Rachunek prawdopodobieństwa i statystyka

termin	**oznaczenie**	**uwagi**
gęstość	$f(x)$	uogólnienie prawdopodobieństwa,
dystrybuanta	$F(x)$	$F(t) = \int_{-\infty}^{t} f(x)\,dx$, ppb zdarzenia $(-\infty, t)$,
kwantyl (rzędu α)	x_α	x_α takie, że $F(x_\alpha) = \alpha$,
kwantyl (rzędu α)	x_α	x_α takie, że $F(x_\alpha) = 1 - \alpha$.

pdf	*probability density function*	gęstość
cdf	*cumulative distribution function*	dystrybuanta
	quantile function	kwantyl

Poniżej — polecenia w języku R[1] oraz w środowisku Octave[2].

Język R

Rozkład	pdf	cdf	kwantyl	generator
normalny	dnorm	pnorm	qnorm	rnorm
Poisson	dpois	ppois	qpois	rpois
χ^2	dchisq	pchisq	qchisq	rchisq
itd...	d$*$	p$*$	q$*$	r$*$

Octave

Rozkład	pdf	cdf	kwantyl	generator
$N(0,1)$	stdnormal_pdf	stdnormal_cdf	stdnormal_inv	stdnormal_rnd
Poisson	poisspdf	poisscdf	poissinv	poissrnd
χ^2	chi2pdf	chi2cdf	chi2inv	chi2rnd
itd...	$*$pdf	$*$cdf	$*$inv	$*$rnd

1 Testowanie hipotez

Test średniej

Hipotezy : $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$, $H_1 : \mu > \mu_0$.

Założenia: wariancja σ^2 znana, obserwacje pochodzą z rozkładu $N(\mu, \sigma^2)$.

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

Test dwóch średnich

Hipotezy : $H_0 : \mu_1 - \mu_2 = \mu_0$, $H_1 : \mu_1 - \mu_2 \neq \mu_0$, $H_1 : \mu_1 - \mu_2 > \mu_0$.

[1] `https://www.r-project.org/`
[2] `https://www.gnu.org/software/octave/`

Założenia: wariancje σ_1^2, σ_2^2 znane, równe lub nie, obserwacje pochodzą z rozkładów $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$.

Statystyka testowa : $Z = \dfrac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$.

Test średniej bez wariancji

Hipotezy : $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$, $H_1 : \mu > \mu_0$.

Założenia: wariancja σ^2 nie jest znana, liczebność próbki n jest mała, obserwacje pochodzą z rozkładu $N(\mu, \sigma^2)$.

Statystyka testowa : $t = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n-1}} \sim t(n-1)$, gdzie $s^2 = \displaystyle\sum_{k=1}^{n} (x_k - \bar{x})^2$.

Test dwóch średnich bez wariancji

Hipotezy : $H_0 : \mu_1 - \mu_2 = \mu_0$, $H_1 : \mu_1 - \mu_2 \neq \mu_0$, $H_1 : \mu_1 - \mu_2 > \mu_0$.

Założenia: wariancje σ_1^2, σ_2^2 nie są znane, lecz równe, obserwacje pochodzą z rozkładów $N(\mu_1, \sigma^2)$, $N(\mu_2, \sigma^2)$.

Statystyka testowa : $t = \dfrac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{s} \sqrt{\dfrac{n_1 n_2}{n_1 + n_2}} \sim t(n_1 + n_2 - 2)$, gdzie $s^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$.

Test wariancji

Hipotezy : $H_0 : \sigma^2 = \sigma_0^2$, $H_1 : \sigma^2 \neq \sigma_0^2$, $H_1 : \sigma^2 > \sigma_0^2$.

Założenia: wariancja σ^2 nie jest znana, obserwacje pochodzą z rozkładu $N(\mu, \sigma^2)$.

Statystyka testowa : $\chi^2 = \dfrac{ns^2}{\sigma_0^2} \sim \chi^2(n-1)$.

Test dwóch wariancji

Hipotezy : $H_0 : \sigma_1^2 = \sigma_2^2$, $H_1 : \sigma_1^2 \neq \sigma_2^2$, $H_1 : \sigma_1^2 > \sigma_2^2$.

Założenia: wariancje σ_1^2, σ_2^2 nie są znane, obserwacje pochodzą z rozkładów $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$.

Statystyka testowa : $F = \dfrac{s_1^2}{s_2^2} \cdot \dfrac{n_2 - 1}{n_1 - 1} \sim F(n_1 - 1, n_2 - 1)$.

Wskaźnik częstości

Hipotezy : $H_0 : p = p_0$, $H_1 : p \neq p_0$, $H_1 : p > p_0$.

Założenia: n duże, tzn. $n > 30$, a także $np_0 \geqslant 5$.

Statystyka testowa : $Z = \dfrac{p - p_0}{\sqrt{p_0(1 - p_0)}} \sqrt{n} \sim N(0, 1)$.

Witold Karczewski