Wiesław W a g n e r (Poznań)

TESTY ZGODNOŚCI Z ROZKŁADEM NORMALNYM DLA PRÓBY PROSTEJ

1. Wstęp

1.1. Wprowadzenie

Rozkład normalny odgrywa podstawową rolę w statystyce matematycznej. Powszechnie stosowany jest do weryfikacji hipotez parametrycznych, gdzie najczęściej a priori zakłada się, że badana zmienna losowa ma rozkład normalny. Założenie to w świetle danych eksperymentalnych nie zawsze jest prawdziwe. Zachodzi stąd potrzeba sprawdzenia, czy rozkład empiryczny badanej zmiennej nie wykazuje nadmiernego odchylenia od rozkładu normalnego. Sprawdzanie zgodności między wspomnianymi rozkładami przeprowadzane jest testami normalności (Pawłowski, 1976, s.223).

Historycznie testy $\sqrt{b_1}$ i b_2 były pierwszymi testami normalności, dalsze wprowadzili: Geary (test g), David, Hartley i Pearson (test u), Kołmogorow (test D) oraz Cramer i von Mises (test CM). Z kolei powstały testy normalności korzystające ze statystyk pozycyjnych. Takie testy podali: Shapiro i Wilk (test W), Shapiro i Francia (test W'), Weisberg i Bingham (test $\widetilde{W}$), Fillibren (test r) oraz D'Agostino (test D_A). Poza wspomnianymi, w literaturze spotyka się kilka innych testów normalności, których obszerny przegląd podany jest w rozdziale 2. Przy prezentacji testów wzięto pod uwagę ich aspekty praktyczne, wskazując źródła zawierające niezbędne tablice wartości krytycznych dla przeprowadzenia wnioskowania statystycznego.

W literaturze polskiej brak jest szerszego opracowania dotyczącego testów normalności, co w poważnym stopniu ogranicza ich szersze zastosowanie. Powstałą lukę mogłaby częściowo wypełnić niniejsza praca.

Bibliografia związana z zagadnieniami testów normalności jest bardzo obszerna. Szersze omówienie literatury podają prace: Hegazy i Green (1975) oraz Pearson i inni (1977).

1.2. Preliminaria.

W pracy zastosowano następującą symbolikę. Litery greckie μ i σ stosujemy dla oznaczenia wartości oczekiwanej i odchylenia standardowego w populacji. Przez X oznaczamy zmienną losową typu ciągłego o rozkładzie prawdopodobieństwa określonym przez dystrybuantę $F(x)$ z parametrami rozkładu: $E(X) = \mu$ i $D^2(X) = \sigma^2$. Ciąg $x_1, x_2, \ldots, x_n$ (krótko ciąg $\{x_i\}$) oznacza próbę prostą złożoną z n obserwacji zmiennej X. Ponadto, $\bar{x}$ oznacza średnią arytmetyczną, S^2 - sumę kwadratów odchyleń, s^2 - wariancję z próby i s - odchylenie standardowe z próby. Ciąg $y_1, y_2, \ldots, y_n$ oznacza ciąg uporządkowany niemalejąco obserwacji $x_1, x_2, \ldots, x_n$, taki, że $y_1 \leq y_2 \leq \ldots \leq y_n$. Fakt, że zmienna X ma rozkład normalny o parametrach μ i σ^2, zapisujemy $X \sim N(\mu, \sigma^2)$. Symbolem $F_n(x)$ oznaczamy dystrybuantę empiryczną

$$F_n(x) = \begin{cases} 0, & \text{dla } x \leq y_1, \\ k/n, & \text{dla } y_k < x \leq y_{k+1}, \quad k = 1, 2, \ldots, n-1, \\ 1, & \text{dla } x > y_n, \end{cases}$$

zaś przez $\Phi(u)$ i $\Phi^{-1}(p)$ dystrybuantę i kwantyl rzędu p-tego standaryzowanego rozkładu normalnego (krótko rozkładu N(0,1)).

Statystyką pozycyjną (krótko s.p.) nazywamy zmienną losową y_k będącą k-tą zmienną w próbie $\{y_i\}$, zaś $\{y_i\}$ ciągiem (μ,σ) -s.p. Jeśli $X \sim N(\mu,\sigma^2)$, to y_k jest normalną s.p. i wtedy $\{y_i\}$ jest ciągiem normalnych s.p. Jeśli $\{y_i\}$ jest ciągiem $N(\mu,\sigma^2)$ - s.p., to $\{u_i\}$, gdzie $u_i = (y_i-\mu)/\sigma$, jest ciągiem N(0,1) - s.p. Dla ostatniego ciągu przyjmujemy następujące oznaczenia dla parametrów rozkładu

$$E(u_i) = m_i, \quad D^2(u_i) = v_{ii}, \quad i = 1,2,\dots,n$$

$$Cov(u_i,u_j) = v_{ij}, \quad i,j = 1,2,\dots,n; \ i \neq j.$$

Wówczas dla $N(\mu,\sigma)$ - s.p., mamy

$$E(y_i) = \mu+\sigma m_i, \quad D^2(y_i) = \sigma^2 v_{ii} \quad i \quad Cov(y_i,y_j) = \sigma^2 v_{ij} .$$

Nadmieńmy, że wartości $m_i = m_{i,n}$, $i = 1,2,\dots,[n/2]$ zostały stablicowane dla różnych n, natomiast $v_{ii} = v_{ii,n}$ oraz $v_{ij} = v_{ij,n}$ dla $i,j = 1,2,\dots,[n/2]$; $i \neq j$ tylko dla $n = 1,2,\dots,20$ (Pearson i Hartley, 1972, tab. 9 i 10).

W końcu oznaczmy przez $NPP[m_i:y_i]$ normalny diagram prawdopodobieństwa, sporządzony w układzie współrzędnych przez zaznaczenie na osi odciętych wartości m_i, a na osi rzędnych wartości y_i.

1.3. Weryfikacja hipotezy o zgodności rozkładu empirycznego z rozkładem normalnym.

Przypuśćmy, że dysponujemy próbą prostą $\{x_i\}$ n obserwacji zmiennej losowej X. Chcemy na podstawie próby $\{x_i\}$ wiedzieć, czy zmienna X ma rozkład prawdopodobieństwa charakteryzujący się pewnymi własnościami, tzn. czy znany jest rozkład hipotetyczny (np. przypuszczamy, że próba $\{x_i\}$ uzyskana została za pomocą niezależnego losowania z populacji o rozkładzie normalnym N(0,1)), czy też jej rozkład należy do określonej klasy rozkładów (np.normalnych).

Rozważmy przypadek, gdy rozkład hipotetyczny jest w zupełności określony za pomocą dystrybuanty $F(x)$ rozkładu normalnego. Chcemy zweryfikować hipotezę statystyczną, że próba $\{x_i\}$ została pobrana z populacji o tym rozkładzie. Przyjmijmy, że sprawdzana hipoteza jest prawdziwa. Wtedy możemy oczekiwać, że dystrybuanta empiryczna $F_n(x)$ będzie przybliżeniem dystrybuanty $F(x)$. Określenie pewnej nieujemnej miary $\mathcal{D}$ odchylenia $F(x)$ i $F_n(x)$ można podać w rozmaity sposób. Jednakże każda miara $\mathcal{D}$ będzie pewną funkcją wartości z próby, a zatem będzie miała określony rozkład z próby. Korzystając z tego rozkładu, określamy prawdopodobieństwo $P(\mathcal{D} > \mathcal{D}_o)$ zdarzenia losowego $\mathcal{D} > \mathcal{D}_o$, gdzie $\mathcal{D}_o$ jest zadaną wielkością. Miara $\mathcal{D}$ jest statystyką z próby, a $\mathcal{D}_o$ wartością krytyczną jej rozkładu przy zadanym poziomie istotności α. Jeżeli w aktualnym przypadku prawdziwa jest nierówność $\mathcal{D} > \mathcal{D}_o$, to wnioskujemy, że odchylenie jest istotne i hipotezę odrzucamy. Z drugiej strony, gdy wystąpi nierówność $\mathcal{D} \leq \mathcal{D}_o$ to uważamy, że odchylenie mogło powstać na skutek wahań losowych i nie mamy podstaw do odrzucenia hipotezy. Test o takim charakterze nazywamy testem zgodności dla hipotez o rozkładzie normalnym (krótko testem normalności).

Formalnie, hipotezę statystyczną o zgodności rozkładu empirycznego z rozkładem normalnym formułujemy następująco. Niech F_N i G oznaczają klasy dystrybuant, odpowiednio rozkładu normalnego i zmiennych losowych mających trzeci różny od zera i skończony czwarty moment centralny. Przyjmujemy, że X jest zmienną losową o nieznanej postaci (tak co do klasy, jak i parametrów) dystrybuanty $F(x)$. Stawiamy hipotezę zerową H_o: $F(x) \in F_N$, że funkcja $F(x)$ należy do klasy F_N, przeciwko hipotezie alternatywnej H_1: $F(x) \in G$, że funkcja $F(x)$ należy do klasy G. Hipotezę H_o uznajemy za prostą (oznaczamy H_o^p), jeżeli klasa F_N ma określone parametry μ i σ, w przeciwnym razie jest ona złożona (oznaczamy H_o^z).

Weryfikację hipotezy H_o przeprowadzamy testami normalności, które można podzielić zależnie od sposobu ich konstrukcji na 4 grupy:

(1) Test χ^2 Pearsona,

(2) Testy oparte na porównaniu dystrybuanty rozkładu empirycznego z dystrybuantą rozkładu normalnego,

(3) Testy oparte na momentach z próby,

(4) Testy oparte na statystykach pozycyjnych.

Ze względu na powszechną znajomość testu χ^2 Pearsona nie będziemy go tutaj omawiać. Dalej w pracy dokonujemy omówienia testów normalności w obrębie każdej z wymienionych grup, podając statystyki na których oparte są testy i sposób w jaki z zaobserwowanej wartości tej statystyki można wnioskować o hipotezie H_o. Dla wielu testów zostały dołączone przykłady liczbowe.

2. Testy dla hipotez o normalności rozkładu

2.1. Testy oparte na porównaniu dystrybuanty rozkładu empirycznego z dystrybuantą rozkładu normalnego.

Niech X będzie zmienną losową o rozkładzie określonym przez dystrybuantę $F(x)$ i niech $\{x_i\}$ stanowi próbę obserwacji zmiennej X. Ponadto niech $\{y_i\}$ stanowi ciąg (μ,σ) - s.p. w próbie $\{x_i\}$. Przyjmujemy, że $z_i = \Phi(u_i)$ dla $i = 1,2,\ldots,n$, gdzie (a) $u_i = (y_i-\mu)/\sigma$, gdy μ i σ znane, (b) $u_i = (y_i - \bar{y})/s$, gdy μ i σ nieznane, (c) $u_i = (y_i - \bar{y})/\sigma$, gdy μ nieznane i σ znane oraz (d) $u_i = (y_i - \mu)/\tilde{s}$, gdy μ znane i σ nieznane, zaś $\tilde{s}^2 = \sum_{i=1}^{n} (y_i - \mu)^2/n$. Wartości dystrybuanty $F_n(x)$ w punkcie y_i oznaczamy przez $F_n(y_i)$. Miarę odległości rozkładu empirycznego od normalnego wyrażamy tutaj funkcją odległości między $F_n(y_i)$ oraz z_i o postaci

$$\mathcal{D} = \max_{1 \le i \le n} \{F_n(y_i) - z_i\}$$

lub w postaci całkowej

$$\mathcal{D} = \int_0^1 [F(y) - F_n(y)]^2 \, dF(y).$$

Niech τ będzie testem normalności o statystyce testowej określonej na mierze $\mathcal{D}$ i niech ponadto $\tau_{\alpha;n}$ będzie wartością krytyczną rozkładu τ dla ustalonych α i n. Hipotezę H_0 odrzucamy na ustalonym poziomie istotności α wtedy i tylko wtedy, gdy $\tau > \tau_{\alpha;n}$, w przeciwnym razie nie ma podstaw do odrzucenia hipotezy H_0.

Dla wielu testów należących do omawianej grupy zostały stablicowane wartości krytyczne (krótko w.k.) dla przypadku (a), tzn. kiedy parametry μ i σ są znane. Stephens (1974) podał w.k. testów przy różnych α, dokonując modyfikacji takiej, aby w.k. nie zależały od wielkości próby. Modyfikację testu τ oznaczamy przez $T(\tau)$, a jego w.k. dla zadanego α przez $T_\alpha(\tau)$.

Poniżej prezentujemy testy: Kołmogorowa, Kołmogorowa-Smirnowa, Kac-Kiefer-Wolfowitza, Kuipera, Watsona, Cramera-von Misesa i Anderson-Darlinga, które zalicza się do dwóch klas testów: Kołmogorowa i Cramera-von Misesa.

T e s t K o ł m o g o r o w a. Statystyka testowa D testu Kołmogorowa jest postaci

$$D = \max(D^+, D^-),$$

gdzie

$$D^+ = \max_{1 \le i \le n} (i/n - z_i) \quad i \quad D^- = \max_{1 \le i \le n} (z_i - (i-1)/n).$$

Do weryfikacji H_o^p można stosować testy oparte na statystykach D^+ lub D^-, zaś dla H_o^z test D. Modyfikacje powyższych testów wraz z w.k. dla $\alpha = 0.05, 0.01$ w przypadkach (a), (b) i (d) są:

Statystyka	$T(D^+)$, $T(D^-)$, $T(D)$	n	$T_{0.05}(\cdot)$	$T_{0.01}(\cdot)$
$D^+(D^-)$	(a) $D^+(\sqrt{n}+0.12+0.11/\sqrt{n})$		1.224	1.518
D	(a) $D(\sqrt{n}+0.12+0.11/\sqrt{n})$		1.358	1.628
	(b) $D(\sqrt{n}-0.01+0.85/\sqrt{n})$		0.895	1.035
	(d) $\sqrt{n}\ D$	10	1.270	1.530
		20	1.290	1.570
		50	1.310	1.595
		100	1.321	1.610
			1.333	1.625

Przykład 1. Wagi mężczyzn w funtach przedstawiały się następująco: 148, 154, 158, 160, 161, 162, 166, 170, 182, 195, 236. Obserwację 236 można uznać za odstającą. Charakterystykami opisowymi próby są: $\bar{y} = 172$, $S^2 = 6225.02$ i $s = 24.95$. Weryfikujemy hipotezę H_o, iż próba z obserwacji wag mężczyzn pochodzi z populacji o rozkładzie normalnym. Przyjmujemy poziom istotności $\alpha = 0.01$.

Obliczenia przeprowadzamy wg schematu

i	y_i	u_i	z_i	$i/n-z_i$	$z_i-(i-1)/n$
1	148	-0.96	0.1685	-0.0775	0.1685
2	154	-0.72	0.2358	-0.0539	0.1449
3	158	-0.56	0.2877	-0.0149	0.1059
4	160	-0.48	0.3156	0.0480	0.0428
5	161	-0.44	0.3300	0.2145	-0.0336
6	162	-0.40	0.3446	0.2008	-0.1099
7	166	-0.24	0.4052	0.2312	-0.1402
8	170	-0.08	0.4681	0.2592	-0.1683
9	182	0.40	0.6554	0.1628	-0.0718
10	195	0.92	0.8212	0.0879	0.0030
11	236	2.56	0.9948	0.0052	0.0857

Mamy zatem $D^+ = 0.2592$ i $D^- = 0.1685$, więc $D = \max(0.2592, 0.1685) = 0.2592$. Ponieważ parametry μ i σ są nieznane, więc obliczamy przekształconą wartość statystyki D wzorem $T(D) = D(\sqrt{n}-0.01+0.85/\sqrt{n}) = 0.924$. Skoro $T(D) = 0.924 < 1.035 = T_{0.01}(D)$, więc nie ma podstaw do odrzucenia hipotezy H_o, iż próba 11 obserwacji pochodzi z populacji o rozkładzie normalnym na poziomie istotności $\alpha = 0.01$.

Pewną odmianą testu D jest test KS Kołmogorowa-Smirnowa oparty na statystyce

$$KS = \max_{1 \leq i \leq n} |i/n - z_i|.$$

Powyższy test może być stosowany nawet dla małych prób, a tablice w.k. dla przypadku (a) są powszechnie znane (patrz, np.Fisz, 1967, tab.VIII), natomiast dla przypadku (b) podał je Lilliefors (1967), a cytuje je także Domański (1979, tab.10).

Inną wersją testu D dla przypadku (b) jest test V_n Kac-Kiefer-Wolfowitza, oparty na statystyce

$$V_n = \sqrt{n} \sup |F_n(y) - \Phi((y-\bar{y})/s|,$$

dla którego w.k. rozkładu V_n podał Soesta (1967).

Kuiper wprowadził test V oparty na statystyce $V = D^+ + D^-$, gdzie statystyki D^+ i D^- są określone jak w teście D. Test V może być stosowany zarówno do prostych jak i złożonych hipotez o normalności. Modyfikacja dla przypadków (a),(b) i (d) wraz z w.k. jest postaci

	T(V)	n	$T_{0.05}(V)$	$T_{0.01}(V)$
(a)	$V(\sqrt{n}+0.155+0.24/\sqrt{n})$		1.747	2.001
(b)	$V(\sqrt{n}+0.05+0.82/\sqrt{n})$		1.489	1.693
(d)	$\sqrt{n}\ V$	10	1.500	1.710
		20	1.535	1.770
		50	1.570	1.810
		100	1.590	1.825
		∞	1.612	1.845

T e s t C r a m e r a - v o n M i s e s a. Jeśli miarę odchylenia dystrybuanty empirycznej $F_n(y)$ od dystrybuanty rozkładu normalnego $F(y)$ wyrazimy w postaci całkowej

$$CM = n \int_0^1 [F_n(y) - F(y)]^2 \, dF(y),$$

to otrzymamy statystykę testu CM Cramera-von Misesa. W praktyce korzysta się z równoważnej do CM statystyki

$$W^2 = 1/(12n) + \sum_{i=1}^{n} [z_i-(2i-1)/(2n)]^2.$$

Test ten stosowany jest do hipotez H_o^p i H_o^z, a jego modyfikacje dla przypadków (a) - (d) wraz z w.k. są postaci

	$T(W^2)$	$T_{0.05}(W^2)$	$T_{0.01}(W^2)$
(a)	$(W^2-0.4/n+0.6/n^2)(1+1/n)$	0.461	0.743
(b)	$W^2(1+0.5/n)$	0.125	0.178
(c)	nie podlega modyfikacji	0.165	0.237
(d)	$n \geq 5$	0.443	0.723

Przykład 2. Korzystając z danych w przykładzie 1, obliczamy wielkości $w_i = z_i - (2i-1)/(2n)$ dla $n = 11$ oraz $i=1,2,\ldots,11$, które są następujące: 0.123, 0.099, 0.060, -0.003, -0.079, -0.155, -0.186, -0.214, -0,227, -0.042, 0.040. Ich suma kwadratów wynosi 0.1564, a stąd $W^2 = 0.1639$. Skoro występuje przypadek (b), więc $T(W^2) = 0.1713$. Ponieważ $T_{0.01}(W^2) = 0.178 > 0.171$, więc nie ma podstaw do odrzucenia hipotezy H_0.

Jedną z wersji testu CM jest test U^2 Watsona, oparty na statystyce

$$U^2 = W^2 - n(\bar{z} - 0.5)^2,$$

gdzie $\bar{z} = (z_1 + \ldots + z_n)/n$. Jego modyfikacje wraz z w.k. są postaci

	$T(U^2)$	$T_{0.05}(U^2)$	$T_{0.01}(U^2)$
(a)	$(U^2-0.1/n+0.1/n^2)(1+8/n)$	0,187	0.267
(b)	$U^2(1+0.5/n)$	0.116	0.163
(c)	nie podlega modyfikacji	0.157	0.227
(d)	$n \geq 5$	0.153	0.221

Przykład 3. Z obliczeń w przykładzie 2 mamy $W^2 = 0.1635$. Sumując elementy z_i, $i = 1,2,\ldots,11$ podane w przykładzie 1, otrzymujemy $z_1 + \ldots + z_{11} = 5.0269$, a stąd $\bar{z} = 0.457$ i $U^2 = 0.143$. Ponieważ $T(U^2) = 0.140 < 0.163 = T_\alpha(U^2)$, więc nie ma podstaw do odrzucenia H_0.

Inną wersją testu CM w postaci całkowej jest test WCM wagowy Cramera-von Misesa, oparty na statystyce

$$WCM = n \int_0^1 [F_n(y) - F(y)]^2 \frac{dF(y)}{F(y)(1-F(y))} ,$$

którą otrzymuje się ze statystyki CM przez nadanie jej wagi $1/F(y)(1-F(y))$. W praktyce korzysta się z modyfikacji testu WCM zwanego testem A^2 Anderson-Darlinga, opartego na statystyce

$$A^2 = -n - (1/n) \sum_{i=1}^{n} (2i-1)[\ln z_i + \ln(1-z_{n-i+1})],$$

która zmodyfikowana dla przypadku (b) wraz z w.k. dla (a)-(d) jest

$T(A^2)$		$T_{0.05}(A^2)$	$T_{0.01}(A^2)$
(a)	$n \geq 5$	2.492	3.857
(b)	$A^2(1+4/n-25/n^2)$	0.787	1.092
(c)	nie podlega modyfikacji	1.105	1.573
(d)	$n \geq 5$	2.323	3.690

Przykład 4. Korzystając z danych w przykładzie 1 oraz podanych tam wartości z_i, obliczamy odpowiednie logarytmy naturalne i sumę występującą po prawej stronie statystyki A^2, otrzymując -131.415, a stąd $A^2 = 0.947$. Ponieważ parametry μ i σ są nieznane, więc $T(A^2) = 1.096$. Skoro $T(A^2) = 1.096 > 1.092 = T_{0.01}(A^2)$, więc hipotezę H_o odrzucamy na poziomie istotności $\alpha = 0.01$. Zauważmy, że tylko test A^2 odrzucił hipotezę H_o, iż próba wag 11 mężczyzn pochodzi z populacji o rozkładzie normalnym. Wspominaliśmy, że w rozważanej próbie występuje obserwacja odstająca $y_{11} = 236$. Wskazuje to na pewien typ odstępstwa od normalności, tzn. próba wykazuje nadmierną skośność wywołaną przez wynik $y_{11} = 236$.

Omawianiem mocy wspomnianych testów zajmujemy się w punkcie 3.

2.2. Testy normalności oparte na momentach z próby.

Niech X będzie zmienną losową o skończonych trzecim μ_3 i czwartym μ_4 momencie centralnym, $\{x_i\}$ próbą prostą obserwacji zmiennej X, zaś m_r, r-tym momentem centralnym z próby

$$m_r = (1/n) \sum_{i=1}^{n} (x_i - \bar{x})^r, \qquad r = 2,3,4,\ldots$$

Testy normalności oparte na statystykach korzystających z momentów m_r związane są następującą własnością. Jeśli $X \sim N(\mu,\sigma^2)$, to $\mu_3 = 0$ i $\mu_4 = 3\sigma^4$, a standaryzowane trzeci i czwarty moment centralny są postaci $\sqrt{\beta_1} = \mu_3/\sigma^3 = 0$ oraz $\beta_2 = \mu_4/\sigma^4 = 3$. Innymi słowy $\sqrt{\beta_1}$ i β_2 są odpowiednio równe zero i 3 dla populacji o rozkładzie normalnym. Testy normalności w omawianej grupie stosowane są do sprawdzania hipotez H_o^z i są niezmiennicze ze względu na zmianę skali i położenie próby $\{x_i\}$.

T e s t $\sqrt{b_1}$ - s t a n d a r y z o w a n y t r z e c i m o m e n t z p r ó b y. Statystyka $\sqrt{b_1}$ testu $\sqrt{b_1}$ jest postaci

$$\sqrt{b_1} = m_3/\sqrt{m_2^3} = \sqrt{n} \sum_{i=1}^{n} (x_i - \bar{x})^3/(S^2)^{3/2},$$

gdzie S^2 jest sumą kwadratów odchyleń. Test $\sqrt{b_1}$ stosowany jest do weryfikacji hipotez H_o^z, przeciwko hipotezie H_1 orzekającej, iż rozkład jest skośny. W.k. rozkładu $\sqrt{b_1}$ dla $n > 25$ podane zostały w zbiorze tablic Pearsona i Hartley'a (1966, tab. 34B), a dla $n \leq 25$ przez Mulhollanda (1977).

T e s t b_2 - s t a n d a r y z o w a n y c z w a r t y m o m e n t z p r ó b y. Statystyka b_2 testu b_2 jest postaci

$$b_2 = m_4/m_2^2 = n \sum_{i=1}^{n} (x_i - \bar{x})^4/(s^2)^2.$$

Test b_2 stosowany jest do weryfikacji hipotez H_o^z, przeciwko hipotezie H_1 orzekającej, iż rozkład jest symetryczny o spłaszczeniu odbiegającym od normalnego. W.k. przy $n \geq 50$ zawiera zbiór tablic wyżej wymienionych autorów.

Przykład 5. Dla danych z przykładu 1 mamy: $n = 11$, $\bar{x} = 172$, $m_2 = 565.91$, $m_3 = 22602.54$, $m_4 = 1598992.13$, a stąd $\sqrt{b_1}=1.679$ i $b_2 = 4.993$. Ponieważ w.k. rozkładu $\sqrt{b_1}$ dla $n = 11$ oraz $\alpha= 0.01$ wynosi 1.899 (Mulholland, 1977, tab.2, s.408) jest większa od 1.679, więc nie ma podstaw do odrzucenia hipotezy H_o.

T e s t K^2. Niech $X(p)$ oznacza kwantyl rzędu p-tego rozkładu zmiennej losowej X. Dla statystyki b_2 określamy prawdopodobieństwo $0 < p < 1$ równe $P(b_2 \leq b_{2,p;n}) = p$, gdzie $b_{2,p;n}$ jest w.k. rozkładu b_2 dla ustalonych p i n, którą można odczytać m.in. z nomogramów danych przez D'Agostino i Pearsona (1973), pozwalających dla danego b_2 odczytać p. Oznaczmy przez $X(b_2)$ kwantyl rzędu p rozkładu N(0,1), czyli $X(b_2) = \Phi^{-1}(p)$. Dla rozkładu $\sqrt{b_1}$ kwantyl $X(\sqrt{b_1})$ określamy wzorem

$$X(\sqrt{b_1}) = \partial \ln\{\sqrt{b_1}/\lambda + [(\sqrt{b_1}/\lambda)^2 + 1]^{1/2}\},$$

gdzie stałe ∂ i $1/\lambda$ są stablicowane przez wspomnianych autorów. Test K^2 oparty jest na statystyce

$$K^2 = X^2(\sqrt{b_1}) + X^2(b_2),$$

która przy prawdziwości hipotezy H_o ma rozkład χ^2 z 2 stopniami swobody. Test K^2 prowadzi do odrzucenia hipotezy H_o na ustalonym poziomie istotności, gdy $K^2 > \chi^2_{\alpha;2}$.

Zaletą testu K^2 jest to, że pozwala on jednocześnie badać odstępstwo od normalności wywołane przez skośność i spłaszczenie. Test o takich własnościach nazywamy testem omnibus. Test K^2 może być stosowany dla prób o liczebnościach $20 \leq n \leq 200$. Zauważmy, że dla dużych prób ($n > 200$) w miejsce K^2 stosujemy

$$\tilde{K}^2 = n[4(\sqrt{b_1})^2 + (b_2 - 3)^2]/24,$$

która ma asymptotyczny rozkład χ^2 z 2 stopniami swobody.

Przykład 6. Dla próby prostej o liczebności $n = 82$ wyznaczono: $\sqrt{b_1} = 0.506$ i $b_2 = 4.621$. Aby obliczyć $X(\sqrt{b_1})$ odczytujemy stałe: $\partial = 3.686$ i $1/\lambda = 1.0785$ (D'Agostino i Pearson, tab.4, s.621), wtedy $X(\sqrt{b_1}) = 1.0795 \ln\{0.506 \cdot 1.0795 + [(0.506 \cdot 1.0795)^2 + 1]^{1/2}\} = 1.9247$. Z kolei z nomogramów odczytujemy dla $b_2 = 4.621$ i $n = 82$ prawdopodobieństwo $p = 0.992$ i następnie $X(b_2) = \Phi^{-1}(0.992) = 2.4089$. Zatem $K^2 = (1.9247)^2 + (2.4089)^2 = 9.5071$. Ponieważ $\chi^2_{0.01;2} = 9.210 < 9.5071$, więc hipotezę H_o odrzucamy. Można sugerować, że odpowiedzialne za odrzucenie H_o jest nadmierne spłaszczenie odbiegające znacznie od liczby 3.

2.3. Testy oparte na statystykach pozycyjnych.

Niech $y_1, y_2, \ldots, y_n$ będzie ciągiem (μ, σ) - s.p. wyznaczonych z próby prostej $\{x_i\}$. Zastosowanie s.p. do konstrukcji testów normalności ujawnia się w tym, że przy prawdziwości hipotezy H_o (tzn. hipotezie, że próba $\{x_i\}$ pochodzi z populacji o rozkładzie normalnym) zmienne losowe y_i mają wartości oczekiwane i wariancje wyrażone znanymi funkcjami liniowymi parametrów μ i σ

(patrz, punkt 1.2), co pozwala zastosować metodę najmniejszych kwadratów do estymacji parametrów μ i σ (Lloyd, 1952). Najogólniej, najlepszy nieobciążony liniowy estymator $\hat{\sigma}$ parametru σ można wyrazić w postaci liniowej kombinacji obserwacji $\{y_i\}$

$$\hat{\sigma} = \sum_{i=1}^{h} d_{i,n} (y_{n-i+1} - y_i),$$

gdzie $h = n/2$ lub $h = (n-1)/2$ zależnie od tego czy n jest liczbą parzystą lub nieparzystą, zaś współczynniki $d_{i,n}$ spełniają pewne żądane własności (np. suma współczynników jest równa zero). Z drugiej strony nieobciążony estymator z próby parametru σ wyraża odchylenie standardowe z próby s. Wówczas kwadrat ilorazu tych dwóch statystyk $(\hat{\sigma}/s)^2$ z pominięciem stałej jest bliski jedności. Na tej zasadzie skonstruowane są testy: W, W', $\tilde{W}$, D_A i r. Formalnie ujmują one informację zawartą w NPP[m_i : y_i], w którym układ punktów (m_i, y_i), $i = 1,2,\ldots,n$ powinien pokrywać się z prostą regresji, jeśli próba $\{x_i\}$ pochodzi z populacji normalnej. Odchylenie od prostej reprezentuje wtedy miarę odstępstwa rozkładu zmiennych y_i od rozkładu normalnego, wyrażoną przez $\hat{\sigma}$.

Na innych zasadach skonstruowano testy g i u. Tutaj statystyki są wyrażone także w postaci ilorazowej, ale za estymator parametru σ jest brane odpowiednio odchylenie przeciętne i rozstęp z próby.

T e s t g G e a r y. Z niezmienniczości testu g na zmianę skali i położenia obserwacji w układzie współrzędnych, można statystykę g tego testu wyznaczyć bezpośrednio z próby $\{x_i\}$ lub $\{y_i\}$. Jest ona ilorazem odchylenia przeciętnego i standardowego z próby i wyraża się wzorem

$$g = \sum_{i=1}^{n} |y_i - \bar{y}| \,/(nS^2)^{1/2}.$$

Rozkład g przy prawdziwości H_o został znaleziony, a jego w.k. zostały podane w zbiorze tablic Pearsona i Hartley'a (1966, tab. 34 A). Test g jest stosowany dla małych prób i weryfikacji H_o przeciwko H_1 orzekającej, że rozkład jest symetryczny.

Przykład 7. Dla danych z przykładu 1 mamy: $n = 11$, $s^2 = 6225.02$ oraz $|y_1-\bar{y}|+\ldots+|y_{11}-\bar{y}| = 194$, zatem $g = 194/(11 \cdot 6225.02)^{1/2} = 0.741$. Ponieważ $g_{0.01;11} = 0.936 > 0.741$, więc hipotezy H_o nie mamy podstaw do odrzucenia na poziomie istotności $\alpha = 0.01$.

T e s t u D a v i d a - H a r t l e y'a - P e a r s o n a. Statystyka u testu u wyraża się ilorazem rozstępu i odchylenia standardowego z próby i jest postaci

$$u = (y_n - y_1)/s.$$

Test u jest zalecany do H_o^z, przeciwko H_1 orzekającej, że rozkład jest symetryczny. W.k. rozkładu u zawarte są we wspomnianym zbiorze tablic (tab.29C).

T e s t T' S p i e g e l h a l t e r a. Zauważmy, że występujące w próbie obserwacje odstające prowadzą do pomniejszenia i powiększenia odpowiednio wartości statystyk g i u. Spiegelhalter (1977) skonstruował omnibus test T' jako kompozycję testów g i u, opartego na statystyce

$$T' = [1/(c_n \cdot u)^{n-1} + 1/g^{n-1}]^{1/(n-1)},$$

gdzie $c_n = (1/2n)(n!)^{1/(n-1)}$. Dla zastosowań praktycznych testu T' podajemy dla niektórych n wartości c_n i $T'_{\alpha;n}$ gdy $\alpha = 0.05, 0.10$:

n	5	10	15	20	50	100
c_n	0.331	0.268	0.245	0.232	0.207	0.197
$T'_{0.05;n}$	1.532	1.453	1.423	1.403	1.337	1.308
$T'_{0.10;n}$	1.512	1.417	1.387	1.369	1.317	1.294

Przykład 8. Dla danych z przykładu 1 mamy: $n=11$, $u = 3.537$ i $g = 0.741$. Wyznaczamy $c_{11} = 0.262$ i $T' = [1/(0.262\cdot 3.527)^{10} + 1/0.741^{10}]^{0.1} = 1.364$. Weryfikujemy hipotezę H_o przy $\alpha = 0.05$. Żądaną wartość krytyczną interpolujemy wg wzoru

$$T'_{0.05;11} = T'_{0.05;10} + \frac{T'_{0.05;10} - T'_{0.05;15}}{10 - 15}(11 - 10) =$$

$$= 1.453 + (1.453 - 1.423)/(-5) = 1.447.$$

Ponieważ $T' = 1.364 < 1.447$, więc nie ma podstaw do odrzucenia hipotezy H_o na poziomie istotności $\alpha = 0.05$.

T e s t W S h a p i r o - W i l k a. Statystyka W testu W jest ilorazem kwadratu najlepszego liniowego nieobciążonego estymatora $\hat{\sigma}$ parametru σ i wariancji z próby. Estymator $\hat{\sigma}$ jest liniową kombinacją s.p. $\{y_i\}$ o postaci

$b = \sum_{i=1}^{h} a_{i,n}(y_{n-i+1} - y_i)$, gdzie $a_{i,n}$ są stałymi spełniającymi warunki: $\sum_{i=1}^{n} a_{i,n} = 0$ i $\sum_{i=1}^{n} a_{i,n}^2 = 1$ i zawarte są w zbiorze tablic Pearsona i Hartley'a (1972, tab.12, także Domański, 1979, tab.11). Statystyka W jest postaci

$$W = b^2/S^2 = [\sum_{i=1}^{h} a_{i,n}(y_{n-i+1} - y_i)]^2 / \sum_{i=1}^{n}(y_i - \bar{y})^2.$$

W.k. rozkładu W dla $3 \leq n \leq 50$ także zawierają się we wspomnianym wyżej zbiorze tablic (tab.14). Test W ma własność testu omnibus. Niskie wartości W wskazują na odstępstwo od normalności. Formalnie, jeśli $W \leq W_{\alpha;n}$, to hipotezę H_o odrzucamy na ustalonym poziomie istotności α.

Przykład 9. Korzystamy z danych zawartych w przykładzie 1. Obliczenia związane z wyznaczaniem b przeprowadzamy wg schematu

i	y_{12-i}	y_i	$a_{i,11}$	$a_{i,11}(y_{12-i}-y_i)$
1	236	148	0.5601	49.2888
2	195	154	0.3315	13.5915
3	182	158	0.2260	5.4240
4	170	160	0.1429	1.4290
5	166	161	0.0695	0.0695
			Suma	69.8028

Skoro $S^2 = 6225.02$ i $b = 69.8028$, to $W = (69.8028)^2/6225.02 = 0.783$. Ponieważ $W = 0.783 < 0.792 = W_{0.01;11}$, więc hipotezę H_o odrzucamy. Zauważmy, że jeśli z próby złożonej z pozostałych 10 obserwacji obliczymy $W = 0.908$, to nie będzie podstaw do odrzucenia hipotezy H_o, gdyż $W_{0.01;10} = 0.781$. Wnioskujemy stąd, że obserwacja 236 jest odpowiedzialna za wystąpienie asymetrii prawostronnej. Test W jak widać jest czuły na odstępstwo od normalności wywołane przez skośność.

T e s t W′ S h a p i r o - F r a n c i a. Statystyka W′ testu W′ jest wyrażona w postaci identycznej jak statystyka W z tym, że współczynniki $a_{i,n}$ są zastąpione przez $c_{i,n} = m_{i,n}/[\sum_{i=1}^{n} m_{i,n}^2]^{1/2}$, które są znormalizowanymi wartościami oczekiwanymi N(0,1) - s.p. Statystyka W′ wyraża się w postaci

$$W' = [\sum_{i=1}^{h} m_{i,n}(y_{n-i+1} - y_i)]^2/(\tilde{c}_n S^2),$$

gdzie $\tilde{c}_n = m^2_{1,n} + \ldots + m^2_{n,n}$. W.k. $W'_{\alpha;n}$ rozkładu W' dla $50 \le n \le 99$ podali Shapiro i Francia (1972).

Przykład 10. Ciężary zebranych owoców z 60 krzewów czerwonej porzeczki w kg przedstawiały się następująco:

7.5	3.8	4.9	7.9	8.5	10.2	11.0	5.8	5.2	9.1
8.3	6.7	8.3	9.5	6.4	5.5	7.3	8.2	11.6	7.1
6.2	8.5	9.5	8.3	5.9	12.6	10.8	4.9	6.3	3.9
8.2	8.6	6.9	7.5	6.9	3.3	5.9	7.3	3.1	7.8
6.3	9.3	7.1	6.6	7.6	7.8	6.9	7.5	10.2	5.4
8.9	7.3	6.5	4.2	8.6	8.1	7.3	5.3	6.9	6.2

Przypuszczamy, że obserwacje ciężaru owoców porzeczki z krzewów pochodzą z populacji o rozkładzie normalnym, przyjmując $\alpha = 0.05$. Charakterystykami opisowymi próby o liczebności $n=60$ są: $\bar{y} = 7.32$, $S^2 = 229.39$, $s = 1.955$, współczynniki skośności $g_1 = 0.198$ i spłaszczenia $g_2 = 0.186$. Odczytujemy ze wspomnianych tablic Pearsona i Hartley'a wartości $m_{i,60}$ dla $i = 1,2,\ldots,60$. Obliczenia prowadzi się wg poniższego schematu ($r_i = y_{n-i+1} - y_i$)

i	y_{61-i}	y_i	r_i	$m_{i,60}$	$r_i m_{i,60}$
2	12.6	3.1	9.5	2.319	22.031
2	11.6	3.3	8.3	1.935	16.061
....	......	......	......	......	......
29	7.3	7.3	0	0.041	0
30	7.3	7.3	0	0.041	0
				Suma	113.913

Zatem $W' = (113.913)^2/(57.3767 \cdot 229.39) = 0.986 > 0.963 = W'_{0.05;60}$, nie ma więc podstaw do odrzucenia hipotezy H_0.

Należy tutaj zaznaczyć, że zarówno przy obliczaniu statystyki W jak i W' trzeba dla powtarzających się kolejnych różnic $y_{n-i+1} - y_i$, brać średnią arytmetyczną z wag dla tych obserwacji.

T e s t $\widetilde{W}$ W e i s b e r g a - B i n g h a m a. Jeśli wielkości $m_{i,n}$ w teście W' zastąpimy przez

$$\widetilde{m}_{i,n} = \Phi^{-1}\left(\frac{i - 3/8}{i + 1/4}\right), \qquad i = 1,2,\ldots,n,$$

to otrzymamy test $\widetilde{W}$ oparty na statystyce

$$\widetilde{W} = [\sum_{i=1}^{n} m_{i,n} y_i]^2/(S^2 \cdot \sum_{i=1}^{n} m_{i,n}^2),$$

dogodny do obliczeń na maszynie cyfrowej, gdyż obliczanie kwantyli $\widetilde{m}_{i,n}$ przeprowadza się standardowym programem. Przeprowadzone badania metodą symulacji Monte Carlo wykazały, iż wartości dystrybuanty testów W' i $\widetilde{W}$ różnią się między sobą nieznacznie. Pozwala to zamiennie stosować test W' lub $\widetilde{W}$.

T e s t D_A D'A g o s t i n o. Test W może być stosowany dla prób o liczebnościach $n \leq 50$. Na podobnych zasadach, ale bez konieczności wyznaczania wag, skonstruowany jest test D_A oparty na statystyce D_A, która jest ilorazem Downtona liniowego nieobciążonego estymatora

$$\sigma = 2\sqrt{\pi} \sum_{i=1}^{n} (i - (n+1)/2) y_i / n(n-1)$$

parametru σ i odchylenia standardowego z próby i wyraża się w postaci

$$D_A = [\sum_{i=1}^{n} i y_i - ((n+1)/2) \sum_{i=1}^{n} y_i]/(n^3 S^2)^{1/2} .$$

Test D_A zalecany jest dla dużych prób $n \geq 50$ i posiada własności testu omnibus. W praktyce dokonuje się modyfikacji statystyki D_A do postaci

$$Y = \sqrt{n}(D_A - 0.282095)/0.029986,$$

dla której zostały wyznaczone w.k. (D'Agostino, 1971, s.343).

Przykład 11. Korzystając z danych zawartych w przykładzie 10, mamy:

$$S^2 = 229.39, \quad \sum_{i=1}^{60} y_i = 438.8, \quad \sum_{i=1}^{60} y_i \cdot i = y_1+2y_2+\ldots+60y_{60} =$$

$= 15340.8$, a stąd $D_A = (15340.8 - 30.5 \cdot 438.8)/(60^3 \cdot 229.39)^{1/2} =$ $= 0.278077$. Z kolei obliczamy $Y=\sqrt{60}(0.278077-0.282095)/0.029986 =$ $= -1.038$. Dla $n = 60$ i $\alpha = 0.01$ mamy dwie w.k.: -3.846 i 1.301. Ponieważ $-3.846 < -1.038 < 1.301$, więc nie ma podstaw do odrzucenia hipotezy H_o.

Test r współczynnika korelacji normalnego diagramu prawdopodobieństwa.

Test r jest skonstruowany na podobnych zasadach jak test W ze zmienionymi wagami. W miejsce $a_{i,n}$ stosujemy mediany $M_{i,n}$ ciągu N(0,1) - s.p., które wyznaczamy wzorem $M_{i,n} = \Phi^{-1}(m_i)$, gdzie $m_n = (0.5)^{1/n}$, $m_1 = 1 - m_n$ oraz $m_i = (i - 0.3175)/(n + 0.365)$, $i = 2,3,\ldots,n-1$. Statystyka r omawianego testu jest określona ilorazem współczynnika korelacji prostoliniowej r w NPP$[M_{i,n} : y_i]$ i odchylenia standardowego z próby i wyraża się wzorem

$$r = \sum_{i=1}^{n} M_{i,n}y_i/[S^2 \cdot \sum_{i=1}^{n} M_{i,n}^2]^{1/2}.$$

W.k. rozkładu r podał Fillibren (1975) dla $3 \leq n \leq 100$. Niskie wartości r wskazują na odstępstwo od rozkładu normalnego. Test r jest stosowany do hipotez H_o^z i jest testem jednostronnym.

Przykład 12. Niech dana będzie próba $n = 7$ obserwacji: 6, 1, -4, 8, -2, 5, 0, dla której $S^2 = 118$. Porządkujemy obserwacje: -4, -2, 0, 1, 5, 6, 8 i wyznaczamy wielkości m_i, $i = 1,2,...,7$, którymi są: $m_1 = 0.0943$, $m_2 = (2 - 0.3175)/(7 + 0.365) = 0.2284$, $m_3 = 0.3642$, $m_4 = 0.5000$, $m_5 = 0.6358$, $m_6 = 0.7716$ i $m_7 = 0.9057$. Korzystając z tablic kwantyli rozkładu $N(0,1)$ (Zieliński, 1972, tab.3) odczytujemy mediany $M_{i,n}$, którymi są: $M_{1,7} = \Phi^{-1}(m_1) = \Phi^{-1}(0.0943) = -1.3149 = -M_{1,7}$, $M_{2,7} = -0.7439 = -M_{6,7}$, $M_{3,7} = -0.3468 = -M_{5,7}$ i $M_{4,7} = 0$. Obliczamy $M_{1,7}^2 + ... + M_{7,7}^2 = 4.8052$ i $y_1 M_{1,7} + ... + y_7 M_{7,7} = 23.4640$, a stąd $r = 23.4640/(4.8052 \cdot 118)^{1/2} = 0.985$. Ponieważ $r = 0.985 > 0.899 = r_{0.05;7}$, więc nie ma podstaw do odrzucenia hipotezy H_o na poziomie istotności $\alpha = 0.05$.

T e s t T H e g a z y - G r e e n a. Test T jest skonstruowany na zasadach testu zgodności. Różnica występuje w tym, że w miejsce wartości $F_n(y_i)$ stosujemy kwantyle $m'_{i,n} = \Phi^{-1}(i/(n+1))$ rozkładu $N(0,1)$. Możliwe są tutaj testy oparte na statystykach, które są odchyleniem przeciętnym, bądź średnim odchyleniem z różnic między wartościami u_i oraz $m'_{i,n}$. Statystyki T_1 i T_2 testu T są postaci:

$$T_1 = \sum_{i=1}^{n} |u_i - m'_{i,n}|/n, \quad T_2 = \sum_{i=1}^{n} (u_i - m'_{i,n})^{2/n} .$$

Dogodnością dla praktyki tych testów jest to, że ich w.k. dla $\alpha = 0.05$, 0.01 można wyznaczyć z prostych wyrażeń, które podajemy poniżej:

Test	α	Wyrażenie	Stałe a,b,c
T_1	0.01	$a + b\ln n + c(\ln n)^2$	0.7105, -0.1751, 0.0108
	0.05	"	0.6027, -0.1481, 0.0090
T_2	0.01	$a + b/n + c/n^2$	0.0178, 2.8736, -8.2894
	0.05	"	0.0126, 1.9227, -5.0067

Przykład 13. Dla danych z przykładu 12 mamy: $\bar{y} = 2$, $n = 7$ oraz $s = 4.1057$. Obliczenia szczegółowe zestawiamy poniżej

i	y_i	u_i	$m'_{i,7}$	$u_i - m'_{i,7}$
1	-4	-1.4613	-1.1503	-0.3110
2	-2	-0.9742	-0.6745	-0.2997
3	0	-0.4871	-0.3186	-0.1685
4	1	-0.2436	0.0000	-0.2436
5	5	0.7307	0.3186	0.4221
6	6	0.9742	0.6745	0.2997
7	8	1.4613	1.1503	0.3110

Stąd otrzymujemy $T_1 = 0.294$ i $T_2 = 0.091$, natomiast w.k. dla $\alpha = 0.05$ są odpowiednio równe

$$T_{1,0.05;7} = 0.6027 - 0.1481 \ln 7 + 0.0090(\ln 7)^2 = 0.349,$$

$$T_{2,0.05;7} = 0.0126 + 1.9227/7 - 5.0067/49 = 0.183.$$

Ponieważ obliczone wartości statystyk nie przekraczają w.k., więc nie mamy podstaw do odrzucenia hipotezy H_o.

3. Ogólna dyskusja testów normalności

Dokonaliśmy w pracy przeglądu różnych testów normalności. Zaliczyliśmy je do trzech podstawowych grup w zależności od sposobu konstrukcji testu. Wśród nich wiele ma własność testu omnibus. Te ostatnie można zalecać, kiedy posiadamy pewne informacje a priori odstępstwa od normalności. Z tej grupy omówiliśmy testy: K^2, T', W, W', D_A oraz r. Wszystkie zostały podane w ostatnich 15 latach. Dotychczas nie skonstruowano testu omnibus w grupie testów opartych na porównaniu dystrybuanty rozkładu empirycznego z normalnym. Wspomniana grupa testów nie ma praktycznie większego znaczenia dla weryfikacji hipotezy H_o (z wyjątkiem testu A^2) i powinna być raczej stosowana jako grupa testów przybliżonych i orientacyjnych. Związane jest to z tym, iż dobre własności od strony mocy wykazują te testy przy określonych wartościach parametrów μ i σ. Jeśli chodzi o test χ^2 - Pearsona, to winien on także być traktowany jako test przybliżony. Zasadniczo winno się korzystać z testów normalności, które są odpowiednie dla złożonych hipotez o normalności.

Wiele testów normalności wykazuje podobne własności. Statystyki na których oparte są testy $\sqrt{b_1}$, b_2, u, g, W, r są niezmiennicze ze względu na zmianę skali i położenia, a zatem są odpowiednie do testowania hipotez H_o^z. Testy D, KS, W^2, U^2, V, A^2 mają całkowicie określone rozkłady przy hipotezie H_o^p i mogą być odpowiednie do testowania prostych hipotez o normalności.

Dla większości testów nie znaleziono dotychczas funkcji gęstości ich statystyk rozkładu. Żądane dla nich w.k. są generowane metodą Monte Carlo. Testy W, W', $\widetilde{W}$ i r mają lewostronne obszary krytyczne i małe wartości ich statystyk wskazują na istotne odchylenie rozkładu zmiennej losowej X od rozkładu normalnego. Pozostałe testy mają prawostronne obszary krytyczne. Niektóre statystyki W, W', $\widetilde{W}$ i r wymagają użycia odpowiednich współczynników dla każdego n.

Zagadnienie mocy testów normalności jest stosunkowo dobrze zbadane (patrz, np. Shapiro i inni, 1968, Stephens, 1974, Giorgi i Cinci, 1975, Pearson i inni, 1977). Mało jest dotąd wyników ogólnych, całkowicie zakończonych i nadających się do stosowania, jak to ma miejsce w teorii testów parametrycznych. Trudności natury matematycznej związane ze znajdowaniem mocy testów są zazwyczaj bardzo duże. Trudne jest także, z punktu widzenia praktycznego ustalenie hipotezy alternatywnej.

Wiadomo, że dla rozkładu normalnego trzeci standaryzowany moment centralny $\sqrt{\beta_1}$ wynosi zero, zaś czwarty standaryzowany moment centralny β_2 wynosi trzy. Odpowiednio można scharakteryzować inne rozkłady przez podanie pary wartości $(\sqrt{\beta_1}, \beta_2)$. Zatem za hipotezy alternatywne wybiera się hipotezy orzekające, że rozkład badanej zmiennej losowej X można zaliczyć do jednej z niżej wymienionych grup w zależności od wartości $\sqrt{\beta_1}$ oraz β_2. Poszczególne grupy przedstawiamy poniżej, podając dla każdej z nich typ rozkładu oraz niektóre rozkłady do nich należące:

Grupa 1: $|\sqrt{\beta_1}| > 0.3$, $\beta_2 > 3.0$; rozkłady asymetryczne z "długim ogonem" - χ^2, logarytmiczno-normalny, niecentralny χ^2, wykładniczy, Weibulla, Pareto.

Grupa 2: $|\sqrt{\beta_1}| > 0.3$, $\beta_2 < 3.0$; rozkłady asymetryczne z "krótkim ogonem" - beta, S_B Johnsona.

Grupa 3: $|\sqrt{\beta_1}| \leq 0.3$, $\beta_2 > 4.5$; rozkłady symetryczne z "długim ogonem" - podwójny χ^2, jednostajny, Cauchy'ego, Laplace'a, Tukey'a, S_U Johnsona, logistyczny.

Grupa 4: $|\sqrt{\beta_1}| \leq 0.3$, $\beta_2 < 2.5$; rozkłady symetryczne z "krótkim ogonem" - beta, podwójny χ^2, S_B Johnsona, Tukey'a.

Grupa 5: $|\sqrt{\beta_1}| \leq 0.3$, $2.5 \leq \beta_2 \leq 4.5$; rozkłady prawie normalne - t-Studenta z 10 stopniami swobody, S_B Johnsona z parametrami $\gamma = 0$ i $\partial = 3$, logistyczny, S_B Johnsona z parametrami $\gamma = 1$, $\partial = 2$, S_U Johnsona z parametrami $\gamma = 0$ i $\partial = 10$, Weibulla przy $k = 2$. Zaliczanie rozkładów do poszczególnych grup dokonujemy przy pomocy wartości $\sqrt{\beta_1}$ oraz β_2, które są znane. Różnią się one znacznie w obrębie tego samego rozkładu ustalonego przy różnych wartościach parametrów, przez które jest on określony. Na przykład dla χ_4^2 mamy $\sqrt{\beta_1} = 1.41$ i $\beta_2 = 6.00$, zaś dla χ_{10}^2 jest $\sqrt{\beta_1} = 0.89$ i $\beta_2 = 4.20$. Z tego względu ten sam rozkład przy różnie określonych wartościach parametrów jest zaliczony do różnych grup.

W tabeli 1 przedstawiamy moce (wyrażone w procentach) niektórych testów normalności dla poziomu istotności $\alpha = 0.05$ i przy wielkości próby prostej $n = 20$, uwzględniając rozkłady alternatywne (krótko r.a.): χ_m^2 z $m = 1,2,4,10$ stopniami swobody, logarytmiczno-normalny $LN(\mu,\sigma)$ o parametrach $\mu = 0$ i $\sigma = 1$, T-Studenta t_2 z dwoma stopniami swobody, Cauchy'ego t_1, beta $B(p,q)$ o parametrach $p = 2$ i $q = 1$, jednostajny $B(1,1)$ oraz Laplace'a.

Procedura testowa dla badania normalności oparta na teście W prawie we wszystkich przedstawionych r.a. w tabeli 1 ma większą moc niż pozostałe testy normalności. W szczególności test W jest czuły na asymetrię z "długim ogonem". Przykładowo dla populacji o rozkładzie χ_{10}^2, χ_4^2 i $LN(0,1)$ wartości mocy dla $\alpha = 0.05$ i $n = 20$ wynosi 29, 50 i 93%. W grupie testów normalności opartych na s.p. test W należy uznać za najlepszy od strony mocy. Winien on być zalecany do weryfikacji hipotezy H_0, jeśli liczebność próby jest nie większa od 50. Podobne własności mocy posiadają testy W' oraz r, które mogą być proponowane dla dużych prób o liczebnościach $50 \leq n \leq 100$.

Testy oparte na momentach z próby wykazują dużą moc przy określonym typie r.a. i tak odpowiednio: test $\sqrt{b_1}$ dla rozkła-

Tabela 1. Empirycznie wyznaczone moce niektórych testów normalności wyrażone w procentach dla $\alpha = 0.05$ oraz $n=20$

Rozkłady alternatywne	Momenty		Testy normalności												
	$\sqrt{\beta_1}$	β_2	χ^2	D	W^2	V	U^2	A^2	b_1	b_2	K^2	W	W'	Y	r
χ^2_1	2.83	15.0	94	86	94	94	93	-	89	53	82	98	94	80	94
χ^2_2	2.00	9.0	33	59	74	71	70	82	74	34	60	84	82	52	82
χ^2_4	1.41	6.0	13	33	45	-	-	-	49	27	60	50	24	24	-
χ^2_{10}	0.89	4.2	7	18	23	-	-	-	29	19	27	29	-	16	-
LN(0,1)	6.19	113.9	95	78	88	84	85	91	89	58	82	93	94	77	94
t_1	0	-	41	86	88	87	88	98	39	81	79	88	91	92	92
B(1,1)	0	1.8	11	12	16	17	18	17	0	29	16	23	4	8	4
B(2,1)	-0.57	2.4	8	-	-	-	-	-	8	13	12	35	-	6	-
Laplace'a	0	6.0	17	22	26	22	25	26	25	27	31	25	33	28	33
t_2	0	-	-	55	-	-	-	-	52	53	54	54	59	56	60

- nie badano

dów asymetrycznych z "długim ogonem" $(\chi_1^2, \chi_2^2, \chi_4^2, LN(0,1))$ oraz test b_2 dla rozkładów symetrycznych z "długim ogonem" $(t_1, B(1,1), t_2)$. Interesującą moc posiada test K^2, zawierającą się między wartościami mocy testów $\sqrt{b_1}$ oraz b_2 przy odpowiednich r.a. Wynika stąd wniosek, że wśród wspomnianych testów należy testy $\sqrt{b_1}$ i b_2 stosować przy określonych r.a., natomiast test K^2 w przypadku niemożliwości określenia r.a.

Testy normalności oparte na mierze zgodności rozkładu empirycznego z rozkładem normalnym wykazują podobne własności mocy, choć wśród nich najlepszym jest test A^2, a najgorszym test D. Testy W^2, V i U^2 mają podobne własności co test A^2.

W świetle wymienionych testów normalności test χ^2 ma znacznie niższą moc przy podanych r.a. Dlatego też nie powinien być on stosowany w praktyce jako test normalności. W miejsce testu χ^2 w przypadku dużych prób $(n > 50)$ bez potrzeby konstruowania szeregu rozdzielczego należy stosować test Y D'Agostino, względnie test W' gdy $50 \le n \le 100$. Zasadniczo dla prób o liczebnościach $n < 50$ należy stosować test W, jeśli nie jest możliwe określenie r.a. dla hipotezy H_1.

Należy zaznaczyć, że moc testu rośnie wraz ze wzrostem liczebności próby n. Przykładowo dla testu W przy $\alpha = 0.05$ i r.a. χ_4^2, χ_{10}^2, $LN(0,1)$ moc kształtuje się następująco:

n	χ_4^2	χ_{10}^2	$LN(0,1)$
10	24	11	60
20	50	29	93
30	71	35	99
40	87	48	100
50	95	56	100

Dlatego przy badaniu normalności rozkładu zmiennej winno się dysponować co najmniej próbą o liczebności $n \ge 30$. Z drugiej strony moc testu maleje wraz z obniżeniem poziomu istotności α.

Przykładowo dla testów Y i W przy r.a. LN(0,1) mamy następujące moce testów wyrażone w procentach:

	0.10		0.05		0.02		0.01	
n	Y	W	Y	W	Y	W	Y	W
10	51	68	42	58	34	45	28	38
20	80	95	75	92	66	86	61	81
30	93	99	90	99	86	97	82	96
40	97	100	94	100	92	99	90	99
50	99	100	98	100	97	100	96	100

Stąd ważnym staje się odpowiedni dobór poziomu istotności α do weryfikacji hipotezy H_o.

Wśród podanych wyżej grup, można by sugerować, iż najlepszymi testami normalności od strony mocy przy odpowiednich liczebnościach prób są:

Grupa 1 - test W, dla $n \leq 50$,
test W', dla $50 < n \leq 100$,
dowolny z testów W^2, V, U^2, dla $n > 100$;

Grupa 2 - test W, dla $n \leq 50$,
test K^2, dla $n > 50$;

Grupa 3 - test r, dla $n \leq 50$,
jeden z testów: W' lub r, dla $50 < n < 100$,
test Y, dla $n \geq 100$;

Grupa 4 - test b_2, dla $n < 20$,
test K^2, dla $20 \leq n \leq 200$;

Grupa 5 - test W, dla $n \leq 20$,
test b_2, dla $20 < n < 50$,
jeden z testów: W' lub r, dla $50 \leq n \leq 100$,
jeden z testów: K^2 lub Y, dla $n > 100$.

Najogólniej procedura weryfikacji hipotezy H_o o rozkładzie normalnym zmiennej losowej X na podstawie próby prostej $\{x_i\}$ pobranej z populacji wg odpowiedniego schematu losowania, winna przebiegać następująco. Wyznaczamy z próby $\{x_i\}$ wielkości $\sqrt{b_1}$ i b_2, będące ocenami $\sqrt{\beta_1}$ i β_2, a następnie wybieramy jedną z wymienionych wyżej grup i wybieramy odpowiedni test normalności w zależności od liczebności próby n. Jeśli znany jest typ odstępstwa od normalności (np. skośność), to wybieramy test o największej mocy odpowiedni do określonej hipotezy H_1 (np. test $\sqrt{b_1}$ do rozkładów asymetrycznych skośnych), a gdy takich informacji a priori brak, to stosujemy jeden z testów omnibus.

Literatura cytowana

D'Agostino, R.B., 1971, An omnibus test of normality for moderate and large size samples. Biometrika 58, str. 341-348.

D'Agostino, R.B., Pearson, E.S., 1973, Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$. Biometrika, 60, str.613-622.

Domański, C., 1979, Statystyczne testy nieparametryczne. PWE, Warszawa.

Filliben, J.L., 1975, The probability Plot Correlation Coefficient Tests for Normality. Technometrics, 17, str.111-117.

Fisz, M., 1967, Rachunek prawdopodobieństwa i statystyka matematyczna. PWN, Warszawa.

Giorgi, G.M., Cinci, S., 1975, Sulle funzioni discriminatorie di alcuni tests univariate di normalita. Studi di Economia, 6, str.100-133.

Hegazy, Y.A.S., Green, J.R., 1975, Some New Godness-of-fit Tests Using Order Statistics. Appl. Statist., 24, str. 299-308.

Lilliefors, H.W., 1967, On the Kołmogorow-Smirnov test for normality with mean and variance unknown. JASA, 62, str. 399-402.

Lloyd, E.H., 1952, Least squares estimation of location and scale parameters using order statistics. Biometrika, 39, str.88-95.

Mulhollander, H.P., 1977, On the distribution of b_1 for samples of size at most 25, with tables. Biometrika, 64, str.401-409.

Pearson, E.S., Hartley, H.O., 1966, Biometrika Tables for Statisticians, Vol.1, Cambridge University Press.

Pearson, E.S., Hartley, H.O., 1972, Biometrika Tables for Statisticians, Vol.2, Cambridge University Press.

Pearson, E.S., D'Agostino, R.B., Bowman, K.O., 1977, Tests for departure from normality: Comparison of powers. Biometrika, 64, str.231-246.

Pawłowski, Z., 1976, Statystyka matematyczna. PWN, Warszawa.

Shapiro, S.S., Francia, R.S., 1972, Approximate analysis of variance test for normality. JASA, 67, str.215-216.

Shapiro, S.S., Wilk, M.B., 1965, An analysis of variance test for normality (complete samples), Biometrika, 52, str. 591-611.

Shapiro, S.S., Wilk, M.B., Chen, M.J., 1968, A comparative study of various tests for normality. JASA, 63, str. 1343-1372.

Stephens, M.A., 1974, EDF statistics for goodness-of-fit and some comparisons. JASA, 69, str.730-737.

Spiegelhalter, D.J., 1977, A test for normality against symmetric alternatives. Biometrika, 64, str.415-418.

van Soest, J., 1967, Some experimental results concerning tests of normality. Statistica Neerlandica, 21.

Zieliński, R., 1972, Tablice statystyczne. PWN, Warszawa.