UWAGA,
TO SĄ MOJE PRZEKONANIA, ZA POPRAWNOŚĆ ROZWIĄZAŃ NIE RĘCZĘ!!
WZORY SĄ NA ŚCIĄDZE, KTÓRĄ NIEDAWNO KTOŚ WRZUCIŁ NA HYDEPARK

Zad 1
Omówić w jaki sposób weryfikuje się poprawność przyjętego modelu regresji prostej
Rozwiązanie
Dana jest dwuwymiarowa zmienna losowa (X,Y) oraz n elementowa próba .
$E(Y|X=x)=bX+a$., czyli zachodzi związek $Y_i=bX_i+a+e_i$ dla i=1...n, gdzie e_i oznacza błąd losowy .
Weryfikacja
Sprawdzamy czy prawdziwe są założenia modelu o normalności rozkładu reszt modelu, sprawdzamy istotność współczynników regresji oraz dopasowanie modelu do danych.
Miarą dopasowania jest wspólczynnik determinacji R^2, określający jaką część zmienności Y można wyjaśnić zmiennością X.
Istotność współczynników regresji można sprawdzić za pomocą procedury ANOVA stawiając hipotezę, że b=0, przy hipotezie alternatywnej, że b<>0.
Normalność reszt – dowolnym testem na normalność rozkładu

{Dla regresji wieloliniowej podobnie, tylko X jest k-wymiarową zmienną losową, współczynnik determinacji zastępujemy skorygowanym współczynnikiem determinacji, a b jest wektorem współczynników. W ANOVA testujemy hipotezę, że b jest wektorem złożonym z samych zer, wobec hipotezy alternatywnej, że istnieje niezerowa współrzędna w b}

zad 2
w całodobowym sklepie spożywczym notowano liczbę sprzedanych butelek w ciągu kolejnych 2 lat. Zakładając, że w modelu nie ma wyraźnego trendu wyznaczyć indeksy sezonowosci i podać ich interpretację.

Lata\kwartały	1	2	3	4
1994	1100	1300	2100	1200
1995	1300	1400	2500	1300

rozwiązanie
mamy cztery podokresy w każdym cyklu, przy czym liczba obserwacji dla każdego podokresu jest równa 2.
Wyznaczamy średnią dla każdego podokresu
$y_1=(1100+1300)/2=1200$
$y_2=(1300+1400)/2=1350$, $y_3=(2100+2500)/2=2300$, $y_4=(1200+1300)/2=1250$
oraz średnią ogólną (ze wszystkich obserwacji) y=1525
wyznaczamy indeksy sezonowości jako ilorazy średnich w podokresach i średniej ogólnej.
$O_i=y_i/y$:
$O_1=1200/1525=0,7869$, $O_2=1350/1525=0,8852$, $O_3=2300/1525=1,5082$,
$O_4=1250/1525=0,8197$
Interpretacja (czy naprawdę o taką chodzi ??)
Najlepszy dla sprzedaży jest trzeci kwartał, najgorszy pierwszy.

Zad3
1. Trzej nauczyciele języka polskiego miało ocenić w skali punktowej 1 – 20 wypracowania czterech uczniów pewnej szkoły. Wyniki przedstawia tabela. Chcemy ocenić, czy wszyscy trzej nauczyciele są tak samo surowi. Jakie narzędzie statystyczne należy użyć do rozwiązania tego problemu i jakie hipotezy należy postawić?