4 i 5. TESTY STATYSTYCZNE

Definicja. Hipotezą statystyczną nazywamy przypuszczenie dotyczące nieznanego rozkładu badanej cechy populacji, o prawdziwości lub fałszywości którego wnioskuje się na podstawie pobranej próbki.

PRZYKŁADY:

1. *Wysuwamy hipotezę, że badana cecha ma rozkład normalny.*
2. *Wiemy, że badana cecha ma rozkład normalny o nieznanej wartości średniej μ i znanym odchyleniu standardowym $\sigma = 1$. Wysuwamy hipotezę, że $\mu = 5$.*
3. *Dane są dwa zbiory obserwacji, np. wysokości plonów uzyskane podczas naważenia nawozem A i nawozem B.*
   (a) *Wysuwamy hipotezę, że oba zbiory można traktować jako pochodzące z populacji o tym samym rozkładzie.*
   (b) *Z wcześniejszych badań wiemy, że zbiory te można traktować jako pochodzące z populacji o rozkładach normalnych odpowiednio $\mathcal{N}(\mu_1, \sigma_1)$ i $\mathcal{N}(\mu_2, \sigma_2)$, gdzie parametry μ_1, μ_2, σ_1 i σ_2 są nieznane, ale takie, że $\sigma_1 = \sigma_2$. Wysuwamy przypuszczenie, że średnia wartość plonów przy nawożeniu nawozem A jest większa niż średnia wartość plonów przy nawożeniu nawozem B (tzn. że $\mu_1 > \mu_2$).*

Definicja. Hipotezy, które dotyczą wyłącznie wartości parametru lub parametrów rozkładu badanej cechy nazywamy hipotezami parametrycznymi. Hipotezy, które nie są hipotezami parametrycznymi nazywamy hipotezami nieparametrycznymi.

W powyższym przykładzie hipotezy 2 i 3(b) są parametryczne, natomiast hipotezy 1 i 3(a) są nieparametryczne.

Definicja. Hipotezą prostą nazywamy hipotezę, która jednoznacznie określa rozkład badanej cechy. Hipotezą złożoną nazywamy hipotezę, która określa całą grupę rozkładów.

Hipoteza z przykładu 2 jest hipotezą prostą. Pozostałe hipotezy są złożone.

W praktyce rozważamy dwie hipotezy: hipotezę zerową (będziemy ją oznaczać H_0) i hipotezę alternatywną (tą będziemy oznaczać H_1). Jeśli odrzucamy hipotezę zerową, to przyjmujemy hipotezę alternatywną i na odwrót.

PRZYKŁAD:

Wiemy, że wysokość plonów przy nawożeniu starą metodą ma rozkład normalny o średniej 5 i odchyleniu standardowym 1: $\mathcal{N}(5, 1)$ i że wysokość plonów przy nawożeniu nową metodą ma rozkład normalny o tym samym odchyleniu standardowym $\sigma = 1$ ale o nieznanej średniej μ: $\mathcal{N}(\mu, 1)$. Chcemy sprawdzić czy nowa metoda zwiększyła średnią wysokość plonów.

W tym celu będziemy testować $H_0 : \mu = 5$ przeciwko $H_1 : \mu > 5$.

Przeprowadzamy eksperyment losowy. Wynik takiego eksperymentu to wektor losowy $\mathbb{X} = (X_1, X_2, \ldots, X_n)$ o wartościach w $\mathbb{R}^n$.

Definicja. Statystyką testową nazywamy funkcję $\delta(X_1, X_2, \ldots, X_n)$, która służy do weryfikacji H_0 przeciwko H_1. Zbiór wszystkich możliwych wartości funkcji δ dzielimy na dwa rozłączne zbiory W i W' takie, że:

- jeśli $\delta(x_1, x_2, \ldots, x_n) \in W$, to H_0 odrzucamy,
- jeśli $\delta(x_1, x_2, \ldots, x_n) \in W'$, to H_0 przyjmujemy.

W nazywamy zbiorem krytycznym testu (zbiorem odrzuceń H_0).

Musimy dobrze skonstruować statystykę testową i rozsądnie dobrać zbiór krytyczny, tak by podejmować decyzje zgodne z rzeczywistością. Jednak, ponieważ decyzje podejmujemy jedynie na podstawie próbki, nie mamy całkowitej informacji o badanej populacji i w związku z tym zawsze jesteśmy narażeni na popełnienie błędu - podjęcie decyzji niezgodnej z rzeczywistością. Dokładniej, możemy popełnić jeden z dwóch błędów:

1. odrzucić H_0 w sytuacji, gdy jest ona prawdziwa (tzw. błąd pierwszego rodzaju);
2. przyjąć H_0 w sytuacji, gdy jest ona fałszywa (tzw. błąd drugiego rodzaju).

Chcielibyśmy by prawdopodobieństwa obu tych błędów były jak najmniejsze. Niestety, gdy przy ustalonej statystyce testowej, zmieniamy W tak by malał błąd pierwszego rodzaju, to błąd drugiego rodzaju rośnie i na odwrót. Postępujemy zatem tak:

- ustalamy z góry maksymalną wartość prawdopodobieństwa błędu pierwszego rodzaju, oznaczamy tą wartość α i nazywamy ją poziomem istotności testu (zwyczajowo przyjmuje się $\alpha = 0,01$ lub $\alpha = 0,05$, czasami $\alpha = 0,1$);

- zbiór krytyczny W wyznaczamy tak by prawdopodobieństwo błędu pierwszego rodzaju nie przekraczało α i by prawdopodobieństwo błędu drugiego rodzaju było możliwie najmniejsze.

Zapisujemy to symbolicznie:

$$\underbrace{P(\delta(X_1, X_2, \ldots, X_n) \in W | H_0)}_{\text{prawdop. błędu pierwszego rodzaju}} \quad \leq \quad \alpha \qquad \text{i}$$

$$\underbrace{P(\delta(X_1, X_2, \ldots, X_n) \notin W | H_1)}_{\text{prawdop. błędu drugiego rodzaju}} \quad - \quad \text{możliwie najmniejsze}$$

Definicja. Moc testu parametrycznego (power of the test) to funkcja zmiennej θ (gdzie θ to badany parametr) dana wzorem

$$\beta(\theta) = P(\underbrace{\delta(X_1, X_2, \ldots, X_n)}_{\text{statystyka testowa}} \in W|\theta) = \text{prawdop. odrzucenia } H_0 \text{ w sytuacji, gdy nieznany parametr przyjmuje wartość } \theta.$$

Jeśli $H_0 : \theta = \theta_0$, to $\beta(\theta_0) = P(\delta(X_1, X_2, \ldots, X_n) \in W|\theta_0) = P(\delta(X_1, X_2, \ldots, X_n) \in W|H_0) =$ prawdopodobieństwo błędu pierwszego rodzaju.

Natomist jeśli $H_1 : \theta = \theta_1$, to wówczas $\beta(\theta_1) = P(\delta(X_1, X_2, \ldots, X_n) \in W|\theta_1) = P(\delta(X_1, X_2, \ldots, X_n) \in W|H_1) = 1 - P(\delta(X_1, X_2, \ldots, X_n) \notin W|H_1) = 1-$ prawdopodobieństwo błędu drugiego rodzaju.

Definicja. Najmniejszy poziom istotności, przy którym zaobserwowana wartość statystyki testowej prowadzi do odrzucenia H_0, nazywamy p-wartością (p-value) przeprowadzonego testu. Tzn.

$$\begin{aligned} p - \text{value} \leq \alpha \quad &\Rightarrow \quad \text{odrzucamy } H_0, \\ p - \text{value} > \alpha \quad &\Rightarrow \quad \text{nie ma podstaw do odrzucenia } H_0 \text{ (przyjmujemy } H_0). \end{aligned}$$

TESTY PARAMETRYCZNE

Podstawowe statystyki próbkowe i funkcje w R wyliczające te statystyki	
średnia z próby: $\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$, > mean(dane)	wariancja z próby: $s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \overline{x})^2$ > var(dane)

Oznaczenia podstawowych kwantyli i funkcje w R wyliczające te kwantyle			
dla rozkładu normalnego	dla rozkładu t-Studenta	dla rozkładu chi-kwadrat	dla rozkładu F-Snedecora
z_α	$t_{\alpha,n}$	$\chi^2_{\alpha,n}$	$F(\alpha, n, m)$
> qnorm(α)	> qt(α, n)	> qchisq(α, n)	qf(α, n, m)

<table>
<tr><th colspan="3">Weryfikacje hipotez dotyczących wartości średniej na poziomie istotności α</th></tr>
<tr><td colspan="3">UWAGA: jeżeli wyznaczone wartości statystyk (Z lub T) należą do odpowiednich zbiorów krytycznych, to H_0 odrzucamy.</td></tr>
<tr><td colspan="3">Model I. $X \sim N(\mu, \sigma), \mu$ - nieznane, σ - znane. Hipoteza zerowa $H_0 : \mu = \mu_0$. Statystyka testowa $Z = \frac{\overline{x}-\mu_0}{\sigma}\sqrt{n}$.</td></tr>
<tr><td>Hipoteza alternatywna
$H_1 : \mu \neq \mu_0$
Zbiór krytyczny
$W = \left(-\infty; -z_{1-\alpha/2}\right\rangle \cup \left\langle z_{1-\alpha/2}; +\infty\right)$</td><td>Hipoteza alternatywna
$H_1 : \mu > \mu_0$
Zbiór krytyczny
$W = \langle z_{1-\alpha}; +\infty)$</td><td>Hipoteza alternatywna
$H_1 : \mu < \mu_0$
Zbiór krytyczny
$W = (-\infty; -z_{1-\alpha}\rangle$</td></tr>
<tr><td colspan="3">Model II (t.test). $X \sim N(\mu, \sigma), \mu$ - nieznane, σ- nieznane. Hipoteza zerowa $H_0 : \mu = \mu_0$. Statystyka testowa $T = \frac{\overline{x}-\mu_0}{s}\sqrt{n}$.</td></tr>
<tr><td>Hipoteza alternatywna
$H_1 : \mu \neq \mu_0$
Zbiór krytyczny
$W = \left(-\infty; -t_{1-\alpha/2,n-1}\right\rangle \cup \left\langle t_{1-\alpha/2,n-1}; +\infty\right)$</td><td>Hipoteza alternatywna
$H_1 : \mu > \mu_0$
Zbiór krytyczny
$W = \langle t_{1-\alpha,n-1}; +\infty)$</td><td>Hipoteza alternatywna
$H_1 : \mu < \mu_0$
Zbiór krytyczny
$W = (-\infty; -t_{1-\alpha,n-1}\rangle$</td></tr>
<tr><td colspan="3">Model III. X ma rozkład dowolny (próba duża $n \geq 100$). Hipoteza zerowa $H_0 : \mu = \mu_0$. Statystyka testowa $Z = \frac{\overline{x}-\mu_0}{s}\sqrt{n}$.</td></tr>
<tr><td>Hipoteza alternatywna
$H_1 : \mu \neq \mu_0$
Zbiór krytyczny
$W = \left(-\infty; -z_{1-\alpha/2}\right\rangle \cup \left\langle z_{1-\alpha/2}; +\infty\right)$</td><td>Hipoteza alternatywna
$H_1 : \mu > \mu_0$
Zbiór krytyczny
$W = \langle z_{1-\alpha}; +\infty)$</td><td>Hipoteza alternatywna
$H_1 : \mu < \mu_0$
Zbiór krytyczny
$W = (-\infty; -z_{1-\alpha}\rangle$</td></tr>
<tr><td colspan="3">Model IV (prop.test). X ma rozkład dwupunktowy $P(X = 1) = p, P(X = 0) = q = 1 - p$, p - nieznane, $n\hat{p} \geq 5$ i $n\hat{q} \geq 5$, gdzie $\hat{p} = \frac{k}{n} = \frac{\text{ilość sukcesów}}{\text{ilość prób}}$, $\hat{q} = 1 - \hat{p}$.
Hipoteza zerowa $H_0 : p = p_0$. Statystyka testowa $Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0\,(1-p_0)}{n}}}$.</td></tr>
<tr><td>Hipoteza alternatywna
$H_1 : p \neq p_0$
Zbiór krytyczny
$W = \left(-\infty; -z_{1-\alpha/2}\right\rangle \cup \left\langle z_{1-\alpha/2}; +\infty\right)$</td><td>Hipoteza alternatywna
$H_1 : p > p_0$
Zbiór krytyczny
$W = \langle z_{1-\alpha}; +\infty)$</td><td>Hipoteza alternatywna
$H_1 : p < p_0$
Zbiór krytyczny
$W = (-\infty; -z_{1-\alpha}\rangle$</td></tr>
<tr><td colspan="3">Jeśli w modelu IV nie jest spełnione założenie, że $n\hat{p} \geq 5$ i $n\hat{q} \geq 5$, to zamiast prop.test używamy testu dokładnego binom.test.</td></tr>
</table>

<table>
<tr><th colspan="3">Weryfikacja hipotezy dotyczącej jednej wariancji na poziomie istotności α</th></tr>
<tr><td colspan="3">UWAGA: jeżeli wyznaczona wartość statystyki χ^2 należy do odpowiedniego zbioru krytycznego, to H_0 odrzucamy.</td></tr>
<tr><td colspan="3">Model. $X \sim N(\mu, \sigma), \mu$ - nieznane, σ - nieznane. Hipoteza zerowa $H_0 : \sigma^2 = \sigma_0^2$. Statystyka testowa $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$.</td></tr>
<tr><td>Hipoteza alternatywna $H_1 : \sigma^2 \neq \sigma_0^2$
Zbiór krytyczny
$W = \left(0, \chi^2_{\alpha/2;n-1}\right\rangle \cup \left\langle \chi^2_{1-\alpha/2;n-1}; +\infty\right)$</td><td>Hipoteza alternatywna $H_1 : \sigma^2 > \sigma_0^2$
Zbiór krytyczny
$W = \left\langle \chi^2_{1-\alpha;n-1}; +\infty\right)$</td><td>Hipoteza alternatywna $H_1 : \sigma^2 < \sigma_0^2$
Zbiór krytyczny
$W = \left(0; \chi^2_{\alpha;n-1}\right\rangle$</td></tr>
</table>

<table>
<tr><th colspan="2">Weryfikacja hipotezy dotyczącej równości dwóch wariancji na poziomie istotności α</th></tr>
<tr><td colspan="2">UWAGA: jeżeli wyznaczona wartość statystyki F należy do zbioru krytycznego W, to H_0 odrzucamy.</td></tr>
<tr><td colspan="2">Model I. (test F: var.test)
$X \sim N(\mu_1, \sigma_1)$, $Y \sim N(\mu_2, \sigma_2)$, $\mu_1, \mu_2, \sigma_1, \sigma_2$ - nieznane; dysponujemy niezależnymi próbami losowymi z tych populacji.
Hipoteza zerowa $H_0 : \sigma_1^2 = \sigma_2^2$. Statystyka testowa $F = s_1^2/s_2^2$ (w liczniku jest większa z wariancji).</td></tr>
<tr><td>Hipoteza alternatywna $H_1 : \sigma_1^2 \neq \sigma_2^2$
Zbiór krytyczny
$W = \langle F(1 - \alpha/2, n_1 - 1, n_2 - 2); +\infty)$</td><td>Hipoteza alternatywna $H_1 : \sigma_1^2 > \sigma_2^2$
Zbiór krytyczny
$W = \langle F(1 - \alpha, n_1 - 1, n_2 - 2); +\infty)$</td></tr>
</table>

Weryfikacje hipotez dotyczących dwóch średnich na poziomie istotności α

UWAGA: jeżeli wyznaczone wartości statystyk (Z lub T) należą do odpowiednich zbiorów krytycznych, to H_0 **odrzucamy**.

Model I. $X \sim N(\mu_1, \sigma_1)$, $Y \sim N(\mu_2, \sigma_2)$, μ_1, μ_2 - nieznane, σ_1, σ_2 - znane; dysponujemy niezależnymi próbami losowymi z tych populacji.
Hipoteza zerowa $H_0 : \mu_1 = \mu_2$. Statystyka testowa $Z = \frac{\overline{x}-\overline{y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$.

Hipoteza alternatywna	Hipoteza alternatywna	Hipoteza alternatywna
$H_1 : \mu_1 \neq \mu_2$	$H_1 : \mu_1 > \mu_2$	$H_1 : \mu_1 < \mu_2$
Zbiór krytyczny	Zbiór krytyczny	Zbiór krytyczny
$W = \left(-\infty; -z_{1-\alpha/2}\right\rangle \cup \left\langle z_{1-\alpha/2}; +\infty\right)$	$W = \langle z_{1-\alpha}; +\infty)$	$W = (-\infty; -z_{1-\alpha}\rangle$

Model II.(unpaired t-test: t.test(...,paired=FALSE,var.equal=TRUE))
$X \sim N(\mu_1, \sigma_1)$, $Y \sim N(\mu_2, \sigma_2)$, μ_1, μ_2 - nieznane, σ_1, σ_2 - nieznane, ale takie, że $\sigma_1 = \sigma_2$; dysponujemy niezależnymi próbami losowymi z tych populacji.
Hipoteza zerowa $H_0 : \mu_1 = \mu_2$. Statystyka testowa $T = \frac{\overline{x}-\overline{y}}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$.

Hipoteza alternatywna	Hipoteza alternatywna	Hipoteza alternatywna
$H_1 : \mu_1 \neq \mu_2$	$H_1 : \mu_1 > \mu_2$	$H_1 : \mu_1 < \mu_2$
Zbiór krytyczny	Zbiór krytyczny	Zbiór krytyczny
$W = \left(-\infty, -t_{1-\alpha/2,n_1+n_2-2}\right\rangle \cup \left\langle t_{1-\alpha/2,n_1+n_2-2}; +\infty\right)$	$W = \langle t_{1-\alpha,n_1+n_2-2}; +\infty)$	$W = (-\infty, -t_{1-\alpha,n_1+n_2-2}\rangle$

Jeśli w **modelu II** nie jest spełnione założenie, że $\sigma_1 = \sigma_2$, to zamiast t.test(...,paired=FALSE,var.equal=TRUE) należy użyć **t.test(...,paired=FALSE,var.equal=FALSE)**

Model III.(paired t-test: t.test(..., paired=TRUE))
$X - Y \sim N(\mu, \sigma)$, μ, σ - nieznane. Dysponujemy parami obserwacji, gdzie pary są wzajemnie niezależne.
Hipoteza zerowa $H_0 : \mu_1 = \mu_2$. Statystyka testowa $T = \frac{\overline{z}}{s_z}\sqrt{n}$, gdzie $z_i = x_i - y_i, i = 1, 2, \ldots, n$.

Hipoteza alternatywna	Hipoteza alternatywna	Hipoteza alternatywna
$H_1 : \mu_1 \neq \mu_2$	$H_1 : \mu_1 > \mu_2$	$H_1 : \mu_1 < \mu_2$
Zbiór krytyczny	Zbiór krytyczny	Zbiór krytyczny
$W = \left(-\infty; -t_{1-\alpha/2,n-1}\right\rangle \cup \left\langle t_{1-\alpha/2,n-1}; +\infty\right)$	$W = \langle t_{1-\alpha,n-1}; +\infty)$	$W = (-\infty; -t_{1-\alpha,n-1}\rangle$

Model IV. Cechy X, Y mają rozkłady dowolne ($n_1 \geq 100$, $n_2 \geq 100$), $\mu_1, \mu_2, \sigma_1, \sigma_2$ - nieznane; dysponujemy niezależnymi próbami losowymi z tych populacji.
Hipoteza zerowa $H_0 : \mu_1 = \mu_2$. Statystyka testowa $Z = \frac{\overline{x}-\overline{y}}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$.

Hipoteza alternatywna	Hipoteza alternatywna	Hipoteza alternatywna
$H_1 : \mu_1 \neq \mu_2$	$H_1 : \mu_1 > \mu_2$	$H_1 : \mu_1 < \mu_2$
Zbiór krytyczny	Zbiór krytyczny	Zbiór krytyczny
$W = \left(-\infty; -z_{1-\alpha/2}\right\rangle \cup \left\langle z_{1-\alpha/2}; +\infty\right)$	$W = \langle z_{1-\alpha}; +\infty)$	$W = (-\infty; -z_{1-\alpha}\rangle$

Model V (prop.test). Cechy X, Y mają rozkłady dwupunktowe, $P(X = 1) = p_1 = 1 - P(X = 0)$, $P(Y = 1) = p_2 = 1 - P(Y = 0)$, p_1, p_2 - nieznane, $n_1\hat{p}_1 \geq 5$ i $n_1(1 - \hat{p}_1) \geq 5$ i $n_2\hat{p}_2 \geq 5$ i $n_2(1 - \hat{p}_2) \geq 5$.
Hipoteza zerowa $H_0 : p_1 = p_2$. Statystyka testowa $Z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\frac{\overline{p}\overline{q}}{n}}}$, gdzie $\hat{p}_1 = \frac{k_1}{n_1}$, $\hat{p}_2 = \frac{k_2}{n_2}$, $\overline{p} = \frac{k_1+k_2}{n_1+n_2}$, $\overline{q} = 1- \overline{p}$, $n = \frac{n_1 n_2}{n_1+n_2}$.

Hipoteza alternatywna	Hipoteza alternatywna	Hipoteza alternatywna
$H_1 : p_1 \neq p_2$	$H_1 : p_1 > p_2$	$H_1 : p_1 < p_2$
Zbiór krytyczny	Zbiór krytyczny	Zbiór krytyczny
$W = \left(-\infty; -z_{1-\alpha/2}\right\rangle \cup \left\langle z_{1-\alpha/2}; +\infty\right)$	$W = \langle z_{1-\alpha}; +\infty)$	$W = (-\infty; -z_{1-\alpha}\rangle$

Jeśli w **modelu V** nie jest spełnione założenie, że n_1, n_2 są wystarczająco duże, to zamiast prop.test należy zastosować dokładny test Fishera **fisher.test** oparty na rozkładzie hipergeometrycznym.