

Laboratorium 7A, 7B

Testowanie hipotez statystycznych

Ogólna charakterystyka laboratorium 7A i 7B:

Celem laboratorium jest budowa modeli statystycznych w oparciu o analizę ANOVA (Analysis of Variance). W modelu ANOVA zmienne zależne są ciągłe, a zmienne niezależne są zmiennymi kategorycznymi.

Polecenia do wykonania w trakcie laboratorium (część 1, lab 7A):

Pewna firma chce zbadać jak wybrany typ reklamy produktu wpływa na sprzedaż.

Analiza jest prowadzona w oparciu o tabele ADS i ADS1 zawierającą następujące informacje:

- **Ad** - typ reklamy
- **Area** – obszar kraju (tylko tabela ADS1)
- **Sales** – wielkość sprzedaży

Wstępna analiza danych.

1. Wyświetl 10 pierwszych obserwacji plików danych Ads i Ads1.
2. Oblicz następujące statystyki dla wszystkich danych i dla poszczególnych typów reklamy: średnia, min, max, oraz odchylenie standardowe (PROC MEANS) i zapisz w pliku.
3. Narysuj wykres typu BOX PLOT pokazujący statystyki sprzedaży dla poszczególnych typów reklamy (PROC SGPLOT). Zinterpretuj otrzymany wykres.

Testowanie hipotezy statystycznej .

Hipoteza zerowa:

$$H_0: S1=S2=S3=S4$$

$$H_1: S1 \neq S2 \text{ } S1 \neq S2 \text{ or } S1 \neq S3 \text{ or } S1 \neq S4 \text{ or } S2 \neq S3 \text{ or } S2 \neq S4 \text{ or } S3 \neq S4$$

S1 – średnia sprzedaż dla reklamy typu: *display*

S2 – średnia sprzedaż dla reklamy typu: *paper*

S3 – średnia sprzedaż dla reklamy typu: *people*

S4 – średnia sprzedaż dla reklamy typu: *radio*

Model podstawowy:

$$M_{ik} = \mu + T_i$$

Weryfikacja wyjść dla danych rzeczywistych:

$$Y_{ik} = M_{ik} + \epsilon_{ik} \quad Y_{ik} = \mu + T_i + \epsilon_{ik}$$

gdzie

M_{ik} - oznacza k-tą wartość zmiennej wyjściowej dla reklamy typu i (dla modelu)

Y_{ik} oznacza k-tą wartość zmiennej wyjściowej dla reklamy typu i

μ jest średnią dla wszystkich obserwacji

T_i jest różnicą między średnią sprzedaży dla wszystkich obserwacji i średnią dla danego typu reklamy i

ϵ_{ik} jest różnicą między wartością rzeczywistą k-tej obserwacji dla i-tej klasy reklamy oraz wartością tej obserwacji uzyskaną z modelu M_{ik}

Model rozszerzony:

$$MR_{jik} = \mu + \alpha_j + T_i$$

Weryfikacja wyjść dla danych rzeczywistych:

$$Y_{jik} = MR_{jik} + \varepsilon_{jik} \quad Y_{jik} = \mu + \alpha_j + T_i + \varepsilon_{jik}$$

gdzie

Y_{jik} oznacza k-tą wartość zmiennej wyjściowej dla reklamy typu i i obszaru j

μ jest średnią dla wszystkich obserwacji

α_j jest różnicą między średnią sprzedaży dla wszystkich obserwacji i średnią dla danego obszaru j

T_i jest różnicą między średnią sprzedaży dla wszystkich obserwacji i średnią dla danego typu reklamy i

ε_{jik} jest błędem dla danej obserwacji

1. Sprawdź, czy hipoteza 0 jest spełniona (PROC GLM). Dokonaj interpretacji wyników procedury GLM
2. Podaj parametry modelu podstawowego i rozszerzonego . Oblicz wyjścia na podstawie modeli.

Punktacja laboratorium (część 7A):

Laboratorium jest oceniane w skali 0-5.

1. Statystyki opisowe – (1 pkt).
2. Sprawdzenie hipotezy (1 pkt).
3. Modele (3 pkt).

Zadania (część 2, lab 7B):

1. Oblicz korzystając z systemu SAS wartości (dane w zbiorze zbior Ads i Ads1)

- $SS_T = \sum_i \sum_j (y_{ij} - \bar{y})^2$ - całkowita suma odchyłeń

- $SS_M = \sum_i n_i (\bar{y}_i - \bar{y})^2$ - odchylenia międzygrupowe

- $SS_E = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2$ - odchylenia w grupie

2. Oblicz:

- $F = MS_M / MS_E$
- $MS_M = SS_M / (k-1)$
- $MS_E = SS_E / (n-k)$

3. Sprawdź hipotezę zerową dla przedziałów ufności: 90, 95, 98
4. Przy testowaniu hipotez należy skorzystać z tabel statystycznych:

http://www.socr.ucla.edu/Applets.dir/F_Table.html

Zadania uzupełniające są oceniane w skali 0-5.

1. Statystyki opisowe – (1 pkt).
2. Sprawdzenie hipotezy (1 pkt).
3. Hipotezy (3 pkt).