# Statystyka i Analiza Danych

## W5: Testowanie hipotez


dr hab. inż. Katarzyna Filipiak, prof. PP

Instytut Matematyki
Politechnika Poznańska


2023/2024

# Hipoteza statystyczna

Hipoteza statystyczna (parametryczna) – przypuszczenie dotyczące parametrów badanej populacji.

Testowanie hipotez – wnioskowanie statystyczne polegające na podjęciu decyzji na podstawie próby czy hipotezę należy odrzucić czy też nie.

| Hipoteza | obustronna | lewostronna | prawostronna |
|---|---|---|---|
| $H_0$ | $=$ | $\geqslant (=)$ | $\leqslant (=)$ |
| $H_1$ | $\neq$ | $<$ | $>$ |
| notacja w R | `two.sided` | `less` | `greater` |

# Przykłady hipotez

1. Lekarze postulują, że dzieci w wieku 10 lat są zbyt otyłe. Wiadomo, że średnia waga zdrowego 10-latka to 37 kg. Sformułuj interesującą hipotezę.

2. Żeby zaplanować produkcję istotna jest znajomość zmienności czasów prodkcji pewnego elementu. Metodę produkcji tego elementu uznaje się za efektywną, jeżeli zmienność (wariancja) czasu trwania procesu jest mniejsza niż 8 godz$^2$. Inżynier twierdzi, że nowa metoda spełnia to założenie. Sformułuj interesującą hipotezę.

3. Meteorolodzy twierdzą, że klimat się zmienia. Na podstawie średniej dobowej temperatury na wiosnę, która powinna wynosić $10^\circ$C, sformułuj odpowiednią hipotezę.

# Przykład - motywacja

W celu zwiększenia produkcji miedzi przez kopalnię *ABC* inżynier szuka nowego sposobu wydobycia miedzi z dużego złoża. W chwili obecnej znaczna część rudy jest odrzucana ze względu na zbyt niską jakość (związaną z bieżącymi cenami miedzi na świecie). Inżynier zainteresowany jest nowym procesem wydobycia opartym na wypłukiwaniu miedzi ze złoża za pomocą bakterii. Efektywność tej metody zależy od odmiany bakterii użytej w procesie oraz od struktury samego złoża.

Ponieważ nowa metoda została już kilkukrotnie z powodzeniem zastosowana, przeprowadzone zostanie badanie statystyczne w celu stwierdzenia, czy metoda ta powinna być stosowana na szeroką skalę.

Pobranych zostanie losowo 100 porcji rudy z nienaruszonego złoża, które następnie potraktowane zostaną obiecującą odmianą bakterii. Z każdej próbki zostanie wyznaczona ostateczna ilość wydobytej miedzi (w funtach na tonę).

# Przykład - c.d.

Sukces nowej metody zależeć będzie od poziomu średniej ilości miedzi, $\mu$, która w końcu zostanie wydobyta z całego złoża.

Na podstawie bieżących cen miedzi oraz kosztów przeprowadzenia nowego procesu za poziom „krytyczny" uznaje się 36 funtów miedzi wydobytej z tony złoża. Sformułuj interesującą badacza hipotezę.

$$H_0: \ \mu \leqslant 36 \qquad\qquad H_1: \ \mu > 36$$

# Test

Statystyka testowa – statystyka, której wartości pozwalają zdecydować o odrzuceniu $H_0$.

Obszar krytyczny (obszar odrzucenia) testu – zbiór tych wartości statystyki testowej, dla których $H_0$ jest odrzucana.

Obszar krytyczny: $R$

# Przykład - statystyka i poziom krytyczny

Decyzja o wdrożeniu nowej metody musi zostać podjęta bez wiedzy, która z hipotez jest prawdziwa. Można tego dokonać wykorzystując średnią z próby, $\overline{X}$ (wysoki poziom $\overline{X}$ wskazywać będzie na efektywność nowej metody). Należy ustalić pewną wartość krytyczną, która będzie służyć jako punkt „zaprzeczenia" między wdrożeniem nowej metody a jej wstrzymaniem.

Wartość krytyczna ustalona została przez specjalistów na 40 funtów miedzi z tony złoża. Wartość ta jest większa niż testowane 36 funtów/tonę i ustalana jest zwykle właśnie w taki sposób, aby leżała poza zasięgiem hipotezy zerowej (zmniejsza to szansę, że nietypowy wynik próby wpłynie na podjęcie złej decyzji).

# Testowanie hipotez – błędy

| Wynik Testu | Stan Natury | |
|---|---|---|
| | $H_0$ prawdziwa | $H_0$ fałszywa |
| Odrzucić $H_0$ | Błąd (I-go rodzaju) | Poprawna |
| Nie odrzucić $H_0$ | Poprawna | Błąd (II-go rodzaju) |

$\alpha =$ prawdopodobieństwo popełnienia błędu I-go rodzaju (poziom istotności)

$\beta =$ prawdopodobieństwo popełnienia błędu II-go rodzaju

$1 - \beta$ – moc testu

# Przykład - błędy

Dla zasad wprowadzonych w przykładzie z miedzią, przy prawdziwości hipotezy zerowej dowolna wartość średnia nie większa niż 36 jest dopuszczalna, np. $\mu = 12{,}3$, $\mu = 27{,}5$, $\mu = 34{,}9$. Wyznaczenie błędów I-go i II-go rodzaju odbywa się zwykle dla najgorszego przypadku, gdy $\mu$ leży dokładnie w granicy oznaczanej przez $\mu_0$ (u nas $\mu_0 = 36$).

Inżynier nie zna wartości $\sigma$ dla nowej metody, jednakże na podstawie swoich kompetencji ustalił, że na wielkość $\sigma$ może wpływać błąd techniczny wagi oraz straty wynikające z wymywania rudy wodą, a więc $\sigma = 25$ (funtów/tonę).

Na podstawie centralnego twierdzenia granicznego $\overline{X} \sim N(\mu; 2{,}5)$.

Wyznacz prawdobodobieństwa popełnienia błędu I-go i II-go rodzaju.

# Przykład - błąd I-go rodzaju

$$\begin{aligned} \alpha &= P(\text{ odrzucenie } H_0 \mid H_0 \text{ jest prawdziwa }) \\ &= P(\overline{X} > 40 \mid \mu \leqslant 36) \end{aligned}$$

Załóżmy, że $\mu = 12{,}3$. Wówczas $\overline{X} \sim N(12{,}3,\ 2{,}5)$ oraz

$$\alpha = P(\overline{X} > 40 \,|\, \mu = 12{,}3) = 1 - F(40) = 1 - \mathtt{pnorm}(40,\ 12.3,\ 2.5) = 0$$

Załóżmy, że $\mu = 32$. Wówczas $\overline{X} \sim N(32,\ 2{,}5)$ oraz

$$\alpha = P(\overline{X} > 40 \,|\, \mu = 32) = 1 - F(40) = 1 - \mathtt{pnorm(40,32,2.5)} = 0{,}00069$$

Załóżmy, że $\mu = 36$. Wówczas $\overline{X} \sim N(36,\ 2{,}5)$ oraz

$$\alpha = P(\overline{X} > 40 \,|\, \mu = 36) = 1 - F(40) = 1 - \mathtt{pnorm(40,36,2.5)} = 0{,}0548 \approx 5\%$$

# Przykład - błąd II-go rodzaju

$$\begin{aligned} \beta &= P(\text{ nie odrzucenie } H_0 \mid H_0 \text{ jest fałszywa }) \\ &= P(\overline{X} \leqslant 40 \mid \mu > 36) \end{aligned}$$

Załóżmy, że $\mu = 37$. Wówczas $\overline{X} \sim N(37,\ 2{,}5)$ oraz

$$\beta = P(\overline{X} \leqslant 40 \,|\, \mu = 37) = F(40) = \texttt{pnorm}(40,\ 37,\ 2.5) = 0{,}8849$$

Załóżmy, że $\mu = 42$. Wówczas $\overline{X} \sim N(42,\ 2{,}5)$ oraz

$$\beta = P(\overline{X} \leqslant 40 \,|\, \mu = 42) = F(40) = \texttt{pnorm}(40,\ 42,\ 2.5) = 0{,}2119$$

Załóżmy, że $\mu = 45$. Wówczas $\overline{X} \sim N(45,\ 2{,}5)$ oraz

$$\beta = P(\overline{X} \leqslant 40 \,|\, \mu = 45) = F(40) = \texttt{pnorm}(40,\ 45,\ 2.5) = 0{,}0228$$

# Procedura testowa – podsumowanie

1. Sformułowanie hipotezy
2. Wybór statystyki testowej $\hat{\theta}$
3. Wyznaczenie obszaru krytycznego $R$
4. Podjęcie decyzji:

   $\hat{\theta} \in R$: odrzucamy hipotezę $H_0$ (przyjmujemy $H_1$)

   $\hat{\theta} \notin R$: brak podstaw do odrzucenia $H_0$
5. Interpretacja:

   $\hat{\theta} \in R$: na poziomie istotności $\alpha$ **dane potwierdzają** hipotezę badacza ($H_1$)

   $\hat{\theta} \notin R$: na poziomie istotności $\alpha$ **dane nie potwierdzają** hipotezy badacza ($H_1$)

# Przykład - decyzja

SCENARIUSZ 1. Przypuśćmy, że ze 100 porcji wydobytej miedzi uzyskano średnio $47{,}5$ funtów/tonę. Ponieważ wartość ta leży w obszarze odrzucenia $H_0$, hipoteza $\mu \leqslant 36$ musi być odrzucona. Inżynier wnioskuje, że wprowadzenie nowej metody jest korzystne i rekomenduje ją zarządowi kopalni. Inżynier ma komfort psychiczny, bo obliczona średnia przekracza założony poziom 36 funtów/tonę (co najmniej czterokrotnie, bo $(47{,}5 - 36)/2{,}5 > 4$).

# Przykład - decyzja

SCENARIUSZ 2. Przypuśćmy, że ze 100 porcji wydobytej miedzi uzyskano średnio $32{,}1$ funtów/tonę. Wartość ta leży w obszarze nie-odrzucenia $H_0$ i inżynier wnioskuje, że nowa metoda jak na razie nie jest ekonomiczna. Inżynier zdaje sobie sprawę, że jest znaczne prawdopodobieństwo, że jego decyzja nie jest poprawna, nie ma jednak możliwości zdobycia tej wiedzy. Problem wdrożenia nowej metody wydobywania miedzi może zostać otwarty ponownie, np. gdy wzrosną ceny miedzi czy też znaleziona zostanie lepsza odmiana bakterii.

# Testowanie hipotez o średniej populacyjnej $\mu$

# Test hipotezy o $\mu$ (1)

Założenie: $X_i \sim N(\mu, \sigma)$, $\sigma$ - znane

Statystyka testowa: $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \underset{H_0}{\sim} N(0,1)$

| $H_0$ | $H_1$ | Obszar krytyczny $R$ zależny od $\alpha$ |
|---|---|---|
| $\mu = \mu_0$ | $\mu \neq \mu_0$ | $(-\infty, -z_{1-\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$ |
| $\mu \leqslant \mu_0$ | $\mu > \mu_0$ | $(z_{1-\alpha}, \infty)$ |
| $\mu \geqslant \mu_0$ | $\mu < \mu_0$ | $(-\infty, -z_{1-\alpha})$ |

$z_\beta -$ kwantyl rozkładu $N(0,1)$

Testowanie hipotez o $\mu$ w R:

`z.test(dane, sigma.x` $= \sigma$, `mu` $= \mu_0$, `alternative="two.sided")`

UWAGA! Wymagany pakiet `BSDA`

# Test hipotezy o $\mu$ (2)

Założenie: $X_i$ o dowolnym rozkładzie, próba **duża** ($n \geqslant 30$)

Statystyka testowa: $$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \underset{H_0}{\sim} N(0,1)$$

| $H_0$ | $H_1$ | Obszar krytyczny $R$ zależny od $\alpha$ |
| --- | --- | --- |
| $\mu = \mu_0$ | $\mu \neq \mu_0$ | $(-\infty, -z_{1-\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$ |
| $\mu \leqslant \mu_0$ | $\mu > \mu_0$ | $(z_{1-\alpha}, \infty)$ |
| $\mu \geqslant \mu_0$ | $\mu < \mu_0$ | $(-\infty, -z_{1-\alpha})$ |

$z_\beta$ – kwantyl rozkładu $N(0,1)$

Testowanie hipotez o $\mu$ w R:

```
zsum.test(mean(dane), sd(dane), length(dane), mu = μ0, alternative="two.sided")
```

UWAGA! Wymagany pakiet `BSDA`

Założenie: $X_i \sim N(\mu, \sigma)$, $\sigma$ - nieznane

Statystyka testowa: $t = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \underset{H_0}{\sim} t_{n-1}$

| $H_0$ | $H_1$ | Obszar krytyczny $R$ zależny od $\alpha$ |
| --- | --- | --- |
| $\mu = \mu_0$ | $\mu \neq \mu_0$ | $(-\infty, -t_{n-1,1-\alpha/2}) \cup (t_{n-1,1-\alpha/2}, \infty)$ |
| $\mu \leqslant \mu_0$ | $\mu > \mu_0$ | $(t_{n-1,1-\alpha}, \infty)$ |
| $\mu \geqslant \mu_0$ | $\mu < \mu_0$ | $(-\infty, -t_{n-1,1-\alpha})$ |

$t_{n-1,\beta} -$ kwantyl rozkładu $t_{n-1}$

Testowanie hipotez o $\mu$ w R:

```
t.test(dane, mu = μ0, alternative="two.sided")
```

UWAGA! Inne opcje `alternative`: `"greater"`, `"less"`

# Przykład 2

Producent opon twierdzi, że średnia żywotność nowo wprowadzonych na rynek opon przekracza 50 000 mil. W celu weryfikacji tego twierdzenia, specjalna agencja pobrała próbę 8 zestawów opon i przetestowała je na drogach. Otrzymano następujące przebiegi (w tys. mil): 53,417; 51; 48,583; 53; 49; 51; 52; 50. Następnie obliczono, że średnia żywotność testowanych opon to 51 000 mil, natomiast odchylenie standardowe żywotności testowanych opon to 1760 mil. Przyjmując poziom istotności 5% i normalność rozkładu żywotności opon zweryfikuj twierdzenie producenta. Zweryfikuj hipotezę dla poziomu istotności 10%.

$$n = 8, \quad \overline{x} = 51\,000, \quad s = 1760, \quad \alpha = 0{,}05, \quad \sigma - \text{nieznane}$$

$$H_0: \ \mu \leqslant 50\,000 \qquad\qquad H_1: \ \mu > 50\,000$$

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{51000 - 50000}{1760/\sqrt{8}} = 1{,}61$$

# Przykład 2 - rozwiązanie

# Przykład 2 - rozwiązanie dla $\alpha = 0{,}10$

# Przykład 2 - p-wartość

$$H_0: \ \mu \leqslant 50\,000 \qquad\qquad H_1: \ \mu > 50\,000$$

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{51000 - 50000}{1760/\sqrt{8}} = 1{,}61$$

$\alpha = 5\%$: $\quad t \notin R \quad$ brak podstaw do odrzucenia $H_0$

$\alpha = 10\%$: $\quad t \in R \quad$ odrzucamy $H_0$

$$0{,}05 < p - \text{value} < 0{,}10$$

$p-$value – zmienna losowa oznaczająca wartość „graniczną”, decydująca o nieodrzuceniu / odrzuceniu hipotezy $H_0$

# Przykład 2 - rozwiązanie w R

$$H_0: \ \mu \leqslant 50\,000 \qquad\qquad H_1: \ \mu > 50\,000$$

```
        One Sample t-test

data: dane
t = 1.607, df = 7, p-value = 0.07604
alternative hypothesis: true mean is greater than 50000
95 percent confidence interval:
 49821.08       Inf
sample estimates:
mean of x
   51000
```

# Procedura testowa w R

1. Sformułowanie hipotezy
2. Wybór funkcji służącej do testowania
3. Wyznaczenie $p$-wartości
4. Podjęcie decyzji:

   $\alpha > p$-wartość: odrzucamy hipotezę $H_0$ (przyjmujemy $H_1$)

   $\alpha < p$-wartość: brak podstaw do odrzucenia $H_0$
5. Interpretacja:

$\alpha > p$-wartość: na poziomie istotności $\alpha$ dane **potwierdzają** hipotezę badacza ($H_1$)

$\alpha < p$-wartość: na poziomie istotności $\alpha$ dane **nie potwierdzają** hipotezy badacza ($H_1$)

# Test a przedział ufności

Przykład: $X_i \sim N(\mu, \sigma)$, $\sigma$ nieznane

Hipoteza: $H_0: \ \mu = \mu_0$ przeciwko $H_1: \ \mu \neq \mu_0$

Obszar odrzucenia $H_0$: $t \in R \quad \Leftrightarrow$

$$t < -t_{n-1,1-\alpha/2} \quad \text{lub} \quad t > t_{n-1,1-\alpha/2}$$

Obszar „akceptacji" $H_0$: $$-t_{n-1,1-\alpha/2} < t < t_{n-1,1-\alpha/2}$$

$$-t_{n-1,1-\alpha/2} < \frac{\overline{X}-\mu_0}{S/\sqrt{n}} < t_{n-1,1-\alpha/2}$$

$$\overline{X} - t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}} < \mu_0 < \overline{X} + t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

# Test a przedział ufności

Obszar „akceptacji” $H_0: \ \mu = \mu_0$:

$$\overline{X} - t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}} < \ \mu_0 \ < \overline{X} + t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

$100(1-\alpha)\%$ przedział ufności dla $\mu$:

$$\overline{X} - t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}} < \ \mu \ < \overline{X} + t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

Jeżeli $\mu_0$ należy do $100(1-\alpha)\%$ przedziału ufności, to na poziomie istotności $\alpha$ nie ma podstaw do odrzucenia hipotezy $H_0: \ \mu = \mu_0$.

## Testowanie hipotez o wariancji populacyjnej $\sigma^2$

# Test hipotezy o $\sigma^2$

Założenie: $X_i \sim N(\mu, \sigma)$, $\sigma$ - nieznane

Statystyka testowa: $$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \underset{H_0}{\sim} \chi^2_{n-1}$$

| $H_0$ | $H_1$ | Obszar krytyczny $R$ zależny od $\alpha$ |
|---|---|---|
| $\sigma^2 = \sigma_0^2$ | $\sigma^2 \neq \sigma_0^2$ | $(0, \chi^2_{n-1,\alpha/2}) \cup (\chi^2_{n-1,1-\alpha/2}, \infty)$ |
| $\sigma^2 \leqslant \sigma_0^2$ | $\sigma^2 > \sigma_0^2$ | $(\chi^2_{n-1,1-\alpha}, \infty)$ |
| $\sigma^2 \geqslant \sigma_0^2$ | $\sigma^2 < \sigma_0^2$ | $(0, \chi^2_{n-1,\alpha})$ |

$\chi^2_{n-1,\beta}$ – kwantyl rozkładu $\chi^2_{n-1}$

Testowanie hipotez o $\sigma^2$ w R:

`sigma.test(dane, sigmasq =` $\sigma_0^2$`, alternative="two.sided")`

UWAGA! Wymagany pakiet `TeachingDemos`

# Przykład 3

Inżynier przypuszcza, że zróżnicowanie (wariancja) czasu naprawy urządzeń jest krótsze niż 25 (godzin$^2$). Pobrano próbę 15 uszkodzonych urządzeń i zmierzono ich czas naprawy. Otrzymano następujące dane: 10, 10, 15, 12, 9, 8, 4, 10, 3, 4, 6, 5, 7, 8, 13, a następnie obliczono, że wariancja czasu naprawy to 12,35 godzin$^2$. Przyjmując poziom istotności 5% zweryfikuj przypuszczenie inżyniera.

$$n = 15, \quad s^2 = 12{,}35, \quad \alpha = 0{,}05$$

# Przykład 3 - rozwiązanie w R

czas = c(10,10,15,12,9,8,4,10,3,4,6,5,7,8,13)

One sample Chi-squared test for variance

data: czas
X-squared = 6.9173, df = 14, p-value = 0.06215
alternative hypothesis: true variance is less than 25
95 percent confidence interval:
 0.00000 26.31913
sample estimates:
var of czas
 12.35238

# Testowanie hipotez o proporcji populacyjnej $p$

# Test hipotezy o $p$

Próba duża ($n > 100$) $\qquad$ $T$ - liczba sukcesów, $\qquad \widehat{p} = \frac{T}{n}$

Statystyka testowa: $\qquad Z = \dfrac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}} \underset{H_0}{\sim} N(0,1)$

| $H_0$ | $H_1$ | Obszar krytyczny $R$ zależny od $\alpha$ |
|---|---|---|
| $p = p_0$ | $p \neq p_0$ | $(-\infty, -z_{1-\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$ |
| $p \leqslant p_0$ | $p > p_0$ | $(z_{1-\alpha}, \infty)$ |
| $p \geqslant p_0$ | $p < p_0$ | $(-\infty, -z_{1-\alpha})$ |

$z_\beta -$ kwantyl rozkładu $N(0,1)$

Testowanie hipotez o $p$ w R:

binom.test $(T$, $n$, p $= p_0$, alternative="two.sided")

prop.test $(T$, $n$, p $= p_0$, alternative="two.sided")

# Przykład 4

Statystyk przypuszcza, że proporcja inżynierów pozostających w swojej profesji po 10 latach od uzyskania tytułu inżyniera różni się od $50\%$. W związku z tym wybrał próbę $200$ inżynierów, którzy obronili pracę w roku 2009 i sprawdził, na jakim stanowisku pracują w roku 2019. Okazało się, że $111$ osób nadal pracowało jako inżynierowie. Przyjmując poziom istotności $5\%$ zweryfikuj przypuszczenie statystyka.

$$n = 200, \quad T = 111, \quad \alpha = 0{,}05$$

$$H_0 : \ p = 0{,}5 \qquad\qquad H_1 : \ p \neq 0{,}5$$

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}}$$

# Przykład 4 - rozwiązanie

# Przykład 4 - rozwiązanie w R

Exact binomial test

data: 111 and 200
number of successes = 111, number of trials = 200, p-value = 0.1374
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
0.4832463 0.6250989
sample estimates:
probability of success
0.555