

Elementarna statystyka

Test Istotności

Alexander Bendikov

Uniwersytet Wrocławski

27 kwietnia 2017

- Wnioskowanie statystyczne:

- 1) **Przedziały ufności:** Używamy ich kiedy naszym celem jest oszacowanie jakiegoś parametru populacji (np. średniej)
- 2) **Testy istotności:** Używamy ich do oceny, czy dane potwierdzają jakieś przypuszczenie odnośnie populacji (hipotezy)

Przykład: Napoje dietetyczne (takie jak Pepsi Zero) używają sztucznych substancji słodzących. Okazuje się, że używane substancje tracą swoje właściwości słodzące wraz z upływem czasu. Producenci kontrolują ten proces chcąc zachować jakość swoich produktów w okresie przydatności do spożycia. Wykwalifikowani testerzy próbują napój dwukrotnie (w odstępie miesiąca), i oceniają „słodkość” w skali od 0 do 10, po czym zapisują różnicę $n_1 - n_2$. Poniżej mamy wyniki 10 testerów:

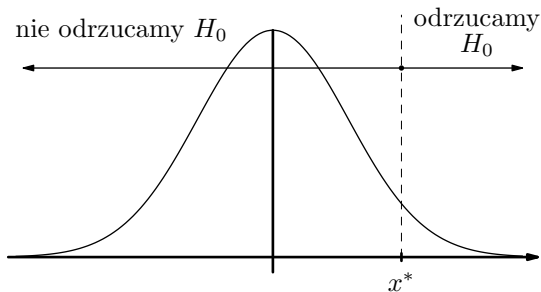
2.0 0.4 0.7 2.0 - 0.4 2.2 - 1.3 1.2 1.1 2.3

Pytanie: Czy te dane potwierdzają przypuszczenie, że napoje dietetyczne tracą swoją „słodkość” z upływem czasu?

Odpowiadamy:

1. Znajdujemy średnią próby $\bar{x} = 1.02$;
2. Niech μ będzie (nieznana) średnią utratą słodyczy w całej partii produkcyjnej (populacji),
Formułujemy hipotezę zerową $H_0 : \mu = 0$ (napój nie traci swojej słodyczy z czasem);
Formułujemy hipotezę alternatywną $H_a : \mu > 0$ (napój traci słodkość z czasem)
3. Założenie: Z wieloletnich doświadczeń wynika, że odchylenie standardowe dla wyników testerów słodkości wynosi $\sigma = 1$. Zakładamy więc, że $\bar{x} \in N(\mu, \frac{\sigma}{\sqrt{n}})$.
4. Przy założeniu hipotezy zerowej H_0 oczekivalibyśmy, że dane będą takie, że średnia próby $\bar{x}$ będzie bliska założonej średniej $\mu = 0$. Idea testu istotności jest wybór takiej krytycznej wartości x^* , że jeżeli $0 \leq \bar{x} \leq x^*$ to uznamy, że *dane potwierdzają hipotezę zerową*, a jeżeli $\bar{x} > x^*$ to uznamy, że *dane nie potwierdzają hipotezy zerowej*, natomiast *potwierdzają hipotezę alternatywną*.

- Zauważmy, że wybór x^* dopuszcza pewną niewielką możliwość błędu: zwykła zmienność losowa może spowodować $\bar{x} > x^*$, i odrzucenie H_0 , pomimo, że hipoteza jest prawdziwa.



W naszym przypadku $\bar{x} = 1,02$, a więc

$$P(\bar{x} > 1,02 \mid \mu = 0) = P\left(\frac{\bar{x}}{1/\sqrt{10}} > \frac{1,02}{1/\sqrt{10}}\right) = P(Z > 3,23) = 6 \cdot 10^{-4}.$$

Wynika stąd, że jeżeli przyjmiemy za $x^* = 1$ (czyli $P(Z > 3,16)$) to odrzucimy H_0 z bardzo małym możliwym błędem:

$$P(\bar{x} > 1,0 \mid \mu = 0) \simeq 6 \cdot 10^{-4}.$$

Ogólnie, używamy następującej terminologii

- z-wartość $= \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}},$
- p-wartość $= P(\bar{x} > x^* \mid \mu = \mu_0).$

5. Istotność statystyczna: W ostatnim kroku oceny danych w kontekście hipotezy H_0 porównujemy wyliczoną p-wartość z wartością ustaloną $\alpha = 0,05, 0,01, 0,001$, którą uznajemy za *decydującą*. Jeżeli p-wartość jest mniejsza lub równa α mówimy, że dane są *statystycznie istotne* na poziomie α .

W naszym przykładzie z testowaniem coli otrzymaliśmy $p = 0,0006$, czyli dane są statystycznie istotne na poziomie $\alpha = 0,001$, ale nie są statystycznie istotne na poziomie $\alpha = 0,0001$. Pobrana próba jest dobrym argumentem za uznaniem $\mu > 0$. Powtórzmy to: otrzymana z próby mała p -wartość (czyli p -wartość mniejsza niż α) jest dobrym statystycznym dowodem przeciwko H_0 .

Rodzaje H_a

Wybieramy pewną wartość μ_0 i formułujemy hipotezę $H_0: \mu = \mu_0$. Hipotezę alternatywną możemy sformułować jako:

1. Hipotezę jednostronną: $H_a: \mu < \mu_0$ lub $H_a: \mu > \mu_0$;
2. Hipotezę dwustronną $H_a: \mu \neq \mu_0$.

Hipoteza jednostronna jest właściwa w przypadku gdy z góry wiemy w którą stronę oczekujemy odchylenia, hipoteza obustronna w przypadku gdy nie wiemy.

Z-test dla średniej populacji

Mamy próbę $x_1, x_2, \dots, x_n \in N(\mu, \sigma)$, przy czym

- σ jest znana
- μ jest nieznana

Przeprowadzamy test hipotezy $H_0 : \mu = \mu_0$ na podstawie próby $x_1, x_2, \dots, x_n$ przy wybranym poziomie istotności α :

1) Ustalamy hipotezy:

$$H_0 : \mu = \mu_0$$

$$H_a : \mu > \mu_0$$

2) Obliczamy wartość $Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ (to jest z-wartość);

3) Ustalamy p -wartość i porównujemy z α ;

4) Wniosek: odrzucamy H_0 , na poziomie istotności α , jeżeli p -wartość jest mniejsza niż α , i odpowiednio przyjmujemy hipotezę H_0 , jeżeli p -wartość jest większa niż α

Przykład: Puszka coli powinna zawierać 300 ml napoju. W rzeczywistości występuje pewna zmienność. Rozkład ilości coli jest zmienną normalną z $\sigma = 2$ ml. Inspektor zmierzył zawartość coli w $n = 6$ puszkach i otrzymał następujące dane:

299,4 297,7 301,0 298,9 300,2 297,0

Czy te dane stanowią przekonujący (na poziomie istotności $\alpha = 0,1$) uzasadnienie tezy, że średnia zawartość coli w puszcze wynosi 300 ml?

Rozwiązanie:

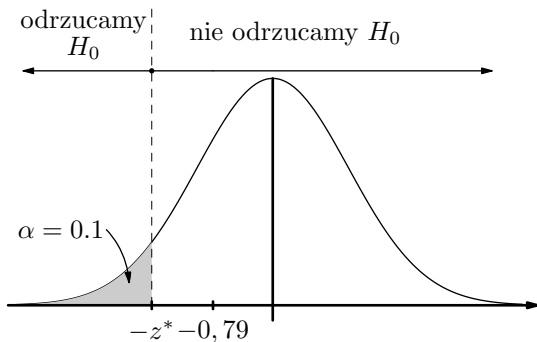
1. $H_0 : \mu = 300 \text{ ml}$

$H_a : \mu < 300 \text{ ml};$

Zakładamy, że hipoteza H_0 jest prawdziwa.

2. $\bar{x} = 299,03$, skoro $\sigma = 3$ jest znana, używamy Z-testu i otrzymujemy

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{299,03 - 300}{3/\sqrt{6}} = -0,79;$$



3. Mamy p -wartość $p = P(Z \leq -0,79) = 0,2148$
4. p -wartość $> \alpha = 0,1$: Próba nie stanowi wystarczającego uzasadnienia (na poziomie istotności $\alpha = 0,1$) odrzucenia hipotezy H_0 na rzecz hipotezy H_a . Akceptujemy hipotezę $H_0 : \mu = 300$ ml.

Przykład: Przyjmijmy, że w przypadku braku specjalnych przygotowań wyniki SATM (Scholastic Assessment Test in Mathematics - amerykański odpowiednik matury z matematyki) mają rozkład normalny z $\mu = 475$ i $\sigma = 100$. Grupa $n = 100$ uczniów przechodzi intensywny kurs przygotowawczy do SATM, którego celem jest uzyskanie lepszych wyników na teście.

- a) Po egzaminie okazuje się, że dla tej grupy uczniów otrzymaliśmy $\bar{x} = 491,4$. Czy ten wynik uzasadnia przypuszczenie (na poziomie istotności 0,05), że kurs istotnie wpływa na wzrost wyniku SATM?
- b) Jeżeli $\bar{x} = 491,5$ to czy taki rezultat jest istotny na poziomie 0,05?

Rozwiązanie:

1. $H_0 : \mu = 475$

$H_a : \mu > 475;$

2. Ponieważ $\sigma = 100$ jest znana, to stosujemy test Z :

a) $z\text{-wartość} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{491,4 - 475}{100 / \sqrt{100}} = 1,64;$

b) $z\text{-wartość} = \frac{491,5 - 475}{100 / \sqrt{100}} = 1,65;$

3. Dla górnego 5% pola mamy wartość krytyczną $z^* = 1,645$. W takim razie:

a) Skoro $z = 1,64 < z^* = 1,645$ nie odrzucamy hipotezy $H_0 : \mu = 475$, dane nie potwierdzają, że kurs wpływa na wzrost wyników.

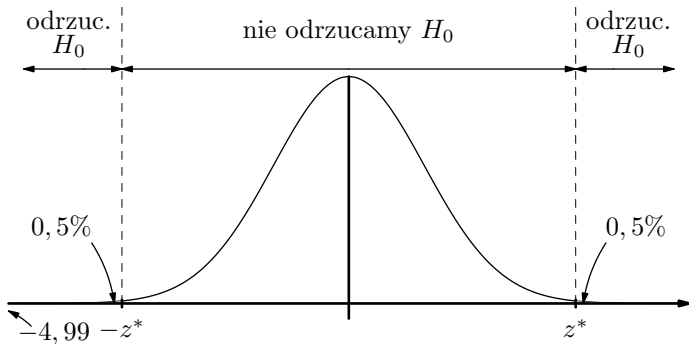
b) W tym przypadku $z = 1,65 > z^* = 1,645$, więc odrzucamy hipotezę H_0 , dane potwierdzają, że kurs podnosi wyniki (na poziomie istotności $\alpha = 0,05$).

Przykład: Laboratorium zlecono sprawdzenie, czy zawartość substancji aktywnej w pewnym leku istotnie wynosi 0,86. Laboratorium dokonuje 3 pomiarów, i otrzymuje wyniki:

$$X: 0,9403 \quad 0,8363 \quad 0,8447 \quad \Rightarrow \quad \bar{x} = 0,8404$$

Zakładamy, że $X \in N(\mu, \sigma)$ z $\sigma = 0,0068$. Czy jest wystarczające uzasadnienie (na poziomie istotności $\alpha = 0,01$), że $\mu \neq 0,86$?

- 1) Hipotezy: $H_0: \mu = 0,86$, $H_a: \mu \neq 0,86$;
- 2) Używamy Z -testu $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{0,8404 - 0,86}{0,0068 / \sqrt{3}} = -4,99$.
- 3) Sprawdzamy p -wartość: $p = P(|Z| > 4,99) = 2,6 \cdot 10^{-7}$
- 4) Skoro $p < \alpha$ to odrzucamy hipotezę H_0 na rzecz H_a na poziomie istotności $\alpha = 0,01$. Wartość $\bar{x} = 0,8404$ jest istotna na poziomie 0,01.



Użycie przedziałów ufności

Przykład: Pan profesor statystyki twierdzi, że średni wiek studentów statystyki to 24 lata. Chcielibyśmy zweryfikować jest twierdzenie, na poziomie istotności 0,05. Losowa próba 116 studentów statystyki miała średnią 22,7 lat, a wiadomo, że odchylenie standardowe σ wynosi 5,68. Przedział ufności, $z^* = 1,96$, $SE = 0.527$, $m = 1,03$, czyli przedział ufności na poziomie $C = 1 - \alpha = 0,95$ to $[21,67, 23,73]$

- 1) Hipotezy: $H_0 : \mu = 24$, $H_a : \mu \neq 24$;
- 2) Z-test: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{22,7 - 24}{5,68 / \sqrt{116}} = -2,47$
- 3) Możemy zastosować następującą regułę: akceptujemy H_0 jeżeli wartość μ należy do przedziału ufności, odrzucamy, jeżeli nie należy (dla H_a dwustronnej). Równoważnie, przyjmujemy, jeżeli $z \in [-z^*, z^*]$, odrzucamy w przeciwnym przypadku. Zauważmy, że jest to dokładnie to samo, co ustalanie p -wartości i porównanie z α . W naszym przypadku $z < -z^*$, więc z nie należy do $z \in [-z^*, z^*]$, i hipotezę H_0 odrzucamy. Po wyznaczeniu przedziału ufności od razu widać, że 24 leży poza nim.

Przyjmujemy $H_0 : \mu = \mu_0$ na poziomie istotności α wtedy i tylko wtedy gdy μ_0 należy do przedziału ufności μ na poziomie ufności $C = 1 - \alpha$. Innymi słowy przyjmujemy $H_0 : \mu = \mu_0$ jeżeli $-z^* < z < z^*$ lub, równoważnie, jeżeli

$$\bar{x} - m < \mu_0 < \bar{x} + m, \quad m = z^* \frac{\sigma}{\sqrt{n}}.$$

W naszym przykładzie hipotezę $H_0 : \mu = 24$ odrzucamy, ale od razu widać, że hipotezę $H_0 : \mu = 23$ przyjęlibyśmy.

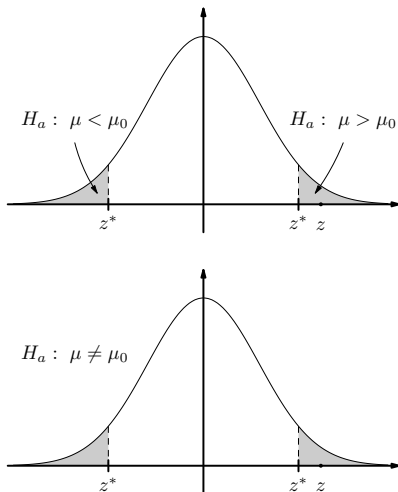
Prawdopodobieństwa błędów I i II rodzaju

Decyzja:

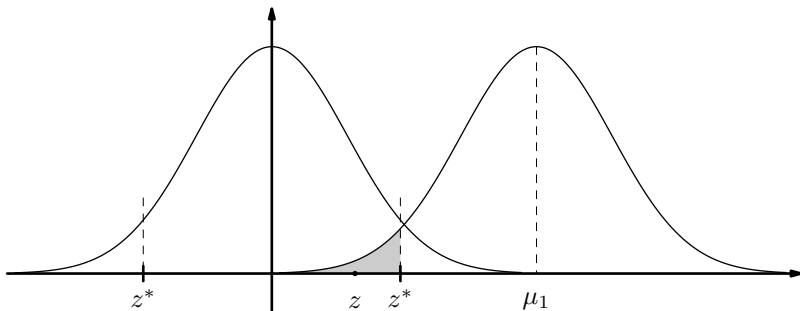
	H_0 prawdziwa	H_a prawdziwa	
H_0 odrzucona	Błąd typu I	OK	← Moc testu (w stronę H_a)
H_0 przyjęta	OK	Błąd typu II	

1. Błąd typu I i istotność

Prawdopodobieństwo błędu typu I to $P(\text{odrzućenie } H_0 | H_0 \text{ prawdziwa})$. Jest to dokładnie istotność testu α dla H_a dwustronnej i $\alpha/2$ dla H_a jednostronnej



2. Błąd typu II



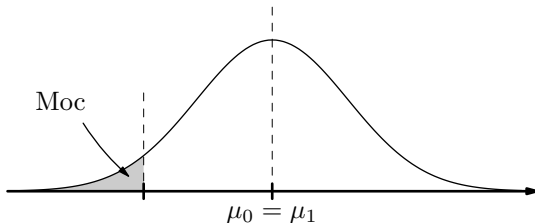
Rysunek: Prawdopodobieństwo błędu typu II ($H_a : \mu = \mu_1$)

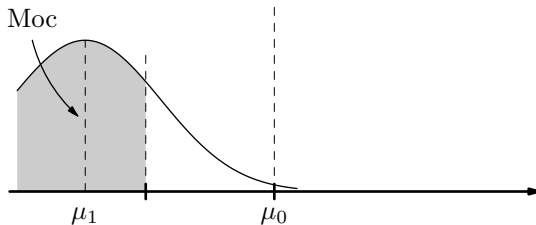
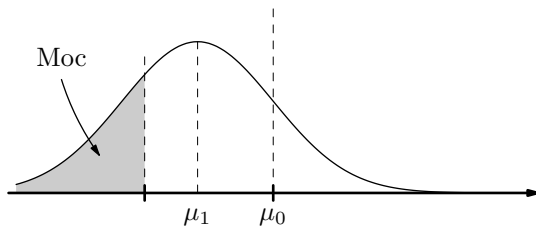
Prawdopodobieństwo błędu typu II to $P(\text{przyjęcie } H_0 | H_a \text{ prawdziwa})$. Jeżeli $\mu = \mu_1$ to jest równe $P(\bar{x} > \mu_0 - z^* \sigma / \sqrt{n} | \mu = \mu_1)$. Prawdopodobieństwo to dąży do 0 gdy $\mu_1 \rightarrow -\infty$.

Przykład: $X \sim N(\mu, \sigma)$, $\sigma = 60$, $n = 840$ i $H_0 : \mu = 275$. Przyjmujemy $\alpha = 0,01$. Mamy

- Typ II ($H_a : \mu = 270$) $P(\bar{x} > 275 - 2,576 \cdot 60/\sqrt{840} \mid \mu = 270) = P(\bar{x} > 269,667 \mid \mu = 270) = P(z > -1.5) = 0.5596$
- Typ II ($H_a : \mu = 268$) $P(z > 0.806) = 0.2119$
- Typ II ($H_a : \mu = 266$) $P(z > 1,77) = 0.0384$

Moc testu = $P(\text{przyjęta } H_a \mid H_a \text{ prawdziwa})$ (zależy od H_a). Załóżmy, że $\mu = \mu_1$ i w teście przyjęliśmy $H_0 : \mu = \mu_0$, $H_a : \mu < \mu_0$





Moc testu istotności mierzy jego zdolność do wykrycia hipotezy alternatywnej.

$$\begin{aligned}\text{Moc } (\mu = \mu_1) &= P(\text{przyjęcie } H_a \mid H_a \text{ prawdziwa} : \mu = \mu_1) \\ &= 1 - P(\text{odrzućenie } H_0 \mid H_a \text{ prawdziwa} : \mu = \mu_1) \\ &= 1 - \text{Typ II } (\mu = \mu_1).\end{aligned}$$

Uwaga: Gdy μ_1 oddala się od μ_0 Moc testu $(\mu = \mu_1)$ rośnie do 1.

Przykład: $H_0 : \mu = 275$, $H_a : \mu < 275$; $X \in N(\mu, \sigma)$,

$\sigma = 60$, $n = 840$, $\alpha = 0,01$. Otrzymujemy:

- $\text{Moc } (\mu = 270) = 1 - 0,5596 = 0.400$
- $\text{Moc } (\mu = 268) = 1 - 0,2119 = 0.798$
- $\text{Moc } (\mu = 266) = 1 - 0,0384 = 0.962$

Przykład: Średnia średnica X „czegoś” powinna wynosić 2,000 mm. Przyjmujemy, że $X \in N(\mu, \sigma)$, gdzie $\sigma = 0,010$. Postanawiamy spróbować odrzucić hipotezę $H_0 : \mu = 2,000$ na poziomie istotności $\alpha = 0,05$.

(1) Pobieramy próbę SRS z $n = 5$ i pewną średnią $\bar{x}$. Ustalamy regułę dla odrzucenia $H_0 : \mu = 2$ na rzecz $H_a : \mu \neq 2$. To jest hipoteza dwustronna, więc przyjmujemy H_0 jeżeli $\mu = 2$ leży w $C = 0,95$ -przedziale ufności, czyli

$\bar{x} \pm m = \bar{x} \pm z^* \sigma / \sqrt{n} = \bar{x} \pm 1,96 \cdot 0,01 / \sqrt{5} = \bar{x} \pm 0,0088$. To jest równoważne

$$2 - 0,0088 < \bar{x} < 2,0088.$$

(2) Znajdujemy błędy typu I i II:

$$P(\text{błąd I typu}) = P(\text{przyjęcie } H_0 \mid H_0 \text{ prawdziwa}) = \alpha = 0,05$$

$$P(\text{błąd II typu} : \mu = 2,015) =$$

$$= P(\text{odrzućenie } H_a \mid H_a \text{ prawdziwa} : \mu = 2,015) =$$

$$= P(1,991 < \bar{x} < 2,001 : \mu = 2,015) =$$

$$= P(-5,32 < z < -1,39) = 0,0823.$$

Przykład: Mamy próbę SRS rozmiaru $n = 9$ z populacji $N(\mu, \sigma)$, gdzie $\sigma = 1$. Chcemy przetestować hipotezę $H_0 : \mu = 0$ względem $H_a : \mu > 0$. Decydujemy, że H_0 odrzucimy, jeżeli $\bar{x} > 0,5$, w przeciwnym przypadku przyjmujemy H_0 .

1. Błąd typu I to odrzucenie H_0 jeżeli H_0 jest prawdziwa.

$$\alpha = P(\bar{x} > 0,5) = P\left(\frac{\bar{x} - 0}{1/3} > \frac{0,5}{1/3}\right) = P(z > 1,5) \approx 0,07$$

2. Błąd typu II ($\mu = 0,7$) to odrzucenie H_a jeżeli H_a jest prawdziwe:

$$\begin{aligned} P(\text{typ II} : \mu = 0,7) &= P(\bar{x} < 0,5 : \mu = 0,7) = \\ &= P\left(\frac{\bar{x} - 0}{1/3} < -0,6\right) = P(z < -0,6) \approx 0,27 \end{aligned}$$

3. Moc ($\mu = 0,7$) $= 1 - 0,27 = 0,73$

4. $P(\text{typ II} : \mu = 1) = P(z < -1,5) = 0,07$

5. Moc ($\mu = 1$) $= 1 - 0,07 = 0,93$.

Przykład: Chirurg przypuszcza, że jeżeli poda swoim pacjentom po operacji witaminę E to przyspieszy pomyślny zrost i powrót do zdrowia. Chcąc potwierdzić swoje przypuszczenia wybiera próbę 38 pacjentów. Wiadomo, że powrót do zdrowia trwa średnio 33 dni (bez witaminy E), z odchyleniem standardowym 5,7 dnia. Pacjenci z tej próby dostaną witaminę E, a lekarze opiekujący się pacjentami po operacji, którzy podejmują decyzje o zakończeniu rekonwalescencji nie wiedzą którzy pacjenci należą do próby. $\bar{x}$ jak zwykle oznacza średni czas powrotu do zdrowia pacjentów wybranych do badania.

- a) Co w tej sytuacji oznacza błąd typu I?
- b) Co w tej sytuacji oznacza błąd typu II?
- c) Przy założeniu, że nowa procedura nie ma wpływu na czas powrotu do zdrowia, jaki będzie (przybliżony) rozkład $\bar{x}$?
- d) Jeżeli $\bar{x} = 31,9$ i nowa procedura nie ma wpływu na czas powrotu do zdrowia, to jakie jest prawdopodobieństwo, że grupa wyzdrowiała tak szybko (albo jeszcze szybciej)?
- e) Czy taki wynik byłby statystycznie istotny?
- f) Przy takim $\bar{x}$ jakie byłyby wnioski odnośnie efektywności procedury?

Ustalamy $H_0 : \mu = 33$, $H_a : \mu < 33$.

- a) Błąd typu I: wniosek, że witamina E pomaga, kiedy w istocie nie pomaga;
- b) Błąd typu II: wniosek, że witamina E nie pomaga, kiedy w istocie pomaga;
- c) $\bar{x} \in N(33, 5, 7/\sqrt{38})$;
- d) Jeżeli H_0 jest prawdziwa, to $z = 31,9 - 33/0,925 = -1,19$. W takim razie $p = P(z < -1,19) = 0,1172$.
- e) Nie. Około 12% wszystkich prób miałyby średnie czasy powrotu do zdrowia takie lub szybsze. Nawet przy $\alpha = 0,1$ te dane nie są statystycznie istotne.
- f) Ten eksperyment nie dostarcza mocnych statystycznych dowodów przeciwko H_0 .