1 Wstęp

1.1 Czym są efekty losowe?

Jednokierunkowa ANOVA
Na poprzednich zajęciach mówiliśmy o modelach liniowych, o jedno- i dwuczynnikowej analizie wariancji. W tych modelach estymowaliśmy nieznane wartości stałych parametrów, W ANOVIE rolę parametrów pełniły efekty różnych poziomów czynnika. W tym przypadku obserwacje zmiennej odpowiedzi możemy zapisać jako: $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$
$i = 1, 2, \ldots, k,$
$j = 1, 2, \ldots, n$
Y_{ij} - j-ta obserwacja na i-tym poziomie czynnika
μ - ogólna wartość średnia
α_i - efekt i-tego poziomu czynnika
ε_{ij} - niezależne zmienne losowe $\sim N(0, \sigma^2)$
Obserwacje Y_{ij}, $j = 1, 2, \ldots, n$ dla każdej ustalonej wartości wskaźnika i tworzą **grupę**.
Efekty α_i są stałymi o nieznanych wartościach, zakłądaliśmy przy tym, że $\sum_{i=1}^{k} \alpha_i = 0$. Bez takiego założenia wielkościμ i α_i nie byłyby określone jednoznacznie.
Dane zgrupowane:

- każda obserwacja należy do jednej grupy i jest tylko jeden czynnik grupujący
- hierarchiczne, zagnieżdżone zbiory danych - bardziej skomplikowane przypadki
- nie możemy zakładać niezależności obserwacji, obserwacje są skorelowane w ramach grup
- efekty losowe - wygodne do modelowania tego typu danych
  Użycie efektów losowych jest popularnym sposobem używanym do modelowania takich danych.

1. **Efekt stały** - nieznana liczba, którą estymujemy z danych
   Są często używane w modelach liniowych i uogólnionych modelach liniowych
2. **Efekt losowy** - zmienna losowa, estymujemy parametry opisujące rozkład efektu losowego

1.2 Przykłady

1. **Z astronomii:**
   1861 r. - astronom Airy sformułował model $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ z losowymi efektami α_i do opisu obserwowanych przez teleskop obiektów astronomicznych.
   Obserwacje Y_{ij} opisywane są mianowicie jako suma parametru stałego μ, określającego średnią wartość obserwacji, losowego efektu i-tej nocy α_i, spowodowanego zmiennyim warunkami atmosferyczymi lub innymi czynnikami losowymi charakterstycznymi (i stałymi) dla danej nocy, oraz błedu losowego ε_{ij} obserwacji Y_{ij} spowodowanego czynnikami losowymi różnymi od czynników stałych dla danej nocy, składających się na efekt α_i.
   Jasne, że potraktowanie efektu czynnika nocy jako stałego, a nie losowego, byłoby bardzo grubym błędem.

2. **Z medycyny:**
   Zazwyczaj interesuje nas lecznie konkretnymi lekami i traktujemy efekty wywołane przez te leki jako stałe. Sensownie jest jednak uznać efekt pacjentów za losowy. Traktujemy pacjentów poddanych leczeniu, jako wybranych losowo z większego zbioru pacjentów, których cechy chcemy estymować. Zwykle nie interesuje nas wyłącznie ta wąska grupa pacjentów, ale cała populacja pacjentów.

Kiedy używać efektów losowych? Czasem w miarę jasne jest, że np. lepiej byłoby użyć efektów losowych, ale w niektórych przypadkach, może być kwestią dyskusji, czy efekty losowe czy stałe będą bardziej odpowiednie. Użycie efektów losowych jest ambitniejsze w takim sensie, że próbujemy powiedzieć coś o szerszej populacji a nie tylko naszej konkretnej próbce.

- gdy efektom danego czynnika nie można przypisać charakteru stałego, należy uznać je za losowe
- gdy chcemy uzyskać informacje o całej populacji, nasze obserwacje traktujemy jako próbkę z całej populacji
- niektórzy statystycy polecają zawsze używanie efektów losowych

1.3 Model mieszany

Z efektami, które rozsądnie jest przedstawić jako losowe, spotykamy sięw praktyce nierzadko, zwłąszcza wtedy, gdy mamy do czynienia ze zjawiskami bardziej złożonymi, niż zjawiska dające sięopisać za pomocą modelu jednokierunkowej klasyfikacji. Jak wiadomo, modele analizy wariancji możemy przedstawić jako modele liniowej analizy regresji z odpowiednio okreśłoną macierzą planu doświadczenia X.

- w modelu mieszanym pojawiają się efekty stałe i efekty losowe
- najprostszy przykład - **model dwukierunkowej klasyfikacji**:

$$y_{ijk} = \mu + \tau_i + \nu_j + \varepsilon_{ijk}$$

gdzie
μ, τ_i - efekty stałe
efekty losowe, i.i.d.: $\nu_j \sim N(0, \sigma_\nu^2)$
$\varepsilon_{ijk} \sim N(0, \sigma^2)$

Przykład - badania biologiczne:
Dla przykładu, w badanich biologicznych często bada sięjakąścehcę potomków zadanej populacji l matek. Powiedzmy, że wartość tej cechy rozsądnie jest uzależnić od płci potomka. Można wówczas zbadać trafność przyjęcia modelu dwukierunkowej klasyfikacji z efektami czynnika płci, jako efektami stałymi oraz efektem j-tej matki, $j = 1, 2, \ldots, l$, jako efektem Z efektami, które rozsądnie jest przedstawić jako losowe, spotykamy sięw praktyce nierzadko, zwłąszcza wtedy, gdy mamy do czynienia ze zjawiskami bardziej złożonymi, niż zjawiska dające sięopisać za pomocąmodelu jednokierunkowej klasyfikacji.
Y_{ijk} oznacza m-tą obserwację na i-tym poziomie czynnika płeć ($m = 2$) i j-tym, losowym poziomie czynnika matka, μ jest ogólną wartością średnią, τ_i jest (stałym) efektem t-tego poziomu czynnika płeć, ν_j jest efektem związanym z j-tą matką, n_{ij} jest liczbą potomków i-tej płci j-tej matki oraz ε_{ijk} są niezależnymi zmiennymi losowymi o rozkładzie normalnym $N(o, \sigma_\varepsilon^2)$. O efektach stałych zakładamy jak w ANOVIE, że sumują się do zera. Dodatkowo zakładamy zwykle, że efekty ν_j są nieskorelowane między sobą i z błedami ε_{ijk}, posiadają zerową wartość oczekiwaną i wspólną wariancją $\sigma_\nu^2 \geqslant 0$, niezależną od wskaźnika j.

Estymacja efektów, przykład
Hipoteza zerowa przy estymacji efektów:

- stałych - w modelu mieszanym chcemy estymować τ_i i testować hipotezę: $H_0 : \tau_i = 0 \ \forall i$
- losowych - $H_0 : \sigma_\nu^2 = 0$

Różnica - w pierwszym przypadku musimy przetestować wiele parametrów efektu stałego, podczas gdy w drugim przypadu estymujemy i testujemy jedynie jeden parametr efektu losowego.
Teraz powiemy więcej o estymacji i testowaniu modeli losowych i mieszanych.

2 Estymacja

Estymacja efektów w modelu losowym jest możliwa przy użyciu:

- klasyfikacji jednokierunkowej (ANOVY)
- metody największej wiarogodności (ML)
- metody największej wiarogodności z restrykcjami (REML)

2.1 Jednokierunkowa klasyfikacja (ANOVA)

Najprostszy model losowy - **model jednokierunkowej klasyfikacji (ANOVA)**:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

$i = 1, \ldots, a$
$j = 1, \ldots, n$
$\alpha_i \sim N(0, \sigma_\alpha^2)$
$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$

- **Komponenty wariancyjne**: σ_α^2, σ_ε^2
- **Wewnątrzgrupowy współczynnik korelacji**: $\rho = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}$

Estymatory

- **Dekompozycja wariancji:**

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i\cdot})^2 + \sum_{i=1}^a \sum_{j=1}^n (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

  $SST = SSE + SSA$

- $E(SSE) = a(n-1)\sigma_\varepsilon^2$
  $E(SSA) = (a-1)(n\sigma_\alpha^2 + \sigma_\varepsilon^2)$
- **Estymatory:**
  $\hat{\sigma}_\varepsilon^2 = SSE/(a(n-1)) = MSE$
  $\hat{\sigma}_\alpha^2 = \frac{SSA/(a-1) - \hat{\sigma}_\varepsilon^2}{n} = \frac{MSA - MSE}{n}$

Wady korzystania z ANOVY:

- estymator wariancji może przyjmować ujemne wartości
- skomplikowane obliczenia w przypadku gdy dane są niezbalansowane

2.2 Metoda największej wiarogodności (ML)

Jak widzimy ANOVA nie jest doskonała. Szukamy metody, która będzie działała dobrze dla danych niezablansowanycy, którą będzie można łatwo zastosować i nie będzie problemów z ujemną wariancją. Te warunki spełnia ML.

- nie ma takich wad jak ANOVA
- musimy założyć, jaki rozkład mają błędy i efekty losowe (zwykle rozkład normalny)
  ML zadziałą też dla innych rozkładów, ale zwykle nie rozważa się innych rozkładów niż rozkład normalny.
- dla modelu wyłącznie z efektami stałymi mamy model:

$$y = X\beta + \varepsilon,\ y \sim N(X\beta, \sigma^2 I)$$

gdzie
X - macierz planu doświadczenia nxp
β - wektor p stałych parametrów
ε - wektor błędów losowych

Model mieszany Ww. model możemy uogólnić na model mieszany. Mając dane wartości efektów losowych możemy modelować odpowiedź y.

- dla modelu mieszanego:

$$y = X\beta + Z\gamma + \varepsilon,\ y|\gamma \sim N(X\beta + Z\gamma, \sigma^2 I)$$

  gdzie
  Z - macierz nxq opisująca wpływ efektów losowych na obserwacje y
  γ - wektor q efektów losowych
  ε - wektor błędów losowych
  Przyjmuje się, iż wartości oczekiwane obydwu wektorów losowych są wektorami zerowymi. O macierzy kowariancji wektora ε zakłada się zwykle, że jest macierzą diagonalną o postaci $\sigma^2 I$, gdzie I jest macierzą jednostkową.

- dalej zakładamy, że $\gamma \sim N(0, \sigma^2 D)$, Na macierz kowariancji wektora efektów losowych γ nie nakłada się w ogólności żadnych ograniczeń (poza oczywistym wymaganiem jej symetryczności i nieujemnej określoności); macierz tę zapisuję się zwykle jako równą $\sigma^2 D$
  wówczas możemy zapisać rozkład y (bezwarunkowy)

$$vary = varZ\gamma + var\varepsilon = \sigma^2 ZDZ^T + \sigma^2 I$$
$$y \sim N(X\beta, \sigma^2(I + ZDZ^T))$$

- przyjmuje się, że wektory losowe γ i ε są niezależne

Przy takich założeniach macierz kowariancji wektora obserwacji Y ma postać $\sigma^2 ZDZ^T + \sigma^2 I$. Najczęściej przyjmuje się ponadto, że zarówno wektor ε jak i wektor γ mają rozkłady normalne (o podanych parametrach)

Estymacja przy użyciu metody największej wiarogodności (ML)
Gydbyśmy znali D, wówczas moglibyśmy estymować β przy użcyiu metody najmniejszych kwadratów. Jednak estmacja komponentów wariancyjnych D jest jednym z celów naszej analizy. W tej sytuacji jedną z metod estymacji, którą możemy tutaj użyć jest metoda największej wiarogdności.

- jeśli przyjmiemy, że $V = I + ZDZ^T$, wówczas łączna gęstość y wynosi:

$$\frac{1}{(2\pi)^{n/2}|\sigma^2 V|^{1/2}} e^{-\frac{1}{2\sigma^2}(y-X\beta)^T V^{-1}(y-X\beta)}$$

- logarytm wiarogodności:

$$l(\beta, \sigma, D|y) = -\frac{n}{2}log2\pi - \frac{1}{2}log|\sigma^2 V| - \frac{1}{2\sigma^2}(y - X\beta)^T V^{-1}(y - X\beta)$$

- estymacji podlega nieznany wektor stałych parametrów modelu β, wariancja σ^2 i macierz komponentów wariancyjnych D

Wszystko wydaje się proste, jednak w praktyce może to stwarzać trudności.

Wady korzystania z metody największej wiarogodności (ML):

- trudności w przypadku bardziej złożonych modeli, wymagających estymacji większej liczby paremetrów efektów losowych pojawiają się trudności.
- problem z estymatorem wariancji, gdy maksimum wiarogodności jest przyjmowane dla ujemnych wartości. Często trzeba w takich sytuacjach przyjąć po prostu, że estymator wariancji wynosi zero (gdy pochodna wiarogodności nie zeruje się np. dla wartości nieujemnych). To komplikuje obliczenia numeryczne.
- estymatory są często dość silnie obciążone (?)

2.3 Metoda największej wiarogodności z restrykcjami (REML)

- nie ma wielu wad, które dotyczyły ANOVY i ML
  próbują rozwiązać problem obciążenia
- dla danych zbalasowanych, estymatory obliczone przy użyciu REML są zwykle takie same jak obliczone przy użyciu ANOVY
- dostajemy estymatory nieobciążone lub o zredukowanym obciążeniu
- idea: bierzemy liniową kombinację X taką że $K^T X = 0$.
  Wówczas: $K^T y \sim N(0, \sigma^2 K^T V K)$,
  następnie możemy korzystać z metody ML (nie mamy już efektów stałych)
  Kiedy już zastymujemy efekty losowe, możemy łatwo estymować parametry efektów stałych

2.4 R

- **REML**
  Teraz zaprezentujemy estymatory uzyskiwane metodą największej wiarogodności. Używamy pakiety lme4 (dopasowuje modele mieszane; wcześniej nlme).
  Model jak widzimy ma efekty stałe i losowe. Efekty losowe tutaj to przecięcie prezentowane przez pierwszą jedynkę w formule modelu. Efekty losowe są reprezentowane przez (1 operator) - zaznaczone jest, że dane są zgrupowane po operatorze, 1 oznacza, że fekt losowy jest stały w każdej grupie.

REML jest domyślną metodą dopasowania (tutaj jest użyty). Widzimy, że daje identyczne estymatory jak ANOVA, $\hat{\sigma}^2 = 0.106$, $\hat{\sigma}^2_\alpha = 0.068$ i $\mu^* = 60.4$. Dla danych niezbalansowanych estymatory REMLA i ANOVY niekoniecznie sąidentyczn.

- **ML**
  Możemy też skorzystać z tradycyjnej metody największej wiarogodności. Jak widzimy wariancja między grupami wynosi 0.0482 i jest mniejsza niż w REMLu, natomiast wariancja wewnątrz grupy wynosi 0.1118 i jest większa niż poprzednio.

3 Testowanie

3.1 Test ilorazu wiarogodności

Używają standardowej teorii wiarogodności, możemy wyprowadzić test, który służy do porównywania dwóch zagnieżdżonych hipotez H_0) i H_1, obliczając test ilorazu wiarogodności.
Do porównywania hipotez H_0 i H_1 używamy testu **ilorazu wiarogodności**:

$$2(l(\hat{\beta}_1, \hat{\sigma}_1, \hat{D}_1|y) - l(\hat{\beta}_0, \hat{\sigma}_0, \hat{D}_0|y)$$

gdzie
$\hat{\beta}_1, \hat{\sigma}_1, \hat{D}_1$ - estymatory największej wiarogodności (MLE) parametrów przy hipotezie zerowej
$\hat{\beta}_0, \hat{\sigma}_0, \hat{D}_0$ - MLE przy hipotezie alternatywnej
Rozkład zerowy tej statystyki testowej to w przybliżeniu rozkład χ^2 z liczbą stopni swobody równą różnicy stopni swobody w modelu zerowym i pierwszym (w wymiarach dwóch przestrzeni parametrów (gdy modele są identyfikowalne)).

Wady testu ilorazu wiarogodności:

- test jest przybliżony i często przybliżenie to jest nienajlepsze
- wymaga wielu dodatkowych założeń, np. parametry przy hipotezie zerowej nie mogą być na brzegu przestrzeni parametrów
  Jako, że w naszym przypadku $H_0 : \hat{\sigma}^2 = 0$, to stanowi to prawdziwy problem. Nawet jeśli te warunki są spełnione, to przybliżenie rozkładu χ^2 jest bardzo kiepskie.

3.2 Testowanie efektów stałych

- jeśli chcemy użyć testu ilorazu wiarogodności do porównania dwóch zagnieżdżonych modeli, które różnią się tylko efektami stałymi, to nie możemy użyć metody REML do ich porównania. Przyczyną jest to, że REML estymuje efekty losowe przez rozważanie liniowych kombinacji danych, które likwidują efekty stałe. Jeśli te efekty stałe zostaną zmienione, wówczas wiarogodności dwóch modeli nie będą bezpośrednio porównywalne.

- Jeśli chcemy korzystać z testu ilorazu wiarogodności do testowania efektów stałych używamy ML.
- p-wartości generowane przez test ilorazu wiarogodności dla efektów stałych są przybliżone i często zbyt małe, przez co wyolbrzymiają wagę niektórych efektów.
- do znalezienia dokładniejszej p-wartości testu iloraz wiarogodności możemy użyć metody **parametryczego bootstrapu**. Generujemy dane (przy założeniu modelu zerowego używając dopasowanych estymatorów przy hipotezie zerowej)a następnie obliczamy test ilorazu wiarogodności dla tak wygenerowanych danych. Powtarzamy to wielokrotnie i używamy otrzymanych wyników do ocenienia istotności obserwowanej statystyki testowej. Metoda ta zostanie zaprezentowana później.
- alternatywnie: możemy użyć standardowych testów F lub t (parametry efektu losowego warunkujemy po estymowanych wartościach)
  Tutaj zakładamy, że kowariancja losowej części modelu, D, jest równa estymowanej wartości i postępujemy tak jak w metodzie najmniejszych kwadratów.

3.3 Testowanie efektów losowych

- hipoteza zerowa zwykle ma postać $H_0 : \sigma^2 = 0$
  problem - to nie jest wnętrze przestrzeni parametrów, a standardowe obliczenie asymptotycznego rozkładu χ^2 dla ilorazu wiarogodności opiera się o hipotezę zerową leżącą we wnętrzu przestrzeni parametrów. To założenie jest złamane, kiedy testujemy, czy wariancja jest równa zero.
- rozkład zerowy jest wtedy w ogólności nieznany, musimy zastosować metody numeryczne (bootstrap), jeśli chcemy dokładniej testować.
- jeśli jednak przyjmiemy, że otrzymujemy rozkład χ^2 ze standordową liczbą stopni swobody, wówczas otrzymana p-wartość będzie zwykle większa niż powinna. To oznacza, że jeśli obserwujemy istotny efekt używając przybliżenia χ^2, to możemy być na 100 % pewni, że ten efekt jest rzeczywiście istotny.
  Jeśli, otrzymujemy wątpliwe wyniki, powinniśmy skorzystać z dokładniejszej, ale czasochłonnej metody bootstrapu.