Testowanie hipotez

Niech $(\mathcal{X}, \mathcal{B}, \mathcal{P}=\{P_\theta: \theta\in\Theta\})$ będzie przestrzenią statystyczną, przy czym

$$\Theta = \Theta_0 \cup \Theta_1 \text{ i } \Theta_0 \cap \Theta_1 = \varnothing.$$

Problem testowania hipotez można sformułować następująco: na podstawie obserwacji $X\in\mathcal{X}$ zweryfikować hipotezę

$$H_0: \theta\in\Theta_0 \qquad \text{wobec alternatywy} \qquad H_1: \theta\in\Theta_1.$$

Problem testowania hipotezy nie jest problemem badawczym lecz decyzyjnym. Możemy podjąć jedną z dwóch możliwych decyzji

- d_o - akceptacja hipotezy H_0
- d_1 - odrzucenia H_0 i akceptacji H_1

Możemy więc podjąć poprawna decyzję lub popełnić jeden z dwóch rodzajów błędów

- **błąd pierwszego rodzaju** $d_1|H_0$ polegający na odrzuceniu H_0 , gdy w rzeczywistości jest ona prawdziwa,
- **błąd drugiego rodzaju** $d_0|H_1$ polegający na akceptacji H_0 , gdy w rzeczywistości jest ona fałszywa.

	d_0	d_1
H_0	Decyzja poprawna	Błąd I rodzaju $d_1\|H_0$
H_1	Błąd II rodzaju $d_0\|H_1$	Decyzja poprawna

Potrzebujemy pewnej procedury, która podpowie jaką podjąć decyzję gdy dysponujemy wynikiem eksperymentu X. Rozwiązaniem problemu (H_0, H_1) testowania hipotezy H_0 przeciwko alternatywie H_1 będzie pewna funkcja $\varphi: \mathcal{X}\to[0,1]$ zwana **zrandomizowanym testem statystycznym**. Dysponując testem statystycznym φ i X:

- z prawdopodobieństwem $1-\varphi(X)$ podejmujemy decyzję d_o - akceptacji hipotezy H_0
- z prawdopodobieństwem $\varphi(X)$ podejmujemy decyzję d_1 - odrzucenia H_0 i akceptacji H_1.

d_0	d_1
$1-\varphi(X)$	$\varphi(X)$

Aby więc podjąć konkretną decyzję należy więc użyć pewnego mechanizmu losowego, który produkuje dwa wyniki z prawdopodobieństwami $\varphi(X)$ i $1-\varphi(X)$ i na tej podstawie podjąć (wylosować) decyzję. Podejście takie jest krytykowane przez praktyków, jako stwarzające możliwość nadużyć w interpretacji danych. Zdecydowanie większe uznanie praktyków znalazły testy niezrandomizowane, dla których zbiór wartości funkcji φ jest zbiorem dwuelementowym $\{0,1\}$. Test taki dzieli przestrzeń prób na dwa rozłączne zbiory:

$C=\varphi^{-1}(\{1\})=\{X\in\mathcal{X}: \varphi(X=1\}$ zwany zbiorem odrzucenia hipotezy H_0 (lub zbiorem krytycznym

i

$\varphi^{-1}(\{0\})=\{X\in\mathcal{X}: \varphi(X)=0\}$ zwany zbiorem akceptacji H_0.

Konstrukcja testu niezrandomizownego jest więc równoważna rozbiciu przestrzeni prób na dwa rozłączne podzbiory : odrzucenia i akceptacji H_0. Jeżeli zaobserwowana wartość zmiennej losowej X wpada do zbioru krytycznego $C=\varphi^{-1}(\{1\})$, to odrzucamy H_0 , w przeciwnym przypadku akceptujemy H_0.

Oczywiście nie każdy test statystyczny jest dobrym testem. Aby móc porównywać testy i w konsekwencji wybrać najlepszy można podobnie jak w problemie estymacji punktowej próbować określić stratę jaką ponosi statystyk podejmując błędną decyzję. Następnym krokiem jest wtedy wyznaczenie średniej straty dla danego testu, czyli wyznaczenie jego funkcji ryzyka. Testy, podobnie jak estymatory, można porównywać, porównując ich funkcje ryzyka. Można tu również wyróżnić :

- podejście globalne (porównywanie funkcji ryzyka) z zawężaniem klasy testów,
- podejście minimaksowe,
- podejście bayesowskie.

Największe uznanie zyskało jednak w teorii testów podejście **Neymana-Pearsona**. W ujęciu tym przypisuje się różną wagę błędom I i II rodzaju. Ważniejszy jest błąd pierwszego rodzaju. Dlatego nakłada się pewne ograniczenie na prawdopodobieństwo błędu I rodzaju i wśród dostępnych testów spełniających to ograniczenie poszukuje się testu, który minimalizuje prawdopodobieństwo błędu II rodzaju.

Def. Funkcją mocy testu φ nazywamy funkcję $\beta_\varphi(\theta)=E_\theta(\varphi(X))$.

Uwaga: dla testu niezrandomizowanego $\beta_\varphi(\theta)=P_\theta(C)$

Łatwo zauważyć, że dla testu niezrandomizowanego mamy:

$\theta\in\Theta_0$ (czyli prawdziwa jest H_0)$\Rightarrow$ $\beta_\varphi(\theta)=P_\theta(d_1\,|H_0)$ (pr. błędu I rodzaju)

Rzeczywiście $\beta_\varphi(\theta)=0\cdot P_\theta(d_0\,|H_0)+1\cdot P_\theta(d_1\,|H_0)=P_\theta(d_1\,|H_0)$

$\theta\in\Theta_1$ (czyli prawdziwa jest H_1)$\Rightarrow$ $\beta_\varphi(\theta)=P_\theta(d_1\,|H_1)=1-P_\theta(d_0\,|H_1)$

(pr. braku błędu II rodzaju- inaczej **moc testu** φ).

Def. **Rozmiarem testu** φ nazywamy liczbę $\sup_{\theta\in\Theta_0}\beta_\varphi(\theta)$ (lub $\sup_{\theta\in\Theta_0}P_\theta(C)$ dla testu niezrandomizowanego).

Def. Mówimy, że test φ hipotezy H_0 przeciwko H_1 jest **testem na poziomie istotności** α , jeżeli $\sup_{\theta\in\Theta_0}\beta_\varphi(\theta)\le\alpha$, czyli rozmiar testu nie przekracza α.

Wśród testów na poziomie istotności α poszukujemy najlepszego w sensie Neymana-Pearsona czyli najmocniejszego. Klasyczny lemat Neymana-Pearsona podaje sposób konstrukcji testu

najmocniejszego dla testowania **hipotezy prostej** H_0: $\theta=\theta_0$ (czyli $\Theta_0=\{\theta_0\}$) przeciwko **prostej alternatywie** H_1: $\theta=\theta_1$ (czyli $\Theta_1=\{\theta_1\}$). Hipoteza prosta specyfikuje więc dokładnie jeden rozkład prawdopodobieństwa określony na przestrzeni prób. Hipoteza H_0 głosi, że prawdziwym rozkładem na przestrzeni prób jest rozkład $P_0=P_{\theta_0}$ a konkurencyjna hipoteza H_1 głosi, że prawdziwym rozkładem na przestrzeni prób jest rozkład $P_1=P_{\theta_1}$. Hipotezę, która nie jest prosta (tzn. specyfikuje przynajmniej dwuelementową rodzinę rozkładów prawdopodobieństwa) nazywamy **hipotezą złożoną**.

Lemat Neymana-Pearsona. Jeżeli rozkłady P_0 i P_1 mają gęstości p_0 i p_1 (w przypadku ciągłym względem miary Lebesguea a w przypadku dyskretnym względem miary liczącej) , to test najmocniejszy na poziomie α hipotezy H_0 przeciwko H_1 istnieje i ma postać

$$\varphi(x)=\begin{cases}1 & \text{gdy} \quad p_1(x)>kp_0(x)\\ \gamma & \text{gdy} \quad p_1(x)=kp_0(x)\\ 0 & \text{gdy} \quad p_1(x)<kp_0(x)\end{cases},$$

gdzie $k\geq 0$ i $\gamma\in<0,1)$ są tak dobranymi stałymi, aby rozmiar testu był równy α

Z postaci funkcji testowej φ wynika, że optymalny test H_0 przeciwko H_1 jest testem zrandomizowanym. W przypadku testowania hipotez dotyczących ciągłych rozkładów prawdopodobieństwa można stałą γ przyjąć równą 0 i otrzymać w ten sposób test niezrandomizowany. W przypadku dyskretnym zwykle nie można znaleźć testu niezrandomizowanego o zadanym rozmiarze. Można jednak znaleźć optymalny test niezrandomizowany o rozmiarze zbliżonym do zadanego. Lemat Neymana-Pearsona podaje porządek w jakim należy włączać poszczególne punkty przestrzeni prób do zbioru krytycznego. Porządek ten wyznacza wartość ilorazu $\frac{p_1(x)}{p_0(x)}$. Im wyższa wartość tego ilorazu tym szybciej punkt x trafia do zbioru krytycznego. Proces ten należy przerwać wówczas, gdy dołączenie następnych punktów do zbioru krytycznego powoduje przekroczenie założonego poziomu istotności.

Zwykle w konstrukcji testu występuje etap pośredni polegający na wyznaczeniu tzw. statystyki testowej T: $\mathcal{X}\ni X\to T(X)$, która przenosi dalsze rozważania z przestrzeni $(\mathcal{X}, \mathcal{B}, \mathcal{P}=\{P_\theta: \theta\in\Theta\})$ do „prostszej" przestrzeni $(R,\mathcal{B}, \mathcal{P}^T=\{P_\theta^T: \theta\in\Theta\})$.

Przykład. Niech $(X_1,...,X_n)$ $(n=9)$ będzie ciągiem **iid** o rozkładzie P z dystrybuantą F. Znaleźć test najmocniejszy na poziomie α dla problemu testowania hipotezy H_0: $P=N(0,1)$ przeciwko alternatywie H_1: $P=N(1,1)$.

Zgodnie z lematem Neymana-Pearsona obszar krytyczny testu najmocniejszego ma postać

$$\{(X_1,...,X_n): \frac{1}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}\sum_{i=1}^{n}(X_i-1)^2} > k\frac{1}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}\sum_{i=1}^{n}X_i^2}\} = \{(X_1,...,X_n): T(X)=\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i>k\} \text{ przy}$$

czym $P_0(\overline{X} > k)=\alpha$. Statystyką testową jest więc w tym przypadku $T(X) = \overline{X}$. Wiadomo, że w przypadku prawdziwości H_0 statystyka $\overline{X}$ ma rozkład $N(0,\frac{1}{3})$ $(\sigma = \frac{1}{3})$. Stąd

$P_0(\overline{X} > k)=\alpha=0{,}05 \Leftrightarrow k = F^{-1}_{N(0,\frac{1}{3})}(1-\alpha) = F^{-1}_{N(0,\frac{1}{3})}(0{,}95) = 0{,}55$. W języku programu STATISTICA $k = \text{vNormal}(1-\alpha;0;1/3)$

W przypadku prawdziwości hipotezy H_1 statystyka $\overline{X}$ ma rozkład $N(1,\frac{1}{3})$. Wobec tego możemy dla zadanego poziomu istotności wyznaczyć moc testu :

$$\text{moc} = P_1(\overline{X} > k)=1-F_{N(1,\frac{1}{3})}(k)=0{,}91.$$

Powtarzając obliczenia dla dowolnego $\alpha\in(0,1)$ możemy wyznaczyć **krzywą mocy** znalezionego testu $H_0: m = 0$ przeciwko $H_1: m = 1$

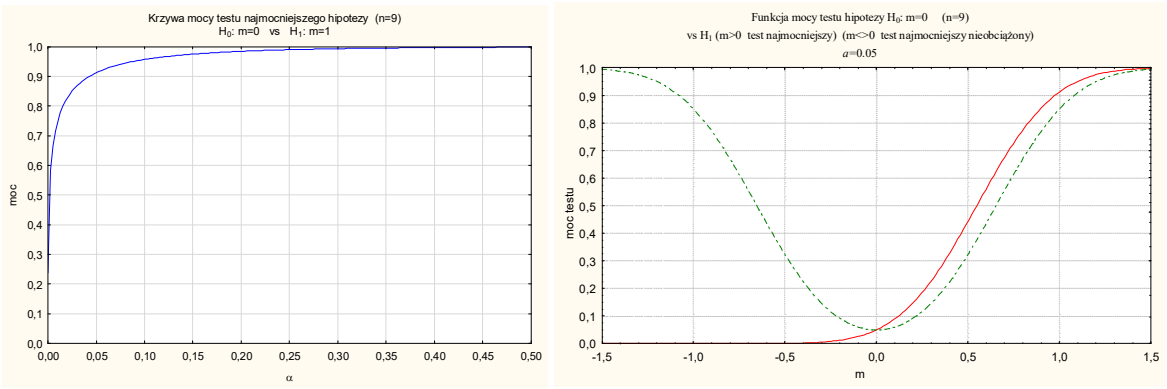
$$\text{moc}(\alpha) = 1-F_{N(0,\frac{1}{3})}(F^{-1}_{N(1,\frac{1}{3})}(1-\alpha)).$$

W języku programu STATISTICA $\text{moc}(\alpha) = 1-\text{INormal}(\text{vNormal}(1-\alpha;0;1/3);1;1/3))$.

Ustalając poziom istotności α=0,05 możemy wyznaczyć funkcję mocy testu moc(m) na poziomie α=0,05 dla różnych alternatyw m. Z konstrukcji testu widać że ma on dokładnie taką samą postać w przypadku testowania hipotezy prostej $H_0: m = 0$ przeciwko złożonej alternatywie $H_1: m > 0$. Funkcja mocy testu, to zależność mocy testu od parametru m specyfikowanego przez hipotezę alternatywną.

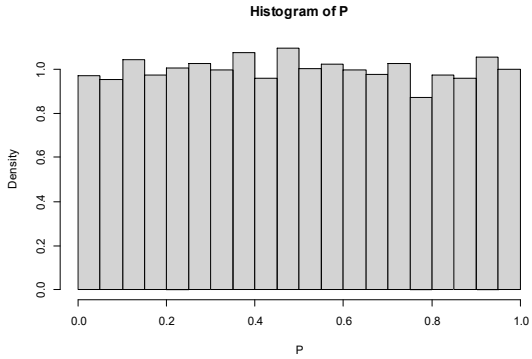
$$\text{moc}(m) = 1-F_{N(m,\frac{1}{3})}(F^{-1}_{N(0,\frac{1}{3})}(0{,}95)).$$

W języku programu STATISTICA moc(m) =1-INormal(VNormal(0,95;0;1/3);m;1/3)

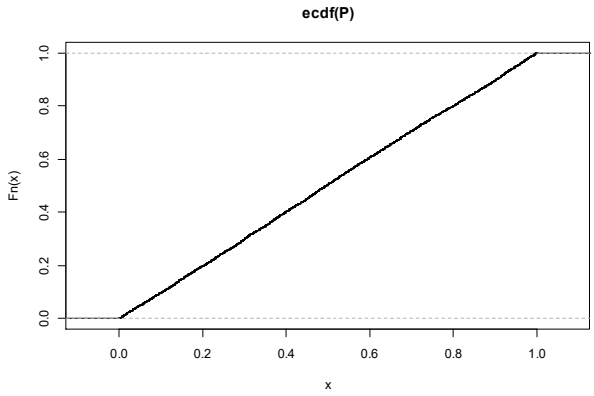


Krzywa mocy (a) i funkcja mocy (b)

```
# rozkład p-wartości dla testu N(0,1) vs N(1,1)
m <- 10000 #liczba generacji
n <- 9 #rozmiar próbki
P <- c() #dynamiczna deklaracja wektora
for (i in 1:m){
  X <- rnorm(n,0,1)
  Tx<-mean(X)
  P[i] <- 1-pnorm(Tx,0,1/sqrt(n))
} # koniec pętli for
hist(P, prob=TRUE) # Analizujemy wyniki:
```




```
#zamiast pętli for
P<-replicate(m, 1-pnorm(mean(rnorm(n,0,1)),0,1/sqrt(n)))
hist(P, prob=TRUE)
#dystrybuanta empiryczna rozkładu p-wartości
plot(ecdf(P))
```



Def. Test φ złożonej hipotezy H_0: $\theta \in \Theta_0$ przeciwko złożonej alternatywie H_1: $\theta \in \Theta_1$ nazywamy testem jednostajnie najmocniejszym na poziomie α, gdy dla każdego testu ψ H_0 przeciwko H_1 na poziomie α prawdziwy jest warunek: $\forall \theta \in \Theta_1 \ \beta_\varphi(\theta) \geq \beta_\psi(\theta)$.

W pewnych (nielicznych) przypadkach można z lematu Neymana Pearsona skonstruować test jednostajnie najmocniejszy złożonej hipotezy przeciwko złożonej alternatywie.

Testy nieobciążone

W sytuacji, gdy nie istnieje test jednostajnie najmocniejszy na poziomie α możemy próbować ograniczyć klasę testów, podobnie jak ograniczyliśmy klasę estymatorów do estymatorów nieobciążonych i być może w tej ograniczonej klasie znajdziemy test jednostajnie najmocniejszy.

Def. Test φ nazywamy testem nieobciążonym na poziomie istotności α, gdy

- $\sup_{\theta \in \Theta_0} \beta_\varphi(\theta) \leq \alpha$, czyli jest on testem na poziomie α
- $\forall\, \theta \in \Theta_1 \qquad \beta_\varphi(\theta) \geq \alpha$

czyli prawdopodobieństwo odrzucenia H_0 gdy jest ona **fałszywa** jest co najmniej takie jak prawdopodobieństwo jej odrzucenia, gdy jest ona **prawdziwa**. Inaczej moc testu ma być co najmniej tak duża jak rozmiar testu.