Pokus - experiment, jehož výsledky se i za stejných podmínek liší.

Hod dvěma mincemi (s/bez rozlišení).

Náhodná veličina - pokus s numerickými výsledky (realizacemi).

Měření intenzity provozu v ramenech křižovatky.

Jev - hodnota nebo interval hodnot, které je možno změřit.

Měřená rychlost automobilu je menší než 50 km/h.

Pravděpodobnost

1. relativní frekvence sledovaného jevu,
2. subjektivní míra důvěry v nastoupení sledovaného jevu, která se většinou opírá o znalosti dříve získané (apriorní informace).

Pr. šestky na kostce. Pr. čekání na autobus.

Soubor (populace, náhodný generátor) je dán množinou přípustných hodnot a pravidly, jak jsou generovány. Tato pravidla jsou vyjádřena pomocí pravděpodobností a jsou pevná (nemění se).

Kostka.

Výběr je množina hodnot naměřených na dané veličině.

$\{5, 1, 3, 1\}$

Náhodný výběr z veličiny X bere v úvahu potenciální opakovatelnost výběru. Jeho složky jsou nezávislé stejně rozdělené náhodné veličiny $[X_1, X_2, \cdots, X_n]$ s realizacemi $[x_1 x_2, \cdots, x_n]$ což jsou jednotlivé výběry.

$$\left\{\begin{array}{c} [5,\ 1,\ 3,\ 1]_1 \\ [2,\ 4,\ 6,\ 3]_2 \\ [5,\ 5,\ 1,\ 3]_3 \\ \cdots \end{array}\right\}$$
$$[X_1, X_2, X_3, X_4]$$

Generátor

diskrétní - pravděpodobnostní funkce

$$P(X = i) = f(x_i)$$

x_i	1	2	3
$f(x_i)$	0.35	0.12	0.53

spojitý - hustota pravděpodobnosti $f(x)$ pro kterou platí

$$P(X \in (a, b)) = \int_a^b f(x)\, dx$$

$f(x) = a \exp(-ax)\,,\ a > 0,\ x \geq 0.$

Popis souboru - distribuce $f(x)$ nebo momenty:

Spojitá náhodná veličina

- střední hodnota

$$E\left[X\right] = \int_{-\infty}^{\infty} x f\left(x\right) dx$$

- rozptyl

$$D\left[X\right] = \int_{-\infty}^{\infty} \left(x - E\left[X\right]\right)^2 f\left(x\right) dx$$

- kovariance

$$C\left[X,Y\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(x - E\left[X\right]\right)\left(y - E\left[Y\right]\right) f\left(x,y\right) dxdy$$

Diskrétní náhodná veličina

- střední hodnota

$$E\left[X\right] = \sum_i x_i f\left(x_i\right)$$

- rozptyl

$$D\left[X\right] = \sum_i \left(x_i - E\left[X\right]\right)^2 f\left(x_i\right)$$

- kovariance

$$C\left[X,Y\right] = \sum_{i,j} \left(x_i - E\left[X\right]\right)\left(y_j - E\left[Y\right]\right) f\left(x_i, y_j\right)$$

Popis výběru

- průměr

$$\bar{x} = \frac{1}{N} \sum_{t=1}^{N} x_t$$

- rozptyl

$$s_x^2 = \frac{1}{N} \sum_{t=1}^{N} \left(x_t - \bar{x}\right)^2$$

- kovariance

$$c_{x,y} = \frac{1}{N} \sum_{t=1}^{N} \left(x_t - \bar{x}\right)\left(y_t - \bar{y}\right)$$

Popis náhodného výběru

- výběrový průměr

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

je náhodná veličina s momenty

$$E\left[\bar{X}\right] = E\left[X\right]$$

$$D\left[\bar{X}\right] = \frac{D\left[X\right]}{N}$$

- výběrový rozptyl

$$D\left[\bar{X}\right]=\frac{1}{N-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$$

- výběrový korelační koeficient

$$r=\frac{c_{x,y}}{\sqrt{s_x^2 s_y^2}}\in\langle -1,1\rangle$$

a pro $r=0$ jsou x a y nekorelované.

Sdružená pravděpodobnost

$$f\left(x,y\right)=P\left(X=x\wedge Y=y\right)$$

$f\left(x,y\right):$

$x\backslash y$	1	2
1	0.2	0.1
2	0.1	0.2
3	0.3	0.1

nebo

z	1	2	3	4	5	6
x	1	1	2	2	3	3
y	1	2	1	2	1	2
	0.2	0.1	0.1	0.2	0.3	0.1

nebo

$$\int_a^b\int_c^d f\left(x,y\right)dxdy=P\left(X\in\langle a,b\rangle\wedge Y\in\langle c,d\rangle\right)$$

$$f\left(x,y\right)=\frac{1}{2\pi\sqrt{\det(r)}}\exp\left\{-\frac{1}{2}\left[\begin{array}{c}x-\mu_x\\ y-\mu_y\end{array}\right]' r^{-1}\left[\begin{array}{c}x-\mu_x\\ y-\mu_y\end{array}\right]\right\}$$

je to pravděpodobnost x a současně y

Marginální pravděpodobnost

$$f\left(x\right)=\int_{-\infty}^{\infty}f\left(x,y\right)dy,\quad f\left(x\right)=\sum_j f\left(x_i,y_j\right)$$

je to pravděpodobnost x a na y nám nezáleží.

Sdružená pr. je shrnuta k jedné ose.

Podmíněná pravděpodobnost

$$f\left(x|y\right)=\frac{f\left(x,y\right)}{f\left(y\right)}$$

je to pravděpodobnost x pro dané $y=$ číslo.

Příklad: máme naměřené veličiny x a y. $f\left(x|y=1\right)$ dostaneme tak, že vybereme jen záznamy, kde $y=1$ a z nich určíme pravděpodobnost x

Podmíněnou pr. spojité n.v. získáme jako
řez sdružené v bodě daném podmínkou.

Pro diskrétní n.v. je

x	1	2	2	1	2	1	1
y	2	1	1	2	2	1	2
		X	X			X	

$$f(x|y=1)=\begin{cases}1/3 & \text{pro } x=1\\ 2/3 & \text{pro } x=2\end{cases}$$

Nezávislost

$$f(x,y)=f(x)\,f(y)$$

místo součinu podmíněné a marginální je tu součin marginál.

Nezávislost: $f(x|y)=f(x)$. Odtud vzorec.

Řetězové pravidlo

$$f(x,y,z)=f(x|y,z)\,f(y|z)\,f(z)$$

$\cdots$ i pro více veličin. Je to zobecnění definice podmíněné pravděpodobnosti.

Bayesův vzorec

$$f(a|b,c)=\frac{f(b|a,c)\,f(a|c)}{f(b|a,c)}\propto f(b|a,c)\,f(a|c)$$

kde a je příčina, b je důsledek a c jsou další přítomné veličiny. $\propto$ je znaménko úměrnosti - rovnost až na normalizační konstantu.

Poznámka: Bayesův vzorec obrací kauzalitu: příčina vyvolá důsledek; ale i z důsledku se dá odhadovat, jaká příčina ho vyvolala.
Je to základ pro odhadování: Parametry modelu mají za následek generovaná data. Ale podle toho, jaká data jsou generována lze usoudit na parametry, které je generovaly.

U hlavního nádraží bývají nehody a potom
na Florenci je malý provoz. Nehoda $\rightarrow$ provoz.
Ale, jestliže jsme na Florenci a vidíme malý provoz,
můžeme očekávat, že u hlavního nádraží je nehoda.

Statistika pro odhad parametru θ je funkce náhodného výběru jejíž realizace ukazují na tento parametr.

Statistika pro odhad μ je výběrový průměr $\bar{X}$
(velký průměr odhadován malým průměrem)

Statistika je náhodná veličina jejíž realizace jsou průměry z jednotlivých výběrů. Jako náhodný veličina má své momenty: pro $\bar{X}$ je ... výše.

Pro výběr z normálního rozdělení s μ a σ^2 je

$$f(\bar{X})=N\left(\mu,\frac{\sigma^2}{n}\right)$$

Vlastnosti statistiky: nestrannost, konzistence, vydatnost (zatím necháme).

Interval spolehlivosti α-IS pro odhad parametru θ je interval, ve kterém parametr leží s pravděpodobností α (nebo: je interval, ve kterém leží $\alpha \cdot 100\%$ všech odhadů).

Test hypotézy

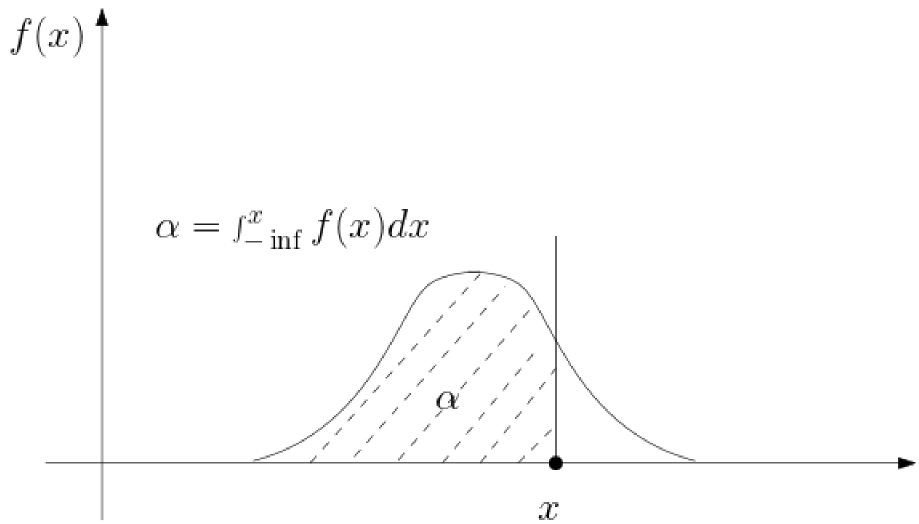
H0	nulová hypotéza		$\mu = \mu_0$	
HA	alternativní hypotéza	$\mu \neq \mu_0$,	$\mu > \mu_0$,	$\mu < \mu_0$
		obou,	pravo,	levo

Na hp odhadové statistiky určíme IS a jeho doplněk je **kritický obor** W.

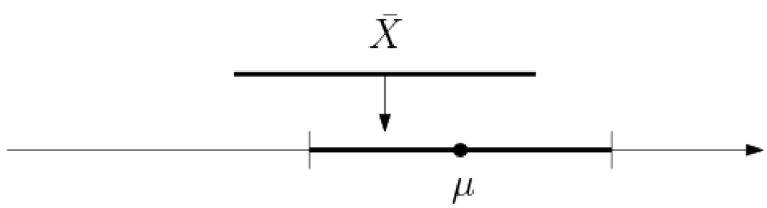
Když hodnota statistiky padne do W tak H0 zamítáme   nebo

Když je p-hodnota malá ($pv < \alpha$) tak taky H0 zamítáme (p-hodnota je dole na obrázku).

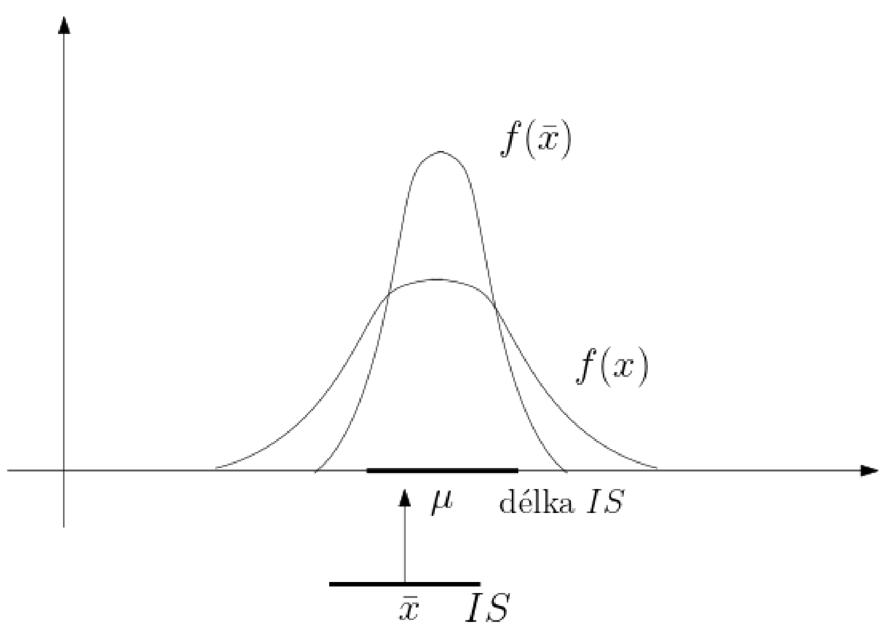
Hustota pravděpodobnosti



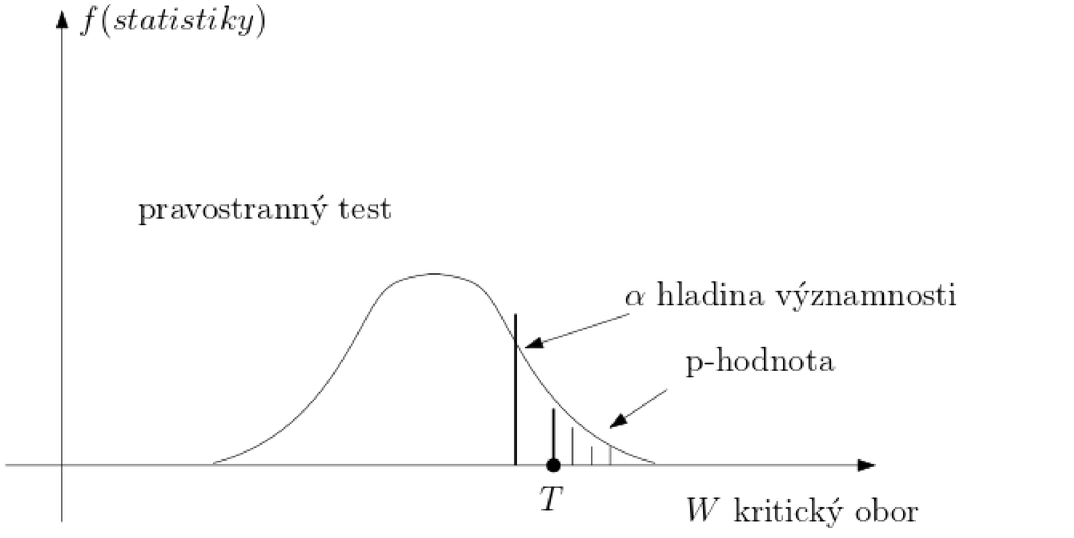
Princip IS



Konstrukce IS



p**-hodnota**


$f(statistiky)$
pravostranný test
α hladina významnosti
p-hodnota
T
W kritický obor