

Analiza Czynnikowa

1. Zmienne wejściowe

Do analizy będzie wykorzystany zbiór danych zawierający **29537 obserwacji** oraz **21 zmiennych**.

Do **zmiennych** należą:

- Q1 Nie potrzebuję cudzych pochwał.
- Q2 Wolałbym lekcję matematyki niż lekcję garncarstwa.
- Q3 Nigdy nie spóźniam się.
- Q4 Nie zawracam sobie głowy czytaniem instrukcji przed złożeniem czegoś.
- Q5 Nie potrafiłbym żyć w nieładzie.
- Q6 Lubię losowość w moim życiu.
- Q7 Lubię zabawy słowne.
- Q8 Kieruję się priorytetami ekonomicznymi.
- Q9 Mam dużo nowych pomysłów.
- Q10 Łatwo się nie stresuję.
- Q11 Rzadko płaczę podczas smutnych filmów.
- Q12 Planuję swoje życie logicznie.
- Q13 Lubię kreatywność.
- Q14 Podejmuję decyzje na podstawie faktów, nie emocji.
- Q15 Tworzę chaos.
- Q16 Łatwo się denerwuję.
- Q17 Jestem romantykiem.
- Q18 Cenię logikę ponad wszystko.
- Q19 Często zapominam odstawić rzeczy na miejsce.
- Q20 Pozostaję spokojny w sytuacjach stresowych.

country Kod kraju (ISO country code)

1.1 Filtrowanie zbioru danych

Usuniemy kraje które posiadają naprawdę znikome liczby obserwacji w celu zwiększenia rzetelności wyników.

Na podstawie kwartyli widać, że 75% krajów nie przekracza 55 obserwacji dlatego próg min. 100 obserwacji uznałem za zadowalający (tym bardziej, że Polska posiada 114 obserwacje a jej wyników analizy jestem ciekawy).

```
> quantile(table(zachowania$country))
      0%      25%      50%      75%     100%
      1.00      2.75     13.50     55.25    16827.00
```

Po filtracji danych do analizy zostaje nadal **26576 obserwacji** rozdzielonych na **23 kraje**.

```
> table(zachowania$country)
 AE   AP   AU   BR   CA   DE   GB   GR   ID   IE   IN   MX   MY   NL   NZ   PH   PK   PL   RO   SE   SG   US   ZA
141  107  852  137 1387  189 1481  143  430  117 1566  100  620  162  197  808  105  114  164  134  281 16827  514
```

2. Sprawdzanie przesłanek ku przeprowadzeniu analizy czynnikowej

2.1 Test sferyczności **Morisa Bartletta**.

Test ma na celu zbadanie macierzy korelacji zmiennych wejściowych.

Hipoteza 0: macierz korelacji zmiennych wejściowych jest macierzą jednostkową

```
> cortest.bartlett(zachowania[, -21])
R was not square, finding R from data
$chisq
[1] 83391.77

$ p.value
[1] 0

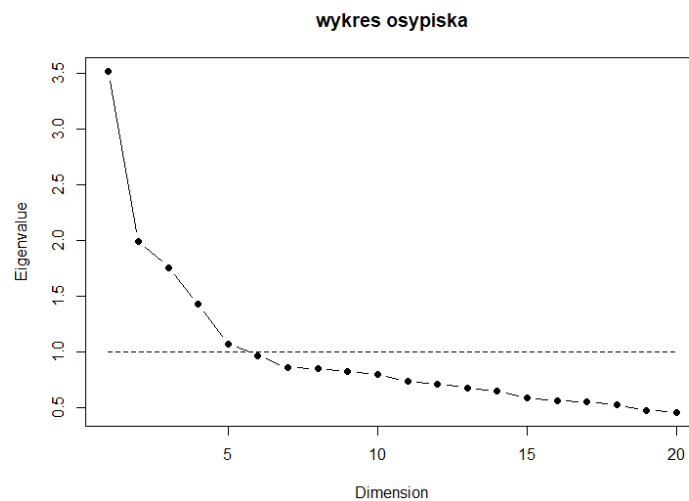
$df
[1] 190
```

p-value = 0 dlatego dla każdego poziomu istotności odrzucamy hipotezę zerową, a więc macierz korelacji nie jest macierzą jednostkową dlatego model czynnikowy jest odpowiedni dla analizowanych zmiennych.

2.2 Test **Kaisera-Meyera-Olkina**

Test jest miarą dopasowania danych do analizy czynnikowej. Współczynnik porównuje korelacje cząstkowe z dwuzmiennowymi współczynnikami korelacji.

```
Overall MSA = 0.79
MSA for each item =
  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10 Q11 Q12 Q13 Q14 Q15 Q16 Q17 Q18 Q19 Q20
0.85 0.84 0.88 0.81 0.77 0.83 0.66 0.86 0.63 0.82 0.82 0.85 0.68 0.82 0.79 0.73 0.73 0.80 0.81 0.77
```

[illegible]

Ponieważ wartości **Cumulative Var** są zbyt niskie (dokonując redukcji tracimy 63% wariancji zmiennych wejściowych) oraz **Mean item complexity** jest wysoki, w celu poprawy wyników wyrzucę zmienną **Q11** oraz **Q8** które posiadają znacznie za duże **com** (jest to 2.5 oraz 3.2 oznacza to, że są znacznie skorelowane z więcej niż dwoma czynnikami).

Po zredukowaniu zmiennych **Mean item complexity** jest bliższy 1, lecz **Cumulative Var** nadal jest na niskim poziomie. Jednak najważniejsze jest to, że żadna ze zmiennych nie posiada wartości **com** znacznie większych od 2.

4.2 Wyniki końcowe

	WLS1	WLS2	WLS4	WLS3	WLS5	h2	u2	com
Q1	0.13	0.33	0.09	-0.09	0.15	0.17	0.83	2.1
Q2	0.41	0.13	0.07	-0.35	0.00	0.31	0.69	2.2
Q3	0.19	0.03	0.33	-0.07	-0.06	0.16	0.84	1.8
Q4	-0.06	0.06	-0.12	0.01	0.33	0.13	0.87	1.4
Q5	0.08	-0.02	0.74	0.01	-0.05	0.55	0.45	1.0
Q6	-0.15	-0.04	-0.19	0.11	0.50	0.33	0.67	1.6
Q7	0.05	0.00	-0.03	0.41	-0.14	0.19	0.81	1.3
Q9	0.04	0.22	-0.02	0.56	0.17	0.39	0.61	1.5
Q10	0.18	0.55	0.02	0.01	0.07	0.34	0.66	1.3
Q12	0.62	0.08	0.30	0.02	-0.17	0.51	0.49	1.7
Q13	-0.12	-0.03	-0.06	0.69	0.14	0.51	0.49	1.2
Q14	0.69	0.26	0.06	-0.05	-0.12	0.56	0.44	1.4
Q15	-0.05	-0.23	-0.60	0.03	0.42	0.59	0.41	2.2
Q16	0.04	-0.72	-0.06	0.05	0.19	0.57	0.43	1.2
Q17	-0.15	-0.14	0.04	0.30	0.03	0.14	0.86	2.0
Q18	0.74	0.10	0.07	-0.04	-0.07	0.56	0.44	1.1
Q19	-0.05	-0.09	-0.57	0.02	0.34	0.45	0.55	1.7
Q20	0.15	0.65	0.02	0.10	-0.01	0.46	0.54	1.2

	WLS1	WLS2	WLS4	WLS3	WLS5
ss loadings	1.76	1.60	1.52	1.21	0.84
Proportion Var	0.10	0.09	0.08	0.07	0.05
Cumulative Var	0.10	0.19	0.27	0.34	0.38
Proportion Explained	0.25	0.23	0.22	0.17	0.12
Cumulative Proportion	0.25	0.49	0.70	0.88	1.00

Mean item complexity = 1.5

- **h2** - zasób zmienności wspólnej pożądane w niej są wysokie wartości.
- **u2** - zasób zmienności swoistej pozostałość po odjęciu zasobu zmienności wspólnej od wariancji całkowitej
- **com** - wartość wskaźnika złożoności, w naszej analizie zmienne średnio posiadają wartość 1.5 (**Mean item complexity** jest na poziomie 1.5 oznacza to, że nie są skorelowane znacznie z więcej niż 2 czynnikami).
- **Cumulative Var** – w naszej analizie przyjmuje niezadowalające wartości (dokonując analizy czynnikowej tracimy 63% zmienności), lecz można to tłumaczyć tym, że dane były zbierane poprzez ankiety internetowe.

Czynniki:

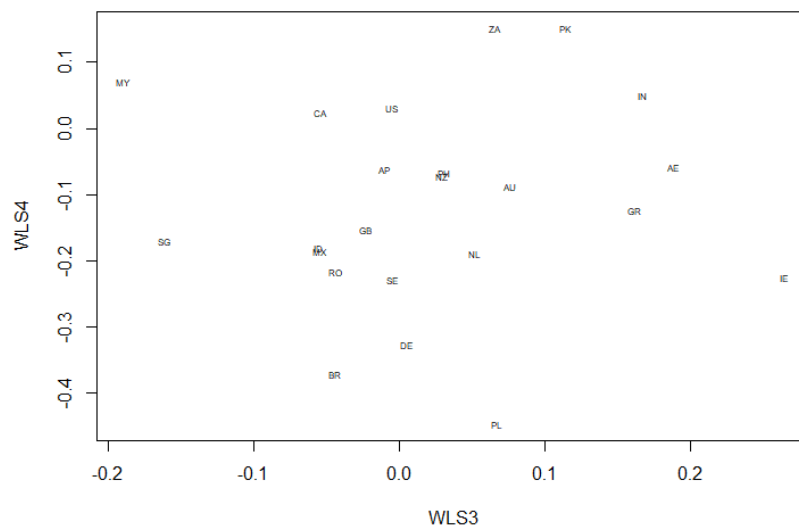
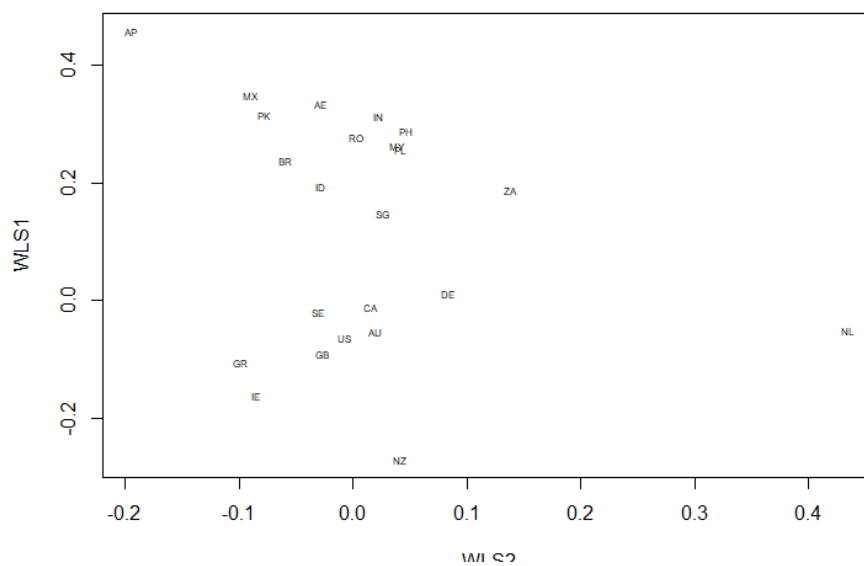
WSL1 – „Ścisłowcy” ludzie którzy cenią logiczne myślenie, podejmują decyzje na podstawie faktów oraz przeważnie lubią matematykę.

WSL2 – „Stoicy” ludzie spokojni, potrafiący zachować zimną krew w sytuacjach stresowych oraz nie potrzebujący cudzych pochwał aby zrozumieć swoją wartość.

WSL4 – „Skrupulatni” lubią porządek, nie tolerują chaosu i nieładu.

WSL3 – „Kreatywni” ludzie o kreatywnym myśleniu, lubiący wykorzystywanie swojej błyskotliwości w grach.

WSL5 – „Chaotyczni” lubiący wydarzenia losowe oraz radzący sobie z problemami swoimi metodami.



Po agregacji danych oraz obliczeniu wartości średnich dla każdego kraju tworzymy wykresy pokazujące jak osobowości kształtują się w danych krajach.

Możemy wyczytać, że np. w Polsce występuje raczej większy odsetek osób „Skrupulatnych” niż „Kreatywnych”