

Zadanie 1

Co oznacza pojęcie "dekompozycja szeregu czasowego"?
Jak się ją przeprowadza? Wymienić i opisać składowe,
na jakie można rozłożyć szereg czasowy.

Dekompozycja szeregu czasowego oznacza rozkład
szeregu na składowe. Wyodrębniamy składowe
szeregu i ugniamy je w sposób liczbowy lub
jakościowy.

Składowe szeregu czasowego:

- 1) Trend (tendencja rozwojowa) - stała tendencja
wzrostowa albo spadkowa - T_+
- 2) Wahania sezonowe - zmiany w pewnych okresach,
sezonach S_+
- 3) Wahania cykliczne - zmiany w pewnych cyklach C_+
- 4) Wahania losowe (przypadkowe) - suma małych
zjawisk ma tendencje okresowe ε_+

Zadanie 2

Poniższa tabela zawiera dane dotyczące średniej zawartości ozonu w atmosferze w kolejnych kwartałach w latach 1999-2000. Wybrać ten szereg średnich ruchomych przyjmując okres o długości 4.

rok	kwartał	ozon
1999	1 I	323
	2 II	360
	3 III	302
	4 IV	284
2000	5 I	320
	6 II	342
	7 III	314
	8 IV	292

$\bar{Y}_{2,5} = 317,25$	$\Rightarrow \bar{Y}_3 = \frac{\bar{Y}_{2,5} + \bar{Y}_{3,5}}{2} = 316,875$
$\bar{Y}_{3,5} = 316,5$	$\Rightarrow \bar{Y}_4 = \frac{\bar{Y}_{3,5} + \bar{Y}_{4,5}}{2} = 314,25$
$\bar{Y}_{4,5} = 312$	$\Rightarrow \bar{Y}_5 = \frac{\bar{Y}_{4,5} + \bar{Y}_{5,5}}{2} = 313,5$
$\bar{Y}_{5,5} = 315$	$\Rightarrow \bar{Y}_6 = \frac{\bar{Y}_{5,5} + \bar{Y}_{6,5}}{2} = 316$
$\bar{Y}_{6,5} = 317$	

Wytyczenie szeregu średnich ruchomych - odfiltrowanie pełnej losowości.

Stosujemy tu scentrowane średnie ruchome

$$\bar{Y}_{2,5} = \frac{Y_I + Y_{II} + Y_{III} + Y_{IV}}{4} = \frac{323 + 360 + 302 + 284}{4} = 317,25$$

$$\bar{Y}_{3,5} = \frac{Y_{II} + Y_{III} + Y_{IV} + Y_I}{4} = \frac{360 + 302 + 284 + 320}{4} = 316,5$$

$$\bar{Y}_{4,5} = \frac{Y_{III} + Y_{IV} + Y_I + Y_{II}}{4} = \frac{302 + 284 + 320 + 342}{4} = 312$$

$$\bar{Y}_{5,5} = \frac{Y_{IV} + Y_I + Y_{II} + Y_{III}}{4} = \frac{284 + 320 + 342 + 314}{4} = 315$$

$$\bar{Y}_{6,5} = \frac{320 + 342 + 314 + 292}{4} = 317$$

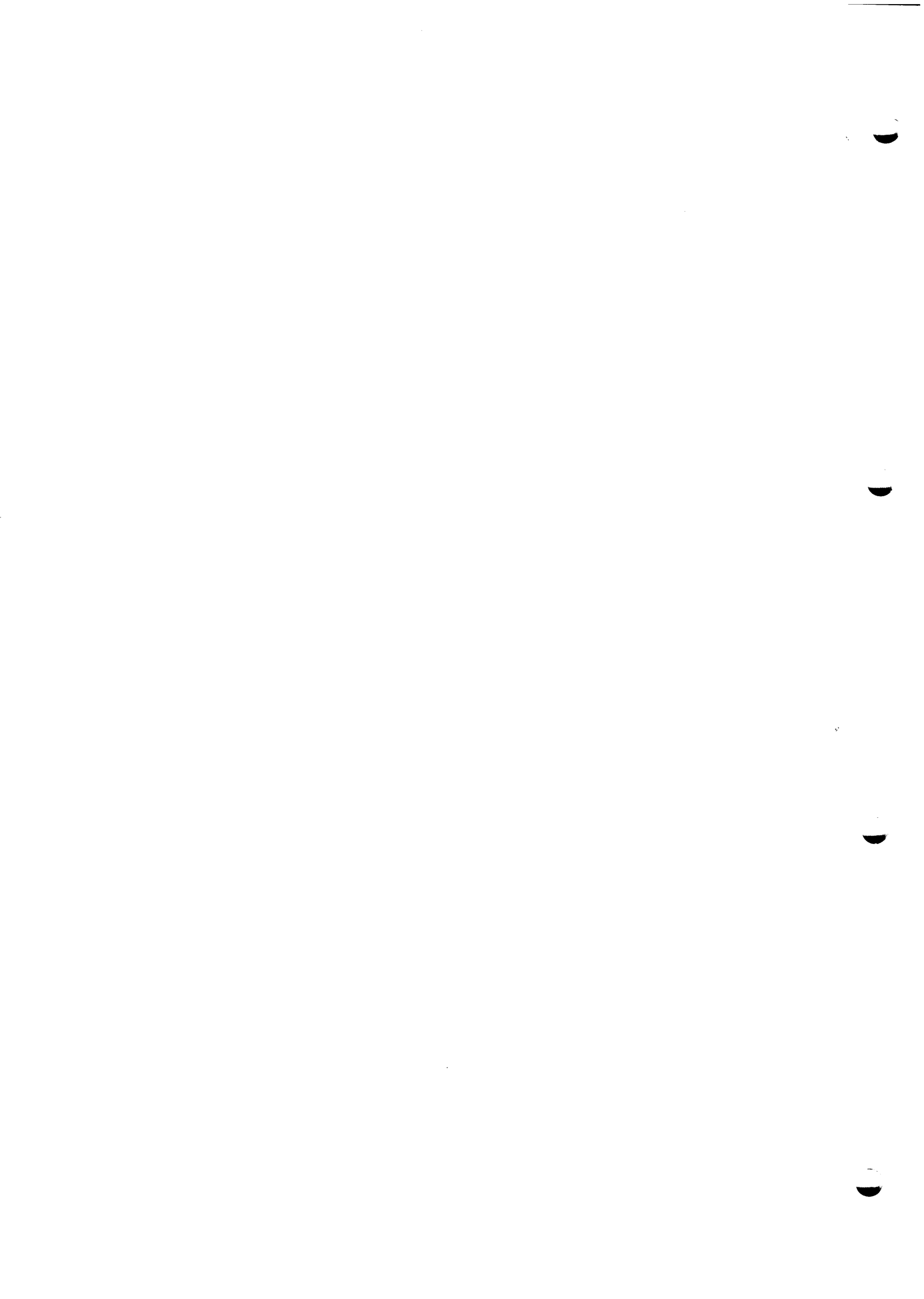
Zadanie 3

Firma produkująca telefaksy postanowiła sprawdzić, czy trzy zespoły elektroniczne A, B i C dają przedstawić taką samą prędkość transmisji danych. Przeprowadzono badanie na trzech rodzajach danych: sam tekst, same ilustracje i tekst z ilustracjami.

Zmierzono kilkunastokrotnie czas transmisji różnymi rodzajami danych używając urządzenia wyposażonego w któryś z wspomnianych rodzajów podrzeczni.

Zaproponować procedurę statystyczną, która materializowałaby się w celu stwierdzenia, czy długość czasu transmisji zależy w istotny sposób od wytych podrzeczni i rodzaju danych. Sformułować hipotezy, które materializowałyby się w celu sformułowania wniosków do poprawnego przeprowadzenia stosowanej procedury

ANALIZA WARIANCJI



Zadanie 4

Badacz przypuszcza, że interesująca go zależność regresyjna można w zadowalający sposób opisać modelem potęgowym $Y = aX^b$. Pokazać, w jaki sposób model ten można sprowadzić do modelu liniowego

$$Y = aX^b \leftarrow \text{model potęgowy}$$

$$\ln Y = \ln a + b \ln X$$

$$Y' = \ln Y$$

$$a' = \ln a$$

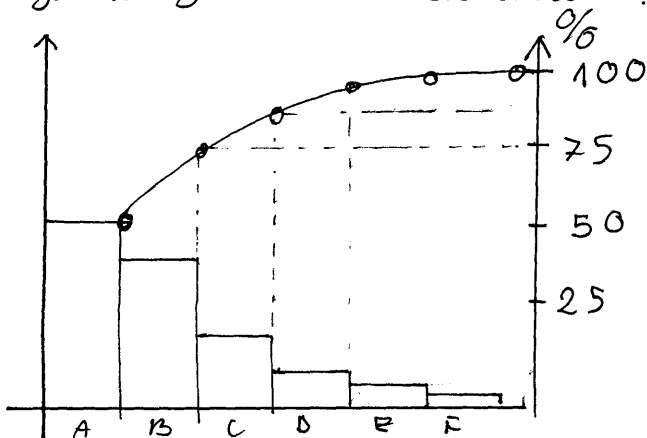
$$X' = \ln X$$

$$\boxed{Y' = a' + bX'} \leftarrow \text{model liniowy}$$

Zadanie 5

Omówić budowę i przeznaczenie karty (wykresu) Pareto

Wykresem Pareto nazywany wykresem stopniowym oparty na uporządkowanych malejąco wartościach liczbowych, sortowanych z poszczególnymi kategoriami. Jest on często wypełniany tzw. krzywą Loblentza, która pokazuje wzrost skumulowanego udziału procentowego wyróżnianych kategorii. Krzywa udziału procentowego ma swoją oddzielną skalę, umieszczoną na osi po prawej stronie wykresu. Wykres Pareto może być konstruowany w oparciu o tablicę rozkładu częstości, która opisuje występowanie analizowanego skutku w różnych obszarach lub z różnych przyczyn. Zatem jest on wdrażaniem uporządkowanego malejąco histogramu. Za pomocą wykresu Pareto można łatwo zweryfikować prawdziwość hipotezy o umiejscowieniu i udziale najważniejszych przyczyn badanego zjawiska. Jeśli znajdziemy ona potwierdzenie, to wykres wyraźnie wskazuje grupy przyczyn, których usunięcie pozwoliłoby uniknąć większości skutków.

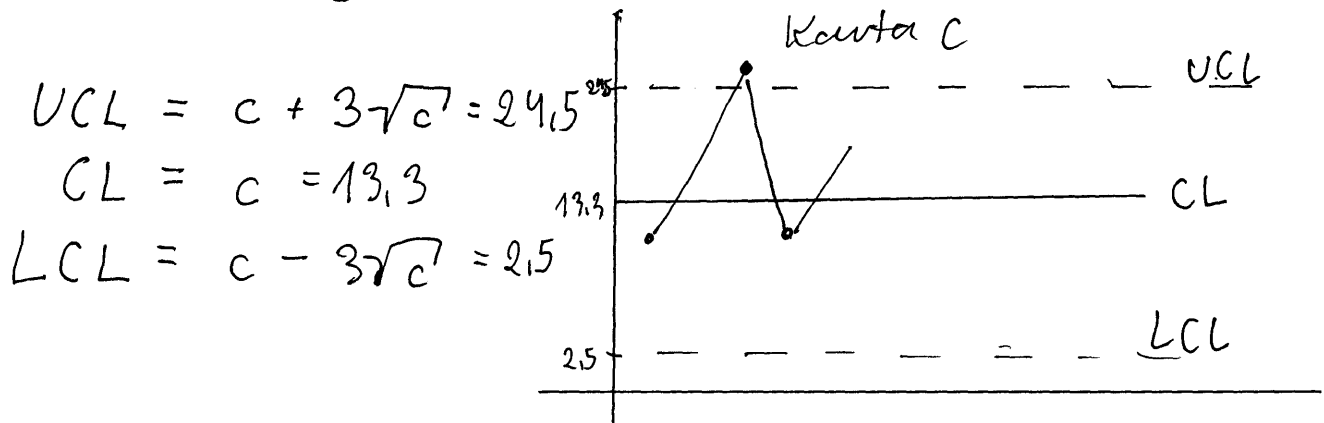


Zadanie 6

Dane są liczby plam na karoserii nowych samochodów: 12, 25, 13, 20, 14, 9, 6, 16, 5

Zaproponować kartę kontrolną, która ma być zastosowana do oceny procesu lakierowania karoserii. Omówić budowę tej karty, sposób jej konstrukcji i zasady postępowania się tą kartą

Karta C - karta liczby niezgodności. Często liczba niezgodności ma rozkład Poissona



UCL - górna granica kontrolna

CL - linia centralna

LCL - dolna linia kontrolna

Zadanie 7

Wyjaśnić, na czym polega kontrola jakości za pomocą jednostopniowego planu badania według oceny alternatywnej.

W tzw. jednostopniowym planie badania, decyzja o przyjęciu bądź odmiesieniu partii podejmowana jest w zależności od tego czy $d > c$, czy też $d \leq c$, gdzie d - liczba elementów wadliwych, c dopuszczalna liczba elementów wadliwych

Zadanie 8

Omów sposób weryfikacji modelu w analizie regresji wielorakiej

Weryfikacja poprawności modelu:

1) Test F (Anova w analizie regresji)

$$\begin{cases} H: \alpha_1 = \alpha_2 = \dots = \alpha_n = 0 & - Y \text{ nie zależy w sposób} \\ & \text{istotny od żadnego } X \\ K: \neg H & - \text{nie wszystkie } \alpha_i \text{ są równe, } Y \text{ zależy} \\ & \text{przynajmniej od jednego } X \end{cases}$$

2) Testy istotności współczynników regresji

$$\begin{cases} H_1: \alpha_1 = 0 \\ K_1: \alpha_1 \neq 0 \end{cases}$$

$$\begin{cases} H_2: \alpha_2 = 0 \\ K_2: \alpha_2 \neq 0 \end{cases}$$

$$\vdots$$

$$\begin{cases} H_k: \alpha_k = 0 \\ K_k: \alpha_k \neq 0 \end{cases}$$

3) Współczynnik determinacji

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Współczynnik determinacji regresji wielorakiej określa jaki procent zmienności funkcji model, jego wartość mówi nam że wrostem liczby zmiennych uwzględnionych w modelu.

$$\bar{R}^2 = 1 - \frac{\frac{SSE}{n-k-1}}{\frac{SST}{n-1}}$$

Skorygowany współczynnik determinacji (koryguje fakt dotarcia zmiennych, które być może nie są istotne (nie wpływają na model))

4) Analiza reszt

- a) losowość - reszty nie powinny zdradzać jakiegokolwiek systematycznego ułożenia
- b) normalność
- c) jednorodności wariacji

Zadanie 9

Poniższy sruwey czasowy to dane kwartalne dotyczące cen baryłki ropy naftowej (w \$) w USA w latach 1984-1985:

36 32 30 33 32 28 27 31

Podać prognozy ceny baryłki ropy na pierwszy kwartał 1986 roku stosując metodę wykładniczą ma wykładniczego. Przyjąć współczynnik wykładniczości = 0,3.

	y	\hat{y}	
	36	36	$\alpha = 0,3$
1984	32	34,8	$\hat{y}_{t+1} = \alpha y_t + (1-\alpha) \hat{y}_t$
	30	44,5	$\hat{y}_{t+1} = 0,3 \cdot 36 + (1-0,3) 36 =$
	33	44,05	$\hat{y}_{t+1} = 10,8 + 0,7 \cdot 36 =$
	32		$\hat{y}_{t+1} = 10,8 + 25,2 = 36$
1985	28		$\hat{y}_2 = 0,3 \cdot 32 + (1-0,3) \cdot 36 =$
	27		$\hat{y}_2 = 9,6 + 25,2 = 34,8$
	31		$\hat{y}_3 = 0,3 \cdot 30 + 0,7 \cdot 34,8 = 9 + 35,5$
			$\hat{y}_3 = 44,5$
			$\hat{y}_4 = 33 \cdot 0,3 + 0,7 \cdot 44,5 = 32,05 +$
			$\hat{y}_4 = 9,9 + 31,15 = 41,05$

Zadanie 10

Opisać sposób weryfikacji poprawności modelu w analizie regresji prostej

Weryfikacja poprawności modelu w analizie regresji prostej:

1) Analiza współczynnika korelacji:

$\begin{cases} H: \rho = 0 & - \text{współczynnik korelacji nie jest istotny} - \text{źle} \\ K: \rho \neq 0 & - \text{współczynnik korelacji jest istotny} - \text{ok} \end{cases}$

2) Analiza współczynnika determinacji:

$R^2 = \frac{SSR}{SST}$ - wyjaśnia jaki procent zmienności

$0 \leq R^2 \leq 1$ jest wyjaśniony przez model
lim R^2 bliżej zbliżony do 1
tym lepiej jest dopasowany model

3) Test F (analiza wariancji w analizie regresji)

$\begin{cases} H: \text{nie ma zależności liniowej} - \text{źle} \\ K: \text{jest zależność liniowa} - \text{ok} \end{cases}$

4) Testy istotności współczynnika regresji:

$$y = a + bx + \varepsilon$$

$\begin{cases} H: b = 0 \leftarrow \text{współczynnik } b \text{ jest nieistotny} - \text{brak zależności liniowej} \\ K: b \neq 0 \leftarrow \text{współczynnik } b \text{ jest istotny} \end{cases}$

$\begin{cases} H: a = 0 \leftarrow \text{model można poprawić usunięciem} \\ K: a \neq 0 \leftarrow \text{dobry model} \end{cases}$

5) Analiza reszt:

- losowość

- normalność reszt

- jednorodność wariancji

Zadanie 11

Badacz przypuszcza, że interesującą go zależność regresyjną można w zadowalający sposób opisać modelem wykładniczym: $Y = \exp(a + bX)$.
Podać, w jaki sposób model ten można sprowadzić do modelu liniowego

$$Y = \exp(a + bX) \leftarrow \text{model wykładniczy}$$

$$X = e^{a + bX}$$

$$\ln Y = a + bX$$

$$\ln Y = Y'$$

$$Y' = a + bX'$$

\leftarrow model liniowy po przekształceniu

$$Y = a + b \underbrace{X}_{X'} + \varepsilon \leftarrow \text{model logarytmiczny}$$

$$X = a + bX' + \varepsilon \leftarrow \text{model liniowy po przekształceniu}$$

$$Y = a + \frac{b}{X} + \varepsilon \leftarrow \text{model odwrotnościowy (arg X)}$$

$$X' = \frac{1}{X} \Rightarrow Y = a + bX' + \varepsilon$$

$$Y = a + bX' + \varepsilon \leftarrow \text{model liniowy po przekształceniu}$$

$$Y = \frac{1}{a + bX + \varepsilon} \leftarrow \text{model odwrotnościowy (arg Y)}$$

$$Y = \frac{1}{Y} \Rightarrow Y' = a + bX + \varepsilon \leftarrow \text{m. liniowy}$$

$$X = \frac{1}{a + \frac{b}{X} + \varepsilon} \leftarrow \text{model podwójnie odwrotnościowy}$$

$$X' = \frac{1}{X}$$

$$Y' = \frac{1}{Y}$$

$$Y' = a + bX' + \varepsilon \leftarrow \text{m. liniowy}$$

Zadanie 12

Redakcja magazynu motoryzacyjnego postanowiła porównać trwałość drogi hamowania marki Polonez wyposażonego w różne rodzaje opon. W tym celu rozpedziła pojazd do prędkości 100 km/h i po czym zmierzyła drogę hamowania. Próby powtórzyła kilkakrotnie dla każdego z badanych rodzajów opon (tzn. Goodyear, Pirelli, Dunlop i Dabica). Zapropozowała procedurę statystyczną, której materiały użył w celu stwierdzenia, czy trwałość drogi hamowania zależy w istotny sposób od rodzaju opon, w jaki sposób zony jest samochód. Sformułowała hipotezy i litery materiały zacytowała i umieszczała zdaniem autora game do poprawnego przeprowadzenia stosownej procedury

Zadanie 13

Producent prętów stalowych na koniec każdej zmiany
wybiera w sposób losowy do sprawdzenia 30 prętów
i notuje liczbę wadliwych prętów w każdej próbie.
Zaproponować kartę kontrolną, którą mógłby się
posłużyć ów producent. Omów budowę tej karty, sposób
jej konstrukcji i zasady postępowania się tą kartą.

KARTA NP - karta liczby jednostek niezgodnych

$$UCL = np + 3\sqrt{np(1-p)}$$

$$CL = np$$

$$LCL = np - 3\sqrt{np(1-p)}$$

Zadanie 14

Podać zastosowanie testu Kruskala - Wallisa i omówić budowę tego testu.

Test Kruskala - Wallisa stosujemy jeśli założenia analizy wariancji nie są spełnione.

Zakładamy, że F_1, F_2, \dots, F_n są rozkładami ciągłymi

$$H: F_1 = F_2 = \dots = F_n$$

$$K: \neg H$$

Zadanie 15

Wyjaśnić termin „współczynnik determinacji” i podać własności tego współczynnika. Dlaczego w niektórych sytuacjach należy podugiwać się skorygowanym współczynnikiem determinacji?

Współczynnikiem determinacji nazywany liczbę:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

R^2 określa jako procent zmienności tłumaczy model regresyjny, jego wartość rośnie wraz ze wzrostem liczby zmiennych uwzględnionych w modelu, jest to miara dopasowania linii regresji do danych.

$$0 \leq R^2 \leq 1$$

im R^2 bliżej 1 tym model lepiej dopasowany. W niektórych sytuacjach stosujemy skorygowany współczynnik ~~reg~~ determinacji aby skorygował fakt dopasowania zmiennych, które być może nie są istotne - nie wpływają na model

$$\bar{R}^2 = 1 - \frac{MSE}{\frac{SST}{n-1}} \quad \bar{R} \in (0, 1)$$

Zadanie 16

Można z kart kontrolnych można stosować w przypadku dysponowania próbami o różnej liczności? —
W jaki sposób obecność prób różniących się liczebnością wpływa na budowę kart kontrolnych?

W przypadku ~~stos~~ dysponowania próbkami o różnej liczności można stosować karty $\bar{X}-S$ oraz karty p, c, np, u

W przypadku próbek o niejednakowej liczności granice kontrolne karty nie są liniami ciągłymi —

Zadanie 17

Opisać metodę estymacji gęstości prawdopodobieństwa metodą estymatorów jądrowych.

Najprostszym estymatorem gęstości rozkładu jest histogram, w którym zasadniczą rolę w jakości estymacji stanowi szerokość klasy. Wada histogramu jest nieciągłość. Niedogodność tę można usunąć stosując estymatory jądrowe postaci:

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^m K\left(\frac{x - X_i}{h}\right)$$

↑ szerokość pasma

Zadanie 18

Pan Antoni poliera co godzinę próbki losową pięciu wałków tłoczonych na tokarce i mierzy ich średnice. Zaproponować kartę kontrolną, która mogłaby się postawić Pan. Antoni w celu stwierdzenia, czy proces tłoczenia wałków jest uregulowany. Omówić budowę tej karty, sposób jej konstrukcji i zasady postępowania się.

Karta $\bar{X} - R$

$$UCL = \bar{\bar{X}} + A_2 n \frac{3}{d_2 \sqrt{n}} \bar{R}$$

$$CL = \bar{\bar{X}}$$

$$LCL = \bar{\bar{X}} - \frac{3}{d_2 \sqrt{n}} \bar{R}$$

karta \bar{X}

karta \bar{R}

$$UCL = D_4 \bar{R}$$

$$CL = \bar{R}$$

$$LCL = D_3 \bar{R}$$

Zadanie 19

Postanowiono porównać długości czasu wymaganego do realizacji pewnego zadania informatycznego prowadzonego na komputerach wyposażonych w różne procesory. W tym celu mianem czas realizacji ustalonego przedsięwzięcia powtórzyliśmy próby kilkakrotnie dla każdego z badanych procesorów. Zaplanować procedurę statystyczną, której materiały użyje w celu stwierdzenia czy długość czasu wymaganego do realizacji danego zadania zależy w istotny sposób od rodzaju procesora, w jaki sposób jest komputer. Sformułować hipotezy, które materiały twierdzą i wymieni założenia wymagane do poprawnego przeprowadzenia stosowanej procedury.

Zadanie 20

W osiedlowym sklepie spożywczym zamówiono litry (w szt.) sprzedanych butelek (0,5l) z napojem Coca-Cola w ciągu dwóch kolejnych lat. Zakładając, że w modelu nie ma uśrednionego trendu, wyznacz indeksy sezonowości i podaj ich interpretację

LATA	KWARTALEY			
	I	II	III	IV
1994	1100	1300	2100	1200
1995	1300	1400	2500	1300

Mamy 4 podokresy w każdym cyklu i mamy czynniki obserwacji dla każdego podokresu jest równa 2. Wyznaczamy średnią dla każdego podokresu

$$Y_1 = \frac{(1100 + 1300)}{2} = 1200$$

$$Y_2 = \frac{1300 + 1400}{2} = 1350$$

$$Y_3 = \frac{2100 + 2500}{2} = 2300$$

$$Y_4 = \frac{1200 + 1300}{2} = 1250$$

Teraz liczymy średnią ze wszystkich średnich

$$Y = \frac{Y_1 + Y_2 + Y_3 + Y_4}{4} = \frac{1200 + 1350 + 2300 + 1250}{4} = 1525$$

Wyznaczamy indeksy sezonowości

$$O_i = Y_i / Y$$

$$O_1 = \frac{1200}{1525} = 0,7869$$

$$O_2 = \frac{1350}{1525} = 0,8852$$

$$O_3 = \frac{2300}{1525} = 1,5082$$

$$O_4 = \frac{1250}{1525} = 0,8197$$

Najlepszy dla sprzedaży jest 3 kwartał, najgorszy 1

Zadanie 21

Treści nauczycieli miało ocenić 20 szkół publicznych 1-20 wypracowania czterech uczniów. Zpracowania były umysły tej są tak samo surowi. Jakże mądrie statystyczne materiały użyć do rozwiązania tego problemu i jakże hipotezy materiały postawić?

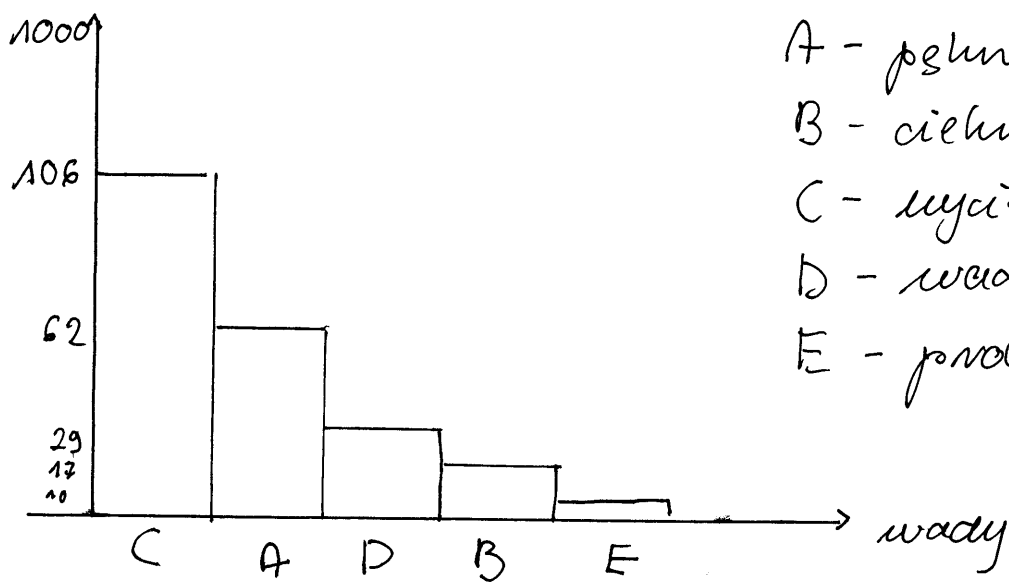
NAUCZYCIEL

A	B	C
19	17	20
20	20	19
10	11	9
14	15	12

Analiza wariancji

Zadanie 22

Wśród 1000 przetestowanych silników samochodowych 62 miało pełnięte obudowy, 17 cięknące łożyska, 106 wycieków oleju, 29 wadliwych cylindrów a 10 problemów z zapłonem. Narysować diagram Pareto dla tych danych i wskazać najistotniejsze problemy w tym procesie produkcyjnym



- A - pełnięte obudowy
- B - cięknące łożyska
- C - wycieki oleju
- D - wadliwe cylindry
- E - problemy z zapłonem

Najistotniejszymi problemami w tym procesie są wycieki oleju i pełnięte obudowy

Zadanie 23

Poniższe dane przedstawiają wielkość sprzedaży dysków komputerowych (w tys. sztuk) w kolejnych 8 latach. Wypracuj model wyładzania rekurencyjnego dla tych danych przyjmując współczynnik wyładzania $\alpha = 0,3$

	1	2	3	4	5	6	7	8
	1992	1993	1994	1995	1996	1997	1998	1999
x	62	57	50	48	52	55	58	61

Konstruujemy szereg s_i pobawiony wahani sezonowych

$$S_i = \alpha \cdot x_i + (1 - \alpha) \cdot S_{i-1} \quad \alpha = 0,3$$

$$S_1 = x_1$$

$$S_1 = \underline{62}$$

$$S_2 = 0,3 \cdot 57 + 0,7 \cdot 62 = 17,1 + 43,4 = \underline{60,5}$$

$$S_3 = 0,3 \cdot 50 + 0,7 \cdot 60,5 = 15 + 42,35 = \underline{57,35}$$

$$S_4 = 0,3 \cdot 48 + 0,7 \cdot 57,35 = \underline{54,545}$$

$$S_5 = 0,3 \cdot 52 + 0,7 \cdot 54,545 = \underline{53,7815}$$

$$S_6 = 0,3 \cdot 55 + 0,7 \cdot 53,7815 = \underline{54,14705}$$

$$S_7 = 0,3 \cdot 58 + 0,7 \cdot 54,14705 = \underline{55,2029}$$

$$S_8 = 0,3 \cdot 61 + 0,7 \cdot 55,2029 = \underline{56,9203}$$

Zadanie 24.

Poniżej dane to średnie roczne ceny wopy naftowej importowanej w kolejnych ośmiu latach. Wyznacz model wykładniczy przez średnie ruchome scentrowane

1973	1974	1975	1976	1977	1978	1979	1980
6	12	13	13	14	21	34	43