Analiza Danych - Podstawy Statystyczne

Weryfikacja hipotez statystycznych

Marek Rupniewski

13 kwietnia 2019

Testowanie hipotez w ujęciu Neymana-Pearsona

Hipotezy

Hipotezy statystyczne są hipotezami dotyczącymi rozkładów prawdopodobieństwa. Wyróżniamy:

- hipotezy proste mówiące, że pewna zmienna ma określony (jednoznacznie) rozkład. Np. moneta jest symetryczna (zmienna opisująca ilość orłów w N rzutach ma rozkład Binom$(N, \frac{1}{2})$),
- hipotezy złożone nie określają jednoznacznie rozkładu prawdopodobieństwa (np. rozkład normalny z nieznaną wariancją) lub nie określają go w ogóle (np. jakiś rozkład nie będący Bern$(\frac{1}{2})$).

Testowanie hipotez w ujęciu Neymana-Pearsona

Mamy do czynienia z dwoma hipotezami:

- H_0 hipotezą zerową, czyli hipotezą, którą w pewnym sensie poddajemy weryfikacji,
- H_1 hipotezą alternatywną, sprzeczną z hipotezą zerową, ale niekoniecznie ją dopełniającą.

Celem jest wskazanie właściwej hipotezy minimalizując przy tym prawdopodobieństwo popełnienia błędu.

Rodzaje błędów

Definicja

Odrzucenie hipotezy H_0, gdy jest ona prawdziwa nazywamy błędem I rodzaju. Prawdopodobieństwo popełnienia takiego błędu na podstawie pewnego testu nazywane jest poziomem istotności tego testu i oznaczane najczęściej symbolem α.

Definicja

Przyjęcie hipotezy H_0, gdy w rzeczywistości nie jest ona prawdziwa nazywamy błędem II rodzaju.
Prawdopodobieństwo popełnienia takiego błędu na podstawie pewnego testu oznaczane jest najczęściej symbolem β. Wielkość $1 - \beta$ nazywana jest mocą testu.

Czym mniejszy poziom istotności testu (α) i czym większa moc tego testu ($1 - \beta$) tym lepiej.

Przykład

Obiekt obserwowany w systemie radarowym może podlegać testowi z hipotezami:
H_0 — obiekt jest pociskiem odrzutowym,
H_1 — obiekt jest pasażerskim odrzutowcem.
Błąd I rodzaju polega na zignorowaniu zagrożenia (potraktowanie pocisku jako odrzutowca),
Błąd II rodzaju polega na fałszywym alarmie (mogącym się zakończyć zestrzeleniem samolotu pasażerskiego).

Przykład

Na podstawie pewnego testu medycznego weryfikuje, czy pacjent cierpi na schorzenie X:
H_1 — tak, cierpi (pozytywny wynik testu),
H_0 — nie (negatywny wynik testu).
Błąd I rodzaju polega na „fałszywym alarmie” (podjęcie niepotrzebnego leczenia mogącego mieć negatywne skutki uboczne),
Błąd II rodzaju polega na zignorowaniu zagrożenia (nie podjęcie terapii gdy jest ona potrzebna).

Definicja

Statystyka testowa (decyzyjna) to statystyka (funkcja próby), na podstawie której weryfikujemy hipotezę. Obszar krytyczny testu, to obszar wartości tej statystyki, który prowadzi do odrzucenia hipotezy zerowej.

W celu zdefiniowania testu statystycznego należy podać statystykę testową oraz obszar krytyczny testu.

Przykład 2 prostych hipotez

W pudełku znajdują się dwie monety: symetryczna oraz taka, że prawdopodobieństwo wypadnięcia orła jest 0.7. Wyciągamy losowo jedną monetę, rzucamy nią 10 razy, notujemy co „wypadło” i na na tej podstawie chcemy weryfikować hipotezę zerową H_0: wylosowaliśmy monetę symetryczną wobec
hipotezy alternatywnej H_1: wylosowaliśmy tę drugą monetę.

Obie hipotezy są proste. Liczba
orłów opisana jest rozkładem Binom$(10, 0.5)$ lub Binom$(10, 0.7)$.

	0	1	2	3	4	5	6	7	8	9	10
$p = .5$	.0010	.0098	.0439	.1172	.2051	.2461	.2051	.1172	.0439	.0098	.0010
$p = .7$	.0000	.0001	.0014	.0090	.0368	.1029	.2001	.2668	.2335	.1211	.0282

Kontynuacja przykładu

k	0	1	2	3	4	5	6	7	8	9	10
$\mathbb{P}(X = k\|p = .5)$	.0010	.0098	.0439	.1172	.2051	.2461	.2051	.1172	.0439	.0098	.0010
$\mathbb{P}(X = k\|p = .7)$	.0000	.0001	.0014	.0090	.0368	.1029	.2001	.2668	.2335	.1211	.0282

Za statystykę decyzyjną weźmy iloraz wiarygodności

$$R = \frac{P(X|H_1)}{P(X|H_0)}.$$

x	0	1	2	3	4	5	6	7	8	9	10
$R(x)$	0.006	0.014	0.033	0.077	0.18	0.42	0.98	2.3	5.3	12.4	28.9

Określamy obszar krytyczny jako $\{R: R > c\}$ (c to tzw. wartość krytyczna).

Biorąc $c = 30$ nie popełnimy błędu I rodzaju.

Biorąc $c = 0$ nie popełnimy błędu II rodzaju.

Biorąc $c = 1$ (odrzucamy H_0 jeśli > 6 orłów) popełnimy błąd I rodzaju z prawd. 0.17, a II rodzaju — z prawd. 0.35.

Biorąc $c = 10$ (odrzucamy H_0 jeśli > 8 orłów) popełnimy błąd I rodzaju z prawd. 0.01, a II rodzaju — z prawd. 0.85.

Testy najmocniejsze

Lemat (Neymana-Pearsona)

Niech H_0 oraz H_1 będą hipotezami prostymi oraz niech dany będzie test oparty na ilorazie wiarygodności (tzn. test odrzucający hipotezę H_0 gdy iloraz wiarygodności jest większy niż pewna stała c) na pewnym poziomie istotności α. Wówczas jest to test o największej mocy spośród wszystkich testów na poziomach istotności nie przekraczających α.

Przykład

Niech $X_1, \ldots, X_n$ będą niezależne o rozkładzie normalnym ze znaną wariancją σ^2 oraz

$$H_0 \colon \mu = \mu_0$$
$$H_1 \colon \mu = \mu_1.$$

Test oparty na ilorazie wiarygodności, to test oparty na średniej $\overline{X}_n$ z próby.

Reguła decyzyjna dla $\mu_1 > \mu_0$:

$$\overline{X}_n > c \quad \Longrightarrow \quad H_1, \quad \text{w przeciwnym wypadku } H_0$$

a dla $\mu_1 < \mu_0$:

$$\overline{X}_n < c \quad \Longrightarrow \quad H_1, \quad \text{w przeciwnym wypadku } H_0.$$

Kontynuacja przykładu z monetą

x	0	1	2	3	4	5	6	7	8	9	10
$R(x)$	0.006	0.014	0.033	0.077	0.18	0.42	0.98	2.3	5.3	12.4	28.9

W związku z dyskretnością zbioru wartości ilorazu wiarygodności, zadając obszary krytyczne postaci $\{R > c\}$, można otrzymać dyskretną liczbę „osiągalnych" poziomów istotności.

Np. test z obszarem krytycznym $\{R > 30\}$ będzie miał zerowy poziom istotności (taki sam poziom dla $\{R > \frac{\mathbb{P}(10|H_1)}{\mathbb{P}(10|H_0)}\}$);

test z obszarem krytycznym $\{R > 28\}$ będzie miał poziom istotności $\alpha \approx 0.001$ (taki sam poziom dla $\{R > \frac{\mathbb{P}(9|H_1)}{\mathbb{P}(9|H_0)}\}$);

test z obszarem krytycznym $\{R > 10\}$ będzie miał poziom istotności $\alpha \approx 0.011$ (taki sam poziom dla $\{R > \frac{\mathbb{P}(8|H_1)}{\mathbb{P}(8|H_0)}\}$);

Czy można skonstruować test na poziomie istotności np. $\alpha = 0.005$?

Testy randomizowane

Test randomizowany to test, który w zależności od wartości statystyki decyzyjnej:

- odrzuca hipotezę zerową (obszar krytyczny),
- nie odrzuca hipotezy zerowej (obszar akceptacji),
- odrzuca hipotezę zerową losowo z zadanym prawdopodobieństwem (na ”granicy obszarów“).

Kontynuacja przykładu z monetą

x	0	1	2	3	4	5	6	7	8	9	10
$p = .5$	.0010	.0098	.0439	.1172	.2051	.2461	.2051	.1172	.0439	.0098	.0010
$p = .7$	.0000	.0001	.0014	.0090	.0368	.1029	.2001	.2668	.2335	.1211	.0282

x	0	1	2	3	4	5	6	7	8	9	10
$R(x)$	0.006	0.014	0.033	0.077	0.18	0.42	0.98	2.3	5.3	12.4	28.9

Skonstruujemy test (randomizowany) z poziomem istotności $\alpha \approx 0.005$.

$$\begin{cases} \text{odrzucamy } H_0 & \text{jeśli } R > \dfrac{\mathbb{P}(9|H_1)}{\mathbb{P}(9|H_0)}, \\ \text{nie odrzucamy } H_0 & \text{jeśli } R < \dfrac{\mathbb{P}(9|H_1)}{\mathbb{P}(9|H_0)}, \\ \text{odrzucamy } H_0 \text{ z pr. } p_* & \text{jeśli } R = \dfrac{\mathbb{P}(9|H_1)}{\mathbb{P}(9|H_0)}. \end{cases}$$

$$0.005 = \alpha = \mathbb{P}(10|H_0) + \mathbb{P}(9|H_0)p_* \approx 0.001 + p_* 0.0098 \Rightarrow p_* \approx 0.4.$$

Niech T będzie statystyką decyzyjną oraz niech obszar krytyczny testu na poziomie istotności α będzie postaci

$$\{T > t_0\},$$

gdzie t_0 dobrane tak by $\mathbb{P}(T > t_0 | H_0) = \alpha$.

Definicja

p-wartością dla zaobserwowanej próby nazywamy minimalną wartość poziomu istotności α, dla której hipoteza zerowa byłaby odrzucona.

p-wartość można interpretować jako prawdopodobieństwo, pod warunkiem H_0, uzyskania wartość statystyki testowej tak samo lub bardziej „ekstremalnej” niż wartość wyznaczona dla zaobserwowanej próby.

Przykład: weryfikacja zdolności nadprzyrodzonych

Osoba twierdząca, że ma nadprzyrodzone zdolności proszona jest o rozpoznanie jednego z 4 kolorów 20 losowo (bez zwracania) wyciągniętych kart (z 52-kartowej talii). T — liczba poprawnie odgadniętych kart.
H_0 : osoba zgaduje,
H_1 : osoba ma szósty zmysł.
H_0 jest prosta (T ma wówczas rozkład Binom$(20, \frac{1}{4})$), a H_1 złożona.

Załóżmy, że osoba trafnie odgadła kolory 9 kart. Hipoteza zerowa zostałaby odrzucona np. przy poziomie istotności $\alpha = 0.05$, a nie zostałaby odrzucona dla $\alpha = 0.01$.
p-wartością dla wyniku tego eksperymentu jest 0.041 ($\mathbb{P}(T \geq 9|H_0) \approx 0.041$).
(dla 10 odgadniętych kart p-wartość wynosiłaby 0.014.)

Testy istotności a przedziały ufności

Zakładamy, że mamy do czynienia z parametryczną rodziną rozkładów z parametrem $\theta \in \Theta$ oraz dysponujemy pewną próbą $X = (X_1, \ldots, X_n)$.

Twierdzenie

Niech $C(X)$ będzie przedziałem ufności dla θ na poziomie ufności $\gamma = 1 - \alpha$. Wówczas obszarem akceptacji dla testu na poziomie istotności α ($H_0: \theta = \theta_0$ wobec $H_1: \theta \neq \theta_0$) jest

$$A(\theta_0) = \{X : \theta_0 \in C(X)\}.$$

Testy istotności a przedziały ufności - przykład 1

Jeśli $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ i σ^2 dane, to przedziałem ufności dla μ na poziomie ufności $1 - \alpha$ jest

$$[\overline{X} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}, \overline{X} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}].$$

Test ($H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$) na poziomie istotności α ma zatem regułę decyzyjną

$$\begin{cases} H_0, & \text{jeśli } \frac{|\overline{X} - \mu_0|}{\sigma/\sqrt{n}} \leq c, \\ H_1, & \text{jeśli } \frac{|\overline{X} - \mu_0|}{\sigma/\sqrt{n}} > c, \end{cases}$$

gdzie $c = z_{1-\alpha/2}$.

Testy istotności a przedziały ufności - przykład 2

Jeśli $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ i σ^2 nie jest znane, to przedziałem ufności dla μ na poziomie ufności $1 - \alpha$ jest

$$\left[\overline{X}_n - \frac{S}{\sqrt{n}} F^{-1}_{t_{n-1}}\left(1 - \alpha/2\right), \overline{X}_n + \frac{S}{\sqrt{n}} F^{-1}_{t_{n-1}}\left(1 - \alpha/2\right)\right].$$

Test ($H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$) na poziomie istotności α ma zatem regułę decyzyjną

$$\begin{cases} H_0, & \text{jeśli } \frac{|\overline{X} - \mu_0|}{S/\sqrt{n}} \leq c, \\ H_1, & \text{jeśli } \frac{|\overline{X} - \mu_0|}{S/\sqrt{n}} > c, \end{cases}$$

gdzie

$$c = F^{-1}_{t_{n-1}}\left(1 - \alpha/2\right).$$

Testowanie zgodności rozkładu

Istota problemu

Dysponujemy próbą $X_1, \ldots, X_n$ i chcemy sprawdzić czy pochodzi ona z danego rozkładu (danego np. funkcją gęstości prawdopodobieństwa lub dystrybuantą), czyli czy rozkład próby jest zgodny z danym rozkładem.

Często o rozkładzie, z którym chcemy sprawdzić zgodność danych, wiemy tylko, że należy do pewnej rodziny (np. rozkładów normalnych). Wówczas najpierw estymujemy parametry rozkładu (np. średnią i wariancję), a następnie badamy zgodność danych (próby) z rozkładem o wyestymowanych parametrach.

Przykład wiodący

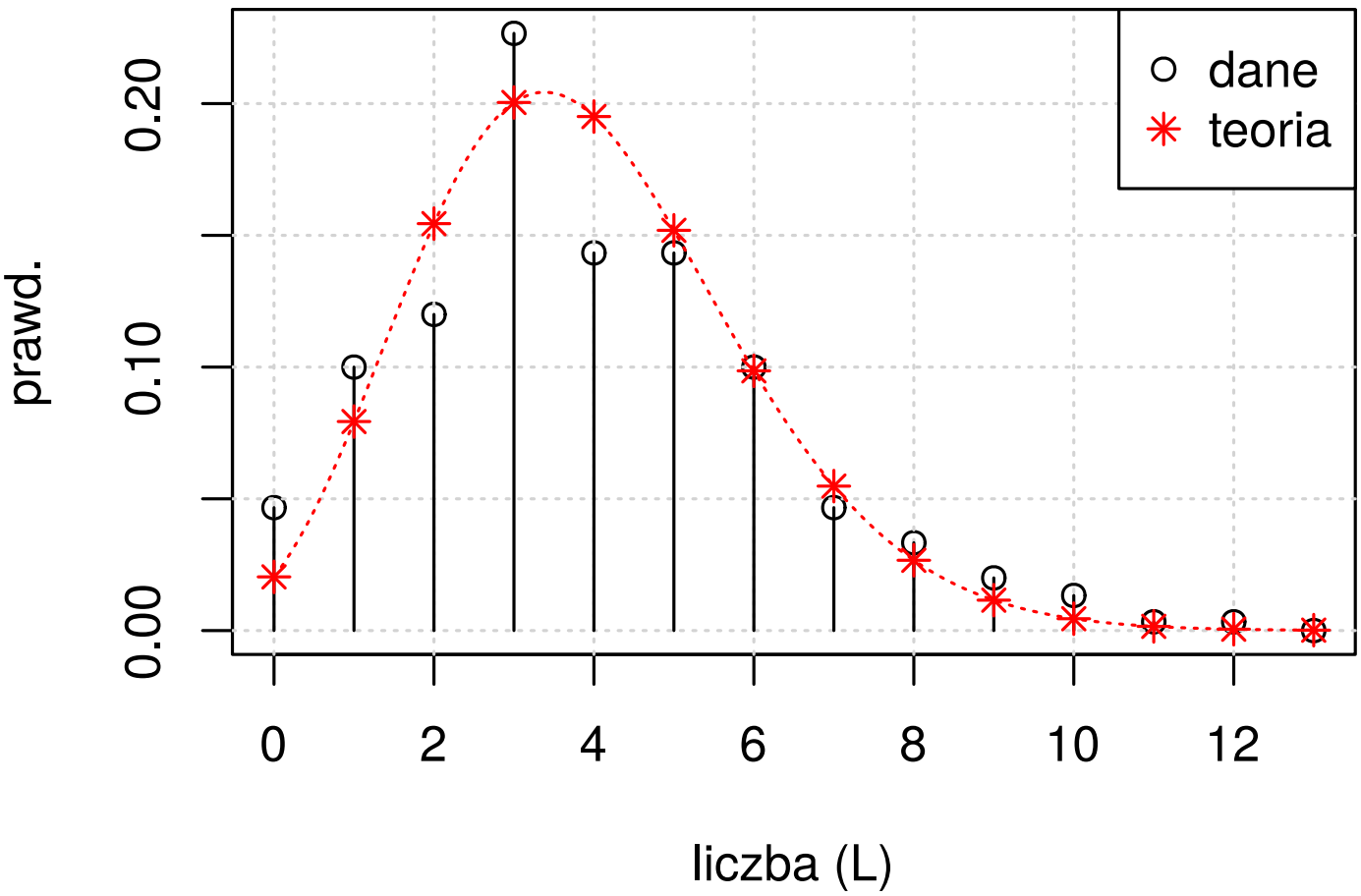
W poniższej tabeli przedstawione są liczby pojazdów skręcających na pewnym skrzyżowaniu w prawo w przeciągu 300 3-minutowych przedziałów czasu. Będziemy badali zgodność tych danych z rozkładem Poissona.

S	0	1	2	3	4	5	6
L	14	30	36	68	43	30	14
S	7	8	9	10	11	12	13+
L	14	10	6	4	1	1	0

S to Liczba „skrętów” w przeciągu 3 minut.
L to liczba 3-minutowych przedziałów z zadaną liczbą „skrętów”.

$$\hat{\lambda} = \frac{0 \times 14 + 1 \times 30 + \cdots + 12 \times 1}{14 + 30 + \cdots + 1} \approx 3.9$$

Wynik estymacji parametru λ



W kierunku testu zgodności χ^2

Rozważmy problem losowego rozmieszczenia n kul w r koszykach, przy założeniu, że prawdopodobieństwo umieszczenia kuli w i-tym koszyku jest równe p_i.

Niech X_k oznacza numer koszyka, do którego wpadła k-ta kula.

$$\mathbb{P}(X_k = i) = p_i, \qquad p_1 + p_2 + \cdots + p_r = 1.$$

Niech ν_i oznacza liczbę kul, które wpadły do i-tego koszyka.

$$\mathbb{E}\nu_i = np_i.$$

Chcemy zbadać jakiego odstępstwa zmiennych ν_i od ich wartości średnich np_i możemy się spodziewać.

Twierdzenie Pearsona

n kul, r koszyków, $p_i = \mathbb{P}(X = i)$, ν_i — liczba kul w i-tym koszyku.

Chcemy zbadać jakiego odstępstwa zmiennych ν_i od ich wartości średnich np_i możemy się spodziewać.

Twierdzenie (Pearsona)

Rozkład zmiennej losowej

$$T = \sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{np_i}$$

zbiega do rozkładu χ^2 *z* $(r-1)$ *stopniami swobody* (χ^2_{r-1}).

Rozkład graniczny zmiennych ν_i

ν_i (ustalone i) można potraktować jako sumę n niezależnych zmiennych o rozkładzie Bernoulliego z prawdopodobieństwem sukcesu $p = p_i$.

Zatem

$$\mathbb{V}\nu_i = n(p_i - p_i^2) = np_i(1 - p_i).$$

W szczególności na mocy centralnego twierdzenia granicznego

$$\frac{\nu_i - np_i}{\sqrt{np_i(1 - p_i)}} \to N(0, 1).$$

Innymi słowy

$$\frac{\nu_i - np_i}{\sqrt{np_i}} \to N(0, 1 - p_i).$$

Test zgodności rozkładu

Rozważamy próbę $X_1, \ldots, X_n$ niezależnych zmiennych losowych o tym samym, dyskretnym rozkładzie.

Oznaczmy przez $B_1, \ldots, B_r$ zbiór wartości jakie mogą przyjmować zmienne X oraz przez $p_1, \ldots, p_r$ prawdopodobieństwa przyjmowania poszczególnych wartości.

Chcemy zbadać, czy próba $X_1, \ldots, X_n$ odpowiada wartościom pewnych ustalonych prawdopodobieństw $p_1^*, \ldots, p_r^*$, czyli czy zachodzi hipoteza:

$H_0 : p_1 = p_1^*, \ldots, p_r = p_r^*$

wobec hipotezy alternatywnej

H_1 : przynajmniej dla jednego i jest $p_i \neq p_i^*$.

$H_0 : p_1 = p_1^*, \ldots, p_r = p_r^*$
H_1 : przynajmniej dla jednego i jest $p_i \neq p_i^*$.

Za statystykę decyzyjną przyjmujemy

$$T = \sum_{i=1}^{r} \frac{(\nu_i - np_i^*)^2}{np_i^*}.$$

Jeśli rzeczywiście $p_i = p_i^*$, $i = 1, \ldots, r$, to na mocy tw. Pearsona

$$T \xrightarrow{d} \chi^2_{r-1}.$$

Gdyby natomiast pewne $p_i \neq p_i^*$, to

$$\frac{\nu_i - np_i^*}{\sqrt{np_i^*}} = \sqrt{\frac{p_i}{p_i^*}} \frac{\nu_i - np_i}{\sqrt{np_i}} + \sqrt{n} \frac{p_i - p_i^*}{\sqrt{p_i^*}}.$$

Zatem wystarczy aby jedno $p_i \neq p_i^*$ aby $T \xrightarrow{n\to\infty} \infty$.

Test zgodności rozkładu χ^2

$H_0 : p_1 = p_1^*, \ldots, p_r = p_r^*$
H_1 : przynajmniej dla jednego i jest $p_i \neq p_i^*$.

Za statystykę decyzyjną przyjmujemy

$$T = \sum_{i=1}^{r} \frac{(\nu_i - np_i^*)^2}{np_i^*}.$$

Reguła decyzyjna w teście zgodności χ^2:

$$\begin{cases} H_0 : T \leq c, \\ H_1 : T > c. \end{cases}$$

Stałą c dobieramy tak, by zapewnić określony poziom istotności α testu. Dla dużych prób można szacować:

$$c \approx F^{-1}_{\chi^2_{r-1}}(1 - \alpha).$$

Kontynuacja przykładu z pojazdami

Mamy $r = 14$ „koszyków” (0 skrętów, 1 skręt, ..., co najmniej 13 skrętów). Chcemy sprawdzić, czy prawdopodobieństwa „wpadnięcia” $n = 300$ „kul” do poszczególnych „koszyków” są równe:

$$p_0^* = \frac{3.9^0}{0!e^{3.9}}, p_1^* = \frac{3.9^1}{1!e^{3.9}}, \ldots, p_{12}^* = \frac{3.9^{12}}{12!e^{3.9}}, p_{13+}^* = 1 - p_0 - \cdots - p_{12}.$$

Wyznaczamy wartość krytyczną testu dla $\alpha = 0.05$:

$$T \sim \chi^2_{13}, \qquad c = F^{-1}_{\chi^2_{13}}(1 - 0.05) \approx 22.4.$$

Wartość statystyki T dla naszych danych: $T \approx 32.6 > c$.
Przy wybranym poziomie istotności hipotezę mówiącą, że próba pochodzi z rozkładu $\text{Pois}(3.9)$ należy odrzucić!
p-wartość dla danego testu jest równa ≈ 0.002 !

Kontynuacja przykładu z pojazdami

Jaką hipotezę sprawdzaliśmy?
Taką, że zaobserwowane liczby skrętów odpowiadają rozkładowi Pois(3.9).

Jak sprawdzić, czy te liczby odpowiadają rozkładowi Pois(λ) dla jakiegokolwiek λ?

Test zgodności χ^2 dla hipotezy złożonej

Fakt

Jeśli w wyjściowej sytuacji z kulami rozważymy hipotezę H_0: każde p_i jest równe $p_i(\theta)$, dla wspólnego $\theta \in \Theta$ wobec hipotezy alternatywnej H_1 przeciwnej do H_0, i jeżeli $\hat{\theta}$ jest estymatorem największej wiarygodności

$$\text{tzn. } \hat{\theta} = \arg\max_{\theta \in \Theta} p_1(\theta)^{\nu_1} \dots p_r(\theta)^{\nu_r},$$

to

$$T = \sum_{i=1}^{r} \frac{(\nu_i - np_i(\hat{\theta}))^2}{np_i(\hat{\theta})} \xrightarrow[n\to\infty]{d} \chi^2_{r-s-1},$$

gdzie s to wymiar przestrzeni parametrów Θ.

Kontynuacja przykładu z pojazdami

Test na poziomie istotności dla hipotezy zerowej: dane „układają się” według pewnego rozkładu Poissona wobec hipotezy alternatywnej: dane nie układają się wg żadnego rozkładu Poissona wygląda podobnie do poprzednio skonstruowanego, gdyż stosowaliśmy go do estymatora największej wiarygodności.

Różnice: $\lambda \in (0, \infty) = \Theta$, zatem $s = \dim \Theta = 1$, rozkład graniczny statystyki T jest zatem rozkładem χ^2 o 12 (a nie 13) stopniach swobody.
p-wartość dla nowego testu wynosi ≈ 0.001!

Średnio raz na tysiąc (!) razy 300-elementowa próba z rozkładu Poissona będzie tak „słabo” lub jeszcze „gorzej” zgodna z rozkładem Poissona niż rozważane w przykładzie dane.

Jak sprawdzać zgodność rozkładu dla ciągłych rozkładów

Rozważmy problem polegający na sprawdzeniu, czy dana próba losowa $X_1, \ldots, X_n$ „pochodzi” z rozkładu ciągłego (np. normalnego $N(\mu, \sigma^2)$ o zadanych parametrach) zadanego pewną dystrybuantą F.

Rozwiązanie 1

Podzielić zbiór wartości, które mogą przyjmować zmienne X, na skończoną liczbę przedziałów. Na podstawie danego (ciągłego) rozkładu wyznaczyć prawdopodobieństwa „wpadnięcia” do każdego z przedziałów. Policzyć ile ze zmiennych X wpada do każdego z przedziałów i przeprowadzić test zgodności χ^2.

W stronę testu Kołmogorowa-Smirnowa

Definicja

Dystrybuanta empiryczna dla próby $X_1, \ldots, X_n$ to dystrybuanta F_n określona wzorem:

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq t),$$

gdzie

$$\mathbb{1}(X_i \leq t) = \begin{cases} 1 & \text{jeśli } X_i \leq t, \\ 0 & \text{jeśli } X_i > t. \end{cases}$$

$X_1, \ldots, X_n$ próba los. z rozkładu o dystrybuancie F.

Fakt

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \xrightarrow[\mathbb{P}]{n \to \infty} 0.$$

Twierdzenie

Rozkład zmiennej losowej

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)|$$

nie zależy od dystrybuanty F (tzn. dla każdej dystrybuanty ciągłego rozkładu i niezależnej próby $X_1, \ldots, X_n$ pochodzącej z tego rozkładu, powyższa zmienna ma taki sam rozkład).

$X_1, \ldots, X_n$ próba los. z rozkładu o dystrybuancie F.

Twierdzenie

$$\mathbb{P}\left(\sqrt{n}\sup_{t\in\mathbb{R}}|F_n(t)-F(t)|\le t\right)\xrightarrow{n\to\infty}H(t)=1+2\sum_{i=1}^{\infty}(-1)^i\mathrm{e}^{-2i^2t^2}.$$

($H(t)$ to dystrybuanta rozkładu Kołmogorowa-Smirnowa).

Test Kołmogorowa-Smirnowa

$X_1, \ldots, X_n$ próba losowa.
F pewna ustalona dystrybuanta rozkładu ciągłego.

H_0: zmienne $X_1, \ldots, X_n$ pochodzą z rozkładu o dystryb. F,
H_1: zmienne $X_1, \ldots, X_n$ nie pochodzą z rozkładu o dystryb. F.

Statystyka decyzyjna: $D_n = \sqrt{n} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)|.$

Reguła decyzyjna:
$$\begin{cases} H_0 \colon D_n \leq c, \\ H_1 \colon D_n > c. \end{cases}$$

Stałą c dobiera się tak, by zapewnić określony poziom istotności α testu (na podstawie rozkładu D_n lub rozkładu granicznego Kołmogorowa-Smirnowa).

Testy normalności

Istnieje wiele testów służących badaniu normalności rozkładu (H_0):

- test D'Agostino na skośność (rozkład normalny ma zerową skośność; duża wartość skośności z próby świadczy na niekorzyść H_0),
- test Anscombe-Glynn-a na kurtozę (rozkład normalny ma kurtozę równą 3; duża odległość kurtozy z próby od tej wartości świadczy na niekorzyść H_0),
- test Jarque-Bera (kombinacja testów skośności i kurtozy),
- test Shapiro-Wilka (oparty na statystykach pozycyjnych).

W pakiecie R dostępne są funkcje: `agostino.test()`, `kurtosis.test()`, `jarque.test()` (biblioteka `moments`) oraz `shapiro.test()`.

Testowanie niezależności

Tabela krzyżowa

Mamy do dyspozycji N-elementową próbę pogrupowaną ze względu na dwie cechy: A oraz B.

Tabela krzyżowa (inne określenia: tab. kontyngencji, rozdzielcza, dwudzielcza):

	$B:1$	$B:2$	$\dots$	$B:J$
$A:1$	n_{11}	n_{12}	$\dots$	n_{1J}
$A:2$	n_{21}	n_{22}	$\dots$	n_{2J}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$A:I$	n_{I1}	n_{I2}	$\dots$	n_{IJ}

$$N = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}, \quad n_{\bigstar j} = \sum_{i=1}^{I} n_{ij}, \quad n_{i\bigstar} = \sum_{j=1}^{J} n_{ij}.$$

Niezależność cech — sformułowanie problemu

Niech p_{ij} oznacza prawdopodobieństwo tego, że element próby ma i-tą wartość cechy A i j-tą wartość cechy B.

$$p_{\bigstar j} = \sum_{i=1}^{I} p_{ij}, \qquad p_{i\bigstar} = \sum_{j=1}^{J} p_{ij}.$$

Niezależność cech oznacza, że dla dowolnych indeksów i oraz j zachodzi równość:

$$p_{ij} = p_{i\bigstar} p_{\bigstar j}.$$

Test niezależności ma na celu sprawdzenie, czy próba świadczy za, czy przeciwko temu, że zachodzą równości j.w. (dla każdego i oraz j).

Weryfikacja niezależności cech polega na sprawdzeniu z jak dużą dokładnością prawdopodobieństwa p_{ij} są równe $p^*_{ij} = \hat{p}_{i\bigstar}\hat{p}_{\bigstar j}$. Metoda największej wiarygodności daje

$$\hat{p}_{i\bigstar} = \frac{n_{i\bigstar}}{N}, \qquad i = 1, 2, \ldots, I,$$

$$\hat{p}_{\bigstar j} = \frac{n_{\bigstar j}}{N}, \qquad j = 1, 2, \ldots, J.$$

Liczba parametrów : $I - 1 + J - 1$.

Rozkład zmiennej losowej

$$T = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(n_{ij} - N\hat{p}_{i\bigstar}\hat{p}_{\bigstar j})^2}{N\hat{p}_{i\bigstar}\hat{p}_{\bigstar j}}$$

zbiega do rozkładu χ^2 z liczbą st. swobody

$$IJ - (I - 1 + J - 1) - 1 = (I - 1)(J - 1).$$

Test niezależności cech — podsumowanie

1. Ustalamy poziom istotności α i wyznaczamy wartość krytyczną testu

$$c = F^{-1}_{\chi^2_{(I-1)(J-1)}}(1-\alpha).$$

2. Estymujemy parametry:

$$\hat{p}_{i\bigstar} = \frac{n_{i\bigstar}}{N}, \quad i = 1, 2, \ldots, I, \qquad \hat{p}_{\bigstar j} = \frac{n_{\bigstar j}}{N}, \quad j = 1, 2, \ldots, J.$$

3. Obliczamy wartość statystyki

$$T = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(n_{ij} - N\hat{p}_{i\bigstar}\hat{p}_{\bigstar j})^2}{N\hat{p}_{i\bigstar}\hat{p}_{\bigstar j}} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(Nn_{ij} - n_{i\bigstar}n_{\bigstar j})^2}{Nn_{i\bigstar}n_{\bigstar j}}.$$

4. Odrzucamy hipotezę zerową (niezależność cech) jeśli $T > c$.

Przykład

48 menadżerów (uczestników szkolenia) analizowało kartoteki osobowe pracowników (24 kartoteki kobiet i 24 kartoteki mężczyzn przygotowane tak by różniły się tylko płcią kandydata).

	Mężczyzn	Kobiet
promocja	21	14
wstrzymanie promocji	3	10

Czy kobiety były dyskryminowane?

```
> chisq.test(matrix(c(21, 3, 14, 10),2,2))

        Pearson's Chi-squared test with Yates' continuity correction

data:  matrix(c(21, 3, 14, 10), 2, 2)
X-squared = 3.7978, df = 1, p-value = 0.05132
```

W stronę testu jednorodności

Tabela krzyżowa:

	$B:1$	$B:2$	$\dots$	$B:J$
$A:1$	n_{11}	n_{12}	$\dots$	n_{1J}
$A:2$	n_{21}	n_{22}	$\dots$	n_{2J}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$A:I$	n_{I1}	n_{I2}	$\dots$	n_{IJ}

Traktujemy każdą kolumnę jako próbę $n_{\bigstar j}$-elementową z rozkładu wielomianowego o prawdopodobieństwach „wpadnięcia elementu do i-tego koszyka" równych π_{ij} ($\mathbb{E}n_{ij} = \pi_{ij}n_{\bigstar j}$).

Badamy, czy rozkłady ze wszystkich kolumn są takie same, czyli czy

$$\pi_{i1} = \pi_{i2} = \cdots = \pi_{iJ} = \pi_i, \qquad i = 1, 2, \dots, I.$$

Test jednorodności

Estymatorem π_i metody n. w. jest $\hat{\pi}_i = \frac{n_{i\bigstar}}{N}$. Statystyka decyzyjna

$$T = \sum_{j=1}^{J}\sum_{i=1}^{I} \frac{(n_{ij} - n_{\bigstar j}\hat{\pi}_i)^2}{n_{\bigstar j}\hat{\pi}_i} = \sum_{j=1}^{J}\sum_{i=1}^{I} \frac{(Nn_{ij} - n_{i\bigstar}n_{\bigstar j})^2}{Nn_{i\bigstar}n_{\bigstar j}}.$$

jest taka jak w teście niezależności cech i tak samo zbiega według rozkładu do $\chi^2_{(I-1)(J-1)}$ (liczbę stopni swobody można wyliczać także jako: $J(I-1) - (I-1) = (I-1)(J-1)$).

Pojęcia test niezależności χ^2 oraz test jednorodności χ^2 używane są wymiennie.

Dwie cechy binarne

	$B:0$	$B:1$
$A:0$	n_{00}	n_{01}
$A:1$	n_{10}	n_{11}

$$N = n_{00} + n_{01} + n_{10} + n_{11},$$
$$n_{\bigstar j} = n_{0j} + n_{1j},$$
$$n_{i\bigstar} = n_{i0} + n_{i1}.$$

Niezależność cech oznacza, że

$$p_{00} = p_{0\bigstar}p_{\bigstar 0},\ p_{01} = p_{0\bigstar}p_{\bigstar 1},\ p_{10} = p_{1\bigstar}p_{\bigstar 0},\ p_{11} = p_{1\bigstar}p_{\bigstar 1},$$

Fakt

W przypadku dwóch cech binarnych niezależność jest równoważna temu, że $p_{00}p_{11} = p_{01}p_{10}$.

Fakt

W przypadku dwóch niezależnych cech binarnych, dla znanych parametrów N *oraz* $n_{\bigstar 0}$ *i* $n_{0\bigstar}$. *Zmienna* n_{00} *ma rozkład hipergeometryczny Hyper*$(N, n_{0\bigstar}, n_{\bigstar 0})$.

Lady Muriel pije herbatę

Lady Muriel twierdziła, że rozpoznaje, czy mleko zostało dodane do filiżanki z herbatą przed, czy po nalaniu herbaty. W eksperymencie podano jej do degustacji 8 filiżanek. Do 4 z nich (losowo) dodano mleko po nalaniu herbaty, a do pozostałych przed. Wynik eksperymentu

	mleko przed	mleko po
Lady twierdzi, że przed	4	0
Lady twierdzi, że po	0	4

Dokładną p-wartość testu można wyznaczyć korzystając z funkcji prawdopodobieństwa rozkładu hipergeometrycznego (tzw. dokładny test Fishera).

Lady Muriel pije herbatę

```
> fisher.test(cbind(c(4,0),c(0,4)))

        Fisher's Exact Test for Count Data

data:  cbind(c(4, 0), c(0, 4))
p-value = 0.02857
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.339059      Inf
sample estimates:
odds ratio
       Inf

> chisq.test(cbind(c(4,0),c(0,4)))

        Pearson's Chi-squared test with Yates' continuity correction

data:  cbind(c(4, 0), c(0, 4))
X-squared = 4.5, df = 1, p-value = 0.03389
```

Porównywanie średnich

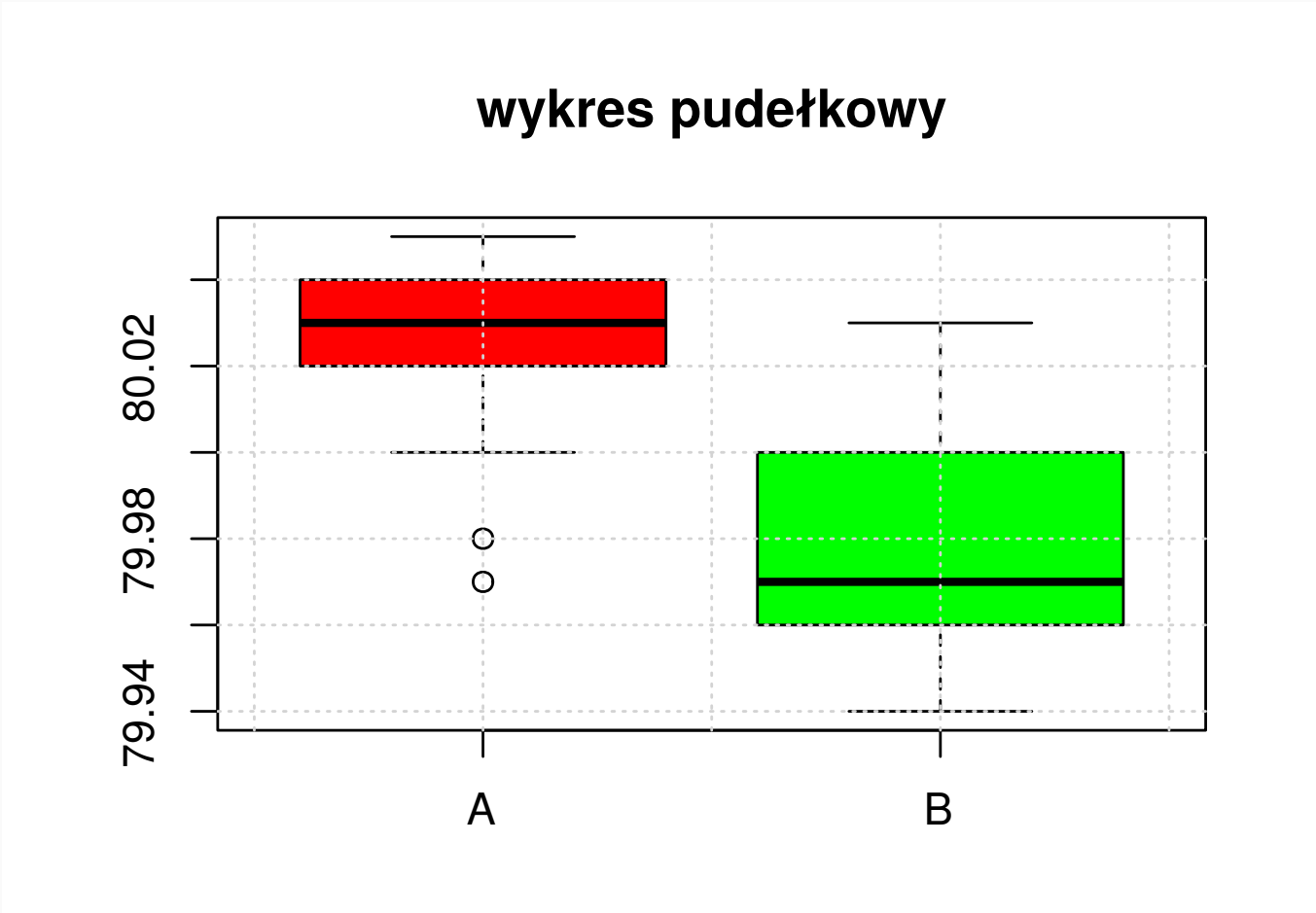
Porównywanie średnich

Mamy do dyspozycji dwie próby $X_1, \ldots, X_n$ oraz $Y_1, \ldots, Y_m$, każda próba z potencjalnie innego rozkładu. Chcemy sprawdzić, czy wartości średnie tych (dwóch!) rozkładów są równe.

Przykład

Dwie metody A i B były użyte do wyznaczenia całkowitego ciepła potrzebnego do ogrzania i stopienia lodu od temperatury –72 °C do wody o temperaturze 0 °C. Wyniki [cal/g]

A	79.98	80.04	80.02	80.04	80.03	80.03	80.04	79.97	80.05	80.03	80.02	80.00	80.02
B	80.02	79.94	79.98	79.97	79.97	80.03	79.95	79.97					



Metody oparte na rozkładzie normalnym

Jeśli próba $X_1, \ldots, X_n$ jest z rozkładu $\mathrm{N}(\mu_X, \sigma^2)$, a niezależna od niej próba $Y_1, \ldots, Y_m$ jest z rozkładu $\mathrm{N}(\mu_Y, \sigma^2)$ (ta sama wariancja!), to

$$\overline{X} - \overline{Y} \sim \mathrm{N}\left(\mu_X - \mu_Y, \sigma^2(n^{-1} + m^{-1})\right).$$

Zazwyczaj wariancja nie jest dana i trzeba ją estymować z próby:

$$s^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{m+n-2}, \quad s_X^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}, \; s_Y^2 = \frac{\sum_{i=1}^m (Y_i - \overline{Y})^2}{m-1}.$$

Zmienna losowa

$$t = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{s\sqrt{n^{-1} + m^{-1}}}$$

ma rozkład t-Studenta z $m + n - 2$ stopniami swobody.

Podsumowanie

Mamy dwie niezależne próby $X_1, \ldots, X_n$ oraz $Y_1, \ldots, Y_m$ z rozkładów, odpowiednio, $\mathrm{N}(\mu_X, \sigma^2)$ oraz $\mathrm{N}(\mu_Y, \sigma^2)$ (μ_X, μ_Y, σ^2 nieznane). Wyznaczamy wartość statystyki decyzyjnej,

$$T = \frac{\overline{X} - \overline{Y}}{s\sqrt{n^{-1} + m^{-1}}}.$$

Dla testu dwustronnego ($H_0 : \mu_X = \mu_Y$ wobec $H_1 : \mu_X \neq \mu_Y$) i poziomu istotności α obszar krytyczny testu jest postaci

$$|T| > c, \qquad c = F^{-1}_{t_{m+n-2}}\left(1 - \frac{\alpha}{2}\right).$$

Dla testu jednostronnego ($H_1 : \mu_X > \mu_Y$) obszar krytyczny:

$$T > c, \qquad c = F^{-1}_{t_{m+n-2}}\left(1 - \alpha\right).$$

Przykład — kontynuacja

A	79.98	80.04	80.02	80.04	80.03	80.03	80.04	79.97	80.05	80.03	80.02	80.00	80.02
B	80.02	79.94	79.98	79.97	79.97	80.03	79.95	79.97					

$$\overline{X} = 80.02, \quad \overline{Y} = 79.98, \quad s_X = 0.024, \quad s_Y = 0.031,$$

$$s = \sqrt{\frac{12s_X^2 + 7s_Y^2}{19}}, \quad s = 0.027,$$

$$T = 3.33.$$

Dla poziomu istotności $\alpha = 0.01$: $c = 2.861$.

p-wartość dla dwustronnego testu mniejsza niż 0.01.

Ale kto powiedział, że wariancje rozkładów X-ów i Y-ów są takie same?

Próby niezależne o nieznanej i być może różnych wariancjach

$$X_i \sim \mathsf{N}(\mu_X, \sigma_X^2), i = 1, \ldots, n, \quad Y_j \sim \mathsf{N}(\mu_Y, \sigma_Y^2), j = 1, \ldots, m.$$

$$s^2 = \frac{s_X^2}{n} + \frac{s_Y^2}{m}, \quad T = \frac{\overline{X} - \overline{Y}}{s}.$$

Statystyka decyzyjna T ma rozkład zbliżony do rozkładu t-studenta z liczbą stopni swobody

$$d \approx \frac{(s^2)^2}{\frac{(s_X^2/n)^2}{n-1} + \frac{(s_Y^2/m)^2}{m-1}}.$$

Dla testu dwustronnego ($H_0 : \mu_X = \mu_Y$ wobec $H_1 : \mu_X \neq \mu_Y$) i poziomu istotności α obszar krytyczny testu jest postaci

$$|T| > c, \qquad c = F_{t_d}^{-1}\left(1 - \frac{\alpha}{2}\right).$$

Przykład — kontynuacja

A	79.98	80.04	80.02	80.04	80.03	80.03	80.04	79.97	80.05	80.03	80.02	80.00	80.02
B	80.02	79.94	79.98	79.97	79.97	80.03	79.95	79.97					

$$\overline{X} = 80.02, \quad \overline{Y} = 79.98, \quad s_X = 0.024, \quad s_Y = 0.031,$$

$$s^2 = \frac{s_X^2}{13} + \frac{s_Y^2}{8} = 0.00017, \quad T = 3.25,$$

$$d = \text{round}(12.03) = 12.$$

Dla poziomu istotności $\alpha = 0.01$: $c = 3.05$.

Przykład obliczeń w R

```
> a=c(79.98,80.04,80.02,80.04,80.03,80.03,80.04,
+ 79.97,80.05,80.03,80.02,80.00,80.02);
> b=c(80.02,79.94,79.98,79.97,79.97,80.03,79.95,79.97);
> t.test(a,b)

Welch Two Sample t-test

data:  a and b
t = 3.2499, df = 12.027, p-value = 0.006939
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.01385526 0.07018320
sample estimates:
mean of x mean of y
 80.02077  79.97875
```

Uwagi o mocy testu porównującego średnie

Moc testu ($H_0 : \mu_X = \mu_Y$ wobec $H_1 : \mu_X \neq \mu_Y$) zależy od

- Różnicy między średnimi $\Delta = |\mu_X - \mu_Y|$ (większa różnica — większa moc),
- Poziomu istotności testu α (większy poziom istotności — większa moc),
- Wariancji prób σ^2 (mniejsza wariancja — większa moc),
- Rozmiarów prób n, m (większe rozmiary — większa moc).

Porównywanie prób „sparowanych” — wstęp

Będziemy rozważać próbę złożoną z par

$$(X_1, Y_1), \ldots, (X_n, Y_n)$$

gdzie niezależne są: X_i od X_j; X_i od Y_j; oraz Y_i od Y_j dla $i \neq j$.

Nie zakłada się natomiast niezależności X_i od Y_i.

Jaki zysk daje parowanie?

Niech

$$\mathbb{E}X_i = \mu_X,\ \mathbb{V}X_i = \sigma_X^2,\ \mathbb{E}Y_i = \mu_Y,\ \mathbb{V}Y_i = \sigma_Y^2, \mathbb{C}(X,Y) = \sigma_{XY}$$

oraz niech $D_i = X_i - Y_i$.

$$\mathbb{E}(\overline{X} - \overline{Y}) = \mathbb{E}\overline{D} = \mu_X - \mu_Y, \quad \mathbb{V}\overline{D} = (\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})/n.$$

Gdyby korelacji między X_i a Y_i nie było, to

$$\mathbb{V}\overline{D} = \mathbb{V}(\overline{X} - \overline{Y}) = (\sigma_X^2 + \sigma_Y^2)/n.$$

Test t-studenta dla par

Jeśli rozkład różnic jest rozkładem normalnym $N(\mu_D, \sigma_D^2)$ (μ_D oraz σ_D^2 nieznane), to zmienna losowa

$$t = \frac{\overline{D} - \mu_D}{s_D/\sqrt{n}}$$

ma rozkład t-Studenta z $n-1$ stopniami swobody.

Test dla hipotezy alternatywnej dwustronnej $\mu_D \neq 0$ ma obszar krytyczny

$$|\overline{D}\sqrt{n}/s_D| > c,$$

gdzie

$$c = F^{-1}_{t_{n-1}}(1 - \frac{\alpha}{2}).$$

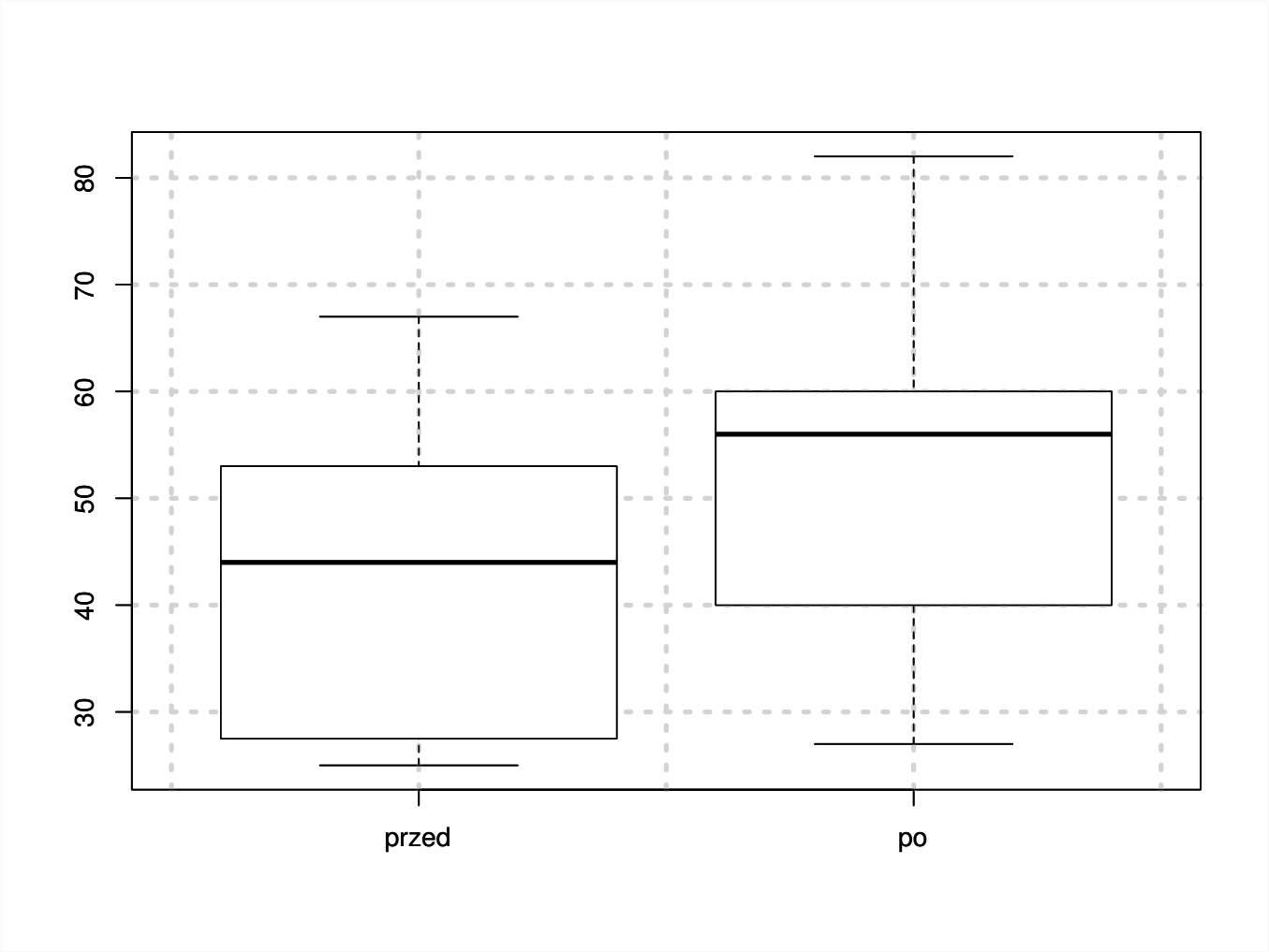
Przykład

W tabeli przedstawiono procentowy udział płytek krwi, które uległy złączeniu w odpowiedzi na odp. stymulację przed i po wypaleniu papierosa przez 11 osób.

przed	25	25	27	44	30	67	53	53	52	60	28
po	27	29	37	56	46	82	57	80	61	59	43

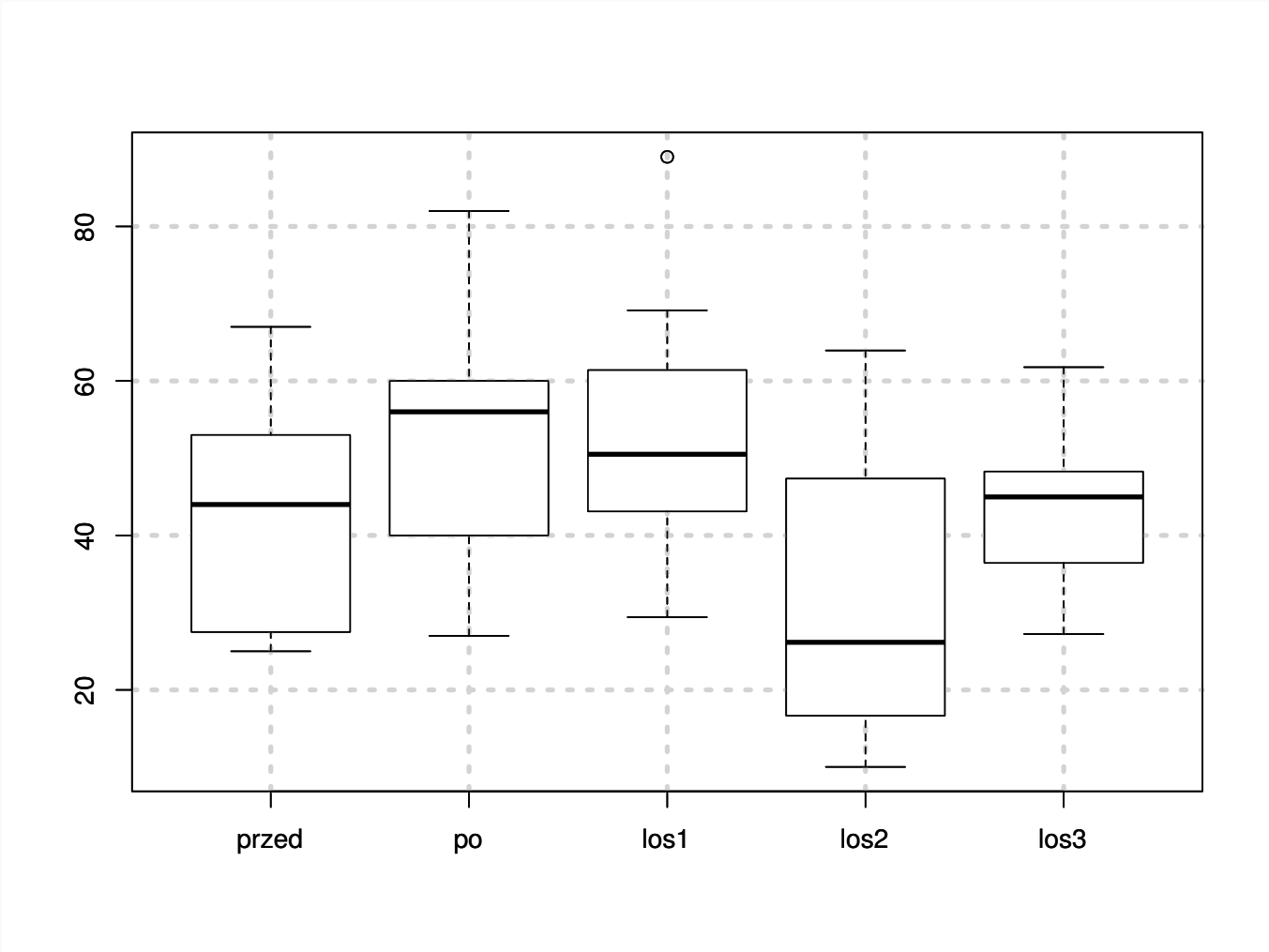
Czy udział płytek, które uległy złączeniu się zwiększył?

Przykład — wykres pudełkowy



Przykład — wykres pudełkowy + symulacje

$\mathrm{los1}$, $\mathrm{los2}$, $\mathrm{los3}$ to wektory losowe o takiej samej liczności jak wektor przed uzyskane z rozkładu normalnego o średniej i wariancji wyestymowanej na podstawie wektora przed.



Przykład — test bez parowania

```
> t.test(po,przed)

      Welch Two Sample t-test

data:  po and przed
t = 1.4164, df = 19.516, p-value = 0.1724
alternative hypothesis: true difference in means is
                        not equal to 0
95 percent confidence interval:
 -4.880458 25.425913
sample estimates:
mean of x mean of y
 52.45455  42.18182
```

Przykład — parowanie

przed	25	25	27	44	30	67	53	53	52	60	28
po	27	29	37	56	46	82	57	80	61	59	43
różnica	2	4	10	12	16	15	4	27	9	−1	15

$$\overline{D} \approx 10.27,\ s_D \approx 4.27.$$

Dla $\alpha = 0.01$, wartość krytyczna testu $c = 3.17$.

Przykład — obliczenia w pakiecie R

```
> t.test(po,przed,paired=TRUE)

        Paired t-test

data:  po and przed
t = 4.2716, df = 10, p-value = 0.001633
alternative hypothesis: true difference in means is
                        not equal to 0
95 percent confidence interval:
  4.91431 15.63114
sample estimates:
mean of the differences
               10.27273
```

Testy nieparametryczne — wstęp

W metodach nieparametrycznych nie zakłada się żadnego konkretnego rozkładu elementów próby.

Test Manna-Whitneya (test sumy rang Wilcoxona)

$X_1, ..., X_n$ próba z pewnego rozkładu F_X,
$Y_1, ..., Y_m$ tzw. próba kontrolna, niezależna od powyższej,
z pewnego rozkładu F_Y.

Chcemy badać, czy (w odpowiednim sensie) wartości X i Y są na podobnym poziomie (z tego samego rozkładu). Np. X_i poziom białych krwinek po kuracji lekiem x, a Y_j poziom tych krwinek w grupie kontrolnej (nie poddawanej kuracji).

Hipoteza zerowa: $H_0: F_X = F_Y$.

Hipoteza alt. jednostronna: $H_1: \mathbb{P}(X > Y) > \mathbb{P}(X < Y)$
lub dwustronna: $H_1: \mathbb{P}(X > Y) \neq \mathbb{P}(X < Y)$.

Idea testu Manna-Whitneya(-Wilcoxona)

1. szeregujemy elementy X_i, Y_j w kolejności rosnącej,
2. sumujemy rangi elementów Y_j (ich numery w uszeregowanym w poprzednim punkcie ciągu),
3. „zbyt mała” lub „zbyt duża” wartość powyższej sumy skłania do odrzucenia hipotezy zerowej (wobec hipotezy alternatywnej dwustronnej).

Przykład

Przykład: $X_1 = 1(1)$, $X_2 = 3(2)$, $Y_1 = 6(4)$, $Y_2 = 4(3)$
(w nawiasach rangi).
Suma rang próby kontrolnej: $R = 4 + 3 = 7$.

Czy to dostatecznie mało/dużo do odrzucenia hipotezy zerowej?

Gdy zachodzi H_0 to rangi przypisane próbie kontrolnej są z równym prawdopodobieństwem równe (u, v) dla każdego $1 \leq u < v \leq m + n$.

Rangi	$(1,2)$	$(1,3)$	$(1,4)$	$(2,3)$	$(2,4)$	$(3,4)$
R	3	4	5	5	6	7

$$\mathbb{P}(R \geq 7) = \frac{1}{6}.$$

Rozkład sumy rang (R) jest stablicowany dla wielu możliwych n i m.

W pakiecie R do przeprowadzania testu Manna-Whitneya-Wilcoxona służy funkcja `wilcox.test()`.

```
> wilcox.test(a,b)

        Wilcoxon rank sum test with continuity correction

data:  a and b
W = 89, p-value = 0.007497
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(a, b) : cannot compute exact p-value with ties
```

Test Wilcoxona dla par

W przypadku, gdy nie mamy podstaw do zakładania, że różnice między wartościami w każdej parze mają rozkład normalny możemy wykorzystać test Wilcoxona:

1. Sortujemy n par według rosnących modułów różnic (między wartościami pary),
2. Nadajemy każdej parze rangę równą pozycji modułu różnicy w uporządkowanym ciągu,
3. Parom o ujemnej różnicy zmieniamy rangi na przeciwne ($x \mapsto -x$),
4. Obliczamy statystykę W_+ równą sumie dodatnich rang,
5. „Zbyt małe" lub „zbyt duże" wartości W_+ świadczą na niekorzyść hipotezy zerowej ($F_X = F_Y$).

$$\mathbb{E}W_+ = \frac{n(n+1)}{4}, \quad \mathbb{V}W_+ = \frac{n(n+1)(2n+1)}{24}.$$

Przykład (koncentracja płytek krwi)

```
> wilcox.test(przed,po,paired=TRUE)

Wilcoxon signed rank test with continuity correction

data:  przed and po
V = 1, p-value = 0.005056
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(przed, po, paired = TRUE) :
  cannot compute exact p-value with ties
```