

Laboratorium statystyczne

Analiza wariancji – porównania wielokrotne + regresja prosta

19. *Reklama*. (reklama.xls)

a) Wykonaj analizę wariancji¹ dla wydatków na produkty mleczne względem zmiennej *TYP_REKLAMY*.

b) Przeprowadź test porównań wielokrotnych Bonferroni, LSD, Tukeya i Scheffego. Skomentuj otrzymane wyniki.

c) Wykonaj analizę wariancji dla wydatków na produkty mleczne względem zmiennej *TYP_REKLAMY* ze zmienną towarzyszącą² *RODZINA_n*

d) Wykonaj analizę wariancji i wybrany test porównań wielokrotnych dla wydatków na produkty mleczne względem zmiennej *MIASTA* i *RODZINA_n*³

e) Wykonaj analizę wariancji dla wydatków na produkty mleczne względem zmiennej *MIASTA* ze zmienną towarzyszącą *RODZINA_n*

20. Dane : *aspiryna* (aspiryna i wino.xls)

Chcemy zbadać, czy czas istotnie wpływa na zawartość substancji aktywnej w tablecie aspiryny. W oparciu o dostępne dane można to zrobić na dwa sposoby: przeprowadzając analizę wariancji zmiennej *skladnik* wg podziału na klasy opisujące upływ czasu (zmienna *czas*) lub budując model regresji zmiennej *skladnik* względem zmiennej⁴ *miesiace*.

a) Przeprowadź na poziomie istotności 0.05 analizę wariancji (ANOVA) zmiennej *skladnik* wg podziału na klasy opisujące upływ czasu (zmienna *czas*), czyli zweryfikuj istotność modelu $M_2 = [\mu_{t_0}, \dots, \mu_{t_0}, \mu_{t_4}, \dots, \mu_{t_4}, \dots, \mu_{t_{24}}, \dots, \mu_{t_{24}}]$ względem modelu

$M_1 = [\mu, \mu, \dots, \mu]$, gdzie⁵ μ_{t_i} oznacza wartość oczekiwaną ilości substancji aktywnej w tablecie, której czas przechowywania mieści się w klasie t_i .

b) Przeprowadź na poziomie istotności 0.05 analizę wariancji dla modelu regresji liniowej zmiennej *skladnik* względem zmiennej *miesiace*, czyli zweryfikuj istotność modelu $M_3 = [\beta_1 0 + \beta_0, \dots, \beta_1 0 + \beta_0, \beta_1 4 + \beta_0, \dots, \beta_1 4 + \beta_0, \beta_1 24 + \beta_0, \dots, \beta_1 24 + \beta_0]$ względem modelu M_1 . Jaki jest współczynnik determinacji dla tego modelu?

c) Przeprowadź *test lack-of-fit*⁶ czyli zweryfikuj istotność modelu M_2 względem modelu M_3 . Wyjaśnij, kiedy można mówić o braku dopasowania regresji i co to praktycznie oznacza.

21. Dane: *wino* (aspiryna i wino.xls)

a) Sprawdź, że model regresji liniowej jest słabo istotnym opisem relacji między spożyciem wina a liczbą zgonów na chorobę wieńcową serca

b) Wybierz metodą strzałki Tukeya⁷ przekształcenia linearyzujące zmienne wino i serce. Na podstawie trzech punktów napisz przybliżone równanie regresji liniowej dla tak przekształconych danych. Opierając się na tym równaniu, powiedz jak odbije się na szansie zgonu z powodu choroby serca dwukrotne zwiększenie konsumpcji wina. Dlaczego w odpowiedzi należy użyć słów *mediana szansy zgonu* ... a nie słów *średnia wartość szansy zgonu* ...?

c) Przekształć odpowiednio zmienne i wykonaj *pełną* analizę regresji liniowej dla tych zmiennych. Czy współczynniki kierunkowe przybliżonego i dokładnego równania regresji różnią się od siebie znacząco?

¹ Analizę wykonaj, sprawdzając założenia ANOVA w razie potrzeby wprowadzając odpowiednią skalę.

² czyli analizę kowariancji

³ traktowanej jako zmienna nominalna

⁴ liczbowej

⁵ Wektory M_1 i M_2 są 42-wymiarowe

⁶ Test braku dopasowania regresji

⁷ Dodatkowe materiały *metoda strzałki Tukeya.pdf* + wykład w8 *regresja nieliniowa*

Laboratorium statystyczne

Analiza wariancji – porównania wielokrotne + regresja prosta