ISSN 1425-7351

INSTYTUT CHEMII I TECHNIKI JĄDROWEJ

INSTITUTE OF NUCLEAR CHEMISTRY AND TECHNOLOGY

WARSZAWA

RAPORTY IChTJ. SERIA B nr 2/96

TEST KOMETKOWY.
2. ANALIZA STATYSTYCZNA WYNIKÓW

Marcin Kruszewski

Warszawa 1996

ZESPÓŁ REDAKCYJNY

dr Wiktor Smułek, Ewa Godlewska, Sylwester Wojtas

WYDAWCA

Instytut Chemii i Techniki Jądrowej
ul. Dorodna 16, 03-195 Warszawa
tel.: (0-22) 11 06 56; telex: 813027 ichtj pl: fax: (0-22) 11 15 32;
e-mail: sekdyrn@orange.ichtj.waw.pl

Raport został wydany w postaci otrzymanej od Autora

Symbol UKD: 57.087

Symbol INIS: C11.00

Słowa kluczowe: ANALIZA STATYSTYCZNA, TEST KOMETKOWY

Test kometkowy. 2. Analiza statystyczna wyników

W pracy opisano niektóre metody statystyczne używane do oceny wyników otrzymanych testem kometkowym. Metody statystyczne opisano na przykładzie pakietu Statistica w oparciu o podręcznik Biostatistical analysis, J.H. Zar, Prentice-Hall International, Inc., 1984.

Comet assay. 2. Statistical analysis of results

Statistical methods commonly used for analysis of results of the comet assay was described, based on the Statistica analytical software.

SPIS TREŚCI

1. ANALIZA POJEDYNCZEJ PRÓBKI

1.1. ANALIZA ROZKŁADU PRÓBKI

Analiza ma na celu określenie czy rozkład próbki jest normalny, oraz określenie parametrów rozkładu (kurtozy i symetrii (skośności)).

Ponieważ parametry badanego rozkładu określa się na podstawie próbki badanej, hipotezę należy sprawdzić testem Shapiro-Wilka lub testem Kołmogonorowa-Smirnowa, z poprawką Lilleforce'a.

H_0: rozkład próbki jest normalny

H_A: rozkład próbki nie jest normalny

Jeżeli $p<0,05$ należy odrzucić H_0 (rozkład próbki nie jest normalny)[1].

1.2. OCENA SKOŚNOŚCI

Skośność (γ_1) jest trzecim momentem średniej ([1], str.81). Statystyczną ocenę skośności prowadzi się w zasadzie dopiero dla dużych próbek $n>150$ (odpowiednia statystyka została opisana w [1], str.118). Jeżeli:

$\gamma_1=0$ to rozkład jest symetryczny;

$\gamma_1<0$ próbka pochodzi z populacji w której średnia jest mniejsza niż mediana;

$\gamma_1>0$ próbka pochodzi z populacji w której średnia jest większa niż mediana.

Dla małych próbek można sprawdzić symetrię rozkładu korzystając z tabeli B.20 w [1]. Wartość (γ_1) obliczoną dla rozkładu porównuje się z wartością w tabeli B.20 dla danego poziomu istotności: $\alpha_{(2)} = 0,05$ dla testu dwustronnego i $\alpha_{(1)} = 0,025$ dla testu jednostronnego.

1.3. OCENA KURTOZY

Kurtoza (γ_2) jest czwarty momentem średniej ([1], str.81). Jej statystyczną ocenę prowadzi się dla próbek $n>1000$ (odpowiednia statystyka została opisana w [1], str.119). Na podstawie obliczonej wartości (γ_2) można dokonać ogólnej oceny kurtozy. Populacje mające:

(γ_2) = 0 są normalne (mezokurtyczne);

(γ_2) < 0 (platykurtyczne) mogą pochodzić z dwóch populacji o takiej samej wariancji ale różnych średnich;

(γ_2) > 0 (mezokurtyczne) mogą pochodzić z dwóch populacji o takiej samej średniej ale różnych wariancjach.

---

[1] We wszystkich przykładach uważa się, że poziom istotności p=0.05 jest wystarczający

2. OKREŚLENIE HOMOGENNOŚCI POŁĄCZONYCH DOŚWIADCZEŃ

Połączenie wyników dwóch doświadczeń i ich wspólna analiza zwiększa moc testu (większe n). Możliwość połączenia wyników dwóch próbek określa się za pomocą testu heterogenności χ^2, opierając się na założeniu, że χ^2 połączonych wyników równe jest sumie χ^2 poszczególnych prób.

H_0: próbki są homogenne

H_A: próbki nie są homogenne

Należy przeprowadzić analizę rozkładu prób testem χ^2 dopasowując je do rozkładu normalnego i zsumować poszczególne wartości, następnie testem χ^2 przeprowadzić analizę rozkładu połączonych prób, dopasowując je do tego samego rozkładu. Od sumy χ^2 dla poszczególnych prób odjąć wartość χ^2 dla prób połączonych. Na podstawie otrzymanej wartości sprawdzić istotność H_0 dla ilości stopni swobody DF = (ilość prób) - 1. Jeżeli $p<0,05$ należy odrzucić H_0 (próbki są heterogenne) i nie można ich połączyć.

3. PORÓWNANIE ŚREDNICH DWÓCH PRÓBEK

3.1. ROZKŁAD NORMALNY

W przypadku potwierdzenia normalności rozkładu należy sprawdzić czy wariancje obu porównywanych prób są równe. Do porównania wariancji używa się testu F Fischera lub testu Lavene'a.

H_0: wariancje są równe

H_A: wariancje nie są równe

Jeżeli $p<0,05$ należy odrzucić H_0 (wariancje nie są równe).

Uwaga: Test F Fischera często jest zawarty w module analizy testem t-Studenta (np. w pakiecie Statistica) i nie ma potrzeby oddzielnego obliczania tego testu.

3.1.1. Wariancje są równe

3.1.1.1. Próby niezależne

W przypadku stwierdzenia równości wariancji do porównania dwóch średnich należy użyć testu t-Studenta dla prób niezależnych.

H_0: Średnie są równe

H_A: Średnie nie są równe

Jeżeli $p<0,05$ należy odrzucić H_0 (średnie nie są równe).

Uwaga: Testem t-Studenta można także sprawdzać hipotezy jednostronne np.

H_0: średnia1> średnia2

H_A: średnia1<średnia2

3.1.1.2. Próby zależne

W przypadku stwierdzenia równości wariancji do porównania dwóch średnich należy użyć testu t-Studenta dla prób zależnych.

H_0: Średnie są równe
H_A: Średnie nie są równe
Jeżeli $p<0,05$ należy odrzucić H_0 (średnie nie są równe).
Uwaga: Test t jest mocny jeżeli ilość obserwacji w próbach jest jednakowa i jednakowe są wariancje prób. Testem t-Studenta można także sprawdzać hipotezy jednostronne np.
H_0: średnia1> średnia2
H_A: średnia1<średnia2

3.1.2. Wariancje nie są równe

Jeżeli n>30 a różnica wariancji nie jest duża i/lub próby nie są równe, to można użyć testu t-Studenta z poprawką na osobną ocenę wariancji lub innych testów np. Cochran-Coxa. Jeżeli różnica wariancji jest duża to lepiej użyć któregoś z testów nieparametrycznych (patrz 3.2.).

3.2. ROZKŁAD NIE JEST NORMALNY

W przypadku stwierdzenia, że rozkład nie jest normalny do porównania dwóch średnich należy użyć testów nieparametrycznych.

3.2.1. Próby niezależne

Do porównania dwóch średnich prób niezależnych należy użyć testu U Mann-Whitney'a (korzystny) lub testu Kołmogonorowa-Smirnowa dla dwóch prób. Korzystnie jest przeprowadzić oba testy, chociaż test Mann-Whitney'a jest silniejszy niż test Kołmogonorowa-Smirnowa dla dwóch prób, to ten ostatni jest bardziej czuły na różnice w kształcie rozkładów.
H_0: Średnie są równe
H_A: Średnie nie są równe
Jeżeli $p<0,05$ należy odrzucić H_0 (średnie nie są równe).

3.2.2. Próby zależne

Do porównania dwóch średnich prób zależnych należy użyć testu znakowego różnic (sign test) lub testu Wilcoxona dla par powiązanych.
H_0: Średnie są równe
H_A: Średnie nie są równe
Jeżeli $p<0,05$ należy odrzucić H_0 (średnie nie są równe).

4. PORÓWNANIE WIELU PRÓB

Najczęściej stosowany sposób porównania wielu prób poprzez wielokrotne porównanie parami testem t-Studenta, nie jest odpowiedni i może prowadzić do mylnej oceny istotności różnicy średnich. Prawdopodobieństwo popełnienia błędu I rodzaju (odrzucenie hipotezy zerowej jeżeli w istocie jest ona prawdziwa) przy porównaniu parami testem t-Studenta 10 średnich wynosi 0,63 przy poziomie ufności $\alpha=0,05$. Porównanie takie powinno się prowadzić metodą jednoczynnikowej analizy wariancji (one-way ANOVA), dwuczynnikowej analizy wariancji (two-way ANOVA) lub wieloczynnikowej analizy wariancji (MANOVA) (dla prób o rozkładzie normalnym) lub ich nieparametrycznymi odpowiednikami, np. ANOVA Kruskal-Wallisa (dla prób o rozkładzie nienormalnym).

4.1. ROZKŁADY NORMALNE I WARIANCJE JEDNAKOWE

Do porównania średnich kilku prób mających normalne rozkłady i jednakowe wariancje używa się jednoczynnikowej analizy wariancji (one-way ANOVA). Jednoczynnikowej analizy wariancji można także użyć, jeżeli różnica między wariancjami nie jest zbyt duża i (1) $n>10$ oraz (2) średnie grup nie są skorelowane z odchyleniem standardowym.

H_0: Średnie są równe

H_A: Średnie nie są równe

Jeżeli $p<0,05$ należy odrzucić H_0 (średnie nie są równe).

Uwaga 1: Istnieją przynajmniej trzy typy analizy ANOVA: Model I i II i III. Analiza typu I powinna być prowadzona jeżeli zakłada się stały wpływ efektu na zmienne, np. wzrastająca dawka czynnika, analiza typu II - kiedy efekt jest rzeczywiście losowy, np. porównanie średnich kontroli. Model III występuje w przypadku gdy działanie czynnika jest mieszane, czyli zachodzi stały i losowy wpływ efektu na zmienne.

Uwaga 2: Przy przeprowadzaniu analizy metodą dwuczynnikowej analizy wariancji korzystnie jest mieć jednakową ilość powtórzeń w każdym punkcie, jednakże możliwe jest też przeprowadzenie takiej analizy przy niejednakowej ilości powtórzeń.

4.2. ROZKŁADY NIENORMALNE I/LUB NIEJEDNAKOWE WARIANCJE

Do porównania średnich kilku prób mających nienormalne rozkłady i/lub niejednakowe wariancje używa się testu Kruskal-Wallisa, zwanego też analizą wariancji według rang. Test ten można też stosować wszędzie tam gdzie stosuje się jednoczynnikowy test ANOVA, wtedy test Kruskal-Wallisa, ma moc 95% testu ANOVA. Testu Kruskal-Wallisa można też użyć do nieparametrycznej analizy dwuczynnikowej

4.3. PORÓWNANIA WIELOKROTNE *POST HOC*

Często po odrzuceniu hipotezy o równości wariancji, zachodzi potrzeba wzajemnego porównania poszczególnych prób (wariancji), w celu określenia, które próby różnią się, a które nie. Chociaż często stosowane jest porównanie parami testem t-Studenta, nie jest to najlepszy sposób takiej analizy. Po odrzuceniu H_0 można przeprowadzić analizę *post hoc*, w celu znalezienia średnich odbiegających. Pakiet Statistika zawiera następujące testy umożliwiające analizę *post hoc*: test LSD (odpowiednik testu t-Studenta), test Scheffe'a, test Neuman-Keulsa, test Duncana, test Tukey'a. Korzystnie jest porównać wyniki kilku testów. Najczęściej używane (i zalecane) są testy Tukey'a (Tukey Honest Significant Difference test) i test Neuman-Keulsa. Testy te powinny być stosowane przede wszystkim w przypadku analizy ANOVA typu I.

Jeżeli próbki nie są równej wielkości można użyć testu Tukey'a dla nierównych wielkości próbek (Tukey HSD test for unequal sample size), zwanego też testem Spjotvoll-Stoline.

Do porównań wielokrotnych prób nieparametrycznych stosuje się odmianę testu Tukey'a, w którym rangi nadaje się sumom, a nie średnim ([1], str. 199).

Pakiet Statistica oferuje też inne rodzaje analizy wariancji, takie jak dwuczynnikowa analiza wariancji, wieloczynnikowa analiza wariancji, czy hierarchiczna analiza wariancji (nested ANOVA), jednakże dokładny opis tych metod wykracza poza ramy tego opracowania. Dokładny opis wymienionych metod zamieszczony jest w [1], str. 206-260.

5. BADANIE ZALEŻNOŚCI MIĘDZY ZMIENNYMI

Na podstawie wyników doświadczalnych można sprawdzić czy zmierzone wielkości zależą od siebie, czy też ich wzajemne powiązanie jest przypadkowe. Różnica między regresją a korelacją polega na tym, że w przypadku regresji zakładamy *a priori*, że Y zależy od X, natomiast badając korelację określamy czy Y zależy od X i czy X zależy od Y.
Po stwierdzeniu, że badane wielkości zależą od siebie (korelacja) można obliczyć krzywą regresji, a więc sposób w jaki Y zmienia się w zależności od X.

5.1. KORELACJA

5.1.1. Zmienne o rozkładzie normalnym

Stopień zależności od siebie dwóch badanych wielkości oznacza się jako współczynnik korelacji r (1), a jego błąd standardowy s_r określony jest wzorem (2), w którym n oznacza liczbę punktów na podstawie których przeprowadzono korelację.

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}, \qquad (1)$$

$$s_r = \sqrt{\frac{1-r^2}{n-2}}. \qquad (2)$$

Za pomocą testu t-Studenta należy sprawdzić istotność statystyczną obliczonego współczynnika korelacji według wzoru (3):

$$t = \frac{r}{s_r}, \qquad (3)$$

w którym r oznacza obliczony współczynnik korelacji, s_r obliczony według wzoru (2) błąd standardowy. Ilość stopni swobody oblicza się $\nu = n - 2$, gdzie n to liczba punktów, na podstawie których przeprowadzono korelację.
H_0: r = 0
H_A: r ≠ 0
Hipotezę zerową H_0: r = 0 odrzuca się na korzyść H_A: r ≠ 0, jeżeli $|t| \geq t_{\alpha(2),\nu}$. Za pomocą testu t-Studenta można także testować hipotezy jednostronne.

5.1.2. Zmienne o rozkładzie nienormalnym

Jeżeli bada się korelację zmiennych o rozkładzie nienormalnym należy zastosować test korelacji rang Spearmana lub analogiczny test Kendala. Odpowiednia statystyka została opisana w [1], str.318.

5.1.3. Porównanie dwóch współczynników korelacji

Program Statistica oferuje także możliwość porównania dwóch współczynników korelacji i określenia istotności różnicy między nimi. Odpowiednia statystyka została opisana w [1], str.313.

5.1.4. Porównanie więcej niż dwóch współczynników korelacji

Do porównania więcej niż dwóch współczynników korelacji wykorzystuje się wzór (4):

$$\chi^2 = \sum_{i=1}^{k}(n_i - 3)z_i^2 - \frac{\left[\sum_{i=1}^{k}(n_i - 3)z_i\right]^2}{\sum_{i=1}^{k}(n_i - 3)}, \qquad (4)$$

w którym n oznacza ilość punktów na podstawie których obliczono korelację, a

$$z = 0{,}5\ln\frac{(1+r)}{(1-r)}. \qquad (5)$$

H_0: $r_1 = r_2 = r_k$
H_A: $r_1 \neq r_2 \neq r_k$
Obliczoną wartość porównuje się z wartością $\chi^2_{\alpha,\nu}$ dla $\nu = k - 1$ stopni swobody (k liczba porównywanych współczynników korelacji). Hipotezę H_0: $r_1 = r_2 = r_k$ (elewacje nie są równe) odrzuca się, jeżeli wartość obliczona jest większa niż niż wartość prawdopodobieństwa $\chi^2_{\alpha,\nu}$. Po odrzuceniu hipotezy o równości współczynników korelacji możliwe jest porównanie *post hoc* poszczególnych współczynników, w celu znalezienia współczynników korelacji różniących się od innych. Odpowiednia statystyka jest opisana w [1], str.317.

5.2. REGRESJA

Po stwierdzeniu zależności pomiędzy zmiennymi, można zbadać w jaki sposób zmienna zależna zależy od zmiennej niezależnej. Do określania zależności między zmiennymi używa się najczęściej metody najmniejszych kwadratów. W tym opracowaniu zostanie omówiona jedynie regresja prostoliniowa, ponieważ jej statystyczne podstawy są najlepiej poznane i jest ona najczęściej używana, jednakże pakiet Statistica oferuje także możliwości stosowania regresji nieliniowej.

Jeżeli zależność pomiędzy zmiennymi jest prostoliniowa, to wyraża ją wzór (6):

$$y = bx + a, \qquad (6)$$

w którym b oznacza współczynnik regresji, odpowiadający kątowi nachylenia prostej regresji (slope), natomiast a odpowiada punktowi przecięcia prostej regresji z osią OY (Y intercept).

5.2.1. Obliczanie kąta nachylenia prostej regresji (b)

Najlepsze przybliżenie wartości b określa wzór (7):

$$b = \frac{\sum xy}{\sum x^2}, \tag{7}$$

w którym

$$\sum xy = \sum x_i y_i - \frac{\left(\sum x_i\right)\left(\sum y_i\right)}{n}, \tag{8}$$

$$\sum x^2 = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}. \tag{9}$$

5.2.2. Błąd standardowy wartości b

Błąd standardowy wartości b określony jest wzorem (10):

$$s_b = \sqrt{\frac{s_{Y\bullet X}^2}{\sum x^2}}, \tag{10}$$

w którym $s_{Y\bullet X}$ oznacza błąd standardowy regresji określony wzorem (11):

$$s_{Y\bullet X} = \sqrt{\frac{tSS - rSS}{n-2}}, \tag{11}$$

w którym

$$rSS = \frac{\left(\sum xy\right)^2}{\sum x^2}, \tag{12}$$

$$tSS = \sum y^2 = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}. \tag{13}$$

Wartość $s_{Y\bullet X}$ jest bezpośrednio związana z współczynnikiem regresji i błędem standardowym y (s_y) i można ją także obliczyć ze wzoru (14)

$$s_{Y\bullet X} = s_y\sqrt{1-r^2}. \tag{14}$$

5.2.3. Punkt przecięcia z osią OY i jego błąd standardowy

Najlepszym przybliżeniem wartości a jest

$$a = \bar{y} - b\bar{x}, \tag{15}$$

w którym $\bar{y}$ i $\bar{x}$ są wartościami średnimi. Błąd standardowy wartości a określony jest wzorem (16):

$$s_a = \sqrt{s_{Y\bullet X}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum x^2}\right]}, \tag{16}$$

w którym $\bar{x}$ jest wartością średnią, a $s_{Y\bullet X}^2$ oznacza błąd standardowy regresji określony wzorem (11) lub (14).

5.2.4. Obliczanie X dla znanego Y

Znając równanie prostej regresji oraz wartość y_i, można na podstawie wzoru (17) obliczyć wartość X (inverse prediction)

$$\hat{x}_i = \frac{y_i - a}{b}, \qquad (17)$$

(1-α) przedziały ufności obliczonego $\hat{x}_i$ nie są symetryczne względem wyznaczonego $\hat{x}_i$ i oblicza się je ze wzoru (18)

$$\bar{x} + \frac{b(y_i - \bar{y})}{K} \pm \frac{t}{K}\sqrt{\left(s^2_{Y \bullet X}\right)\left[\frac{(y_i - \bar{y})^2}{\sum x^2} + K\left(1 + \frac{1}{n}\right)\right]}, \qquad (18)$$

w którym $t = t_{\alpha(2),(n-2)}$, $K = b^2 - t^2\left(s_b^2\right)$, s_b określone jest wzorem (10).

5.2.5. Sprawdzenie istotności obliczonego współczynnika nachylenia prostej regresji

Za pomocą testu t-Studenta można sprawdzić istotność statystyczną obliczonego współczynnika nachylenia prostej regresji.

H_0: $b_1 = 0$

H_A: $b_1 \neq 0$

Odpowiednia statystyka opisana jest wzorem (19):

$$t = \frac{b_1}{s_b}, \qquad (19)$$

w którym s_b jest błędem standardowym obliczonym z wzoru (10). Obliczoną wartość porównuje się z wartością testu t dla $n-2$ stopni swobody.

5.2.6. Porównanie dwóch współczynników nachylenia prostych regresji

Dwa kąty nachylenia można porównać testem t-Studenta. Statystyka jest opisana wzorem:

$$t = \frac{b_1 - b_2}{s_{b_1 - b_2}}, \qquad (20)$$

w którym

$$s_{b_1 - b_2} = \sqrt{\frac{\left(s^2_{Y \bullet X}\right)_p}{\left(\sum x^2\right)_1} + \frac{\left(s^2_{Y \bullet X}\right)_p}{\left(\sum x^2\right)_2}}, \qquad (21)$$

a

$$\left(s^2_{Y \bullet X}\right) = \frac{(rSS)_1 + (rSS)_2}{(rDF)_1 + (rDF)_2}. \qquad (22)$$

Indeksy dolne *1* i *2* odnoszą się do analizowanych prostych regresji. Otrzymaną wartość porównuje się z krytyczną wartością t dla $v = n_1 + n_2 - 4$ stopni swobody. Jeżeli współczynniki nachylenia prostej nie są równe, można obliczyć punkt przecięcia krzywych $X_I Y_I$:

$$X_I = \frac{a_2 - a_1}{b_1 - b_2}, \qquad (23)$$

i

$$Y_l = a_1 + b_2 X_l = a_2 + b_1 X_l. \qquad (24)$$

Natomiast jeżeli H_0: $b_1=b_2$ to można obliczyć wartość wspólnego współczynnika nachylenia prostej regresji b_c wyrażoną wzorem (25):

$$b_c = \frac{\left(\sum xy\right)_1 + \left(\sum xy\right)_2}{\left(\sum x^2\right)_1 + \left(\sum x^2\right)_2} = \frac{\left(\sum x^2\right)_1 b_1 + \left(\sum x^2\right)_2 b_2}{\left(\sum x^2\right)_1 + \left(\sum x^2\right)_2}. \qquad (25)$$

5.2.7. Porównanie dwóch elewacji

Po odrzuceniu H_0: $b_1=b_2$ można przyjąć, że badane populacje są różne. Jednakże, jeżeli kąty nachylenia prostych są takie same, to proste te mogą mieć różny punkt przecięcia z osią OY (elewacja). Do określenia, czy elewacja obu prostych regresji jest taka sama używa się testu t-Studenta :

$$t = \frac{\left(\overline{Y}_1 - \overline{Y}_2\right) - b_C\left(\overline{X}_1 - \overline{X}_2\right)}{\sqrt{MS_C\left[\frac{1}{n_1} + \frac{1}{n_2} + \frac{\left(\overline{X}_1 - \overline{X}_2\right)^2}{A_C}\right]}}, \qquad (26)$$

a

$$MS_C = \left(s^2_{Y \bullet X}\right)_C = \frac{SS_C}{DF_C}, \qquad (27)$$

$$SS_C = C_C = \frac{B_C^2}{A_C}, \qquad (28)$$

$$A_C = \left(\sum x^2\right)_1 + \left(\sum x^2\right)_2, \qquad (29)$$

$$B_C = \left(\sum xy\right)_1 + \left(\sum xy\right)_2, \qquad (30)$$

$$C_C = \left(\sum y^2\right)_1 + \left(\sum y^2\right)_2. \qquad (31)$$

Otrzymaną wartość t porównuje się z wartością t w tablicach dla $\nu = DF_C$ stopni swobody. Jeżeli dwie elewacje prostych regresji nie różnią się znamiennie to można obliczyć wspólną dla obu prostych elewację:

$$a_C = \frac{n_1\overline{Y}_1 + n_2\overline{Y}_2}{n_1 + n_2} - b_C\frac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2}. \qquad (32)$$

Jeżeli stwierdzono, że obie proste regresji mają takie same kąty nachylenia i elewacje to wspólna prosta regresji dla obu populacji wyników będzie miała wzór:

$$Y_i = a_C + b_C X_i. \qquad (33)$$

5.2.8. Porównywanie punktów na prostych regresji

Jeżeli stwierdzono, że dwie proste regresji mają takie same współczynniki nachylenia i takie same elewacje, to proste te są takie same i ich punkty nie różnią się ze statystycznego punktu widzenia. Jeżeli dwie proste regresji mają takie same współczynniki nachylenia, ale różne elewacje, to proste te są równoległe, a punkty na ich leżące różnią się od siebie. Jeżeli dwie proste regresji mają różne współczynniki nachylenia, to proste te przecinają się i można sprawdzić, które z punktów tych prostych różnią się. Porównania dokonuje się testem t-Studenta:

$$t = \frac{\hat{Y}_1 - \hat{Y}_2}{\sqrt{\left(s^2_{Y \cdot X}\right)_p \left[\frac{1}{n_1} + \frac{1}{n_2} + \frac{\left(X - \overline{X}_1\right)^2}{\left(\sum x^2\right)_1} + \frac{\left(X - \overline{X}_2\right)^2}{\left(\sum x^2\right)_2}\right]}}. \quad (34)$$

Otrzymaną wartość t porównuje się z wartością t w tablicach dla $v = n_1 + n_2 - 4$ stopni swobody. Testu należy jednak używać z ostrożnością ponieważ zakłada, że każda z przewidywanych wartości Y związana jest z taką samą wariancją. Test jest najsilniejszy, jeżeli obie proste mają takie same X i n.

5.2.9. Porównanie więcej niż dwóch współczynników nachylenia prostych regresji

Do porównania więcej niż dwóch współczynników prostych prostych regresji używa się procedury znanej jako analiza kowariancji.

H_0: $\beta_1 = \beta_2 = \beta_k$

H_A: $\beta_1 \neq \beta_2 \neq \beta_k$

$$F = \frac{\left(\frac{cSS - pSS}{k - 1}\right)}{\frac{pSS}{DF_p}}, \quad (35)$$

w którym

$$cSS = \sum_{i=1}^{k} C_i - \frac{\left(\sum_{i=1}^{k} B_i\right)^2}{\sum_{i=1}^{k} A_i}, \quad (36)$$

$$pSS = \sum_{i=1}^{k} SS_i , \quad (37)$$

$$DF_p = \sum_{i=1}^{k} (n_i - 2) = \sum_{i=1}^{k} n - 2k . \quad (38)$$

Jeżeli stwierdzono, że kąty nachylenia prostych regresji nie różnią się, to można obliczyć wspólny kąt nachylenia prostych regresji:

$$b_c = \frac{\sum_{i=1}^{k} \left(\sum xy\right)_i}{\sum_{i=1}^{k} \left(\sum x^2\right)_i}. \quad (38)$$

5.2.10. Porównanie *post hoc* więcej niż dwóch prostych regresji

Jeżeli porównanie kilku kątów nachylenia prostych regresji wykazało, że nie są one równe, to do porównania poszczególnych par kątów nachylenia prostych regresji należy użyć testu Tukey'a lub jeżeli porównuje się z kontrolą testu Dunnett'a lub Scheffe'a. Odpowiednia statystyka opisana jest w [1], str.302.

5.2.11. Porównanie więcej niż dwóch elewacji

Do porównania więcej niż dwóch elewacji używa się procedury znanej jako analiza kowariancji.

H_0: $\beta_1 = \beta_2 = \beta_k$

H_1: $\beta_1 \neq \beta_2 \neq \beta_k$

$$F = \frac{\left(\dfrac{tSS - cSS}{k-1}\right)}{\dfrac{cSS}{DF_c}}, \tag{40}$$

w którym

$$tSS = C_t - \frac{B_t^2}{A_t}, \tag{41}$$

$$cSS = \sum_{i=1}^{k} C_i - \frac{\left(\sum_{i=1}^{k} B_i\right)^2}{\sum_{i=1}^{k} A_i}, \tag{42}$$

$$DF_C = \sum_{i=1}^{k} n_i - k - 1. \tag{43}$$

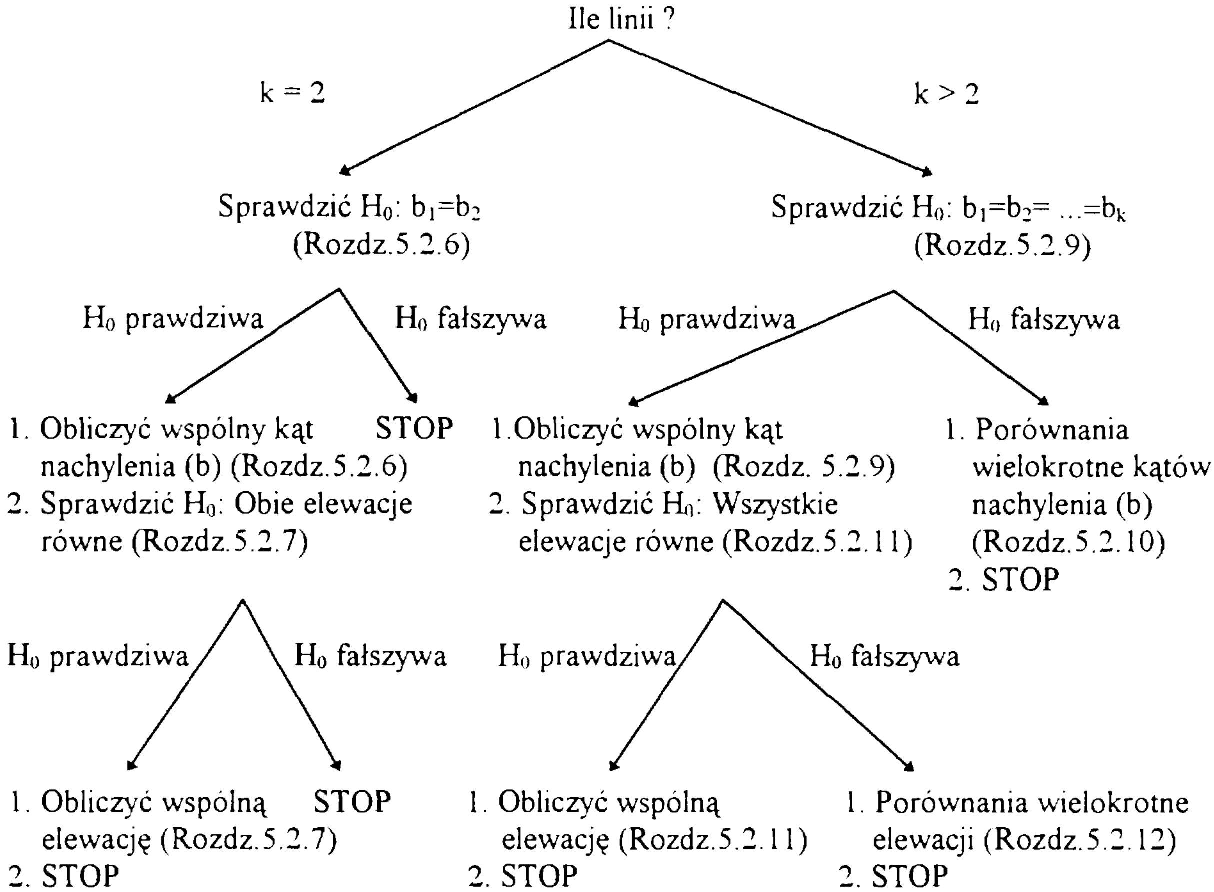
We wzorze tSS oznacza całkowitą sumę kwadratów (total Sum of Squares), a zmienne A_t, B_t *i* C_t, oznaczają odpowiednio Σx^2, Σxy, Σy^2 obliczone dla wszystkich punktów regresji traktowanych jakby należały do jednej prostej.

5.2.12. Porównanie *post hoc* więcej niż dwóch elewacji prostych regresji

Jeżeli porównanie kilku elewacji prostych regresji wykazało, że nie są one równe, to do znalezienia najbardziej odbiegających elewacji prostych regresji można użyć testu Tukey'a lub jeżeli porównuje się z kontrolą testu Dunnett'a lub Scheffe'a. Odpowiednia statystyka opisana jest w [1], str.303.

5.3. DIAGRAM PORÓWNANIA PROSTYCH REGRESJI

Rysunek przedstawia schemat blokowy ilustrujący sposób porównania prostych regresji.



Rys. Schemat blokowy ilustrujący sposób porównania prostych regresji.

6. LITERATURA

[1]. Zar J.H.: Biostatistical analysis. Prentice-Hall International, Inc., 1984.