

# Analiza wariacji

- Podział zaobserwowanej zmienności (wariancji) na zmienność między grupami i w obrębie grup

- Pozwala na ocenę istotności różnic wielu średnich, hipoteza zerowa:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_t$$

- **Założenia:**

- Zmienne objaśniające są niezależne – losowy dobór do grup
- Zmienna zależna ma rozkład normalny w populacji
  - dopuszczalne są niewielkie odstępstwa
  - często badane są wyłącznie reszty (czynnik losowy)
- Wariancje w grupach są jednorodne (homogeniczność wariancji)

# Analiza wariacji

- Model liniowy analizy wariacji:

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

- Gdzie:
  - $y_{ij}$  – j-ta obserwacja z i-tej grupy
  - $\mu$  – średnia wartość cechy w populacji
  - $\alpha_i$  – efekt i-tej grupy
  - $e_{ij}$  – błąd czyli efekt związany ze zmiennością osobniczą, przypadkową, niewyjaśnioną modelem, może być również błędem pomiaru

# Analiza wariancji - założenia

- Testowanie jednorodności wariancji – test Test Leven'a
  - Istotny statystycznie wynik wskazuje na różnice w zmienności pomiędzy grupami



Założenia do analizy wariancji nie są spełnione

SPSS: Analiza → porównywanie średnich →  
Jednoczynnikowa ANOVA → opcje → jednorodność  
wariancji

# Brak jednorodności wariancji

- Co możemy zrobić:
  - Sprawdzamy czy nie ma obserwacji odstających w grupach
  - Dokonać transformacji zmiennej zależnej np. logarytmicznej
  - Wykluczyć grupę, w której wariancja różni się od pozostałych
  - Zastosować test nieparametryczny
  - Korekta Welcha – korekta liczby stopni swobody w teście t-studenta dla  $n$  prób niezależnych

SPSS: Analiza → porównywanie średnich →  
Jednoczynnikowa ANOVA → opcje → Welch

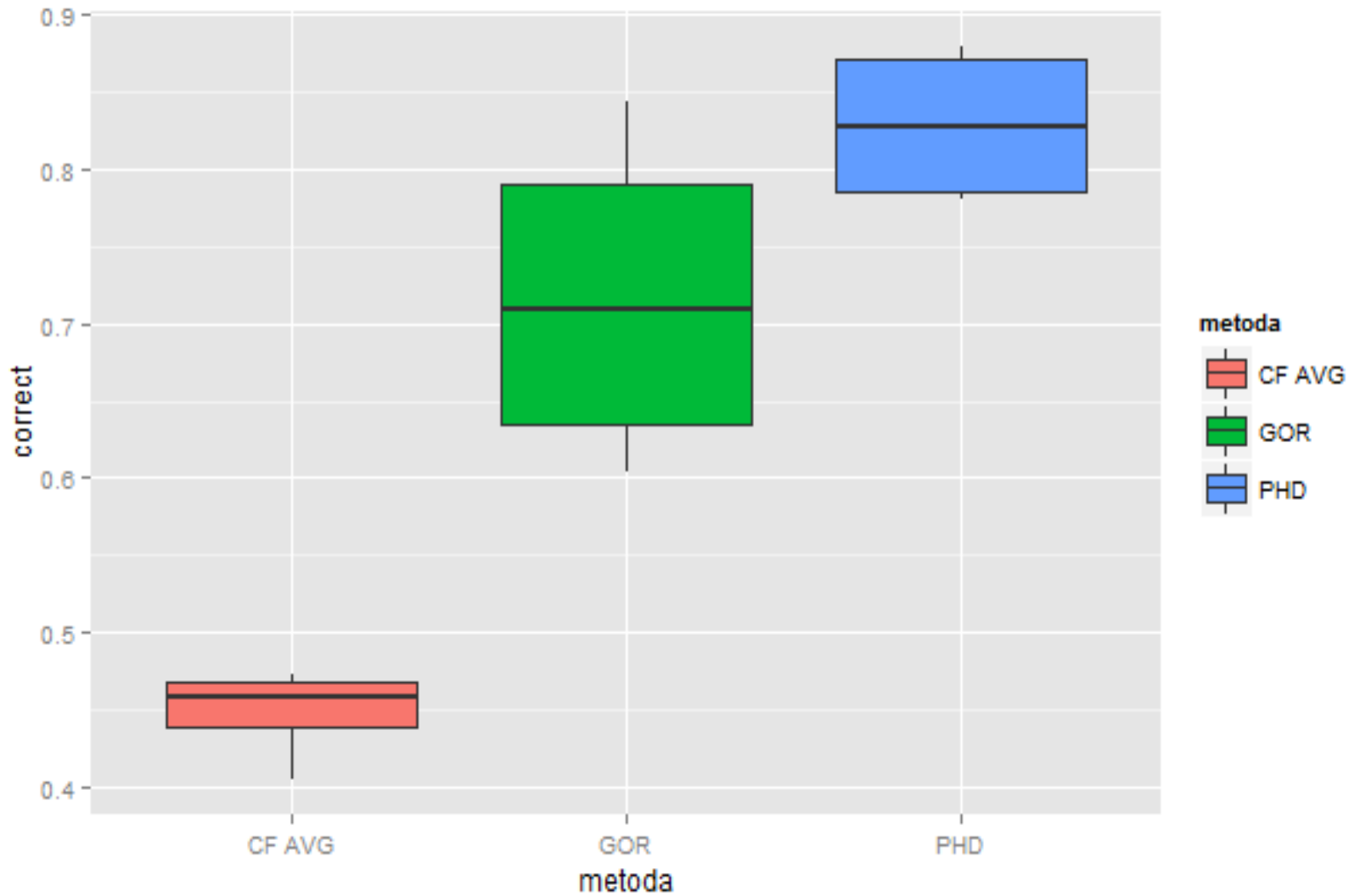
# Analiza wariacji

- Czy metoda wyznaczania struktury drugorzędowej białka ma wpływ na dokładność?

Białko	Metoda	Dokładność
Ubikwityna	CF AVG	0.467
Ubikwityna	GOR	0.645
Ubikwityna	PHD	0.868
DeoxyHb	CF AVG	0.472
DeoxyHb	GOR	0.844
DeoxyHb	PHD	0.879
Rab5c	CF AVG	0.405
Rab5c	GOR	0.604
Rab5c	PHD	0.787
Prealbumina	CF AVG	0.449
Prealbumina	GOR	0.772
Prealbumina	PHD	0.780

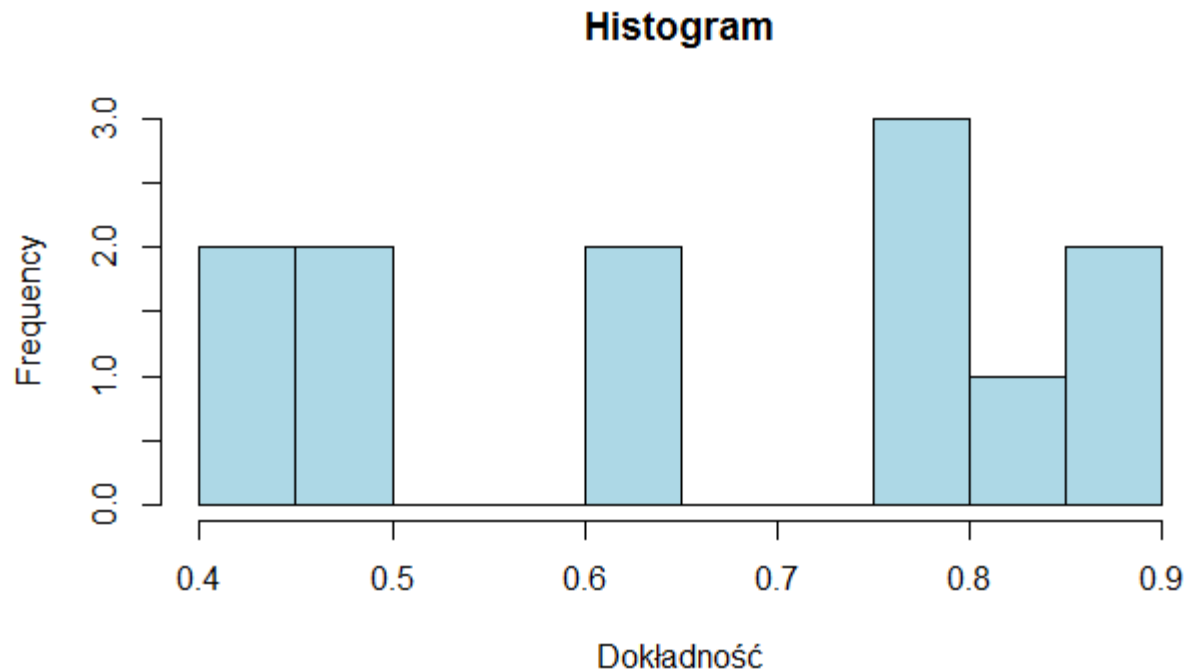
Przykład opisany szczegółowo w książce:  
Seefeld K., Linder E. 2007.  
*Statistics Using R with Biological Examples*

# Wykres skrzynkowy



# Analiza wariancji

## – Rozkład zmiennej zależnej



W praktyce sprawdzamy nie samą zmienną ale rozkład standaryzowanych reszt modelu

# Testy Post-Hoc

- Testy post-hoc wykonujemy, **kiedy różnice pomiędzy grupami są istotne**. Najczęściej stosowane testy:
  - Test Tukeya (UIR)
  - LSD Fishera (NIR)

Inne:

- Test Sheffego
- Test Duncana
- Korekta na wielokrotne testowanie np. Bonferroniego



# Testy Post-Hoc

- Test Tukeya
  - UIR – test Uczciwie Istotnych Różnic
  - HSD – Honestly Significant Differences

Powinien być stosowany jedynie dla **zrównoważonego układu doświadczenia** – podobna liczba obserwacji we wszystkich grupach – w układzie niezrównoważonym brak kontroli błędu pierwszego rodzaju

# Testy Post-Hoc

- LSD Fishera
  - NIR - Najmniejsza Istotna Różnica
  - LSD – Least Squared Difference

## **Nie zakłada się równoliczności grup**

- Polega na wykonaniu  $k(k-1)/2$  testów t-studenta i zastosowaniu korekty na liczbę przeprowadzonych testów

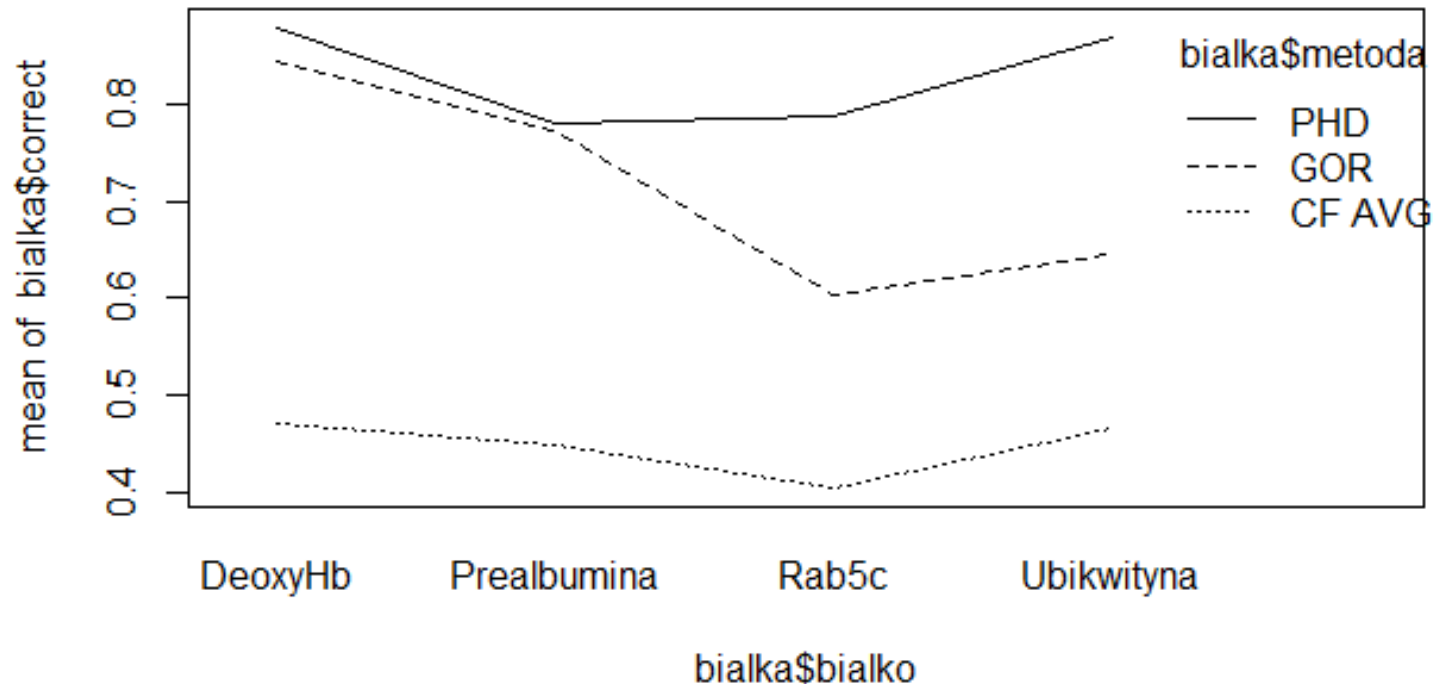
# Jednoczynnikowa vs wieloczynnikowa ANOVA

- Istnieje możliwość przeprowadzenia jednoczynnikowej analizy wariancji oddzielnie dla wszystkich zmiennych objaśniających
- Wady takiego postępowania:
  - Utrata informacji o zależnościach między zmiennymi objaśniającymi
  - Większa wariancja – trudniej stwierdzić istotność niektórych zmiennych
- Zmiennych objaśniających nie powinno być zbyt dużo:
  - Wraz ze wzrostem liczby zmiennych maleje dokładność oceny efektów modelu
  - Idealna sytuacja: min 30 obserwacji na każdą kombinację czynników
- Jednoczynnikowa ANOVA jest dobrym rozwiązaniem do oszacowania efektów jeżeli występuje istotna interakcja

# Interakcje

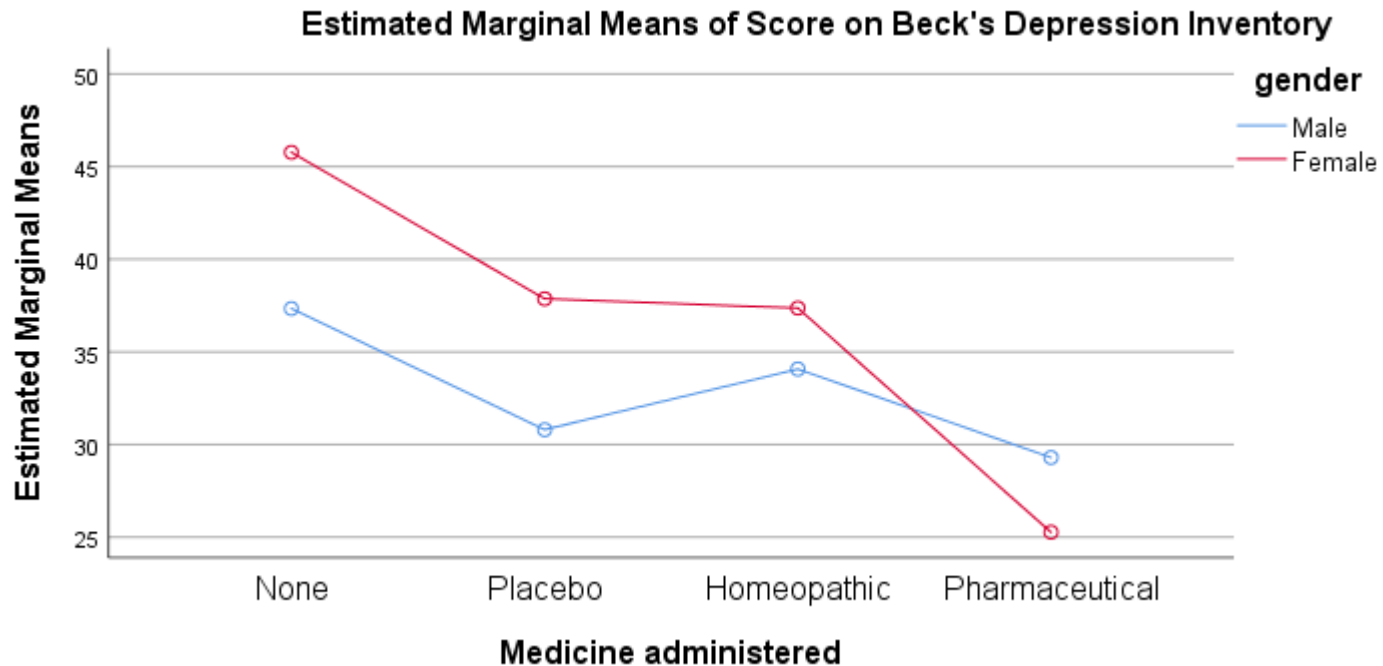
- O interakcjach mówimy jeżeli współdziałanie czynników ze sobą przynosi efekt odmienny niż każdego z osobna
- Wpływ czynników nie musi być istotny statystycznie żeby istotny był wpływ interakcji
- UWAGA: Dodanie interakcji zmienia interpretacje głównych czynników. W przypadku wykrycia statystycznie istotnej interakcji należy powtórzyć analizę testując każdy z czynników osobno

# Jak stwierdzić że występuje interakcja?



Interakcja nie występuje. Linie nie przecinają się.  
Uwaga - W tym przypadku brakuje nam danych –  
widzimy że metoda pomiaru może mieć znaczenie dla  
prealbuminy – linie nie są równoległe

# Jak stwierdzić że występuje interakcja?

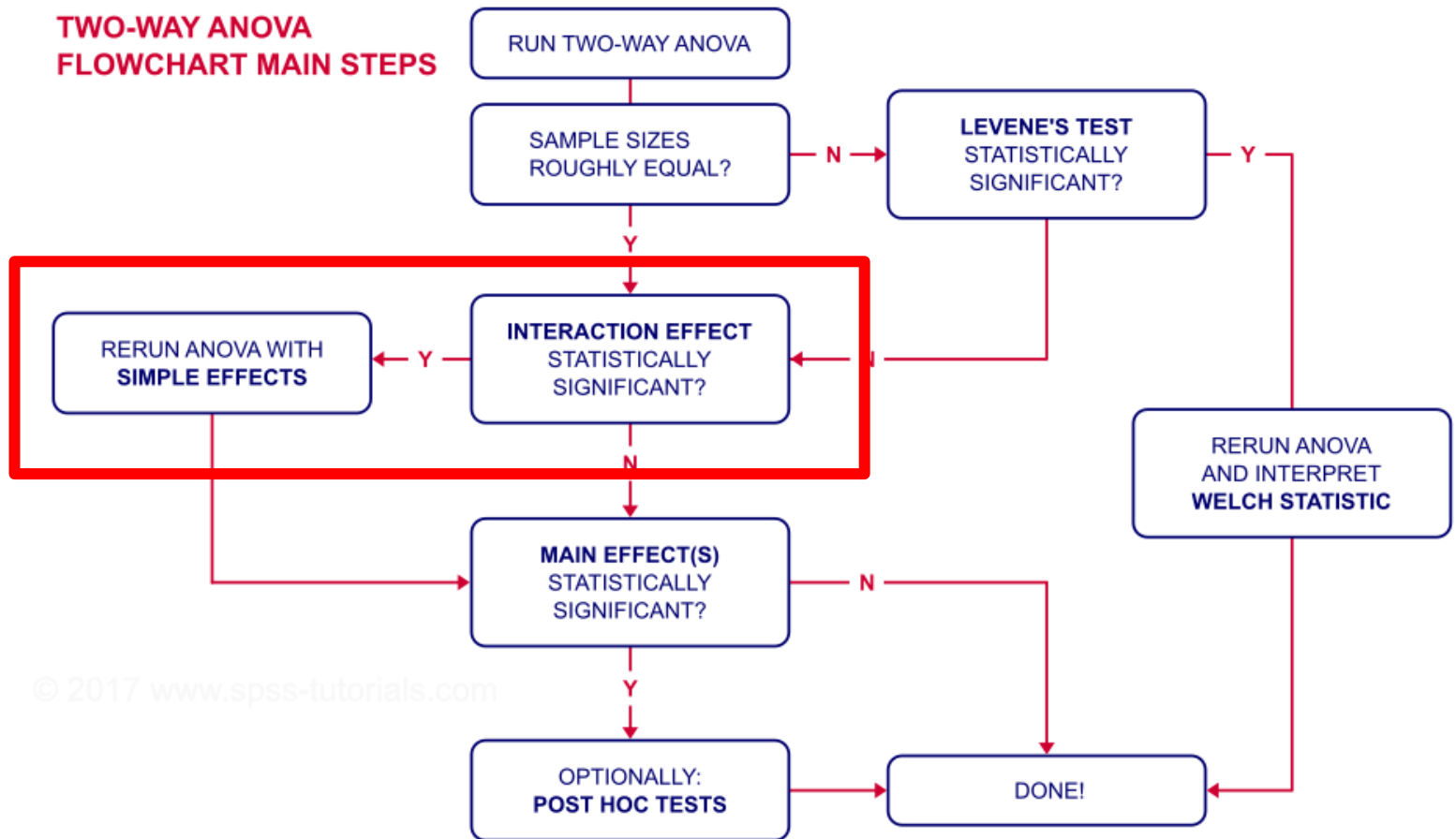


Źródło: <https://www.spss-tutorials.com/spss-two-way-anova-interaction-significant/>

Interakacja występuje. Co to znaczy?

# Model dwuczynnikowy

## TWO-WAY ANOVA FLOWCHART MAIN STEPS



© 2017 www.spss-tutorials.com