35
3.5. Modele zewnętrzne wraz z wyestymowanymi współczynnikami . . . . .	37





# Spis tabel

2.1. Opis argumentów funkcji <code>plspm()</code> . . . . .	19
2.2. Opis funkcji służących do odczytu wyników z obiektu klasy „ <code>plspm</code> ”. . . . .	20
2.3. Opis argumentów funkcji <code>plspm.groups()</code> . . . . .	20
2.4. Opis funkcji służących do odczytu wyników z obiektu klasy „ <code>plspm.groups</code> ”. . . . .	20
A.1. Opis zmiennych mierzalnych z bloku 2. . . . .	43
A.2. Wyniki ankiety dotyczące bloku 2. Zmienne jakościowe. . . . .	43
A.3. Wyniki ankiety dotyczące bloku 2. Zmienna ilościowa. . . . .	43
A.4. Opis zmiennych mierzalnych z bloku 1. . . . .	44
A.5. Wyniki ankiety dotyczące bloku 1. . . . .	44
A.6. Opis zmiennych mierzalnych z bloku 3. . . . .	45
A.7. Wyniki ankiety dotyczące bloku 3. . . . .	45
A.8. Opis zmiennych mierzalnych z bloku 4. . . . .	46
A.9. Wyniki ankiety dotyczące bloku 4. Zmienna jakościowa. . . . .	46
A.10. Wyniki ankiety dotyczące bloku 4. Zmienna ilościowa. . . . .	46
A.11. Opis zmiennych mierzalnych z bloku 5. . . . .	47
A.12. Wyniki ankiety dotyczące bloku 5. Zmienne jakościowe. . . . .	47
A.13. Wyniki ankiety dotyczące bloku 5. Zmienne ilościowa. . . . .	47
A.14. Opis zmiennych mierzalnych z bloku 6. . . . .	48
A.15. Wyniki ankiety dotyczące bloku 6. . . . .	48



# Wprowadzenie

Modelowanie ścieżek jest to rozszerzenie regresji o dodatkowe zależności pomiędzy zmiennymi objaśniającymi. Polega ono na dopasowaniu wielu powiązanych ze sobą równań regresji. Takie podejście staje się coraz bardziej popularne w naukach społecznych oraz szeroko pojętej działalności biznesowej i marketingowej.

Celem pracy jest omówienie metody modelowania ścieżek za pomocą częściowych najmniejszych kwadratów (ang. *partial least squares path modelling*– PLS-PM) oraz zaprezentowanie analizy danych rzeczywistych w środowisku  $\mathcal{R}$ .

Praca podzielona jest na trzy rozdziały. W pierwszej części przedstawione są podstawowe informacje teoretyczne na temat modeli PLS-PM oraz porównywania grup. Drugi rozdział zawiera opis funkcji, stosowanych do estymacji ich współczynników w środowisku  $\mathcal{R}$ . Przede wszystkim zaprezentowano funkcję `plspm()` oraz sposoby porównywania grup opisane w części pierwszej. Ostatni rozdział przedstawia analizę danych rzeczywistych wykonaną przy użyciu technik opisanych w rozdziale drugim. Wykorzystano dane zebrane podczas badania „Diagnoza społeczna” w celu uzyskania odpowiedzi na pytanie, czy poziom szczęścia mężczyzn i kobiet jest znacząco różny w polskim społeczeństwie.

Składu całej pracy dokonano w systemie  $\text{\LaTeX}$ .

## 0.1. Oznaczenia

- Równania matematyczne zapisane są na środku kursywą–

$$f(x_1, x_2) = 0.$$

- Nazwy zmiennych zapisane są również kursywą–  $(x_1, x_2)$ .
- Kod z programu  $\mathcal{R}$  został zapisany czcionką o stałej szerokości i otoczony ramką–

```
> f(x1,x2)
0
```

- Nazwy pakietów środowiska  $\mathcal{R}$  zapisane są czcionką o stałej szerokości– `plspm`.
- Nazwy funkcji środowiska  $\mathcal{R}$  również są zapisane czcionką o stałej szerokości, jednak na końcu nazwy zawsze znajduje się nawias okrągły– `plspm()`.
- Nazwy angielskie umieszczone są w nawiasach i zapisane kursywą po skrócie „ang.”– zmienne ukryte (ang. *latent variables*).



# Rozdział 1

## Wstęp teoretyczny

W niniejszym rozdziale zamieszczono podstawowe informacje na temat modelowania ścieżek przy pomocy metody częściowych najmniejszych kwadratów (ang. *partial least squares path modelling*), w skrócie PLS-PM. Ta część została podzielona na dwie składowe. Pierwsza stanowi opis podstawowych pojęć oraz algorytm estymacji współczynników modelu. Druga prezentuje dwie metody prowadzenia analizy porównawczej grup.

### 1.1. Informacje wstępne

#### 1.1.1. Rodzaje zmiennych

PLS-PM bardzo często wykorzystywany jest do oszacowania różnego rodzaju indeksów, które odnoszą się do pojęć abstrakcyjnych, czyli takich, których nie można bezpośrednio zmierzyć. Często spotykanymi przykładami tego typu są np. poziom satysfakcji, motywacji, sukcesu czy użyteczności. Przykładowo, mimo, że można zapytać respondenta jak duży sukces odniósł podczas edukacji w liceum, nie ma możliwości dokonania jednoznacznego, systematycznego i precyzyjnego pomiaru tej wartości. Niemniej jednak, można dokonać próby oszacowania tej wielkości poprzez zadanie respondentowi odpowiednich pytań, które odnoszą się do osiągnięć z liceum. Można uzyskać informację np. o średniej z ocen na świadectwie, o posiadanie tytułu laureata czy finalisty olimpiady przedmiotowej lub też przeanalizować oceny z matury. Te wskaźniki bardzo łatwo precyzyjnie wyrazić za pomocą liczb.

W modelach PLS-PM wyróżnia się dwa rodzaje zmiennych:

- ukryte (ang. *latent variables, constructs, composites, factors*),
- obserwowane (ang. *manifest variables, indicators, items*).

Zmienne ukryte to takie, których nie można (lub też wymaga to niezwykle dużego nakładu pracy) ani zaobserwować, ani zmierzyć bezpośrednio. Służą do zredukowania ilości zmiennych oraz pomagają tłumaczyć zależności między wybranymi zmiennymi widocznymi. Są one bardzo często spotykane w naukach społecznych (np. inteligencja, zaangażowanie, pewność siebie), ekonomii (np. użyteczność, innowacyjność, poziom rozwoju), ale także w naukach ścisłych (np. urodzaj gleby).

Zmienne obserwowane to takie, które można stosunkowo łatwo precyzyjnie wyrazić w wartościach liczbowych. Mogą mieć zarówno charakter ilościowy, jak i jakościowy. Zawierają informacje odnośnie zmiennych ukrytych.

Ze względu na kierunek przyczynowości wyróżniane są dwa typy zmiennych obserwowanych:

- kształtujące (ang. *formative indicators*) – formują zmienną ukrytą,

- odzwierciedlające (ang. *reflective indicators*)– są formowane przez zmienną ukrytą.

Dla przykładu, gdy rozpatrzy się zmienną ukrytą– urodzaj gleby, wtedy zmiennymi kształtującymi mogą być np. zawartość potasu, kwasowość, zaś zmiennymi odzwierciedlającymi np. wielkość plonu czy jego jakość.

### 1.1.2. Oznaczenia

Do końca tego rozdziału przyjęto następujące oznaczenia:

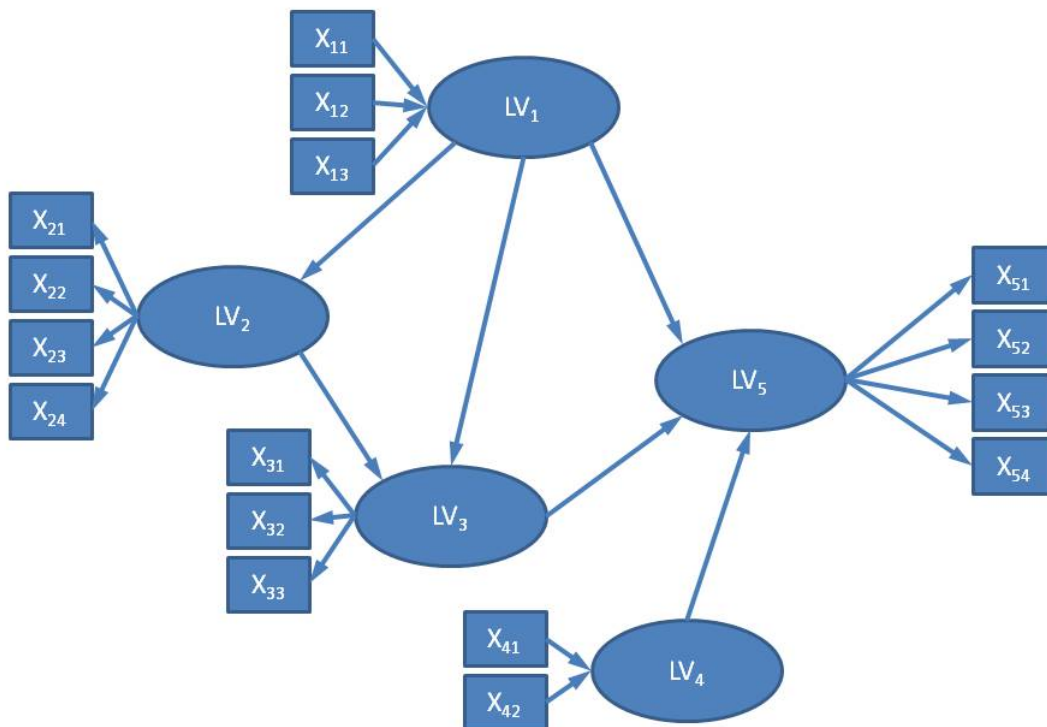
$X$ – blok danych zawierający  $n$  obserwacji i  $p$  zmiennych obserwowanych,  $X \in M_{n \times p}(\mathbb{R})$ ,

$X_j$ – blok danych odnoszący się do jednej zmiennej ukrytej,  $X_j$  są rozłączne i  $X = X_1, X_2, \dots, X_j$ ,

$X_{jk}$ –  $k$ -ta zmienna obserwowana odnosząca się do  $j$ -tej zmiennej ukrytej,

$LV_j$ – zmienna ukryta odnosząca się do bloku  $X_j$ ,

$\widehat{LV}_j$ – estymator  $LV_j$ .



Rysunek 1.1: Przykładowy pełny model PLS-PM [opracowanie własne].

### 1.1.3. Model wewnętrzny

Model wewnętrzny opisuje zależności pomiędzy zmiennymi ukrytymi. Łączy on ze sobą poszczególne bloki, tworząc usystematyzowaną strukturę. Konstrukcja modelu powinna być poparta teorią opisującą modelowane zjawisko. Z punktu widzenia PLS-PM spełnione muszą być następujące cztery założenia:

1. Zależność liniowa:

$$LV_j = \beta_{j0} + \sum_{i \rightarrow j} \beta_{ji} LV_i + \epsilon_j,$$

gdzie sumowanie odbywa się po wszystkich zmiennych ukrytych, które objaśniają  $LV_j$ , a  $\epsilon_j$  oznacza  $j$ -ty błąd losowy, czyli tą część zmienności  $LV_j$ , która nie da się wytłumaczyć za pomocą zmiennych objaśniających.

2. Warunek na predyktory:

$$\mathbb{E}(LV_j | LV_i) = \beta_0 + \sum_{i \rightarrow j} \beta_{ji} LV_i,$$

gdzie sumowanie odbywa się po wszystkich zmiennych ukrytych, które objaśniają  $LV_j$ .

3. Brak korelacji:

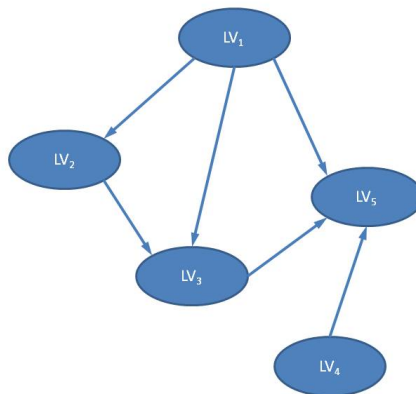
$$\text{cov}(LV_j, \epsilon_i) = \text{cov}(\epsilon_j, \epsilon_i) = 0,$$

gdzie  $i \neq j$ .

4. Graf skierowany ilustrujący model nie może zawierać cykli.

**Definicja 1** Graf skierowany to graf zawierający krawędzie skierowane, czyli takie, po których ruch może odbywać się tylko w jednym kierunku.

**Definicja 2** Cykl to ścieżka, w której powtarzającym się wierzchołkiem jest jej początek będący również jej końcem.



Rysunek 1.2: Przykładowy model wewnętrzny PLS-PM [opracowanie własne].

O modelu wewnętrznym można również myśleć jako o kilku powiązanych ze sobą regresjach liniowych. Ich liczba jest równa, mówiąc obrazowo, ilości zmiennych ukrytych, do których zostały poprowadzone strzałki określające kierunki zależności. W przykładzie naszkicowanym na rysunku 1.2 na stronie 13 można wskazać następujące trzy powiązane ze sobą regresje:  $LV_2$  na  $LV_1$ ,  $LV_3$  na  $LV_1$ ,  $LV_2$  oraz  $LV_5$  na  $LV_1$ ,  $LV_3$ ,  $LV_4$ .

#### 1.1.4. Model zewnętrzny

Model zewnętrzny odnosi się do zależności między pewną liczbą zmiennych obserwowanych i jedną zmienną ukrytą. Podobnie jak w przypadku struktury modelu wewnętrznego, przyporządkowanie zmiennych mierzalnych do poszczególnych zmiennych ukrytych powinno być poparte stosowną wiedzą z zakresu badanego zjawiska.

W związku z tym, że istnieją dwa typy zmiennych mierzalnych, czyli kształtujące i odzwierciedlające<sup>1</sup>, wyróżnia się również dwa typy modeli zewnętrznych. Pierwszy z nich to model zawierający zmienne odzwierciedlające daną zmienną ukrytą, można mówić również o bloku zmiennych odzwierciedlających (ang. *reflective block*). Drugi natomiast zawiera zmienne kształtujące daną zmienną ukrytą, czyli jest to blok zmiennych kształtujących (ang. *formative block*). Należy przyjąć następujące założenia dotyczące modeli wewnętrznych:

1. Zależność liniowa:

$$\begin{aligned} LV_j &= \lambda_{0j} + \lambda_{jk}X_{jk} + \epsilon_j && \text{dla modelu ze zmiennymi kształtującymi,} \\ X_{jk} &= \lambda_{0jk} + \lambda_{jk}LV_j + \epsilon_{jk} && \text{dla modelu ze zmiennymi odzwierciedlającymi,} \end{aligned}$$

gdzie  $\lambda_{jk}$  są współczynnikami modeli zewnętrznych.

2. Warunek na predyktory:

$$\begin{aligned} \mathbb{E}(LV_j|X_{jk}) &= \lambda_{0j} + \lambda_{jk}X_{jk} && \text{dla modelu ze zmiennymi kształtującymi,} \\ \mathbb{E}(X_{jk}|LV_j) &= \lambda_{0jk} + \lambda_{jk}LV_j && \text{dla modelu ze zmiennymi odzwierciedlającymi.} \end{aligned}$$

3. Warunek na estymatory:

$$\widehat{LV_j} = \sum_k w_{jk}X_{jk}$$

oznacza, że zmienne ukryte, których nie można bezpośrednio zmierzyć są estymowane jako kombinacje liniowe zmiennych obserwowanych przyporządkowanych do ich bloku. Dzięki temu można wyliczyć ich wartości, co jest niezbędne do wyliczenia współczynników modelu wewnętrznego.

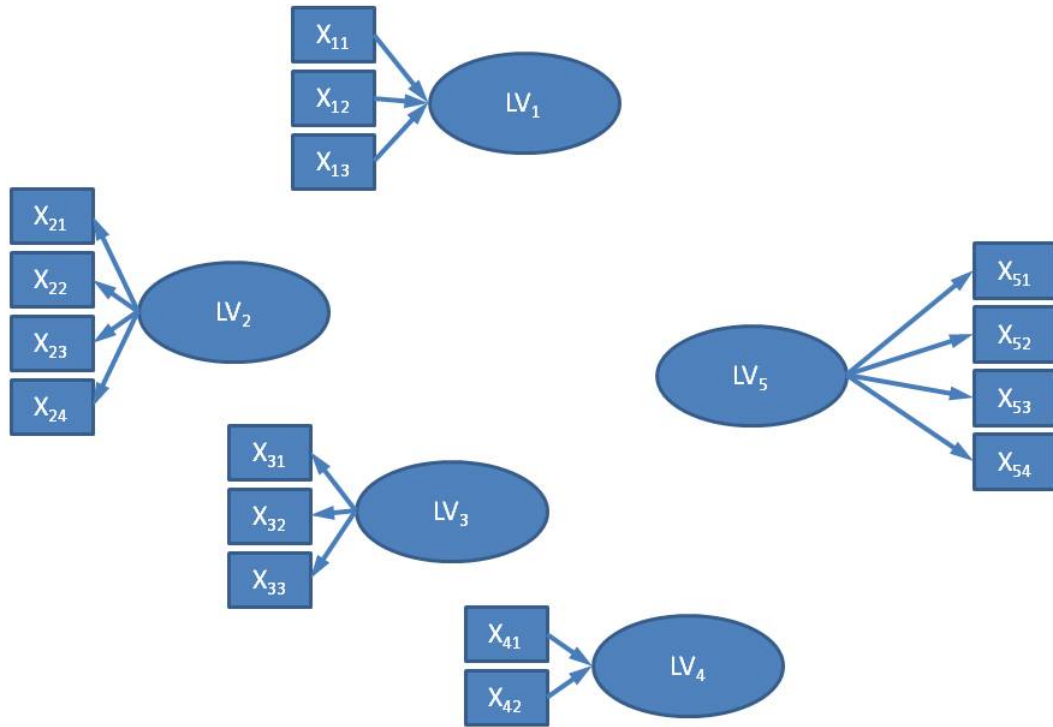
#### 1.1.5. Algorytm estymacji współczynników

Aby możliwa była estymacja parametrów modelu PLS-PM, należy oprócz obserwacji zmiennych mierzalnych posiadać również informacje odnośnie struktury modelu wewnętrznego, przyporządkowania konkretnych zmiennych widocznych do danego bloku oraz typów bloków. Najważniejszymi informacjami uzyskiwanymi w wyniku estymacji są wagi w modelach zewnętrznych, a co za tym idzie, wartości zmiennych ukrytych oraz zależności między nimi.

---

<sup>1</sup>obydwa typy opisano w podrozdziale „Rodzaje zmiennych”





Rysunek 1.3: Przykładowe modele wewnętrzne PLS-PM. Blok pierwszy oraz czwarty zawiera zmienne kształtujące, natomiast pozostałe bloki są typu odzwierciedlającego [opracowanie własne].

Opisany poniżej algorytm estymacji współczynników modelu PLS-PM został zaczerpnięty z podręcznika autorstwa Gaston'a Sanchez'a[4]. Procedura składa się z 3 etapów:

1. wyliczenie wag niezbędnych do przedstawienia zmiennych ukrytych jako kombinacji liniowych zmiennych obserwowanych,
2. wyliczenie współczynników  $\beta$  w modelu wewnętrznym,
3. wyliczenie współczynników  $\lambda$  w modelach zewnętrznych.

Etap pierwszy jest najbardziej złożony pod względem obliczeniowym. Jego zadaniem jest przedstawienie niemierzalnej, hipotetycznej wartości zmiennych ukrytych za pomocą liczb. Składa się on z następującej iteracyjnej procedury:

Krok 0. Należy wyznaczyć w sposób arbitralny wagi początkowe, które następnie będą odpowiednio modyfikowane. Można przyjąć, że wszystkie wagi są na początku równe 1, czyli  $w_{jk} = 1$ .

Krok 1. Z otrzymanych wag w poprzednim kroku należy wyliczyć wartości zmiennych ukrytych, korzystając z równania:

$$LV_j = \sum_k w_{jk} X_{jk}.$$

Krok 2. W tym momencie przechodzi się do modelu wewnętrznego i wylicza współczynniki korelacji  $e_{ij}$  według wzoru:

$$e_{ij} = \begin{cases} \text{cor}(LV_j, LV_i) & \text{dla } LV_j \text{ i } LV_i \text{ sąsiadujących ze sobą,} \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Krok 3. Wylicza się nowe wartości zmiennych ukrytych na podstawie modelu wewnętrznego:

$$\widetilde{LV}_j = \sum_i e_{ij} LV_i.$$

Krok 4. Oblicza się nowe wartości wag, korzystając z typowego wzoru dla metody najmniejszych kwadratów:

$$\begin{aligned} \tilde{w}_{jk} &= \left( \widetilde{LV}_j^T \widetilde{LV}_j \right)^{-1} \widetilde{LV}_j^T X_{jk} & \text{dla modelu ze zmiennymi odzwierciedlającymi,} \\ \tilde{w}_j &= (X_j^T X_j)^{-1} X_j^T \widetilde{LV}_j & \text{dla modelu ze zmiennymi kształtującymi.} \end{aligned}$$

Krok 5. Po każdej iteracji należy porównać nowe wagi z wcześniej wyliczonymi. Gdy są wystarczająco bliskie sobie (można przyjąć, że dla każdego  $j$  oraz  $k$   $|\tilde{w}_{jk} - w_{jk}| < 10^{-5}$ ), należy przerwać algorytm, w przeciwnym razie trzeba powtórzyć kroki od 1 do 5, gdzie w kroku 1 korzystamy z nowych wag wyliczonych w poprzedniej iteracji.

Etap drugi polega na wyliczeniu estymatorów modelu wewnętrznego. W tym celu należy skorzystać ze wzoru:

$$\hat{\beta}_{ji} = \left( \widetilde{LV}_i^T \widetilde{LV}_i \right)^{-1} \widetilde{LV}_i^T \widetilde{LV}_j.$$

Etap trzeci polega na wyliczeniu współczynników modeli zewnętrznych przy wykorzystaniu, tak jak powyżej, klasycznej metody najmniejszych kwadratów.

Zastosowanie powyższego algorytmu estymacji współczynników modelu pozwoliło wyliczyć wartości zmiennych ukrytych, czyli osiągnąć główny cel stosowania PLS-PM oraz współczynniki zarówno modelu wewnętrznego jak i zewnętrznego. Procedura dotycząca wyliczania wag może wydawać się niezwykle uciążliwa do policzenia „na kartce”. Rozdział drugi został poświęcony opisowi oprogramowania umożliwiającego estymację modelu właśnie na podstawie powyższego algorytmu.

## 1.2. Analiza porównawcza grup

Bardzo często po wykonaniu modelu występuje potrzeba porównania go z innymi w celu uzyskania bardziej szczegółowych informacji.

Można wyróżnić cztery główne obszary, w których modele PLS-PM mogą się różnić:

- zależności przyczynowo-skutkowych pomiędzy zmiennymi ukrytymi,
- wielkości współczynników w modelu wewnętrznym,
- dobór zmiennych widocznych,
- charakterystyka rozkładów zmiennych ukrytych.

Aby systematyczne badanie różnic było możliwe, należy zdecydować się na porównywanie modeli różniących się w tylko jeden z wymienionych powyżej sposobów. Najczęściej dokonuje się analizy wielkości współczynników w modelu wewnętrznym, gdyż to właśnie badanie liniowych zależności między poszczególnymi zmiennymi jest głównym celem rozpatrywanej metody modelowania.

Analizę porównawczą najłatwiej jest prowadzić dla dwóch grup. Gdy występuje potrzeba rozpatrywania większej liczby, należy dokonać kilku porównań, gdzie w pierwszym porównaniu rozpatrujemy podział na pierwszą grupę oraz pozostałe, w drugim – na drugą i pozostałe itd.

W tej części zostaną omówione dwie metody repróbkiwania (ang. *resampling*) – bootstrap test t oraz procedura permutacji.

### 1.2.1. Bootstrap test t

To podejście polega na przydzieleniu obserwacji do poszczególnych grup, a następnie poddaniu każdej z nich procedurze bootstrapu. Do tak utworzonych próbek typu bootstrap należy zastosować test t-Studenta.

Załóżmy, że są dwie grupy:  $G_1$  i  $G_2$ , ich liczebności wynoszą odpowiednio  $n_1$  i  $n_2$ . Wartości badanego współczynnika dla tych grup to  $\beta^{G_1}$  i  $\beta^{G_2}$ . Aby sprawdzić, czy współczynniki są różne w sposób znaczący, należy postępować w następujący sposób:

1. wyestymować dwa modele w celu uzyskania wartości  $\beta^{G_1}$  i  $\beta^{G_2}$ ,
2. utworzyć próbę typu bootstrap z każdej z grup,
3. dla każdej próbki ponownie wyestymować modele,
4. po wszystkich repróbkiowaniach obliczyć odchylenia standardowe –  $SE_{G_1}$  i  $SE_{G_2}$ ,
5. obliczyć statystykę testową t.

Aby policzyć wartość statystyki testowej należy skorzystać ze wzoru:

$$t = \frac{\beta^{G_1} - \beta^{G_2}}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \left(\frac{(n_1-1)^2}{n_1+n_2-2} SE_{G_1}^2 + \frac{(n_2-1)^2}{n_1+n_2-2} SE_{G_2}^2\right)}},$$

gdzie  $t \sim t(n_1 + n_2 - 2)$ , czyli  $t$  ma rozkład t-Studenta o  $n_1 + n_2 - 2$  stopniach swobody przy założeniu, że hipoteza zerowa jest spełniona. Odchylenia standardowe  $SE_{G_1}$  i  $SE_{G_2}$  oblicza się ze wzorów:

$$SE_{G_1} = \sqrt{\frac{\sum_{i=1}^{n_1} (\beta_i^{G_1} - \bar{\beta}^{G_1})^2}{n_1}}, \quad SE_{G_2} = \sqrt{\frac{\sum_{i=1}^{n_2} (\beta_i^{G_2} - \bar{\beta}^{G_2})^2}{n_2}},$$

gdzie  $\beta_i^{G_1}$  oraz  $\beta_i^{G_2}$  oznaczają wartości współczynnika  $\beta$  odpowiednio w pierwszej i drugiej grupie przy i-tym repróbkiowaniu, natomiast  $\bar{\beta}^{G_1}$  oraz  $\bar{\beta}^{G_2}$  to średnie z odpowiednio  $\beta_i^{G_1}$  i  $\beta_i^{G_2}$ .

Gdy  $|t| < t_{1-\alpha/2}(n_1 + n_2 - 2)$ , wtedy nie ma podstaw do odrzucenia hipotezy mówiącej o równości współczynników na poziomie istotności  $\alpha$ . W przeciwnym wypadku należy odrzucić hipotezę zerową. Stosując test t-Studenta, nie wolno zapomnieć o założeniach, na których jest on oparty: rozkład normalny danych oraz zbliżona liczebność grup. Są to dość restrykcyjne założenia. Na przykład we wszelkiego rodzaju badaniach ankietowych respondenci mają tendencję do udzielania skrajnych odpowiedzi, co skutkuje rozkładem charakteryzującym się znaczną skośnością. Drugie założenie, mówiące o podobnej liczebności próbek w grupach, uniemożliwia również wiele porównań.

### 1.2.2. Procedura permutacji

Drugą metodą przeprowadzania porównania grup jest procedura permutacji. W odróżnieniu od metody bootstrap, w tym przypadku próbki pochodzą z obserwacji przyporządkowanych w sposób losowy do grup, a nie zastąpionych przez wylosowanie ze zwracaniem.

Procedura ta polega na zbadaniu rozkładu statystyki testowej, będącej różnicą między badanymi współczynnikami. Po obliczeniu odpowiedniego kwantyla rozkładu, należy porównać go z wartością statystyki testowej wyliczonej dla interesujących badacza grup.

Test permutacyjny opisany poniżej służy do zweryfikowania, czy różnica między współczynnikami jest na tyle duża, że należy odrzucić hipotezę mówiącą o identyczności współczynników.

Oznaczenia przyjęto takie same jak w opisie poprzedniej metody. Są dwie grupy:  $G_1$  i  $G_2$  o liczebnościach odpowiednio  $n_1$  i  $n_2$ , wartości współczynników w tych grupach to  $\beta^{G_1}$  i  $\beta^{G_2}$ . Szczegółowa procedura wygląda następująco:

1. wyestymować dwa modele w celu uzyskania wartości  $\beta^{G_1}$  i  $\beta^{G_2}$ ,
2. wyliczyć statystykę testową jako różnicę między współczynnikami, czyli jest ona równa  $\tilde{s} = |\beta^{G_1} - \beta^{G_2}|$ ,
3. połączyć obserwacje z obu grup,
4. wyliczyć rozkład empiryczny  $s$ :
  - dokonać losowego przyporządkowania obserwacji do dwóch grup o liczebnościach  $n_1$  i  $n_2$
  - wyliczyć współczynniki  $\beta_i^{G_1}$  i  $\beta_i^{G_2}$  oznaczające wartość współczynnika  $\beta^{G_1}$  i  $\beta^{G_2}$  przy  $i$ -tej permutacji.
  - wyliczyć  $s_i = |\beta_i^{G_1} - \beta_i^{G_2}|$
5. po wykonaniu pewnej liczby permutacji należy sprawdzić, czy  $\tilde{s} < s_{1-\alpha}$  czyli, czy początkowa statystyka testowa jest większa niż kwantyl empiryczny rzędu  $1 - \alpha$ ,
6. jeżeli  $\tilde{s} < s_{1-\alpha}$ , to nie ma podstaw do odrzucenia hipotezy zerowej mówiącej o identyczności współczynników w grupach na poziomie istotności  $\alpha$ , w przeciwnym razie należy odrzucić tę hipotezę.

Metoda ta w odróżnieniu od poprzedniej nie wymaga żadnych założeń odnośnie rozkładu czy charakteru populacji. Dzięki temu może być stosowana wszędzie tam, gdzie założenia testu  $t$  okazały się zbyt restrykcyjne.

Zarówno bootstrap test  $t$ , jak i procedura permutacji zostały zaimplementowane w pakiecie `plspm` w środowisku  $\mathcal{R}$ . Następny rozdział został poświęcony właśnie opisowi najważniejszych funkcji tego pakietu, służących do estymacji współczynników modeli PLS-PM oraz przeprowadzania analizy grup.

W rozdziale trzecim w celu dokonania analizy grup zastosowano zarówno test  $t$ , jak i procedurę permutacji.

## Rozdział 2

# Opis funkcji w pakiecie $\mathcal{R}$

Ta część pracy została poświęcona przedstawieniu funkcji służących do estymacji modelu ścieżek przy pomocy metody najmniejszych kwadratów oraz jego dalszej analizy w środowisku  $\mathcal{R}$ . Komendy opisane w tym rozdziale opierają się na teorii nakreślonej w poprzednim paragrafie. Posłużą one do analizy danych rzeczywistych zawartej w części następnej. Stosowanie niżej opisanych funkcji jest możliwe jedynie po zainstalowaniu dodatkowego pakietu `plspm`, który został utworzony oraz szczegółowo opisany przez G. Sanchez'a[4, 5].

### 2.1. Funkcja `plspm()`

Funkcja `plspm()` jest podstawową funkcją pakietu `plspm`. Służy zasadniczo do estymacji współczynników oraz wag w modelach zewnętrznych na podstawie algorytmu opisanego w rozdziale pierwszym. Poniżej przedstawiono listę argumentów opisywanej funkcji:

```
plspm(data, inner.matrix, outer.list, modes, boot.val = FALSE, br = NULL,...).
```

Argument	Opis
<code>data</code>	Macierz liczbowa zawierająca zmienne mierzalne.
<code>inner.matrix</code>	Macierz zero-jedynkowa dolnotrójkątna odpowiadająca strukturze modelu wewnętrznego.
<code>outer.list</code>	Wektor przyporządkowujący zmienne mierzalne do konkretnych zmiennych ukrytych. Jego wymiar musi być równy wymiarowi <code>inner.matrix</code>
<code>modes</code>	Wektor wskazujący typ bloku. "A" oznacza blok odzwierciedlający, natomiast "B" – blok kształtujący.
<code>boot.val</code>	Domyślna wartość to wszystkie bloki odzwierciedlające. Wartość logiczna mówiąca, czy stosować bootstrap. Domyślna wartość to <code>FALSE</code> .
<code>br</code>	Liczba wskazująca liczbę powtórzeń bootstrapu. Może przyjmować wartość z zakresu 100-1000. Domyślna wartość to 100

Tabela 2.1: Opis argumentów funkcji `plspm()`.

Poza argumentami opisanymi w tabeli 2.1, funkcja ta posiada kilka dodatkowych, mniej istotnych z punktu widzenia modelu przedstawionego w następnym rozdziale, argumentów określających między innymi maksymalną liczbę iteracji algorytmu oraz warunek zbieżności<sup>1</sup>.

<sup>1</sup>Szczegółowe informacje o algorytmie estymacji współczynników zostały zamieszczone w rozdziale pierwszym

Przykładowy schemat użycia funkcji `plspm()` można znaleźć w rozdziale trzecim w części poświęconej estymacji współczynników modelu. W celu odczytania wyników należy na obiekcie klasy `pls` zastosować funkcje określone w tabeli 2.2.

Argument	Opis
<code>\$outer.mod</code>	Charakterystyka modeli zewnętrznych, między innymi wagi zmiennych mierzalnych.
<code>\$inner.mod</code>	Charakterystyka modeli wewnętrznych, między innymi współczynniki oraz $R^2$ .
<code>\$outer.cor</code>	Korelacja pomiędzy zmiennymi mierzalnymi, a zmiennymi ukrytymi.
<code>\$effects</code>	Całkowita siła oddziaływania między zmiennymi ukrytymi.
<code>\$boot</code>	Informacje na temat bootstrapu.

Tabela 2.2: Opis funkcji służących do odczytu wyników z obiektu klasy „`plspm`”.

## 2.2. Funkcja `plspm.groups()`

Funkcja opisana w tej sekcji służy do analizy polegającej na porównywaniu współczynników modelu wewnętrznego w grupach. Poniżej została przedstawiona lista argumentów tej funkcji:

```
plspm.groups(pls, group, method = "bootstrap", reps = NULL).
```

Argument	Opis
<code>pls</code>	Obiekt klasy „ <code>plspm</code> ” utworzony przy pomocy funkcji <code>plspm()</code> .
<code>group</code>	Faktor przyjmujący dwie wartości, będący kluczem podziału na grupy.
<code>method</code>	Metoda służąca do porównywania współczynników. Możliwe wartości to „ <code>bootstrap</code> ” – metoda bootstrapu i „ <code>permutation</code> ” – metoda permutacji <sup>2</sup> . Domyślnie stosowany jest bootstrap.
<code>reps</code>	Liczba wskazująca liczbę powtórzeń bootstrapu lub permutacji. Domyślna wartość to 100.

Tabela 2.3: Opis argumentów funkcji `plspm.groups()`.

Przykładowy schemat użycia funkcji `plspm.groups` został zaprezentowany w rozdziale trzecim. Aby odczytać wyniki należy na obiekcie klasy `plspm.groups` zastosować funkcje określone w tabeli 2.4.

Argument	Opis
<code>\$test</code>	Współczynniki w modelu wewnętrznym w podziale na grupy oraz w ujęciu globalnym wraz z różnicą między nimi, p-value dotyczące testu t lub permutacji, informacja, czy współczynniki są różne na poziomie istotności 5%
<code>\$global</code>	Współczynniki oraz $R^2$ modelu wewnętrznego w ujęciu globalnym.
<code>\$group1</code>	Współczynniki oraz $R^2$ modelu wewnętrznego w pierwszej grupie.
<code>\$group2</code>	Współczynniki oraz $R^2$ modelu wewnętrznego w drugiej grupie.

Tabela 2.4: Opis funkcji służących do odczytu wyników z obiektu klasy „`plspm.groups`”.

## 2.3. Funkcje graficzne

Pakiet `pls` daje możliwość automatycznej wizualizacji niektórych rezultatów otrzymanych przy pomocy funkcji `pls`. Służą do tego dwie bliźniacze funkcje:

- Rysowanie struktury modelu wewnętrznego– w celu przedstawienia zależności między zmiennymi ukrytymi można zastosować funkcję:

```
innerplot(pls,...),
```

gdzie `pls` jest obiektem klasy „`pls`”. Funkcja ta posiada również całą gamę argumentów, które precyzują wygląd rysunku: położenie grotów strzałek, grubość strzałek, kolory i wiele innych.

- Rysowanie struktury modelu zewnętrznego– w celu przedstawienia zależności między zmiennymi ukrytymi można zastosować funkcję:

```
outerplot(pls, what="loadings",...),
```

gdzie, tak jak w przypadku poprzedniej funkcji, `pls` jest obiektem klasy „`pls`” oraz występuje wiele innych argumentów modyfikujących rysunek. W odróżnieniu od funkcji `innerplot()`, `outerplot()` posiada argument przyjmujący dwie wartości: „`loadings`” oraz „`weights`”. Pierwszy z nich powoduje naniesienie na diagram wartości korelacji między poszczególnymi zmiennymi mierzalnymi a zmienną ukrytą, natomiast druga z nich pokazuje wartości wag, za pomocą których jest liczona wartość zmiennej ukrytej.





## Rozdział 3

# Analiza danych rzeczywistych

### 3.1. Wstęp

Niniejszy rozdział jest poświęcony analizie danych rzeczywistych zebranych w roku 2011 na użytek badania „Diagnoza społeczna. Warunki i jakość życia Polaków.” [8]. Jest to badanie panelowe prowadzone od 2000 roku z częstotliwością 2–3 lata. Głównym zadaniem „Diagnozy społecznej” jest uzupełnienie wskaźników instytucjonalnych, czyli np. danych makroekonomicznych, o opinie i zachowania obywateli w celu wykonania bardziej rzetelnej analizy jakości życia w Polsce. Pytania zadawane przez ankietowanych [11], poza zagadnieniami ekonomicznymi, odnoszą się również do takich dziedzin życia jak: ochrona zdrowia, edukacja, sposoby radzenia sobie z kłopotami, styl życia, stres, zachowania patologiczne, uczestnictwo w kulturze, korzystanie z nowoczesnych technologii. Zatem cały projekt ma charakter interdyscyplinarny.

Celem analizy danych [12] zaprezentowanej w tym rozdziale jest wskazanie sfer życia, które w najistotniejszy sposób wpływają na poziom szczęścia oraz wyestymowanie tego poziomu. Następnie zostanie porównany poziom szczęścia kobiet i mężczyzn oraz jego struktura poprzez porównanie współczynników modeli wewnętrznych wyliczonych odpowiednio dla jednej i drugiej płci.

### 3.2. Struktura modelu

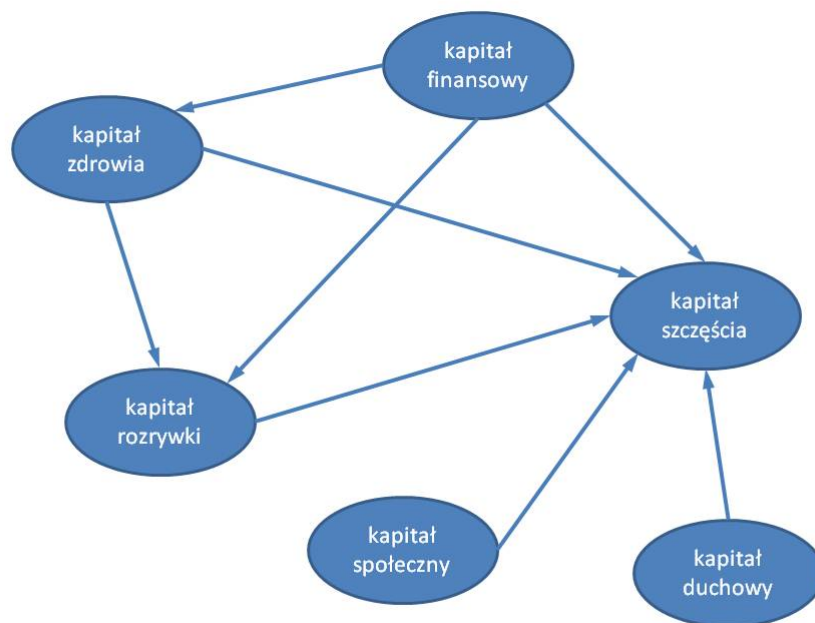
Encyklopedia PWN [2] definiuje szczęście jako:

„1) pomyślny bilans doświadczeń życiowych, powodzenie w realizacji celów życiowych; zdobycie najwyższej cenionych społecznie dóbr i przymiotów; w węższym znaczeniu — sprzyjający los, pomyślny zbieg okoliczności w jakiejś konkretnej dziedzinie czy przedsięwzięciu; 2) chwilowy stan emocjonalnej euforii, odczucie najwyższej radości, doznanie maksymalnej przyjemności; 3) trwałe zadowolenie z życia połączone z pogodą ducha i optymizmem, poznawcza ocena własnego życia jako udanego, wartościowego, sensownego.”

W niniejszej pracy szczęście będzie rozumiane w znaczeniu trzecim, czyli jako „trwałe zadowolenie z życia”. Według amerykańskiego psychologa Abrahama Maslow’a o poziomie szczęścia<sup>1</sup> decyduje to, czy dana osoba ma zaspokojone potrzeby opisane w tzw. „piramidzie Maslow’a” [1]. Według tej teorii, potrzeby życiowe dzielą się na pięć grup. Podstawowa kategoria to potrzeby fizjologiczne, następnie bezpieczeństwo, miłość i przynależność, szacunek

---

<sup>1</sup>rozumianym jako zadowolenie z życia



Rysunek 3.1: Model wewnętrzny [opracowanie własne].

i uznanie oraz potrzeba samorealizacji. Piramida Maslow’a została zilustrowana na rysunku 3.2 na stronie 28.

### 3.2.1. Model wewnętrzny

Do kompozycji modelu wewnętrznego wykorzystano sześć zmiennych ukrytych. Pierwsza z nich to kapitał szczęścia (*kapitał.szczęścia*) czyli zmienna endogeniczna, która jest najbardziej istotna z punktu widzenia modelu. Na kapitał szczęścia wpływa pięć kolejnych zmiennych ukrytych:

- Kapitał finansowy (*kapitał.finansowy*)– obecnie zamożność gwarantuje zaspokojenie potrzeb takich jak: jedzenie, picie (potrzeby fizjologiczne), wygodę (potrzeba bezpieczeństwa), a czasem również posiadanie własnej wartości, poważanie (potrzeba szacunku i uznania). Należy jednak zaznaczyć, że wiele spośród potrzeb nakreślonych przez Maslow’a nie da się kupić np. miłość (potrzeba miłości i przynależności), wolność od strachu (poczucie bezpieczeństwa).
- Kapitał zdrowia (*kapitał.zdrowia*)– brak zdrowia fizycznego może wpływać nie tylko na niemożność zaspokojenia takich potrzeb jak sen czy innych potrzeb fizjologicznych. W przypadku poważnej, długotrwałej choroby osoba może zostać wyalienowana z życia społecznego (potrzeba miłości i przynależności) lub też, jak to ma miejsce np. wśród zarażonych wirusem HIV, wiązać się z utratą własnej wartości i zaufania do siebie (potrzeba szacunku i uznania). Brak zdrowia psychicznego (np. depresja) może wiązać się z brakiem wiary we własny potencjał.
- Kapitał duchowy (*kapitał.duchowy*)– sfera sakralna odpowiada u ludzi wierzących za

posiadanie celów (potrzeba samorealizacji), poczucie własnej wartości (potrzeba szacunku i uznania).

- Kapitał rozrywki (*kapitał.rozrywki*)– ten blok odpowiada przede wszystkim za relacje międzyludzkie na podłożu towarzyskim czyli determinuje poziom zaspokojenia potrzeb szacunku i uznania oraz miłości i przynależności.
- Kapitał społeczny (*kapitał.społeczny*)– aktywność społeczna wiąże się z posiadaniem celów, spełnianiem własnego potencjału (potrzeba samorealizacji) oraz kompetencjami i poważaniem społecznym (potrzeba szacunku i uznania).

Oczywiście istnieje jeszcze bardzo wiele innych determinantów szczęścia np. sytuacja zawodowa czy samorealizacja poprzez działalność naukową. Jednak należy pamiętać, że model jest przybliżeniem rzeczywistości i nigdy nie uda się wskazać wszystkich zmiennych kształtujących poziom szczęścia. Również dla zachowania przejrzystości zdecydowano się poprzestać na blokach opisanych powyżej.

Modelowanie ścieżek pozwala zaobserwować również inne zależności pomiędzy zmiennymi ukrytymi. W związku z tym postanowiono wzbogacić strukturę modelu o następujące zależności:

- Kapitał finansowy → kapitał zdrowia– prawo Engel’a[10] mówi, że wraz ze wzrostem dochodu następuje ponadproporcjonalny wzrost wydatków na ochronę zdrowia. Ta zasada została wielokrotnie potwierdzona w bardzo licznych badaniach empirycznych w wielu krajach na przestrzeni ponad stu lat. Zwiększenie wydatków na zdrowie podnosi kapitał zdrowia, czyli kapitał finansowy kształtuje kapitał zdrowia.
- Kapitał finansowy → kapitał rozrywki– rozrywka zaliczana jest do dóbr luksusowych, czyli takich, których elastyczność dochodowa popytu[6] jest większa niż 1, co oznacza, że wraz ze wzrostem dochodu wydatki na rozrywkę rosną szybciej niż następuje wzrost dochodu. Wzrost wydatków na rozrywkę oznacza wzrost kapitału rozrywki, czyli kapitał finansowy kształtuje kapitał rozrywki.
- Kapitał zdrowia → kapitał rozrywki– choroba w znacznej większości przypadków powoduje w pierwszej kolejności rezygnację z rozrywki, niezależnie czy rozpatruje się jedynie tygodniową gorączkę, czy też poważny uszczerbek na zdrowiu taki jak inwalidztwo. Zatem kapitał zdrowotny kształtuje kapitał rozrywki.

Zarówno kapitał społeczny jak i duchowy zdaje się nie być kształtowany przez zdrowie czy też zasobność finansową. Istnieje wiele przykładów ludzi aktywnych społecznie, którzy są nieuleczalnie chorzy. Patrząc np. na związki zawodowe nie można również twierdzić, że kapitał finansowy kształtuje kapitał społeczny. Sfera duchowa jest również całkowicie niezależna. Struktura modelu wewnętrznego została przedstawiona na rysunku 3.1 na stronie 24.

### 3.2.2. Opis bloku pierwszego

Blok pierwszy określa zmienną ukrytą *kapital.szczescia* przy pomocy pięciu zmiennych mierzalnych:

- ocena dotychczasowego życia ( $s1$ ),
- pragnienie życia ( $s2$ ),
- poczucie osamotnienia ( $s3$ ),
- poczucie bycia kochanym i darzonym zaufaniem ( $s4$ ),
- myśli samobójcze ( $s5$ ).

Dokładne informacje na temat zmiennych zostały zamieszczone w tabelach A.4 i A.5 na stronie 44. Wszystkie zmienne w tym bloku są typu odzwierciedlającego.

### 3.2.3. Opis bloku drugiego

Blok drugi określa zmienną ukrytą *kapital.finansowy* przy pomocy dwóch zmiennych mierzalnych:

- stabilność dochodów ( $f6$ ),
- wielkość dochodów miesięcznych netto (*dochod*),
- zadowolenie z sytuacji finansowej ( $f7$ ).

Dokładne informacje na temat zmiennych zostały zamieszczone w tabelach A.1, A.2 i A.3 na stronie 43. Wszystkie zmienne w tym bloku są typu odzwierciedlającego.

### 3.2.4. Opis bloku trzeciego

Blok trzeci określa zmienną ukrytą *kapital.zdrowia* przy pomocy trzech zmiennych mierzalnych:

- występowanie dolegliwości fizycznych ( $z8$ ),
- trudności w codziennym funkcjonowaniu ( $z9$ ),
- zadowolenie ze stanu zdrowia ( $z10$ ).

Dokładne informacje na temat zmiennych zostały zamieszczone w tabelach A.6 i A.7 na stronie 45. Wszystkie zmienne w tym bloku są typu odzwierciedlającego.

### 3.2.5. Opis bloku czwartego

Blok czwarty określa zmienną ukrytą *kapital.duchowy* przy pomocy dwóch zmiennych mierzalnych:

- najważniejsze wartości ( $d11$ ),
- miesięczna liczba nabożeństw ( $d12$ ).

Dokładne informacje na temat zmiennych zostały zamieszczone w tabelach A.8, A.9 i A.10 na stronie 46. Wszystkie zmienne w tym bloku są typu odzwierciedlającego.

### 3.2.6. Opis bloku piątego

Blok piąty określa zmienną ukrytą *kapital.rozrywki* przy pomocy sześciu zmiennych mierzalnych:

- sport i rekreacja (*r13*),
- liczba przyjaciół (*r14*),
- wyjścia do kina, teatru, na koncert (*r15*),
- wyjścia do pubu, restauracji, kawiarni (*r16*),
- liczba spotkań towarzyskich (*r17*),
- zadowolenie z grona przyjaciół (*r18*).

Dokładne informacje na temat zmiennych zostały zamieszczone w tabelach A.11, A.12 i A.13 na stronie 47. Wszystkie zmienne w tym bloku są typu odzwierciedlającego.

### 3.2.7. Opis bloku szóstego

Blok szósty określa zmienną ukrytą *kapital.spoleczny* przy pomocy czterech zmiennych mierzalnych:

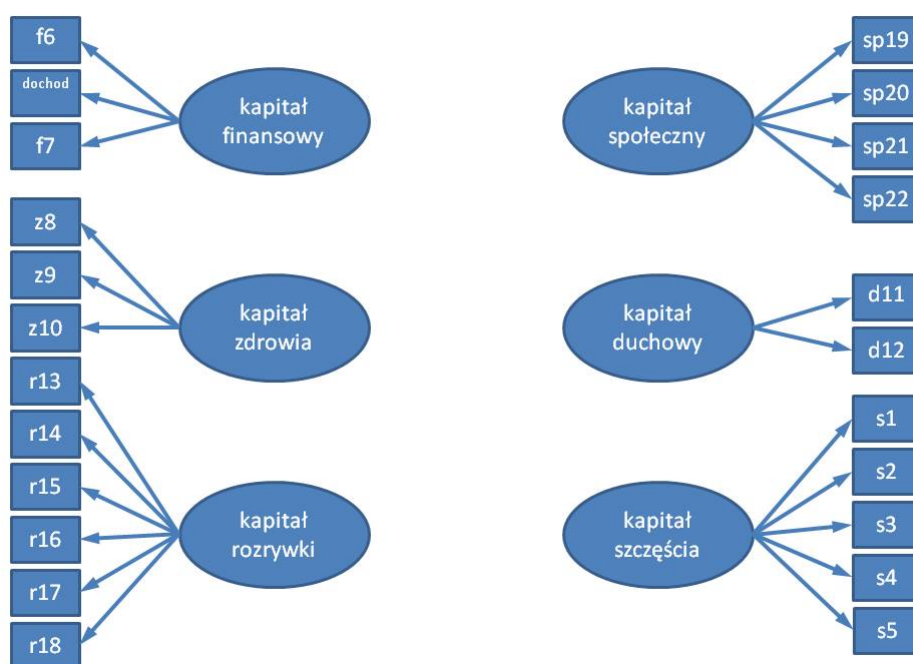
- zaangażowanie w działalność na rzecz lokalnej społeczności (*sp19*),
- uczestnictwo w zebraniach publicznych (*sp20*),
- wolontariat (*sp21*),
- członkostwo w organizacjach (*sp22*).

Dokładne informacje na temat zmiennych zostały zamieszczone w tabelach A.14 i A.15 na stronie 48. Wszystkie zmienne w tym bloku są typu odzwierciedlającego.

Struktury modeli zewnętrznych zostały przedstawione na rysunku 3.3 na stronie 28.



Rysunek 3.2: Piramida Maslow'a[1] [opracowanie własne].



Rysunek 3.3: Modele zewnętrzne. *s1*– ocena dotychczasowego życia, *s2*– pragnienie życia, *s3*– poczucie osamotnienia, *s4*– poczucie bycia kochanym i darzonym zaufaniem, *s5*– myśli samobójcze, *f6*– stabilność dochodów, *dochod*– dochód netto, *f7*– zadowolenie z sytuacji finansowej, *z8*– występowanie dolegliwości fizycznych, *z9*– trudności w codziennym funkcjonowaniu, *z10*– zadowolenie ze stanu zdrowia, *d11*– najważniejsze wartości, *d12*– miesięczna liczba nabożeństw, *r13*– sport i rekreacja, *r14*– liczba przyjaciół, *r15*– wyjścia do kina, teatru, na koncert, *r16*– wyjścia do pubu, restauracji, kawiarni, *r17*– liczba spotkań towarzyskich, *r18*– zadowolenie z grona przyjaciół, *sp19*– zaangażowanie w działalność na rzecz społeczności lokalnej, *sp20*– uczestnictwo w zebraniach publicznych, *sp21*– wolontariat, *sp22*– członkostwo w organizacji [opracowanie własne].

### 3.3. Wstępna analiza danych

Dane[12] zostały pozyskane z portalu SmarterPoland.pl[3]. Po usunięciu wszystkich rekordów, w których znajdowały się braki w zmiennych użytych w modelu, otrzymano 17251 obserwacji. Następnie dokonano przekształcenia danych na dane liczbowe oraz zbadano znak współczynnika korelacji zmiennych pochodzących z tego samego bloku:

```
> cor(DANE1[, 1:5])
      s1      s2      s3      s4      s5
s1 1.0000000 0.3717215 0.3467195 0.2947015 0.2027691
s2 0.3717215 1.0000000 0.3532022 0.2904985 0.3277589
s3 0.3467195 0.3532022 1.0000000 0.3543625 0.2403104
s4 0.2947015 0.2904985 0.3543625 1.0000000 0.2125090
s5 0.2027691 0.3277589 0.2403104 0.2125090 1.0000000
> cor(DANE1[, 6:8])
      f6      dochod      f7
f6 1.0000000 0.1223978 0.3070141
dochod 0.1223978 1.0000000 0.2353601
f7 0.3070141 0.2353601 1.0000000
> cor(DANE1[, 9:11])
      z8      z9      z10
z8 1.0000000 0.7595224 0.5463306
z9 0.7595224 1.0000000 0.5737971
z10 0.5463306 0.5737971 1.0000000
> cor(DANE1[, 12:13])
      d11      d12
d11 1.0000000 0.1262789
d12 0.1262789 1.0000000
> cor(DANE1[, 14:19])
      r13      r14      r15      r16      r17      r18
r13 1.00000000 0.09730727 0.19794087 0.22839170 0.2413527 0.11740769
r14 0.09730727 1.00000000 0.04873300 0.09941927 0.1459825 0.14522579
r15 0.19794087 0.04873300 1.00000000 0.36905866 0.3076829 0.09229124
r16 0.22839170 0.09941927 0.36905866 1.00000000 0.5201598 0.10540448
r17 0.24135265 0.14598252 0.30768294 0.52015984 1.0000000 0.16088021
r18 0.11740769 0.14522579 0.09229124 0.10540448 0.1608802 1.00000000
> cor(DANE1[, 20:23])
      sp19      sp20      sp21      sp22
sp19 1.00000000 0.4382456 0.4691754 0.3832976
sp20 0.4382456 1.0000000 0.3750390 0.3525683
sp21 0.4691754 0.3750390 1.0000000 0.3691184
sp22 0.3832976 0.3525683 0.3691184 1.0000000
```

Wszystkie zmienne w poszczególnych blokach są dodatnio skorelowane. Jest to warunek konieczny do poprawnego wyliczenia modelu przy pomocy zaprezentowanych w rozdziale drugim funkcji. Tak przygotowane dane posłużyły do estymacji parametrów modelu.

### 3.4. Estymacja parametrów modelu

Po wyczyszczeniu danych z niepełnych obserwacji oraz dokonaniu wstępnej analizy danych przystąpiono do zastosowania funkcji `plspm`. Zgodnie z informacjami zawartymi w rozdziale drugim, wskazano strukturę modeli zewnętrznych oraz modelu wewnętrznego. Na podstawie opisu zawartego w podrozdziale *Struktura modelu*, została utworzona macierz charakteryzująca model wewnętrzny:

```
> kapital.finansowy=c(0,0,0,0,0,0)
> kapital.szczescia=c(1,1,1,1,1,0)
> kapital.zdrowia=c(1,0,0,0,0,0)
> kapital.duchowy=c(0,0,0,0,0,0)
> kapital.rozrywki=c(1,1,0,0,0,0)
> kapital.spoleczny=c(0,0,0,0,0,0)
> inner=rbind(kapital.finansowy,kapital.zdrowia,kapital.duchowy,
kapital.rozrywki,kapital.spoleczny,kapital.szczescia)
> colnames(inner)=rownames(inner)
> inner
```

	kapital.finansowy	kapital.zdrowia	kapital.duchowy
kapital.finansowy	0	0	0
kapital.zdrowia	1	0	0
kapital.duchowy	0	0	0
kapital.rozrywki	1	1	0
kapital.spoleczny	0	0	0
kapital.szczescia	1	1	1

	kapital.rozrywki	kapital.spoleczny	kapital.szczescia
kapital.finansowy	0	0	0
kapital.zdrowia	0	0	0
kapital.duchowy	0	0	0
kapital.rozrywki	0	0	0
kapital.spoleczny	0	0	0
kapital.szczescia	1	1	0

Następnie przyporządkowano zmienne mierzalne do poszczególnych bloków oraz określono typy bloków:

```
> outer=list(6:8,9:11,12:13,14:19,20:23,1:5)
> modes=c("A","A","A","A","A","A")
```

Po przygotowaniu wszystkich argumentów wykonano estymację modelu:

```
> pls=plspm(DANE1,inner,outer,modes)
```



### 3.5. Diagnostyka modelu

Weryfikacja adekwatności struktury modelu zawierającego zmienne mierzalne typu odzwierciedlającego, zgodnie z opisem zawartym w czwartym rozdziale podręcznika [4], powinna odnosić się do trzech aspektów.

Po pierwsze, należy pamiętać, że zmienne ukryte można interpretować jako pewnego rodzaju uogólnienie zmiennych mierzalnych opisujących tę zmienną. Zatem, wszystkie zmienne mierzalne należące do jednego bloku powinny być ze sobą wspólnie powiązane. Gdy jedna się zwiększa, to również pozostałe powinny ulec zwiększeniu i na odwrót. Mówiąc krócej, powinny być ze sobą ściśle skorelowane.

Z drugiej strony, zmienne mierzalne z danego bloku muszą wykazywać taką samą zależność (być ściśle skorelowane), nie tylko między sobą nawzajem, ale również ze zmienną ukrytą, którą opisują. To znaczy, że gdy wartość zmiennej ukrytej rośnie, to również zmienne mierzalne rosną i odwrotnie.

Ostatnią rzeczą konieczną do zweryfikowania jest sprawdzenie, czy zmienne mierzalne z danego bloku są najsilniej skorelowane ze zmienną ukrytą, którą opisują. Nie może być takiej sytuacji, że na przykład korelacja pomiędzy zmienną  $f6$  mówiącą o stabilności dochodów jest silniej skorelowana z kapitałem zdrowia niż z kapitałem finansowym.

Do weryfikacji pierwszego założenia służy tak zwana jednowymiarowość zmiennych mierzalnych (ang. *unidimensionality of indicators*). Nazwa pochodzi od interpretacji geometrycznej, która mówi, że zmienne mierzalne opisujące tę samą zmienną ukrytą (mierzące ten sam aspekt) powinny wskazywać ten sam, jeden kierunek. Jednowymiarowość można łatwo odczytać z obiektu klasy `pls` utworzonego we wcześniejszym podrozdziale:

```
> pls$unidim
```

	Type.measure	MVs	C.alpha	DG.rho	eig.1st	eig.2nd
kapital.finansowy	Reflective	3	0.4606296	0.7353201	1.451458	0.8827718
kapital.zdrowia	Reflective	3	0.8342506	0.9010511	2.258200	0.5024065
kapital.duchowy	Reflective	2	0.2242409	0.7205236	1.126279	0.8737211
kapital.rozrywki	Reflective	6	0.5874674	0.7390498	2.064529	0.9608537
kapital.spoleczny	Reflective	4	0.7255383	0.8294486	2.196639	0.6600892
kapital.szczescia	Reflective	5	0.6812549	0.7970491	2.208470	0.8405899

Pierwsza kolumna mówi o typie bloku, następna o liczbie zmiennych mierzalnych zawartych w danym bloku. Następnie są podane wartości dwóch współczynników. Pierwszy z nich to  $\alpha$  Cronbach'a, druga to  $\rho$  Dillon'a-Goldstein'a. Obie służą do weryfikacji jednowymiarowości. Jednak pierwsza z nich może być stosowana wyłącznie, gdy zmienne są znormalizowane, czyli nie w tym przypadku. Więcej informacji na temat tych współczynników można znaleźć w podręczniku [13] na stronie 50. Autorzy, zarówno [13] jak i [4] uznają, że gdy wartość  $\rho$  Dillon'a-Goldstein'a jest większa niż 0,7, to można przyjąć, że zmienne mierzalne cechuje jednowymiarowość. Do weryfikacji tej cechy może posłużyć również wyliczenie wartości własnych macierzy korelacji zmiennych mierzalnych należących do tego samego bloku. Wartości pierwszej wartości własnej powinny być większe niż 1, zaś drugiej powinny być mniejsze niż 1. Zarówno pierwsza, jak i druga metoda potwierdza jednowymiarowość zmiennych mierzalnych należących do tych samych bloków.

Do weryfikacji drugiego założenia zostanie przeprowadzona analiza korelacji między zmiennymi ukrytymi i opisującymi je zmiennymi mierzalnymi. Wartości szukanego współczynnika korelacji są zawarte w kolumnie drugiej:

```

> pls\$outer.mod
\kapital.finansowy
      weights std.loads communal redundan
f6      0.1980    0.7734    0.5981         0
dochod  0.4178    0.7140    0.5098         0
f7      0.7304    0.8895    0.7913         0

\kapital.zdrowia
      weights std.loads communal redundan
z8      0.3372    0.8561    0.7330    0.0451
z9      0.3295    0.8678    0.7530    0.0463
z10     0.4917    0.8650    0.7483    0.0460

\kapital.duchowy
      weights std.loads communal redundan
d11     0.4751    0.7789    0.6067         0
d12     0.8220    0.8820    0.7779         0

\kapital.rozrywki
      weights std.loads communal redundan
r13     0.3065    0.7447    0.5546    0.0435
r14     0.1562    0.7344    0.5393    0.0164
r15     0.2299    0.7138    0.5095    0.0387
r16     0.2564    0.8042    0.6467    0.0535
r17     0.2327    0.8197    0.6719    0.0563
r18     0.5351    0.8795    0.7735    0.0677

\kapital.spoleczny
      weights std.loads communal redundan
sp19    0.3164    0.7603    0.5781         0
sp20    0.3712    0.7509    0.5639         0
sp21    0.3116    0.7293    0.5319         0
sp22    0.3523    0.7194    0.5176         0

\kapital.szczescia
      weights std.loads communal redundan
s1      0.4362    0.7639    0.5835    0.1306
s2      0.3561    0.7397    0.5471    0.1224
s3      0.2722    0.7698    0.5926    0.1004
s4      0.2035    0.7748    0.6003    0.0739
s5      0.2019    0.7157    0.5122    0.0595

```

Wartości większe niż 0,7 są uznawane jako wystarczające. W następnej kolumnie znajdują się współczynniki korelacji podniesione do kwadratu. Ta wartość mówi, jaka część zmienności zmiennej mierzalnej została wytłumaczona poprzez zmienność zmiennej ukrytej. Ta kwestia została bardziej szczegółowo opisana na stronie 62 podręcznika [4]. Zatem wartości korelacji większe niż 0,7 oznaczają, że ponad połowa ( $0,7^2 \approx 0,5$ ) zmienności została wytłumaczona zmiennością zmiennej ukrytej.

Ostatnią kwestią, którą należy zbadać jest powiązanie zmiennych mierzalnych ze zmiennymi ukrytymi opisywanymi w pozostałych blokach (ang. *cross-loadings*) W tym celu trzeba zbadać korelację pomiędzy zmiennymi mierzalnymi i poszczególnymi zmiennymi ukrytymi:

```
> pls\$outer.cor
\ $kapital.finansowy
      kapital.finansowy kapital.zdrowia kapital.duchowy
f6          0.7734          0.0070          0.0971
dochod      0.7140          0.1523         -0.0511
f7          0.8895          0.2505          0.0683
      kapital.rozrywki kapital.spoleczny kapital.szczescia
f6          0.0286         -0.0446          0.2021
dochod      0.1861          0.1397          0.1635
f7          0.2655          0.0566          0.3611
\ $kapital.zdrowia
      kapital.finansowy kapital.zdrowia kapital.duchowy
z8          0.1432          0.8512         -0.0838
z9          0.1435          0.8639         -0.0789
z10         0.2657          0.8709         -0.0240
      kapital.rozrywki kapital.spoleczny kapital.szczescia
z8          0.2451         -0.0018          0.2343
z9          0.2278         -0.0214          0.2423
z10         0.3414          0.0375          0.3490
\ $kapital.duchowy
      kapital.finansowy kapital.zdrowia kapital.duchowy
d11         0.0546         -0.0334          0.7782
d12         0.0721         -0.0601          0.8824
      kapital.rozrywki kapital.spoleczny kapital.szczescia
d11         0.0179          0.0928          0.0497
d12         0.0067          0.1333          0.0862
\ $kapital.rozrywki
      kapital.finansowy kapital.zdrowia kapital.duchowy
r13         0.0986          0.2034         -0.0034
r14         0.0750          0.0301          0.0937
r15         0.0964          0.1384         -0.0238
r16         0.0814          0.1743         -0.0859
r17         0.0748          0.1685         -0.0284
r18         0.2664          0.2671          0.0546
      kapital.rozrywki kapital.spoleczny kapital.szczescia
r13         0.7275          0.1943          0.1242
r14         0.7441          0.1368          0.1355
r15         0.7922          0.1094          0.0805
r16         0.8169          0.1078          0.0858
r17         0.8040          0.1729          0.0895
r18         0.8132          0.0628          0.3019
\ $kapital.spoleczny
```

	kapital.finansowy	kapital.zdrowia	kapital.duchowy
sp19	0.0215	0.0054	0.1015
sp20	0.0211	0.0009	0.0937
sp21	0.0175	0.0316	0.0877
sp22	0.0556	-0.0013	0.1689
	kapital.rozrywki	kapital.spoleczny	kapital.szczescia
sp19	0.1531	0.7601	0.0555
sp20	0.1305	0.7514	0.0653
sp21	0.1735	0.7289	0.0546
sp22	0.1504	0.7194	0.0618
\ \$kapital.szczescia			
	kapital.finansowy	kapital.zdrowia	kapital.duchowy
s1	0.3672	0.2767	0.0824
s2	0.2330	0.2706	0.0537
s3	0.2345	0.2253	0.0542
s4	0.1728	0.1045	0.0798
s5	0.1555	0.1641	0.0405
	kapital.rozrywki	kapital.spoleczny	kapital.szczescia
s1	0.2632	0.0830	0.7615
s2	0.2322	0.0798	0.7361
s3	0.1604	0.0163	0.7746
s4	0.1361	0.0294	0.7808
s5	0.1073	0.0302	0.7157

Weryfikacja adekwatności doboru zmiennych mierzalnych polega na sprawdzeniu, czy korelacja pomiędzy daną zmienną mierzalną a zmienną ukrytą, którą określa, jest większa niż jej korelacja z innymi zmiennymi ukrytymi. Przykładowo, korelacja między zmienną *s5* a *kapital.szczescia* to 0,7157, zaś korelacja *s5* z pozostałymi zmiennymi ukrytymi nie przekracza 0,17. W przypadku rozpatrywanego modelu wszystkie zmienne są adekwatnie użyte.

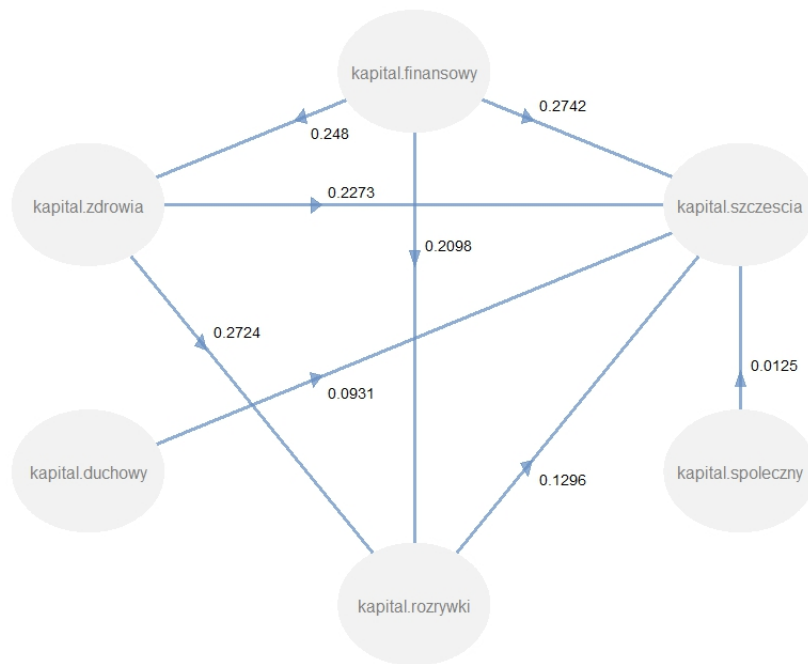
### 3.6. Prezentacja wyników

Głównym celem opisaney w tym rozdziale analizy jest wskazanie determinantów szczęścia oraz oszacowanie jego wartości. Aby zrealizować te cele należy poznać wartości współczynników modelu. Najpierw rozpatrzono współczynniki modelu wewnętrznego:

> pls\ \$effects				
	relationships	dir.effects	ind.effects	tot.effects
1	kapital.finansowy->kapital.zdrowia	0.24800672	0.00000000	0.24800672
2	kapital.finansowy->kapital.rozrywki	0.20977915	0.06755200	0.27733115
3	kapital.finansowy->kapital.szczescia	0.27419137	0.09231516	0.36650653
4	kapital.zdrowia->kapital.rozrywki	0.27237970	0.00000000	0.27237970
5	kapital.zdrowia->kapital.szczescia	0.22725216	0.03531316	0.26256531
6	kapital.duchowy->kapital.szczescia	0.09310232	0.00000000	0.09310232
7	kapital.rozrywki->kapital.szczescia	0.12964680	0.00000000	0.12964680
8	kapital.spoleczny->kapital.szczescia	0.01250715	0.00000000	0.01250715

Pierwsza kolumna wskazuje, z zależnością między którymi zmiennymi ukrytymi ma się do czynienia. Następnie opisany jest efekt bezpośredni równy współczynnikom modelu wewnętrznego oraz efekt pośredni, czyli iloczyn efektów bezpośrednich leżących na pośredniej ścieżce. W ostatniej kolumnie jest suma efektów pośrednich i bezpośrednich. Wyniki przedstawiono również w postaci rysunku 3.4 na stronie 35, który został wykonany przy użyciu polecenia:

```
> innerplot(pls,arr.pos=0.35,arr.width=0.4)
```



Rysunek 3.4: Diagram przedstawiający model wewnętrzny wraz z wyestymowanymi współczynnikami [opracowanie własne].

Do uzyskania pełni informacji odnośnie modelu potrzeba jeszcze poznać wagi, które posłużyły do wyznaczenia zmiennych ukrytych jako kombinacji liniowych zmiennych mierzalnych:

```
> pls$out.weights
      f6      dochod      f7      z8      z9      z10
0.1980279 0.4178485 0.7303846 0.3372309 0.3294739 0.4917463
      d11      d12      r13      r14      r15      r16
0.4750559 0.8220086 0.3064563 0.1562353 0.2299015 0.2563999
      r17      r18      sp19      sp20      sp21      sp22
0.2326878 0.5351178 0.3164359 0.3711676 0.3116136 0.3522641
      s1      s2      s3      s4      s5
0.4362228 0.3560904 0.2721815 0.2034829 0.2018904
```

Wyniki przedstawiono również w postaci rysunku 3.5 na stronie 37, który został wykonany przy użyciu polecenia:

```
> outerplot(pls,what="weights",arr.width=0.2)
```

### 3.7. Analiza grup

Po zaprezentowaniu wyników analizy poziomu szczęścia dla reprezentantów całego społeczeństwa, ciekawe się wydaje przeprowadzenie podobnej analizy dla wybranych grup społecznych w celu weryfikacji, czy różne grupy społeczne odczuwają podobny, czy też znacząco różny poziom zadowolenia z życia. W niniejszym podrozdziale zostanie zweryfikowana hipoteza, czy mężczyźni i kobiety są tak samo szczęśliwe, a jeśli nie, to która z tych grup cechuje się wyższym poziomem szczęścia. Spośród 17251 obserwacji, 8668 to mężczyźni, zaś 8583 to kobiety. Dane zostały podzielone oraz wyestymowano parametry modeli przy wykorzystaniu następujących komend:

```
> P=factor(p)
> mezczyzni=DANE1[DANE1$P==1,]
> kobiety=DANE1[DANE1$P==2,]
> pls_mezczyzni=plspm(mezczyzni,inner,outer,modes)
> pls_kobiety=plspm(kobiety,inner,outer,modes)
```

Należy pamiętać, że zmienna, która stanowi klucz podziału musi być zmienną jakościową, czyli faktorem (ang. *factor*). Po estymacji parametrów modeli wyliczono kapitał szczęścia dla wszystkich obserwacji oraz osobno dla kobiet i mężczyzn:

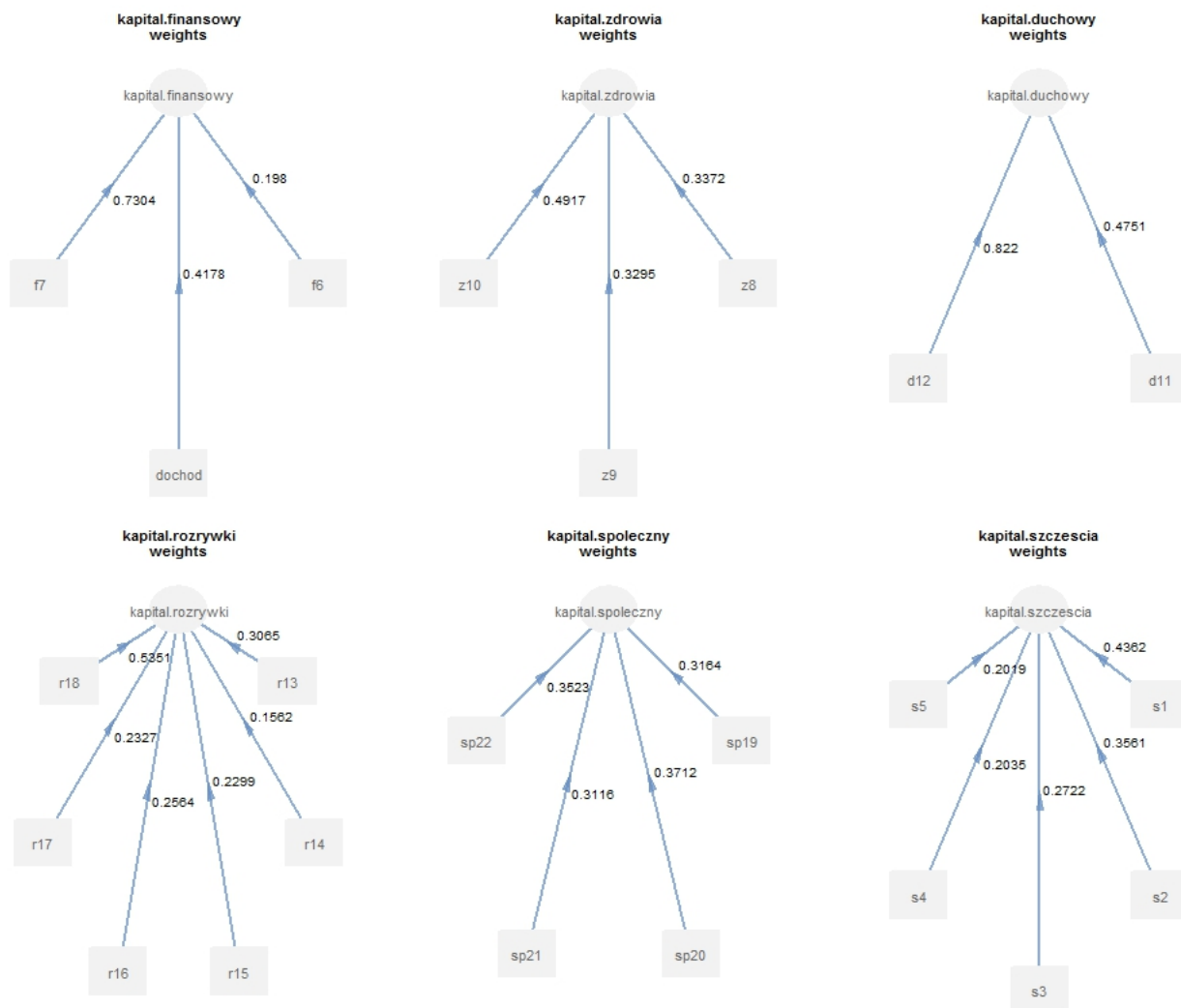
```
> summary(pls$scores[,6])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.9847  5.359   6.119   5.877   6.587   7.190
> summary(pls_mezczyzni$scores[,6])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.9847  5.3660  6.1290  5.8910  6.5960   7.1900
> summary(pls_kobiety$scores[,6])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.019   5.313   6.107   5.862   6.577   7.150
```

W celu zweryfikowania hipotezy odnośnie wpływu płci na poziom szczęścia wykorzystano test Wilcoxon'a [15]:

```
> wilcox.test(pls_mezczyzni$scores[,6],pls_kobiety$scores[,6])

Wilcoxon rank sum test with continuity correction

data:  pls_mezczyzni$scores[, 6] and pls_kobiety$scores[, 6]
W = 37522406, p-value = 0.3218
alternative hypothesis: true location shift is not equal to 0
```



Rysunek 3.5: Diagram przedstawiający modele zewnętrzne wraz z wyestymowanymi wagami. *s1*– ocena dotychczasowego życia, *s2*– pragnienie życia, *s3*– poczucie osamotnienia, *s4*– poczucie bycia kochanym i darzonym zaufaniem, *s5*– myśli samobójcze, *f6*– stabilność dochodów, *dochod*– dochód netto, *f7*– zadowolenie z sytuacji finansowej, *z8*– występowanie dolegliwości fizycznych, *z9*– trudności w codziennym funkcjonowaniu, *z10*– zadowolenie ze stanu zdrowia, *d11*– najważniejsze wartości, *d12*– miesięczna liczba nabożeństw, *r13*– sport i rekreacja, *r14*– liczba przyjaciół, *r15*– wyjścia do kina, teatru, na koncert, *r16*– wyjścia do pubu, restauracji, kawiarni, *r17*– liczba spotkań towarzyskich, *r18*– zadowolenie z grona przyjaciół, *sp19*– zaangażowanie w działalność na rzecz społeczności lokalnej, *sp20*– uczestnictwo w zebraniach publicznych, *sp21*– wolontariat, *sp22*– członkostwo w organizacji [opracowanie własne].

P-value w teście Wilcoxon'a wynosi 0,3218, zatem nie ma podstaw do odrzucenia hipotezy zerowej (na poziomie istotności 5%) mówiącej o równości rozkładów obu grup.

W celu przeprowadzenia bardziej wnikliwej analizy różnic pomiędzy grupami, dokonano również weryfikacji hipotezy mówiącej, że współczynniki modelu estymowanego w grupach nie będą istotnie różne. W tym celu przeprowadzono procedurę permutacji oraz bootstrapu, które zostały opisane w pierwszym rozdziale, przy użyciu funkcji `plspm.groups()`:

```
> pls.group.perm=plspm.groups(pls,DANE1\ $P,method="permutation")
> pls.group.perm
GROUP COMPARISON IN PLS-PM FOR PATH COEFFICIENTS

Scale of Data:          TRUE
Weighting Scheme:       centroid
Selected method:        permutation
Number of replicates:   100

\ $test
```

	global	group.1	group.2
kapital.finansowy->kapital.zdrowia	0.2480	0.2446	0.2515
kapital.finansowy->kapital.rozrywki	0.2098	0.2110	0.2083
kapital.finansowy->kapital.szczescia	0.2742	0.2772	0.2708
kapital.zdrowia->kapital.rozrywki	0.2724	0.2623	0.2824
kapital.zdrowia->kapital.szczescia	0.2273	0.2219	0.2330
kapital.duchowy->kapital.szczescia	0.0931	0.0973	0.0886
kapital.rozrywki->kapital.szczescia	0.1296	0.1333	0.1259
kapital.spoleczny->kapital.szczescia	0.0125	0.0094	0.0177

	diff.abs	p.value	sig.05
kapital.finansowy->kapital.zdrowia	0.0069	0.5743	no
kapital.finansowy->kapital.rozrywki	0.0027	0.9703	no
kapital.finansowy->kapital.szczescia	0.0064	0.6931	no
kapital.zdrowia->kapital.rozrywki	0.0201	0.2079	no
kapital.zdrowia->kapital.szczescia	0.0112	0.4950	no
kapital.duchowy->kapital.szczescia	0.0088	0.4851	no
kapital.rozrywki->kapital.szczescia	0.0074	0.6238	no
kapital.spoleczny->kapital.szczescia	0.0082	0.5842	no

```

Inner models in the following objects:
\ $global
\ $group1
\ $group2
```

Pierwsza kolumna zawiera współczynniki dla całej populacji, następne dla grupy mężczyzn i kobiet, dalej podano różnicę między współczynnikami, p-value testu oraz czy odrzucić hipotezę zerową, która mówi, że różnica między współczynnikami jest statystycznie nieistotna na poziomie istotności 5%. Powyższy test nie wskazał współczynników, których wartości są istotnie różne na poziomie istotności 5%.



Dla pewności zastosowano jeszcze test bootstrapowy:

```
> pls.group.boot=plspm.groups(pls,DANE1\ $P,method="bootstrap")
> pls.group.boot
GROUP COMPARISON IN PLS-PM FOR PATH COEFFICIENTS

Scale of Data:          TRUE
Weighting Scheme:       centroid
Selected method:        bootstrap
Number of replicates:   100

\ $test
```

	global	group.1	group.2	diff.abs
kapital.finansowy->kapital.zdrowia	0.2480	0.2446	0.2515	0.0069
kapital.finansowy->kapital.rozrywki	0.2098	0.2110	0.2083	0.0027
kapital.finansowy->kapital.szczescia	0.2742	0.2772	0.2708	0.0064
kapital.zdrowia->kapital.rozrywki	0.2724	0.2623	0.2824	0.0201
kapital.zdrowia->kapital.szczescia	0.2273	0.2219	0.2330	0.0112
kapital.duchowy->kapital.szczescia	0.0931	0.0973	0.0886	0.0088
kapital.rozrywki->kapital.szczescia	0.1296	0.1333	0.1259	0.0074
kapital.spoleczny->kapital.szczescia	0.0125	0.0094	0.0177	0.0082

	t.stat	deg.fr	p.value	sig.05
kapital.finansowy->kapital.zdrowia	0.3312	17249	0.3702	no
kapital.finansowy->kapital.rozrywki	0.0706	17249	0.4719	no
kapital.finansowy->kapital.szczescia	0.3867	17249	0.3495	no
kapital.zdrowia->kapital.rozrywki	1.4949	17249	0.0675	no
kapital.zdrowia->kapital.szczescia	0.9033	17249	0.1832	no
kapital.duchowy->kapital.szczescia	0.5605	17249	0.2876	no
kapital.rozrywki->kapital.szczescia	0.5628	17249	0.2868	no
kapital.spoleczny->kapital.szczescia	0.6534	17249	0.2568	no

```
Inner models in the following objects:
\ $global
\ $group1
\ $group2
```

Pierwsza kolumna zawiera współczynniki dla całej populacji, następne dla grupy mężczyzn i kobiet, dalej podano różnicę między współczynnikami, wartość statystyki testowej t, liczbę stopni swobody, p-value testu oraz czy odrzucić hipotezę zerową, która mówi, że różnica między współczynnikami jest statystycznie nieistotna na poziomie istotności 5%. Powyższy test również nie wskazał współczynników, których wartości są istotnie różne na poziomie istotności 5%.

### 3.8. Podsumowanie

Niniejszy rozdział zawiera analizę danych rzeczywistych przy wykorzystaniu metody modelowania ścieżek przy pomocy częściowych najmniejszych kwadratów. Polegała ona na wyestymowaniu poziomu szczęścia Polaków oraz zbadania jego determinantów. W tym celu przeanalizowano 17251 obserwacji zebranych w roku 2011 na potrzeby badania „Diagnoza Społeczna” [8]. Konstrukcja modelu opiera się na teorii potrzeb ludzkich Maslow’a [1]. Weryfikacja adekwatności doboru zmiennych została przeprowadzona zgodnie ze wskazówkami zawartymi w czwartym oraz piątym rozdziale podręcznika [4]. Obliczenia zostały wykonane w środowisku  $\mathcal{R}$  przy wykorzystaniu pakietu `plspm` opisanego w drugim rozdziale. Estymacja współczynników modelu pozwoliła na wyliczenie wartości szczęścia dla każdej z obserwacji. Zbadano również poziom szczęścia w podziale na płeć. Test Wilcoxon’a nie wykazał różnic w rozkładzie szczęścia u kobiet i u mężczyzn na poziomie istotności 5%. Także porównanie współczynników w modelach wewnętrznych przy pomocy metod permutacyjnej oraz testu  $t$ , opisanych w pierwszym rozdziale, nie wykazało różnic między determinantami szczęścia obu płci.

## Rozdział 4

# Zakończenie

W pracy omówiono podstawowe informacje na temat modelowania ścieżek przy pomocy częściowych najmniejszych kwadratów.

Rozdział pierwszy zawiera opis podstawowych pojęć związanych z omawianą metodą modelowania. Zostały przedstawione własności zarówno modelu wewnętrznego, jak i modeli zewnętrznych. Następnie zaprezentowano algorytm służący do estymacji współczynników modelu. Druga część pierwszego rozdziału demonstruje dwa sposoby porównywania grup. Pierwszy z nich to zastosowanie testu  $t$ , a drugi – testu permutacyjnego.

W rozdziale drugim przedstawiono najważniejsze funkcje pakietu `plspm`, który służy do estymacji współczynników modeli ścieżek przy pomocy częściowych najmniejszych kwadratów w środowisku  $\mathcal{R}$ . Omówiono funkcję `plspm()`, która służy do wyliczenia parametrów modelu oraz funkcję `plspm.groups()` służącą do porównywania współczynników w dwóch modelach o tej samej strukturze. Ostatni paragraf tego rozdziału poświęcono na opis dwóch funkcji graficznych, które pozwalają na przedstawienie wartości współczynników na diagramie ukazującym strukturę modelu.

Rozdział trzeci zawiera analizę danych rzeczywistych[12] pozyskanych w 2011 roku na potrzeby badania „Diagnoza społeczna. Warunki i jakość życia Polaków.”[8], którego celem było oszacowanie poziomu szczęścia rozumianego jako zadowolenie z życia oraz zweryfikowanie hipotezy mówiącej że poziom szczęścia kobiet i mężczyzn w Polsce jest taki sam. W celu wyliczenia tej wartości nakreślono strukturę modelu na podstawie teorii potrzeb ludzkich Maslow’a[1]. Następnie zbadano znak korelacji zmiennych oraz dokonano estymacji modelu przy użyciu funkcji `plspm()`. Po oszacowaniu parametrów i diagnostyce modelu dokonano analizy grup, gdzie kryterium podziału była płeć. Zastosowanie testu Wilcoxon’a pozwoliło stwierdzić, że rozkład poziomu szczęścia kobiet i mężczyzn nie różni się w sposób statystycznie istotny. Następnie, zastosowanie testu permutacyjnego oraz testu  $t$  nie wykazało istotnych różnic pomiędzy współczynnikami modelu wewnętrznego estymowanego osobno dla mężczyzn i kobiet.



## Dodatek A

# Pytania oraz wyniki ankiety

nazwa	pytanie w ankiecie	odpowiedź
<i>f6</i>	W ostatnich miesiącach czuł Pan/Pani, że Pana/Pani źródło dochodów jest niestale i niepewne.	1– często, ..., 3– nigdy
<i>dochod</i>	Jaki jest Pana/Pani miesięczny dochód netto?	wielkość dochodu netto
<i>f7</i>	W jakim stopniu zadowolony jest Pan/Pani z sytuacji finansowej w swojej rodzinie?	1– bardzo niezadowolona, ..., 6– bardzo zadowolona

Tabela A.1: Opis zmiennych mierzalnych z bloku 2.

nazwa	wartość	liczba wystąpień	częstość wystąpień
<i>f6</i>	1	2064	.120
	2	6871	.398
	3	8316	.482
<i>f7</i>	1	1353	.078
	2	2611	.151
	3	2754	.160
	4	5718	.331
	5	4162	.241
	6	653	.038

Tabela A.2: Wyniki ankiety dotyczące bloku 2. Zmienne jakościowe.

nazwa	min	kwantyl 0.25	mediana	kwantyl 0.75	max
<i>dochod</i>	0	900	1352	2000	40000

Tabela A.3: Wyniki ankiety dotyczące bloku 2. Zmienna ilościowa.

nazwa	pytanie w ankiecie	odpowiedź
s1	Jak ocenia Pan/Pani swoje całe dotychczasowe życie, czy mógłby Pan/Pani powiedzieć, że było?	1– okropne, ..., 7– wspaniałe
s2	Jak silne w tych dniach jest Pana/Pani pragnienie życia?	1– w ogóle nie chce mi się żyć, ..., 10– bardzo mocno chce mi się żyć
s3	Czy czuje się Pan/Pani osamotniony, mimo że tego nie chce?	1– tak, 2– nie
s4	Czy czuje się Pan/Pani kochana i darzona zaufaniem?	1– nie, 2– tak
s5	Jak często w minionych miesiącach zdarzało się Panu/Pani być tak załamanym, że myślał Pan/Pani o samobójstwie?	1– bardzo często, ..., 4– nigdy

Tabela A.4: Opis zmiennych mierzalnych z bloku 1.

nazwa	wartość	liczba wystąpień	częstość wystąpień
s1	1	43	.002
	2	134	.008
	3	864	.050
	4	2890	.168
	5	6164	.357
	6	6634	.385
	7	522	.030
s2	1	55	.003
	2	50	.003
	3	170	.010
	4	272	.016
	5	902	.052
	6	1052	.061
	7	1747	.101
	8	3058	.177
	9	2968	.172
	10	6977	.404
s3	1	3553	.206
	2	13698	.794
s4	1	1350	.078
	2	15901	.922
s5	1	0	.000
	2	3691	.021
	3	1465	.085
	4	15417	.894

Tabela A.5: Wyniki ankiety dotyczące bloku 1.

nazwa	pytanie w ankiecie	odpowiedź
z8	W ostatnich miesiącach odczuwał Pan/Pani dolegliwości fizyczne, takie jak łamanie w kościach, trudności z oddychaniem itp., które utrudniały Panu/Pani wchodzenie do domu, chodzenie po schodach itp.	1– często, ..., 3– nigdy
z9	W ostatnich miesiącach problemy ze zdrowiem utrudniały Panu/Pani wykonywanie codziennych zajęć lub branie udziału w innych zajęciach.	1– często, ..., 3– nigdy
z10	W jakim stopniu zadowolony jest Pan/Pani ze stanu swojego zdrowia?	1– bardzo niezadowolona, ..., 6– bardzo zadowolona

Tabela A.6: Opis zmiennych mierzalnych z bloku 3.

nazwa	wartość	liczba wystąpień	częstość wystąpień
z8	1	3281	.190
	2	8338	.483
	3	5632	.326
z9	1	2523	.146
	2	8073	.468
	3	6655	.386
z10	1	940	.054
	2	2060	.119
	3	2325	.135
	4	5469	.317
	5	5332	.309
	6	1125	.065

Tabela A.7: Wyniki ankiety dotyczące bloku 3.

nazwa	pytanie w ankiecie	odpowiedź
<i>d11</i>	Co według Pana/Pani jest ważniejsze w życiu?	1– przyjemność, dostatek, brak stresu, 2–poczucie sensu, osiągnięcie ważnych celów, mimo trudności, bólu i wyrzeczeń
<i>d12</i>	Jak często przeciętnie w ciągu miesiąca bierze Pan/Pani udział w nabożeństwach lub innych spotkaniach o charakterze religijnym?	liczba nabożeństw

Tabela A.8: Opis zmiennych mierzalnych z bloku 4.

nazwa	wartość	liczba wystąpień	częstość wystąpień
<i>d11</i>	1	7389	.428
	2	9862	.572

Tabela A.9: Wyniki ankiety dotyczące bloku 4. Zmienna jakościowa.

nazwa	min	kwantyl 0.25	mediana	kwantyl 0.75	max
<i>d12</i>	0	0	3	4	31

Tabela A.10: Wyniki ankiety dotyczące bloku 4. Zmienna ilościowa.



nazwa	pytanie w ankiecie	odpowiedź
<i>r13</i>	Czy uprawia Pan/Pani aktywnie którąś z form sportu, ćwiczeń fizycznych?	1– nie, 2– tak
<i>r14</i>	Ile osób zalicza Pan/Pani do grona swoich przyjaciół?	liczba przyjaciół
<i>r15</i>	Ile razy w minionym miesiącu był Pan/Pani w kinie, teatrze lub na koncercie?	liczba wizyt
<i>r16</i>	Ile razy w minionym miesiącu był Pan/Pani w restauracji, kawiarni, pubie?	liczba wizyt
<i>r17</i>	Ile razy w minionym miesiącu był Pan/Pani na spotkaniu towarzyskim?	liczba wizyt
<i>r18</i>	W jakim stopniu zadowolony jest Pan/Pani ze stosunków z grupą przyjaciół?	1– bardzo niezadowolona, ..., 6– bardzo zadowolona

Tabela A.11: Opis zmiennych mierzalnych z bloku 5.

nazwa	wartość	liczba wystąpień	częstość wystąpień
<i>r13</i>	1	11821	.685
	2	5430	.315
<i>r18</i>	1	137	.008
	2	483	.028
	3	1000	.058
	4	6089	.353
	5	8091	.470
	6	1451	.084

Tabela A.12: Wyniki ankiety dotyczące bloku 5. Zmienne jakościowe.

nazwa	min	kwantyl 0.25	mediana	kwantyl 0.75	max
<i>r14</i>	0	3	5	10	99
<i>r15</i>	0	0	0	0	30
<i>r16</i>	0	0	0	1	30
<i>r17</i>	0	0	1	2	30

Tabela A.13: Wyniki ankiety dotyczące bloku 5. Zmienne ilościowa.

nazwa	pytanie w ankiecie	odpowiedź
<i>sp19</i>	Czy w ciągu ostatnich lat zdarzało się, że angażował się Pan/Pani w działania na rzecz społeczności lokalnej?	1– nie, 2– tak
<i>sp20</i>	Czy w ostatnim roku był Pan/Pani na jakimś zebraniu publicznym?	1–nie, 2– tak
<i>sp21</i>	Czy w ostatnim roku wykonywał Pan/Pani nieodpłatnie jakąś pracę lub świadczyła jakieś usługi dla osób spoza rodziny bądź na rzecz organizacji społecznej?	1– nie, 2– tak
<i>sp22</i>	czy jest Pan/Pani członkiem organizacji, stowarzyszeń, partii, komitetów, rad, grup religijnych, związków lub kół?	0– nie, 1– tak, jednej, 2– tak, dwóch, 3–tak, więcej niż dwóch

Tabela A.14: Opis zmiennych mierzalnych z bloku 6.

nazwa	wartość	liczba wystąpień	częstość wystąpień
<i>sp19</i>	1	14348	.832
	2	2903	.168
<i>sp20</i>	1	12856	.745
	2	4395	.255
<i>sp21</i>	1	13760	.798
	2	3491	.202
<i>sp19</i>	0	14526	.842
	1	2157	.125
	2	409	.024
	3	159	.009

Tabela A.15: Wyniki ankiety dotyczące bloku 6.

# Bibliografia

- [1] A. H. Maslow (1943): *A theory of human motivation*, Psychological Review, 50(4)
- [2] Encyklopedia PWN <http://www.encyklopedia.pwn.pl> [20.05.2013]
- [3] Fundacja Naukowa SmarterPoland.pl <http://www.smarterpoland.pl> [15.04.2013]
- [4] G. Sanchez: *PLS Path Modeling with  $\mathcal{R}$* , <http://www.gastonsanchez.com> [10.02.2013]
- [5] G. Sanchez, L. Trinchera, *Tools for Partial Least Squares Path Modeling*, <http://cran.r-project.org> [22.05.2013]
- [6] H. R. Varian (2012): *Mikroekonomia. Kurs średni– ujęcie nowoczesne*
- [7] H. Wold (1980): *Theory and Application of Partial Least Squares* <http://www.nber.org> [25.07.2013]
- [8] J. Czapiński, T. Panek (2011): *Diagnoza społeczna 2011*. <http://www.diagnoza.com> [20.05.2013]
- [9] Ł. Komsta: *Wprowadzenie do środowiska  $\mathcal{R}$* , <http://cran.r-project.org> [15.03.2013]
- [10] N. L. Johnson, S. Kotz (1997): *Leading Personalities in Statistical Sciences from the Seventeenth Century to the Present*
- [11] Rada Monitoringu Społecznego (2011): *Diagnoza społeczna, kwestionariusze 2011*. <http://www.diagnoza.com> [21.05.2013]
- [12] Rada Monitoringu Społecznego (2011): *Diagnoza społeczna: zintegrowana baza danych*. <http://www.diagnoza.com> [20.04.2013]
- [13] V. E. Vinzi, W. W. Chin, J. Henseler, H. Wang (2010): *Handbook of Partial Least Squares Concepts, Methods and Applications* <http://www.springer.com> [10.07.2013]
- [14] *Ważniak* <http://wazniak.mimuw.edu.pl> [26.07.2013]
- [15] W. Niemiński (2011): *Statystyka* [10.10.2013]