



StatSoft

Nebojte se p-hodnot!

Cíl článku je jednoduchý a víc než zřejmý. Chceme, aby všichni pochopili význam p-hodnoty.

P-hodnota je pojem, který spadá do oblasti testování hypotéz. Jistě jste se s touto hodnotou ve výstupech z testů již mnohokrát setkali (malou ukázkou některých výstupů ve STATISTICE vidíte vpravo). Ve světě počítačů je výstup zahrnující p-hodnotu již standardem a často bývá vyžadován i jako výstup v odborných publikacích. Zhodnocení velikosti p-hodnoty je jednou z možností, které říkají, zda je výsledek testu významný nebo nikoliv.

t-test

chí-kvadrát test

Shapiro-Wilkův test

Test nulovosti korelačního koeficientu

| Proměnná | Průměr medium | Průměr high | t | sv | p |
|-------------------------|---------------|-------------|----------|-----|----------|
| Systolic Blood Pressure | 138.5130 | 142.4052 | -1.70210 | 344 | 0.089641 |

| Statist. | Chí-kvadr. | sv | p |
|-------------------|------------|------|-----------|
| Pearsonův chí-kv. | 24.82981 | df=4 | p= .00005 |
| M-V chí-kvadr. | 25.81010 | df=4 | p= .00003 |

| Kategorie | Četnost | Kumulativní četnost | Rel.četn. (platných) | Kumul. % (platných) | Rel.četn. všech | Kumul. % všech |
|----------------------|---------|---------------------|----------------------|---------------------|-----------------|----------------|
| 80.0000<x<=100.0000 | 0 | 0 | 0.00000 | 0.0000 | 0.00000 | 0.0000 |
| 100.0000<x<=120.0000 | 81 | 81 | 17.53247 | 17.5325 | 17.49460 | 17.4946 |
| 120.0000<x<=140.0000 | 221 | 302 | 47.83550 | 65.3680 | 47.73218 | 65.2268 |
| 140.0000<x<=160.0000 | 96 | 398 | 20.77922 | 86.1472 | 20.73434 | 85.9611 |
| 160.0000<x<=180.0000 | 48 | 446 | 10.38961 | 96.5368 | 10.36717 | 96.3283 |
| 180.0000<x<=200.0000 | 8 | 454 | 1.73160 | 98.2684 | 1.72786 | 98.0562 |
| 200.0000<x<=220.0000 | 8 | 462 | 1.73160 | 100.0000 | 1.72786 | 99.7840 |
| ChD | 1 | 463 | 0.21645 | | 0.21598 | 100.0000 |

| Proměnná | Tobacco intake (kg) | LDL Cholesterol | Adiposity | Stress Type-A behavior |
|------------------------|---------------------|-----------------|-----------|------------------------|
| Tobacco intake (kg) | | 0.001 | 0.000 | 0.754 |
| LDL Cholesterol | 0.001 | | 0.000 | 0.345 |
| Adiposity | 0.000 | 0.000 | | 0.355 |
| Stress Type-A behavior | 0.754 | 0.345 | 0.355 | |

Každý statistický test je založen na nějaké testové statistice, tedy číslu spočtenému z dat, podle kterého se rozhodujeme, jestli test vyšel významně nebo nikoli (o principu testování hypotéz si přečtete [zde](#), příklady testování naleznete [zde](#) nebo [zde](#)). U testu máme zvolenu hladinu alfa (chybu prvního druhu), kterou chceme dodržet, podle této hodnoty se typicky vytvoří interval spolehlivosti (pokud je alfa 0,05, pak se jedná o 95 procentní interval, o intervalech si můžete přečíst [zde](#) a [zde](#)). A my typicky porovnáme, jestli spočtená hodnota statistiky leží nebo neleží v tomto intervalu. A teď, co s tím má společného ta p-hodnota. Uveďme si nejdříve definici:

P-hodnota je nejmenší hladina, na které zamítáme.

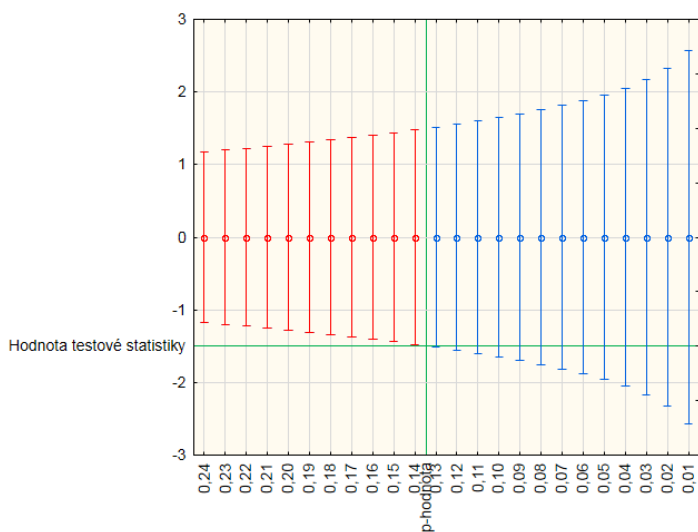
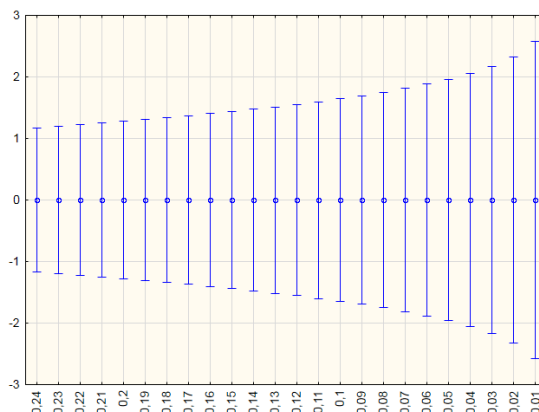
Lidsky řečeno se dá definice p-hodnoty převyprávět následovně:

Pravděpodobnost výsledků, které ještě více svědčí proti H_0 .

Dosažená hladina testu.

P-hodnota je největší hladina, na které nezamítáme.

Když si to ukážeme na příkladu s intervalem spolehlivosti. Pravděpodobnost, že hodnota padne do intervalu je $1-\alpha$. Čím menší α máme, tím větší je interval (α se zmenšuje, pravděpodobnost se zvyšuje, tedy i interval musí být širší, abychom měli větší jistotu, že tam hodnota padne) – viz obrázek vpravo. P-hodnota je pak hodnota α , pro které je interval přesně tak dlouhý, aby jeho okraj byl na naměřené hodnotě testové statistiky. Pokud se podíváme na intervaly pro různé hladiny na obrázku dole, vidíme, že pokud nám tedy vyjde testová statistika rovna například $-1,5$, pak p-hodnota je $0,1336$. Intervaly s hladinou větší než $0,1336$ hodnotu $-1,5$ neobsahují, zatímco hladiny menší ano.

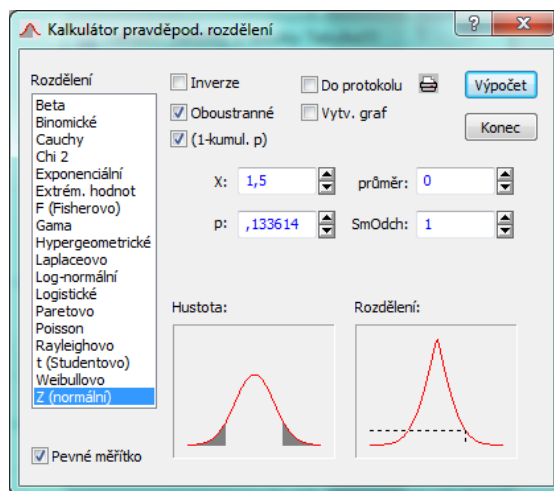


Pokud budeme mít hladinu testu α rovno $0,05$, pak nezamítáme, pokud bychom si ji ale z nějakého zvláštního důvodu potřebovali nastavit na $0,15$, pak bychom již zamítli nulovou hypotézu.

Pro nejklaštější testování tedy stačí p-hodnotu porovnat s hodnotou $0,05$ (nejběžnější hladina testu) a pokud je p-hodnota menší, pak je jasné (viz obrázek), že interval testovou statistiku nepokrývá (jelikož interval pro $0,05$ je užší než pro p-hodnotu) a tedy zamítáme nulovou hypotézu. Pokud by p-hodnota byla větší než $0,05$, pak nulovou hypotézu nezamítáme. Tuto poučku jste jistě již slyšeli:

| | | |
|--------------|-----|----------------------------------|
| $\alpha > p$ | ... | zamítáme H_0 ve prospěch H_A |
| $\alpha < p$ | ... | nezamítáme H_0 |

Poznámka: S příchodem počítačů se výpočet p-hodnot značně zjednodušil. Bude zde popsán postup, jak vypočítat p-hodnotu z příkladu (obrázek intervalů): Pro úlohu jsme použili jednoduše normální rozdělení se střední hodnotou 0 a rozptylem 1 (jinak řečeno testová statistika by měla předpokládané rozdělení normální s parametry 0 a 1). Intervaly pokrývají hodnotu z daného rozdělení s pravděpodobností $1-\alpha$. Pokud bychom chtěli spočítat p-hodnotu pro nějakou hodnotu, pak můžeme použít například pravděpodobnostní kalkulátor. Nastavení pro danou úlohu vidíte vpravo.



Čím blíže středu intervalu je statistika, tím větší je p-hodnota, krajním případem je $p=1$, kdy je testová statistika přesně ve středu intervalu pro teoretické rozdělení. Čím dále je statistika od středu intervalu, tím menší je p-hodnota.

Výhoda p-hodnoty

Tedy pokud nám stačí se pouze rozhodnout, zda vyšel test statisticky významně, pak p-hodnota nám říká vše potřebné a to navíc nezávisle na tom, jakou si zvolíme hladinu alfa, dává nám ihned informaci zároveň pro všechny hladiny. Pokud je p-hodnota menší než zvolená hladina testu (α), tak zamítáme, pokud ne, tak nezamítáme.

Poznamenejme, že zásadní je vědět, co daný test má potvrdit či vyvrátit, jinak řečeno, jak zní nulová hypotéza. Pokud například děláme testy normality, pak nulová hypotéza zní, že rozdělení je normální, pokud tedy zamítáme hypotézu, pak data normální nejsou. Podle p-hodnoty klasicky, jako u každého jiného testu, se rozhodneme o zamítnutí podle pravidla výše.

P-hodnota v softwaru STATISTICA

Prakticky pro jakýkoli test, ať už je to testování středních hodnot, normality nebo přítomnosti odlehých pozorování, všechny testy ve výstupu budou obsahovat informaci o p-hodnotě. Buď přímo ve výsledkové tabulce nebo v záhlaví tabulky, kde se vypisují hlavní výsledky analýzy.

| Statist. : Blood Pressure Level(3) x Tobacco | |
|--|------------------------|
| Statist. | Chi-kvadr. sv p |
| Pearsonův chí-kv. | 24,82981 df=4 p=,00005 |
| M-V chí-kvadr. | 25,81610 df=4 p=,00003 |

| Tabulka četností Systolic Blood Pressure (HeartDisease) Shapiro-Wilk W= 97,91 p=,00000 | | | | | | |
|---|---------|---------------------|-----------------------|---------------------|------------------|----------------|
| Kategorie | Četnost | Kumulativní četnost | Rel. četn. (platných) | Kumul. % (platných) | Rel. četn. všech | Kumul. % všech |
| 80,00000<<=100,0000 | 0 | 0 | 0,00000 | 0,0000 | 0,00000 | 0,0000 |
| 100,00000<<=120,0000 | 81 | 81 | 17,53247 | 17,5325 | 17,49460 | 17,4946 |
| 120,00000<<=140,0000 | 221 | 302 | 47,83550 | 65,3680 | 47,73218 | 65,2268 |
| 140,00000<<=160,0000 | 96 | 398 | 20,77922 | 86,1472 | 20,73434 | 85,9611 |
| 160,00000<<=180,0000 | 48 | 446 | 10,38961 | 96,5368 | 10,36717 | 96,3283 |
| 180,00000<<=200,0000 | 8 | 454 | 1,73160 | 98,2684 | 1,72786 | 98,0562 |
| 200,00000<<=220,0000 | 8 | 462 | 1,73160 | 100,0000 | 1,72786 | 99,7840 |
| ChD | 1 | 463 | 0,21645 | | 0,21598 | 100,0000 |

Závěrem

P-hodnota je jedním z nejdůležitějších pojmů teorie odhadu, je potřeba vědět, co znamená a jak se podle její velikosti zařadit při vyhodnocování testů. Pochopit p-hodnotu kompletně a ne jen slepě využívat poučky, to byl cíl našeho miničlánečku. Snad se to povedlo a už se p-hodnoty nebudete nikdy bát.