Interval spolehlivosti

Motivace

Interval spolehlivosti nebo také intervalový odhad můžeme vnímat jako jakési rozšíření bodového odhadu. Připomeňme, že bodový odhad $\widehat{\vartheta}$ neznámého parametru, označme ho ϑ, jsme získali na základně náhodného výběru $X_1, \cdots, X_n$, který pocházel z rozdělení $L(\vartheta)$. Takto jsme dostali konkrétní bodový odhad ovšem nemohli jsme nic říci o jeho přesnosti.
Naším cílem bude nyní najít intervalový odhad parametru ϑ, tedy najít interval, ve kterém se bude parametr ϑ vyskytovat s určitou pravděpodobností, tuto pravděpodobnost nazýváme **spolehlivost**. Doplněk spolehlivosti se nazývá **riziko** označme ho α, riziko nejčastěji volíme 0,05 méně často potom 0,1 nebo 0,001.

Nový pojem: Interval spolehlivosti

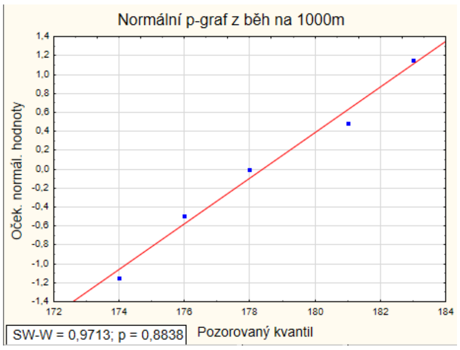
Mějme $X_1, \cdots, X_n$ náhodný výběr, který pochází z rozdělení $L(\vartheta)$, dále označme $h(\vartheta)$ parametrickou funkci, $\alpha \in (0, 1)$, $D = D(X_1, \cdots, X_n)$, $H = H(X_1, \cdots, X_n)$ jsou statistiky, potom:

- interval (D, H) nazýváme oboustranný interval spolehlivosti pro parametrickou funkci $h(\vartheta)$, jestliže pro každě ϑ z parametrického prostoru platí: $P(D < h(\vartheta) < H) \geq 1 - \alpha$.
- interval (D, ∞) nazýváme levostranný interval spolehlivosti pro parametrickou funkci $h(\vartheta)$, jestliže pro každě ϑ z parametrického prostoru platí: $P(D < h(\vartheta) < \infty) \geq 1 - \alpha$
- interval $(-\infty, H)$ nazýváme pravostranný interval spolehlivosti pro parametrickou funkci $h(\vartheta)$, jestliže pro každě ϑ z parametrického prostoru platí: $P(-\infty < h(\vartheta) < H) \geq 1 - \alpha$

Ilustrujme si to nyní na následujícím příkladu. Představme si, že sledujeme výkonost atleta, před tím než ho pošleme na závody bychom rádi věděli, jaký čas můžeme očekávat, že zaběhne. Podívali jsme se tedy na jeho výsledky z 5ti tréningů. Hodnoty si zaznamenáme do tabulky:

	1 běh na 1000m
1	176
2	183
3	178
4	174
5	181

Předpokládejme, že výkonost atleta se řídí normálním rozdělením.



Nejdříve najdeme bodový odhad střední hodnoty (naše parametrická funkce tedy bude $h(\vartheta) = \mu$), bodovým odhadem střední hodnoty je výběrový průměr $M = \frac{1}{n}\sum_{i=1}^{n} X_i$, vypočteme jeho realizaci a získáme m=178,4.
Nyní by nás tedy zajímalo, „jak přesný“ tento odhad je, vypočítáme tedy interval, ve kterém se skutečný parametr $h(\vartheta)$ nachází s pravděpodobností 0,95. K tomu potřebujeme znát bodový odhad parametrické funkce $h(\vartheta)$.Tento odhad již známe a dále najít vhodnou statistiku označme ji T, která vznikne transformací $\widehat{h(\vartheta)} = M$ a pomocí které budeme schopni interval spolehlivosti vyjádřit. Hledáním této statistiky se nebudeme v tomto textu zabývat (přesný postup najdete např. v knize Průvodce základními statistickými metodami). My si zde ukážeme, jak pro tento případ najít IS pomocí softwaru Statistika.

Na záložce „Statistiky“ / zvolíme „Základní statistiky/tabulky“ / „Popisné statistiky“ / klikneme na „Proměnné“ vybereme „běh na 1000m“ / záložka „Detailní výsledky“ zaklikneme „Průměr“, „Meze spolehl. prům.“ / klikneme na „Výpočet“

	Popisné statistiky		
Proměnná	**Průměr**	Int. spolehl. -95,000%	Int. spolehl. 95,000%
běh na 1000m	178,4000	173,8718	182,9282

Tedy s pravděpodobností 95% můžeme očekávat, že střední hodnota výkonu atleta bude v rozmezí 173,8718 do 182,9282. Něco podobného bychom dokázali říci již z pohledu do tabulky. Představme si nyní, že jsme atleta sledovali po dalších 5 tréningů a on svoje předchozí časy zopakoval.
Nyní vidíme, že průměrný zaběhnutý čas zůstal stejný, ale intervalový odhad se zpřesnil.

	1 běh na 1000m
1	176
2	183
3	178
4	174
5	181
6	176
7	183
8	178
9	174
10	181

	Popisné statistiky		
Proměnná	**Průměr**	Int. spolehl. -95,000%	Int. spolehl. 95,000%
běh na 1000m	178,4000	175,9404	180,8596