CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Electrical Engineering
Department of Cybernetics

Testování hypotéz

Petr Pošík

Části dokumentu jsou převzaty (i doslovně)
z *Mirko Navara: Pravděpodobnost a matematická statistika*,
`https://cw.felk.cvut.cz/lib/exe/fetch.php/courses/a6m33ssl/pms_print.pdf`
s laskavým svolením autora.

Testování hypotéz a jeho metodika

Testování hypotéz

Mnoho statistiků trpí nejrůznějšími psychickými poruchami, protože v mládí bylo mnoho jejich hypotéz zamítnuto.

:-)

Motivační příklad: Jasnovidec?

Experiment, kterým chceme ověřit, zda osoba má (větší či menší) jasnovidecké schopnosti:

- Máme balíček karet o 4 barvách (káry, piky, ...) se stejným počtem karet každé barvy.
- Vytáhneme z balíčku 1 kartu a položíme ji lícem dolů. Testovaná osoba řekne svůj tip, jakou barvu karta má. Zkontrolujeme, zda byl odhad správný. Kartu vrátíme do balíčku, balíček zamícháme.
- Opakujeme n-krát, např. 25krát.
- Výsledek experimentu: testovaná osoba se "trefila" v X případech z n.

Jaké rozhodovací pravidlo byste použili?

Motivační příklad: Jasnovidec?

Experiment, kterým chceme ověřit, zda osoba má (větší či menší) jasnovidecké schopnosti:

- Máme balíček karet o 4 barvách (káry, piky, ...) se stejným počtem karet každé barvy.
- Vytáhneme z balíčku 1 kartu a položíme ji lícem dolů. Testovaná osoba řekne svůj tip, jakou barvu karta má. Zkontrolujeme, zda byl odhad správný. Kartu vrátíme do balíčku, balíček zamícháme.
- Opakujeme n-krát, např. 25krát.
- Výsledek experimentu: testovaná osoba se "trefila" v X případech z n.

Jaké rozhodovací pravidlo byste použili?

- Řekli byste, že osoba je jasnovidec, když se trefila v 25 případech z 25?

Motivační příklad: Jasnovidec?

Experiment, kterým chceme ověřit, zda osoba má (větší či menší) jasnovidecké schopnosti:

- Máme balíček karet o 4 barvách (káry, piky, ...) se stejným počtem karet každé barvy.
- Vytáhneme z balíčku 1 kartu a položíme ji lícem dolů. Testovaná osoba řekne svůj tip, jakou barvu karta má. Zkontrolujeme, zda byl odhad správný. Kartu vrátíme do balíčku, balíček zamícháme.
- Opakujeme n-krát, např. 25krát.
- Výsledek experimentu: testovaná osoba se "trefila" v X případech z n.

Jaké rozhodovací pravidlo byste použili?

- Řekli byste, že osoba je jasnovidec, když se trefila v 25 případech z 25?
- Řekli byste, že osoba je jasnovidec, když se trefila v 6 nebo 7 případech z 25?

Motivační příklad: Jasnovidec?

Experiment, kterým chceme ověřit, zda osoba má (větší či menší) jasnovidecké schopnosti:

- Máme balíček karet o 4 barvách (káry, piky, ...) se stejným počtem karet každé barvy.
- Vytáhneme z balíčku 1 kartu a položíme ji lícem dolů. Testovaná osoba řekne svůj tip, jakou barvu karta má. Zkontrolujeme, zda byl odhad správný. Kartu vrátíme do balíčku, balíček zamícháme.
- Opakujeme n-krát, např. 25krát.
- Výsledek experimentu: testovaná osoba se "trefila" v X případech z n.

Jaké rozhodovací pravidlo byste použili?

- Řekli byste, že osoba je jasnovidec, když se trefila v 25 případech z 25?
- Řekli byste, že osoba je jasnovidec, když se trefila v 6 nebo 7 případech z 25?
- Co když se trefila v 10, 15, 20 případech?

Motivační příklad: Jasnovidec?

Experiment, kterým chceme ověřit, zda osoba má (větší či menší) jasnovidecké schopnosti:

- Máme balíček karet o 4 barvách (káry, piky, ...) se stejným počtem karet každé barvy.
- Vytáhneme z balíčku 1 kartu a položíme ji lícem dolů. Testovaná osoba řekne svůj tip, jakou barvu karta má. Zkontrolujeme, zda byl odhad správný. Kartu vrátíme do balíčku, balíček zamícháme.
- Opakujeme n-krát, např. 25krát.
- Výsledek experimentu: testovaná osoba se "trefila" v X případech z n.

Jaké rozhodovací pravidlo byste použili?

- Řekli byste, že osoba je jasnovidec, když se trefila v 25 případech z 25?
- Řekli byste, že osoba je jasnovidec, když se trefila v 6 nebo 7 případech z 25?
- Co když se trefila v 10, 15, 20 případech?
- Jak postupovat systematicky?

Motivační příklad: Jasnovidec?

Experiment, kterým chceme ověřit, zda osoba má (větší či menší) jasnovidecké schopnosti:

- Máme balíček karet o 4 barvách (káry, piky, ...) se stejným počtem karet každé barvy.
- Vytáhneme z balíčku 1 kartu a položíme ji lícem dolů. Testovaná osoba řekne svůj tip, jakou barvu karta má. Zkontrolujeme, zda byl odhad správný. Kartu vrátíme do balíčku, balíček zamícháme.
- Opakujeme n-krát, např. 25krát.
- Výsledek experimentu: testovaná osoba se "trefila" v X případech z n.

Jaké rozhodovací pravidlo byste použili?

- Řekli byste, že osoba je jasnovidec, když se trefila v 25 případech z 25?
- Řekli byste, že osoba je jasnovidec, když se trefila v 6 nebo 7 případech z 25?
- Co když se trefila v 10, 15, 20 případech?
- Jak postupovat systematicky?

Osoba může být v jednom ze dvou skutečných "stavů":

1. **obyčejný člověk**, pak pravděpodobnost, že kartu určí správně, je $p = p_0 = \frac{1}{4}$, nebo
2. **jasnovidec**, pak $p > \frac{1}{4}$.

Motivační příklad: Jasnovidec?

Experiment, kterým chceme ověřit, zda osoba má (větší či menší) jasnovidecké schopnosti:

- Máme balíček karet o 4 barvách (káry, piky, ...) se stejným počtem karet každé barvy.
- Vytáhneme z balíčku 1 kartu a položíme ji lícem dolů. Testovaná osoba řekne svůj tip, jakou barvu karta má. Zkontrolujeme, zda byl odhad správný. Kartu vrátíme do balíčku, balíček zamícháme.
- Opakujeme n-krát, např. 25krát.
- Výsledek experimentu: testovaná osoba se "trefila" v X případech z n.

Jaké rozhodovací pravidlo byste použili?

- Řekli byste, že osoba je jasnovidec, když se trefila v 25 případech z 25?
- Řekli byste, že osoba je jasnovidec, když se trefila v 6 nebo 7 případech z 25?
- Co když se trefila v 10, 15, 20 případech?
- Jak postupovat systematicky?

Osoba může být v jednom ze dvou skutečných "stavů":

1. **obyčejný člověk**, pak pravděpodobnost, že kartu určí správně, je $p = p_0 = \frac{1}{4}$, nebo
2. **jasnovidec**, pak $p > \frac{1}{4}$.

Výsledkem vašeho testu mohou být 2 rozhodnutí:

1. prohlásíte osobu za "obyčejného" člověka, nebo
2. prohlásíte osobu za jasnovidce (zamítnete "obyčejnost").

Motivační příklad: Jasnovidec? (Pokr.)

Použijme test: *Osobu označíme za jasnovidce (zamítneme "obyčejnost"), trefí-li se v 25 případech z 25.*

- Můžeme s tímto pravidlem označit obyčejného člověka za jasnovidce?

Motivační příklad: Jasnovidec? (Pokr.)

Použijme test: *Osobu označíme za jasnovidce (zamítneme "obyčejnost"), trefí-li se v 25 případech z 25.*

- Můžeme s tímto pravidlem označit obyčejného člověka za jasnovidce? Ano.
- Jaká je pravděpodobnost, že se to stane?

Motivační příklad: Jasnovidec? (Pokr.)

Použijme test: *Osobu označíme za jasnovidce (zamítneme "obyčejnost"), trefí-li se v 25 případech z 25.*

- Můžeme s tímto pravidlem označit obyčejného člověka za jasnovidce? Ano.
- Jaká je pravděpodobnost, že se to stane?

$$P[\text{zamítáme "obyčejnost"}|\text{ve skutečnosti "obyčejný"}] = P[X = 25|p = p_0] = \left(\frac{1}{4}\right)^{25} \doteq 10^{-15}$$

Motivační příklad: Jasnovidec? (Pokr.)

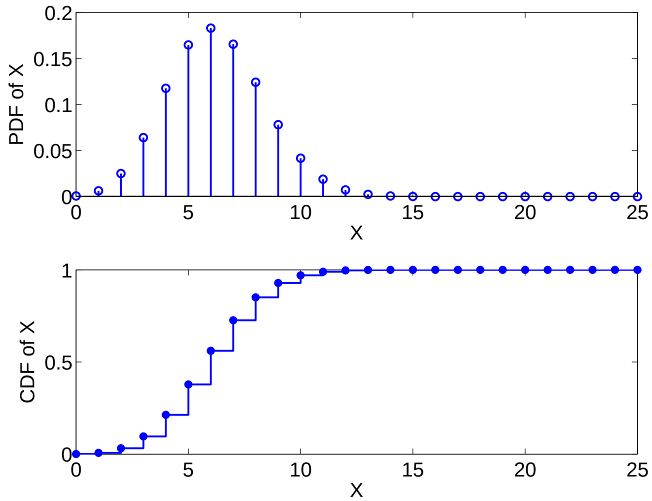
Použijme test: *Osobu označíme za jasnovidce (zamítneme "obyčejnost"), trefí-li se v 25 případech z 25.*

- Můžeme s tímto pravidlem označit obyčejného člověka za jasnovidce? Ano.
- Jaká je pravděpodobnost, že se to stane?

$$P[\text{zamítáme "obyčejnost"}|\text{ve skutečnosti "obyčejný"}] = P[X = 25|p = p_0] = \left(\frac{1}{4}\right)^{25} \doteq 10^{-15}$$

Pro obecný test: *Osobu označíme za jasnovidce (zamítneme "obyčejnost"), trefí-li se alespoň v c případech z n.*

- Jaké rozdělení má X pro "obyčejného" člověka?



Motivační příklad: Jasnovidec? (Pokr.)

Použijme test: *Osobu označíme za jasnovidce (zamítneme "obyčejnost"), trefí-li se v 25 případech z 25.*

- Můžeme s tímto pravidlem označit obyčejného člověka za jasnovidce?   Ano.
- Jaká je pravděpodobnost, že se to stane?

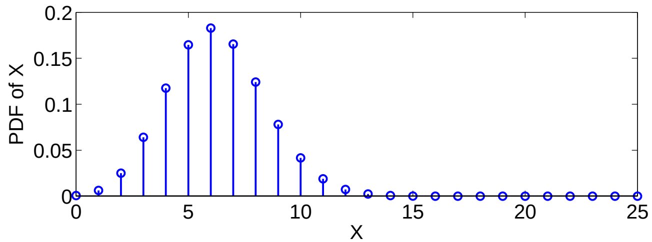
$$P[\text{zamítáme "obyčejnost"}|\text{ve skutečnosti "obyčejný"}] = P[X = 25|p = p_0] = \left(\frac{1}{4}\right)^{25} \doteq 10^{-15}$$

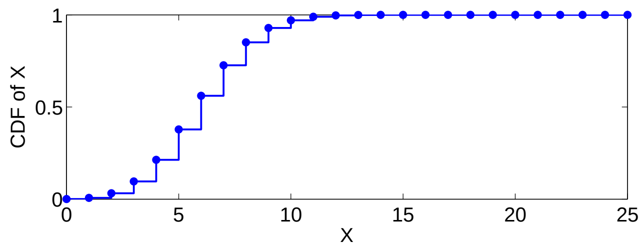
Pro obecný test: *Osobu označíme za jasnovidce (zamítneme "obyčejnost"), trefí-li se alespoň v c případech z n.*

- Jaké rozdělení má X pro "obyčejného" člověka?
  $X \sim \mathrm{Bi}(n, p_0)$
- Pravděpodobnost, že obyčejného člověka označíme za jasnovidce

$$P[\text{zam. "obyč"}|\text{skut. "obyč"}] = P[X \geq c|p = p_0] =$$

$$= \sum_{i=c}^{n} p_{\mathrm{Bi}(n,p_0)}(i) = 1 - F_{\mathrm{Bi}(n,p_0)}(c-1)$$





Motivační příklad: Jasnovidec? (Pokr.)

Použijme test: *Osobu označíme za jasnovidce (zamítneme "obyčejnost"), trefí-li se v 25 případech z 25.*

- Můžeme s tímto pravidlem označit obyčejného člověka za jasnovidce? Ano.
- Jaká je pravděpodobnost, že se to stane?

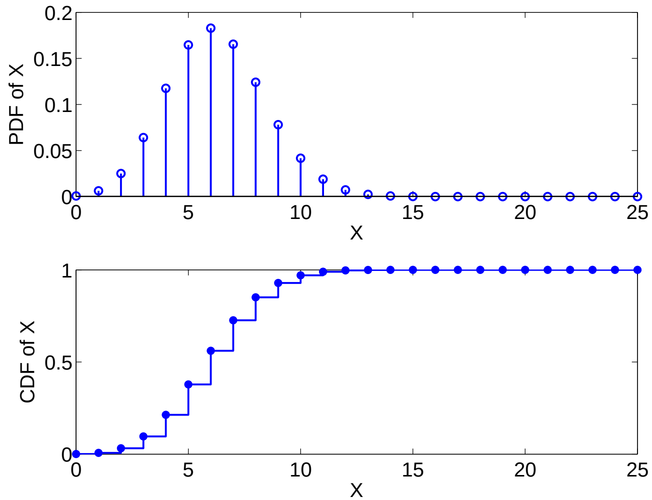
$$P[\text{zamítáme "obyčejnost"}|\text{ve skutečnosti "obyčejný"}] = P[X = 25|p = p_0] = \left(\frac{1}{4}\right)^{25} \doteq 10^{-15}$$

Pro obecný test: *Osobu označíme za jasnovidce (zamítneme "obyčejnost"), trefí-li se alespoň v c případech z n.*

- Jaké rozdělení má X pro "obyčejného" člověka?
  $X \sim \text{Bi}(n, p_0)$
- Pravděpodobnost, že obyčejného člověka označíme za jasnovidce

$$P[\text{zam. "obyč"}|\text{skut. "obyč"}] = P[X \geq c|p = p_0] =$$

$$= \sum_{i=c}^{n} p_{\text{Bi}(n,p_0)}(i) = 1 - F_{\text{Bi}(n,p_0)}(c-1)$$



c	8	9	10	11	12	13	14	15	20	25
$P[X \geq c\|p = p_0]$	0.27	0.15	0.07	0.03	0.01	0.003	0.0009	0.0002	10^{-8}	10^{-15}

- Pravděpodobnost, že by "obyčejný" člověk (náhodou) uhodl 14 nebo více karet z 25 je cca 1/1000.
- Stačí vám to jako důkaz toho, že onen člověk není "obyčejný" (že je jasnovidec)?

Pojmy

Hypotéza H je tvrzení o vlastnostech rozdělení pravděpodobnosti pozorované náhodné veličiny X s distribuční funkcí $F_X(x, \boldsymbol{\theta})$ nebo náhodného vektoru (X, Y) se sdruženou distribuční funkcí $F_{XY}(x, y|\boldsymbol{\theta})$, např:

- Střední hodnota poloměru vyráběných hřídelí se shoduje s nominální hodnotou.
- Úmrtnost při stejném druhu operace na 3 různých pracovištích je stejná.
- Pacienti s AIDS mají červené krvinky menší než zdraví lidé.
- Rozdělení výšky dospělých mužů v ČR je normální.
- Používání léku XY zvyšuje riziko nežádoucích účinků alespoň o 5 %.

Pojmy

Hypotéza H je tvrzení o vlastnostech rozdělení pravděpodobnosti pozorované náhodné veličiny X s distribuční funkcí $F_X(x, \boldsymbol{\theta})$ nebo náhodného vektoru (X, Y) se sdruženou distribuční funkcí $F_{XY}(x, y|\boldsymbol{\theta})$, např:

- Střední hodnota poloměru vyráběných hřídelí se shoduje s nominální hodnotou.
- Úmrtnost při stejném druhu operace na 3 různých pracovištích je stejná.
- Pacienti s AIDS mají červené krvinky menší než zdraví lidé.
- Rozdělení výšky dospělých mužů v ČR je normální.
- Používání léku XY zvyšuje riziko nežádoucích účinků alespoň o 5 %.

Test statistické hypotézy je matematický postup, jímž ověřujeme hypotézu.

- **Nulová hypotéza** H_0 je ta, jejíž platnost ověřujeme.
- **Alternativní hypotézu** H_A stavíme proti H_0.

Pojmy

Hypotéza H je tvrzení o vlastnostech rozdělení pravděpodobnosti pozorované náhodné veličiny X s distribuční funkcí $F_X(x, \boldsymbol{\theta})$ nebo náhodného vektoru (X, Y) se sdruženou distribuční funkcí $F_{XY}(x, y|\boldsymbol{\theta})$, např:

- Střední hodnota poloměru vyráběných hřídelí se shoduje s nominální hodnotou.
- Úmrtnost při stejném druhu operace na 3 různých pracovištích je stejná.
- Pacienti s AIDS mají červené krvinky menší než zdraví lidé.
- Rozdělení výšky dospělých mužů v ČR je normální.
- Používání léku XY zvyšuje riziko nežádoucích účinků alespoň o 5 %.

Test statistické hypotézy je matematický postup, jímž ověřujeme hypotézu.

- **Nulová hypotéza** H_0 je ta, jejíž platnost ověřujeme.
- **Alternativní hypotézu** H_A stavíme proti H_0.

POZOR: *Nelze prokázat platnost statistické hypotézy!* Na základě realizace náhodného výběru buď

- *lze rozhodnout, že hypotéza H_0 není věrohodná* (pak **zamítáme** H_0 a vědomě podstupujeme malé riziko chybného rozhodnutí), nebo
- **nelze zamítnout** H_0, ale v tom případě nevíme, zda je to proto, že
  - H_0 skutečně platí, nebo
  - jen nemáme dostatek dat (dostatečný rozsah výběru) na její zamítnutí.

Průběh testu

Typický postup při testu hypotézy:

1. Formulujeme H_0 a H_A.
2. Zvolíme vhodnou **testovou statistiku** T (**kritérium**) takovou, že
   - její hodnoty co nejtěsněji souvisejí s platností hypotézy H_0,
   - její realizaci jsme schopni spočítat z realizace náhodného výběru a
   - její rozdělení za předpokladu platnosti H_0 známe nebo jsme schopni odvodit.
3. Zvolíme práh $\kappa \in \mathbb{R}$ (**kritickou hodnotu**), pomocí něhož dokážeme rozhodnout o H_0:

$$\text{pro } T > \kappa \text{ zamítáme } H_0,$$
$$\text{pro } T \leq \kappa \text{ nezamítáme } H_0.$$

4. Získáme realizaci náhodného výběru $\boldsymbol{x}$ a provedeme test.

Průběh testu

Typický postup při testu hypotézy:

1. Formulujeme H_0 a H_A.
2. Zvolíme vhodnou **testovou statistiku** T (**kritérium**) takovou, že
   - její hodnoty co nejtěsněji souvisejí s platností hypotézy H_0,
   - její realizaci jsme schopni spočítat z realizace náhodného výběru a
   - její rozdělení za předpokladu platnosti H_0 známe nebo jsme schopni odvodit.
3. Zvolíme práh $\kappa \in \mathbb{R}$ (**kritickou hodnotu**), pomocí něhož dokážeme rozhodnout o H_0:

   pro $T > \kappa$ zamítáme H_0,
   pro $T \leq \kappa$ nezamítáme H_0.
4. Získáme realizaci náhodného výběru $\boldsymbol{x}$ a provedeme test.

Test hypotézy lze považovat za *binární klasifikátor*: jeho úkolem je buď

- zařadit realizaci $\boldsymbol{x}$ do „normální" populace (nezamítne H_0), nebo
- zařadit realizaci $\boldsymbol{x}$ do „anomální" populace (zamítne H_0 ve prospěch H_A).

Průběh testu

Typický postup při testu hypotézy:

1. Formulujeme H_0 a H_A.
2. Zvolíme vhodnou **testovou statistiku T** (**kritérium**) takovou, že
   - její hodnoty co nejtěsněji souvisejí s platností hypotézy H_0,
   - její realizaci jsme schopni spočítat z realizace náhodného výběru a
   - její rozdělení za předpokladu platnosti H_0 známe nebo jsme schopni odvodit.
3. Zvolíme práh $\kappa \in \mathbb{R}$ (**kritickou hodnotu**), pomocí něhož dokážeme rozhodnout o H_0:

   pro $T > \kappa$ zamítáme H_0,
   pro $T \leq \kappa$ nezamítáme H_0.

4. Získáme realizaci náhodného výběru $\boldsymbol{x}$ a provedeme test.

Test hypotézy lze považovat za *binární klasifikátor*: jeho úkolem je buď

- zařadit realizaci $\boldsymbol{x}$ do „normální“ populace (nezamítne H_0), nebo
- zařadit realizaci $\boldsymbol{x}$ do „anomální“ populace (zamítne H_0 ve prospěch H_A).

Testová statistika *T je náhodnou veličinou* (je to funkce náhodného výběru). Rozhodování se proto neobejde bez chyb.

Chyby I. a II. druhu

Stav světa	Rozhodnutí (výsledek testu) H_0 nezamítnuta	H_0 zamítnuta
H_0 platí	OK True negative (TN)	**Chyba I. druhu** False positive (FP)
H_0 neplatí	**Chyba II. druhu** False negative (FN)	OK True positive (TP)

Chyby I. a II. druhu

Stav světa	Rozhodnutí (výsledek testu) H_0 nezamítnuta	 H_0 zamítnuta
H_0 platí	OK True negative (TN)	**Chyba I. druhu** False positive (FP)
H_0 neplatí	**Chyba II. druhu** False negative (FN)	OK True positive (TP)

Říkáme, že výsledek testu je **pozitivní**, když můžeme **zamítnout** H_0, a **negativní**, když nemůžeme.

Chyby I. a II. druhu

Stav světa	Rozhodnutí (výsledek testu) H_0 nezamítnuta	H_0 zamítnuta
H_0 platí	OK True negative (TN)	**Chyba I. druhu** False positive (FP)
H_0 neplatí	**Chyba II. druhu** False negative (FN)	OK True positive (TP)

Říkáme, že výsledek testu je **pozitivní**, když můžeme **zamítnout** H_0, a **negativní**, když nemůžeme.

Chyba I. druhu: zamítneme H_0, která ve skutečnosti platí. Pravděpodobnost chyby I. druhu $\alpha(\kappa)$ je nerostoucí funkce prahu κ a nazýváme ji **hladinou významnosti testu**.

Chyby I. a II. druhu

Stav světa	Rozhodnutí (výsledek testu): H_0 nezamítnuta	H_0 zamítnuta
H_0 platí	OK True negative (TN)	**Chyba I. druhu** False positive (FP)
H_0 neplatí	**Chyba II. druhu** False negative (FN)	OK True positive (TP)

Říkáme, že výsledek testu je **pozitivní**, když můžeme **zamítnout** H_0, a **negativní**, když nemůžeme.

Chyba I. druhu: zamítneme H_0, která ve skutečnosti platí. Pravděpodobnost chyby I. druhu $\alpha(\kappa)$ je nerostoucí funkce prahu κ a nazýváme ji **hladinou významnosti testu**.

Chyba II. druhu: nezamítneme H_0, která neplatí. Pravděpodobnost chyby II. druhu $\beta(\kappa)$ je neklesající funkcí prahu κ. Číslo $1 - \beta$ pak nazýváme **sílou testu**. Hodnotu β lze stanovit jen tehdy, známe-li rozdělení T za předpokladu platnosti H_A!

Chyby I. a II. druhu

Stav světa	Rozhodnutí (výsledek testu) H_0 nezamítnuta	H_0 zamítnuta
H_0 platí	OK True negative (TN)	**Chyba I. druhu** False positive (FP)
H_0 neplatí	**Chyba II. druhu** False negative (FN)	OK True positive (TP)

Říkáme, že výsledek testu je **pozitivní**, když můžeme **zamítnout** H_0, a **negativní**, když nemůžeme.

Chyba I. druhu: zamítneme H_0, která ve skutečnosti platí. Pravděpodobnost chyby I. druhu $\alpha(\kappa)$ je nerostoucí funkce prahu κ a nazýváme ji **hladinou významnosti testu**.

Chyba II. druhu: nezamítneme H_0, která neplatí. Pravděpodobnost chyby II. druhu $\beta(\kappa)$ je neklesající funkcí prahu κ. Číslo $1-\beta$ pak nazýváme **sílou testu**. Hodnotu β lze stanovit jen tehdy, známe-li rozdělení T za předpokladu platnosti H_A!

V lékařské literatuře se často mluví o následujících veličinách:

- **Specificita** (jinak také *true negative rate, TNR*):

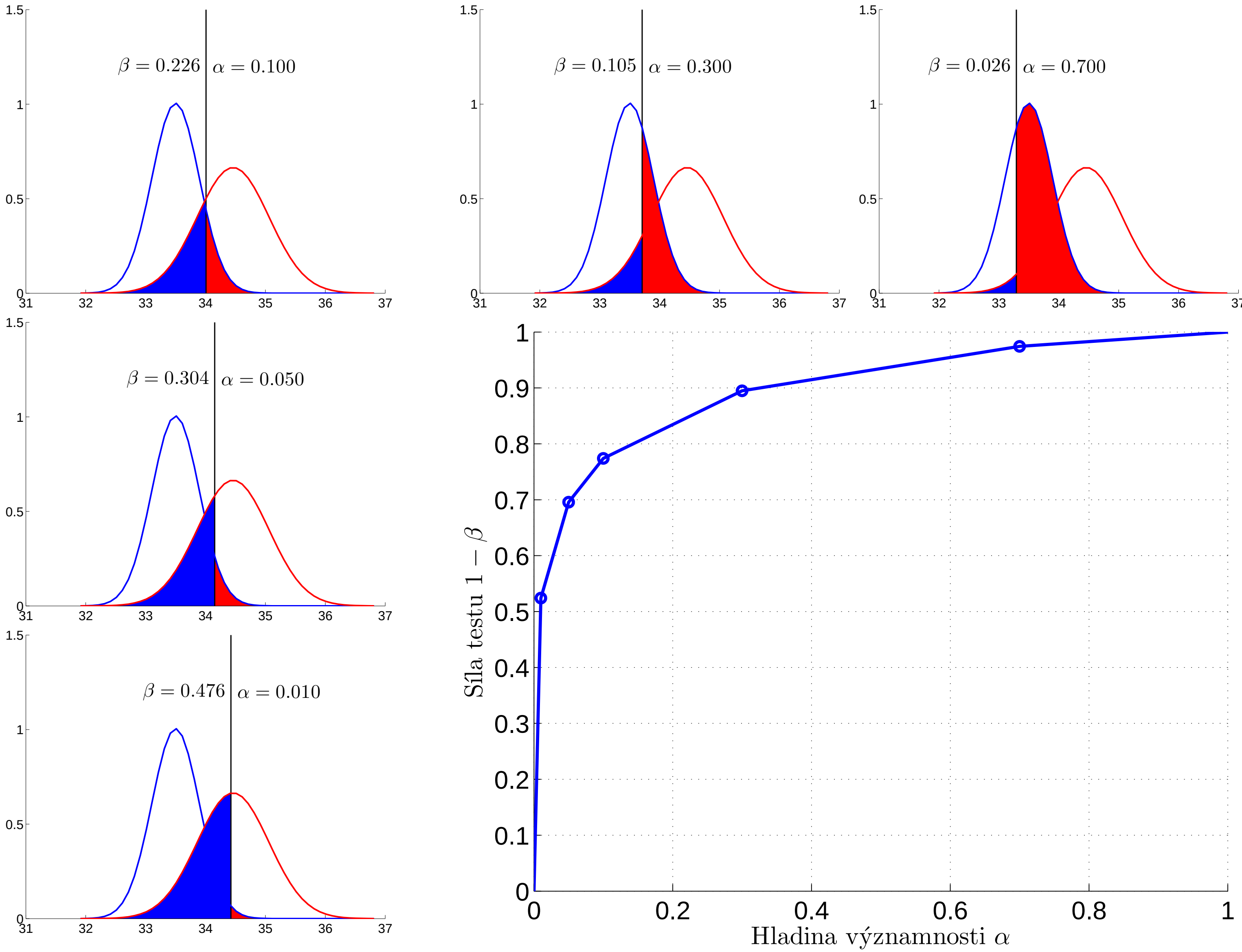
$$spec = TNR = \frac{p_{TN}}{p_{TN} + p_{FP}} = P[H_0 \text{ není zamítnuta} | H_0 \text{ platí}] = 1 - \alpha.$$

- **Senzitivita** (jinak také *true positive rate, TPR,* nebo *recall*):

$$senz = TPR = \frac{p_{TP}}{p_{TP} + p_{FN}} = P[H_0 \text{ je zamítnuta} | H_0 \text{ neplatí}] = 1 - \beta.$$

ROC křivka

Reciever operating characteristic (ROC): závislost významnosti testu α (vodorovná osa) a síly testu $1-\beta$ (svislá osa). Parametrem křivky je kritická hodnota κ.



Volba kritické hodnoty

Volbou kritické hodnoty κ snižujeme riziko jedné chyby, ale zároveň zvyšujeme riziko druhé.

- Hodnotu κ obecně volíme tak, abychom se přiblížili bezchybné klasifikaci ($\alpha = 0, \beta = 0$), tj. bodu (0,1) v grafu ROC.
- Jedinou možností, jak snížit riziko obou chyb zároveň, je získat více dat!

Volba kritické hodnoty

Volbou kritické hodnoty κ snižujeme riziko jedné chyby, ale zároveň zvyšujeme riziko druhé.

- Hodnotu κ obecně volíme tak, abychom se přiblížili bezchybné klasifikaci ($\alpha = 0, \beta = 0$), tj. bodu (0,1) v grafu ROC.
- Jedinou možností, jak snížit riziko obou chyb zároveň, je získat více dat!

Možná kritéria pro volbu prahu κ:

- $\alpha(\kappa) = \beta(\kappa)$
- $\min_\kappa(\alpha(\kappa) + \beta(\kappa))$
- $\min_\kappa e(\alpha(\kappa), \beta(\kappa))$, např. $\min_\kappa(a\alpha(\kappa) + b\beta(\kappa))$, tj. minimalizace výplatní funkce,
- $\alpha(\kappa) =$ předem zvolená malá hodnota.

Volba kritické hodnoty

Volbou kritické hodnoty κ snižujeme riziko jedné chyby, ale zároveň zvyšujeme riziko druhé.

- Hodnotu κ obecně volíme tak, abychom se přiblížili bezchybné klasifikaci ($\alpha = 0, \beta = 0$), tj. bodu (0,1) v grafu ROC.
- Jedinou možností, jak snížit riziko obou chyb zároveň, je získat více dat!

Možná kritéria pro volbu prahu κ:

- $\alpha(\kappa) = \beta(\kappa)$
- $\min_\kappa(\alpha(\kappa) + \beta(\kappa))$
- $\min_\kappa e(\alpha(\kappa), \beta(\kappa))$, např. $\min_\kappa(a\alpha(\kappa) + b\beta(\kappa))$, tj. minimalizace výplatní funkce,
- $\alpha(\kappa) =$ předem zvolená malá hodnota.

Většinou se používá poslední možnost:

- nalezení κ je nejsnazší,
- nepotřebujeme znát rozdělení anomální skupiny (tj. rozdělení, pokud H_0 neplatí).

Volba kritické hodnoty

Volbou kritické hodnoty κ snižujeme riziko jedné chyby, ale zároveň zvyšujeme riziko druhé.

- Hodnotu κ obecně volíme tak, abychom se přiblížili bezchybné klasifikaci ($\alpha = 0, \beta = 0$), tj. bodu (0,1) v grafu ROC.
- Jedinou možností, jak snížit riziko obou chyb zároveň, je získat více dat!

Možná kritéria pro volbu prahu κ:

- $\alpha(\kappa) = \beta(\kappa)$
- $\min_\kappa(\alpha(\kappa) + \beta(\kappa))$
- $\min_\kappa e(\alpha(\kappa), \beta(\kappa))$, např. $\min_\kappa(a\alpha(\kappa) + b\beta(\kappa))$, tj. minimalizace výplatní funkce,
- $\alpha(\kappa) =$ předem zvolená malá hodnota.

Většinou se používá poslední možnost:

- nalezení κ je nejsnazší,
- nepotřebujeme znát rozdělení anomální skupiny (tj. rozdělení, pokud H_0 neplatí).

Kritická hodnota testu κ se volí tak, aby chyba I. druhu nastávala s pravděpodobností α nebo menší (nelze-li dosáhnout rovnosti). Tradičně se volí $\alpha = 0.05$ nebo $\alpha = 0.01$.

- Hodnoty $T > \kappa$ (odpovídají výsledkům málo pravděpodobným za předpokladu platnosti H_0) považujeme za **statisticky významné** (a zamítáme H_0).
- Hodnoty $T \leq \kappa$ (odpovídají výsledkům, jejichž pravděpodobnost není dostatečně malá při platnosti H_0) **nejsou statisticky významné** (a H_0 proto nezamítáme, ani nepotvrzujeme).

Formulace hypotéz

Obvyklým cílem testování hypotéz je stanovit, zda nějaká spekulativní hypotéza o pozorovaném jevu má podporu v datech. O H_0 se předpokládá, že platí, dokud data neposkytnou dostatek „důkazů“ pro její zamítnutí ve prospěch H_A. Proto obvykle:

- H_0 vyjadřuje shodu s předpoklady, neexistenci rozdílu mezi skupinami, neexistenci vlivu, nezávislost, apod. (protože v těchto případech jsme obvykle schopni odvodit rozdělení testové statistiky), zatímco
- H_A vyjadřuje naši spekulativní hypotézu, pro níž hledáme podporu v datech; vyjadřuje odchylku od předpokladů, významný rozdíl, existenci vlivu, závislosti, apod.

Formulace hypotéz

Obvyklým cílem testování hypotéz je stanovit, zda nějaká spekulativní hypotéza o pozorovaném jevu má podporu v datech. O H_0 se předpokládá, že platí, dokud data neposkytnou dostatek „důkazů“ pro její zamítnutí ve prospěch H_A. Proto obvykle:

- H_0 vyjadřuje shodu s předpoklady, neexistenci rozdílu mezi skupinami, neexistenci vlivu, nezávislost, apod. (protože v těchto případech jsme obvykle schopni odvodit rozdělení testové statistiky), zatímco
- H_A vyjadřuje naši spekulativní hypotézu, pro níž hledáme podporu v datech; vyjadřuje odchylku od předpokladů, významný rozdíl, existenci vlivu, závislosti, apod.

H_0 i H_A mohou být

- **jednoduchá hypotéza**, jíž odpovídá jediná hodnota parametru, nebo
- **složená hypotéza**, jíž odpovídá více hodnot parametru.

U složené hypotézy H požadujeme, aby pravděpodobnost chyby I. druhu byla nejvýše α přes všechny hodnoty parametru vyhovující H.

Formulace hypotéz

Obvyklým cílem testování hypotéz je stanovit, zda nějaká spekulativní hypotéza o pozorovaném jevu má podporu v datech. O H_0 se předpokládá, že platí, dokud data neposkytnou dostatek „důkazů“ pro její zamítnutí ve prospěch H_A. Proto obvykle:

- H_0 vyjadřuje shodu s předpoklady, neexistenci rozdílu mezi skupinami, neexistenci vlivu, nezávislost, apod. (protože v těchto případech jsme obvykle schopni odvodit rozdělení testové statistiky), zatímco
- H_A vyjadřuje naši spekulativní hypotézu, pro níž hledáme podporu v datech; vyjadřuje odchylku od předpokladů, významný rozdíl, existenci vlivu, závislosti, apod.

H_0 i H_A mohou být

- **jednoduchá hypotéza**, jíž odpovídá jediná hodnota parametru, nebo
- **složená hypotéza**, jíž odpovídá více hodnot parametru.

U složené hypotézy H požadujeme, aby pravděpodobnost chyby I. druhu byla nejvýše α přes všechny hodnoty parametru vyhovující H.

H_0 a H_A se často formulují tak, že nejsou navzájem svými negacemi a nepokrývají celý prostor možných hodnot parametru $\Longrightarrow$ chaos. Snadno se mu vyhneme, když budeme **formulovat nulovou hypotézu jako negaci alternativní.**

- Je-li $H_A : \theta > c$, nevolíme $H_0 : \theta = c$, ale raději $H_0 : \theta \leq c$.
- Největší riziko chyby I. druhu obvykle odpovídá případu $\theta = c$, takže postup testu je stejný.

Klasický a alternativní postup testu

Předpoklad: Máme testovou statistiku T, která roste s parametrem θ (jehož hodnota je předmětem testu) a má známé rozdělení při platné H_0.

Klasický a alternativní postup testu

Předpoklad: Máme testovou statistiku T, která roste s parametrem θ (jehož hodnota je předmětem testu) a má známé rozdělení při platné H_0.

Klasický test hypotézy: Typický postup:

1. Zvolíme požadovanou hladinu významnosti α.
2. Ze znalosti rozdělení T pro případ, že platí H_0, určíme kritickou hodnotu κ tak, aby $P[T > \kappa] \leq \alpha$, tj. κ je příslušný kvantil rozdělení T: $\kappa = q_T(1-\alpha)$.
3. Porovnáme realizaci t a kritickou hodnotu κ a zamítneme H_0, pokud $t > \kappa$.

Klasický a alternativní postup testu

Předpoklad: Máme testovou statistiku T, která roste s parametrem θ (jehož hodnota je předmětem testu) a má známé rozdělení při platné H_0.

Klasický test hypotézy: Typický postup:

1. Zvolíme požadovanou hladinu významnosti α.
2. Ze znalosti rozdělení T pro případ, že platí H_0, určíme kritickou hodnotu κ tak, aby $P[T > \kappa] \leq \alpha$, tj. κ je příslušný kvantil rozdělení T: $\kappa = q_T(1 - \alpha)$.
3. Porovnáme realizaci t a kritickou hodnotu κ a zamítneme H_0, pokud $t > \kappa$.

Ekvivalentní postup: H_0 zamítáme, pokud hodnota parametru θ platná pro H_0 nepadne do $(1 - \alpha)$ intervalu spolehlivosti.

Klasický a alternativní postup testu

Předpoklad: Máme testovou statistiku T, která roste s parametrem θ (jehož hodnota je předmětem testu) a má známé rozdělení při platné H_0.

Klasický test hypotézy: Typický postup:

1. Zvolíme požadovanou hladinu významnosti α.
2. Ze znalosti rozdělení T pro případ, že platí H_0, určíme kritickou hodnotu κ tak, aby $P[T > \kappa] \leq \alpha$, tj. κ je příslušný kvantil rozdělení T: $\kappa = q_T(1 - \alpha)$.
3. Porovnáme realizaci t a kritickou hodnotu κ a zamítneme H_0, pokud $t > \kappa$.

Ekvivalentní postup: H_0 zamítáme, pokud hodnota parametru θ platná pro H_0 nepadne do $(1 - \alpha)$ intervalu spolehlivosti.

Alternativní postup: zjištění mezní hladiny významnosti, při níž by pozorovaná hodnota t byla kritická, tj.

1. Ze znalosti rozdělení T zjistíme pravděpodobnost, s jakou statistika T nabývá hodnot ještě extrémnějších než t za předpokladu, že platí H_0. Této pravděpodobnosti říkáme **dosažená hladina významnosti** a obvykle se značí P (nebo p-value, p-hodnota).
2. Dosaženou hladinu významnosti P
   - prostě zveřejníme (aby si každý udělal závěr sám), nebo
   - porovnáme se stanovenou požadovanou hladinou významnosti a zamítneme H_0, pokud $P < \alpha$.

Klasický a alternativní postup testu

Předpoklad: Máme testovou statistiku T, která roste s parametrem θ (jehož hodnota je předmětem testu) a má známé rozdělení při platné H_0.

Klasický test hypotézy: Typický postup:

1. Zvolíme požadovanou hladinu významnosti α.
2. Ze znalosti rozdělení T pro případ, že platí H_0, určíme kritickou hodnotu κ tak, aby $P[T > \kappa] \leq \alpha$, tj. κ je příslušný kvantil rozdělení T: $\kappa = q_T(1 - \alpha)$.
3. Porovnáme realizaci t a kritickou hodnotu κ a zamítneme H_0, pokud $t > \kappa$.

Ekvivalentní postup: H_0 zamítáme, pokud hodnota parametru θ platná pro H_0 nepadne do $(1 - \alpha)$ intervalu spolehlivosti.

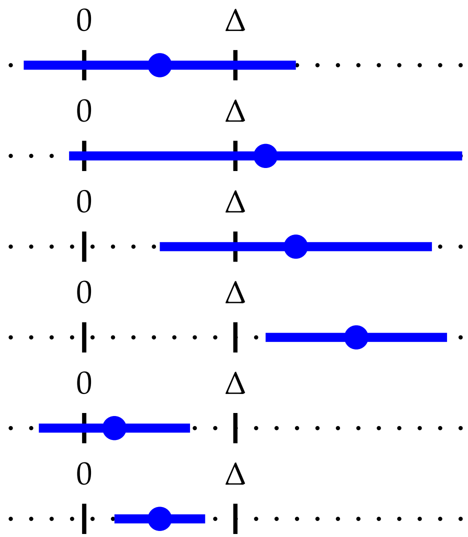
Alternativní postup: zjištění mezní hladiny významnosti, při níž by pozorovaná hodnota t byla kritická, tj.

1. Ze znalosti rozdělení T zjistíme pravděpodobnost, s jakou statistika T nabývá hodnot ještě extrémnějších než t za předpokladu, že platí H_0. Této pravděpodobnosti říkáme **dosažená hladina významnosti** a obvykle se značí P (nebo p-value, p-hodnota).
2. Dosaženou hladinu významnosti P
   - prostě zveřejníme (aby si každý udělal závěr sám), nebo
   - porovnáme se stanovenou požadovanou hladinou významnosti a zamítneme H_0, pokud $P < \alpha$.

Dosaženou hladinu významnosti P lze volně interpretovat jako *míru naší důvěry v platnost H_0.* (Čím nižší, tím je výsledek významnější a tím větší máme “právo” H_0 zamítnout.)

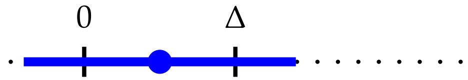
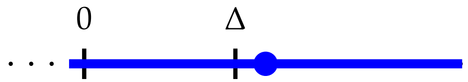
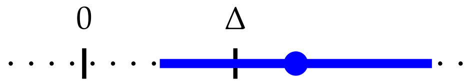
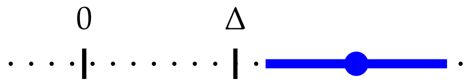
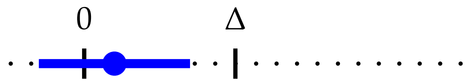
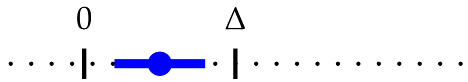
Statistická významnost vs. faktická významnost

- I sebemenší odchylka od předpokladů se s dostatečným rozsahem výběru ukáže jako statisticky významná.
- Označme Δ jistou minimální odchylku, která pro nás bude už fakticky významná.

Interval spolehlivosti	Významnost statistická	skutečná
		

Statistická významnost vs. faktická významnost

- I sebemenší odchylka od předpokladů se s dostatečným rozsahem výběru ukáže jako statisticky významná.
- Označme Δ jistou minimální odchylku, která pro nás bude už fakticky významná.

Interval spolehlivosti	Významnost statistická	skutečná
	Ne	Možná
		
		
		
		
		

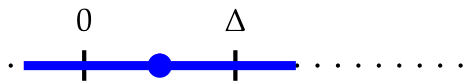
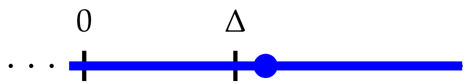
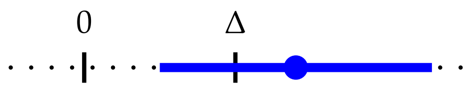
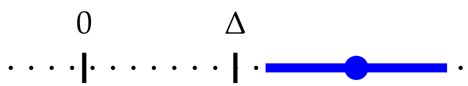
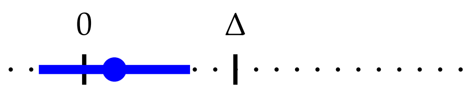
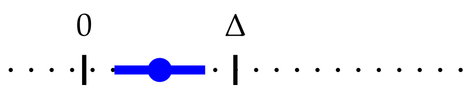
Statistická významnost vs. faktická významnost

Testování hypotéz a jeho metodika
- Jasnovidec?
- Pojmy
- Postup
- Chyby
- ROC křivka
- Volba κ
- Formulace hypotéz
- p-hodnota
- Významnost?

Kuchařka pro testování hypotéz

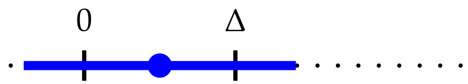
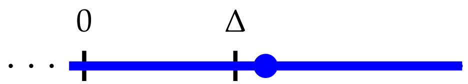
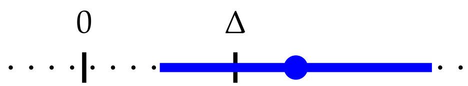
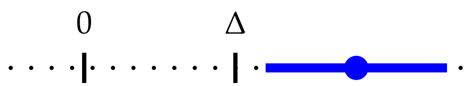
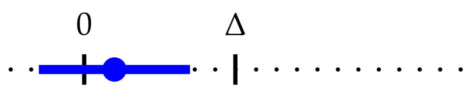
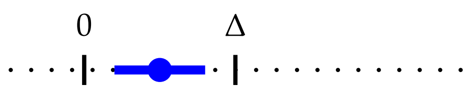
Neparametrické testy

- I sebemenší odchylka od předpokladů se s dostatečným rozsahem výběru ukáže jako statisticky významná.
- Označme Δ jistou minimální odchylku, která pro nás bude už fakticky významná.

Interval spolehlivosti	Významnost statistická	Významnost skutečná
	Ne	Možná
	Ne	Možná
		
		
		
		

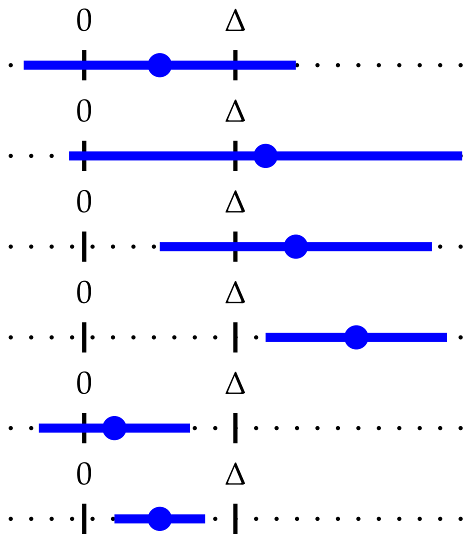
Statistická významnost vs. faktická významnost

- I sebemenší odchylka od předpokladů se s dostatečným rozsahem výběru ukáže jako statisticky významná.
- Označme Δ jistou minimální odchylku, která pro nás bude už fakticky významná.

Interval spolehlivosti	Významnost statistická	Významnost skutečná
	Ne	Možná
	Ne	Možná
	Ano	Možná
		
		
		

Statistická významnost vs. faktická významnost

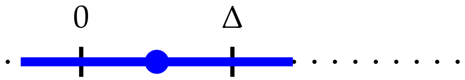
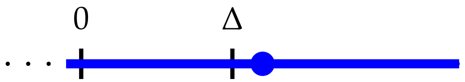
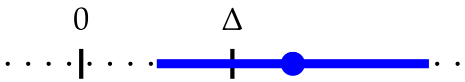
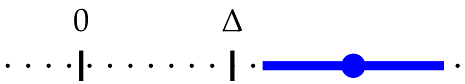
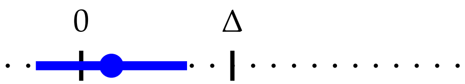
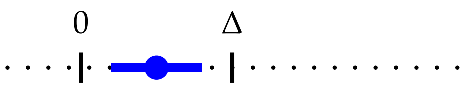
- I sebemenší odchylka od předpokladů se s dostatečným rozsahem výběru ukáže jako statisticky významná.
- Označme Δ jistou minimální odchylku, která pro nás bude už fakticky významná.



Interval spolehlivosti	Významnost statistická	Významnost skutečná
	Ne	Možná
	Ne	Možná
	Ano	Možná
	Ano	Ano

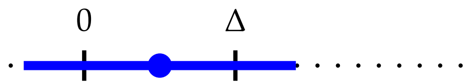
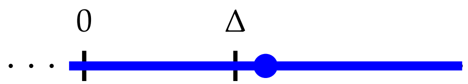
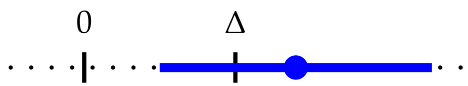
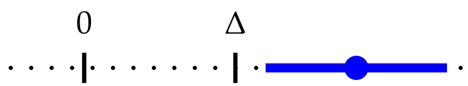
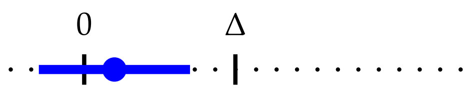
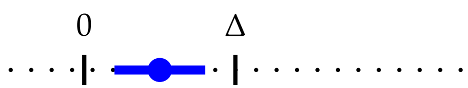
Statistická významnost vs. faktická významnost

- I sebemenší odchylka od předpokladů se s dostatečným rozsahem výběru ukáže jako statisticky významná.
- Označme Δ jistou minimální odchylku, která pro nás bude už fakticky významná.

Interval spolehlivosti	Významnost statistická	Významnost skutečná
	Ne	Možná
	Ne	Možná
	Ano	Možná
	Ano	Ano
	Ne	Ne
		

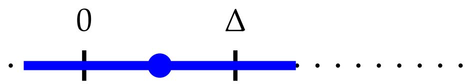
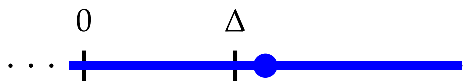
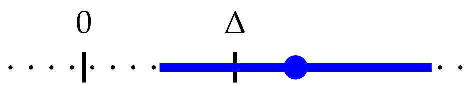
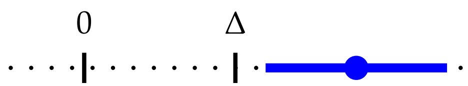
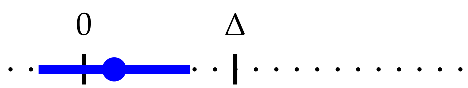
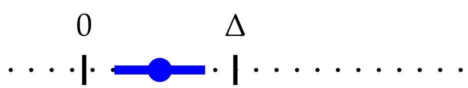
Statistická významnost vs. faktická významnost

- I sebemenší odchylka od předpokladů se s dostatečným rozsahem výběru ukáže jako statisticky významná.
- Označme Δ jistou minimální odchylku, která pro nás bude už fakticky významná.

Interval spolehlivosti	Významnost statistická	Významnost skutečná
	Ne	Možná
	Ne	Možná
	Ano	Možná
	Ano	Ano
	Ne	Ne
	Ano	Ne

Statistická významnost vs. faktická významnost

- I sebemenší odchylka od předpokladů se s dostatečným rozsahem výběru ukáže jako statisticky významná.
- Označme Δ jistou minimální odchylku, která pro nás bude už fakticky významná.

Interval spolehlivosti	Významnost statistická	Významnost skutečná
	Ne	Možná
	Ne	Možná
	Ano	Možná
	Ano	Ano
	Ne	Ne
	Ano	Ne

- Zdaleka ne každý efekt, který je statisticky významný, je významný i reálně.
- Pojem *statistická významnost* je tak lépe chápat jako *statistickou rozeznatelnost*.

Kuchařka pro testování hypotéz

Typický tvar testu

Realizaci t testovací statistiky T, která roste s parametrem θ a pro H_0 má známé rozdělení, porovnáme s kvantily příslušného rozdělení a H_0 zamítneme při extrémních hodnotách (nepravděpodobných při platnosti H_0).

H_0	H_A	H_0 zamítáme, když	dosažená významnost P
$\theta \le c$	$\theta > c$	$t > q_T(1-\alpha)$	$1 - F_T(t)$
$\theta \ge c$	$\theta < c$	$t < q_T(\alpha)$	$F_T(t)$
$\theta = c$	$\theta \ne c$	$t > q_T(1-\frac{\alpha}{2})$ nebo $t < q_T(\frac{\alpha}{2})$	$2\min(F_T(t), 1 - F_T(t))$

Typický tvar testu

Realizaci t testovací statistiky T, která roste s parametrem θ a pro H_0 má známé rozdělení, porovnáme s kvantily příslušného rozdělení a H_0 zamítneme při extrémních hodnotách (nepravděpodobných při platnosti H_0).

H_0	H_A	H_0 zamítáme, když	dosažená významnost P
$\theta \leq c$	$\theta > c$	$t > q_T(1-\alpha)$	$1 - F_T(t)$
$\theta \geq c$	$\theta < c$	$t < q_T(\alpha)$	$F_T(t)$
$\theta = c$	$\theta \neq c$	$t > q_T(1-\frac{\alpha}{2})$ nebo $t < q_T(\frac{\alpha}{2})$	$2\min(F_T(t), 1 - F_T(t))$

V literatuře se často setkáme i s následujícími formulacemi hypotéz, které se ale řeší stejně jako první dva výše uvedené případy:

H_0	H_A
$\theta = c$	$\theta > c$
$\theta = c$	$\theta < c$

Recept: Test střední hodnoty $N(\mu, \sigma^2)$ při známém σ^2

Realizaci testové statistiky

$$t = \frac{\bar{x} - c}{\sigma}\sqrt{n}$$

porovnáme s kvantily *normovaného normálního rozdělení*:

H_0	H_A	H_0 zamítáme, když	dosažená významnost P
$\mu \le c$	$\mu > c$	$t > \Phi^{-1}(1-\alpha)$	$1 - \Phi(t)$
$\mu \ge c$	$\mu < c$	$t < \Phi^{-1}(\alpha)$	$\Phi(t)$
$\mu = c$	$\mu \ne c$	$t > \Phi^{-1}(1-\frac{\alpha}{2})$ nebo $t < \Phi^{-1}(\frac{\alpha}{2})$	$2(1 - \Phi(\lvert t \rvert))$

Recept: Test střední hodnoty $N(\mu, \sigma^2)$ při neznámém σ^2

Realizaci testové statistiky

$$t = \frac{\bar{x} - c}{s_x}\sqrt{n}$$

porovnáme s kvantily *Studentova rozdělení s $n-1$ stupni volnosti*:

H_0	H_A	H_0 zamítáme, když	dosažená významnost P
$\mu \le c$	$\mu > c$	$t > q_{t(n-1)}(1-\alpha)$	$1 - F_{t(n-1)}(t)$
$\mu \ge c$	$\mu < c$	$t < q_{t(n-1)}(\alpha)$	$F_{t(n-1)}(t)$
$\mu = c$	$\mu \ne c$	$t > q_{t(n-1)}(1-\frac{\alpha}{2})$ nebo $t < q_{t(n-1)}(\frac{\alpha}{2})$	$2(1 - F_{t(n-1)}(\lvert t \rvert))$

Příklad: Oboustranný test střední hodnoty

Zadání: 216 pacientům jsme změřili koncentraci bílkovin v krevním séru: $\bar{x} = 34.46$ g/l, $s_x = 5.835$ g/l. Ověřte hypotézu, že střední hodnota koncentrace bílkovin u pacientů tohoto typu je $\mu_0 = 33.5$ g/l, proti možnosti, že se od této hodnoty liší.

Příklad: Oboustranný test střední hodnoty

Zadání: 216 pacientům jsme změřili koncentraci bílkovin v krevním séru: $\bar{x} = 34.46$ g/l, $s_x = 5.835$ g/l. Ověřte hypotézu, že střední hodnota koncentrace bílkovin u pacientů tohoto typu je $\mu_0 = 33.5$ g/l, proti možnosti, že se od této hodnoty liší.

Poznámka: Protože neznáme skutečný rozptyl σ^2, měli bychom použít Studentovo rozdělení. Protože ale máme dostatečně velký vzorek, můžeme jej aproximovat normálním rozdělením.

Příklad: Oboustranný test střední hodnoty

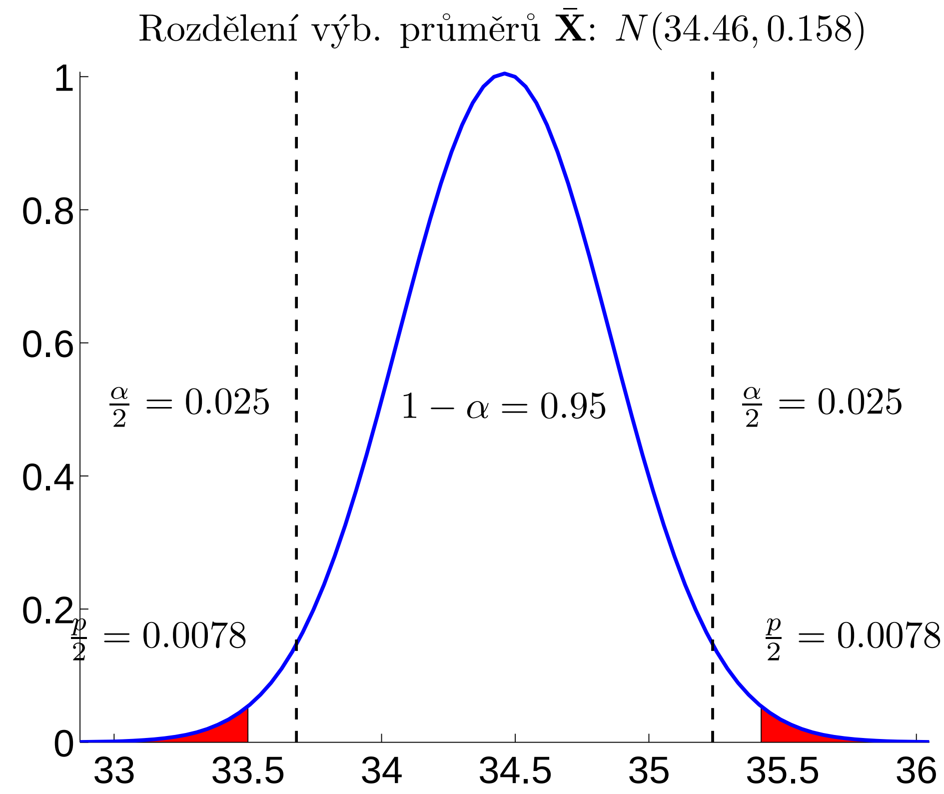
Zadání: 216 pacientům jsme změřili koncentraci bílkovin v krevním séru: $\bar{x} = 34.46$ g/l, $s_x = 5.835$ g/l. Ověřte hypotézu, že střední hodnota koncentrace bílkovin u pacientů tohoto typu je $\mu_0 = 33.5$ g/l, proti možnosti, že se od této hodnoty liší.

Poznámka: Protože neznáme skutečný rozptyl σ^2, měli bychom použít Studentovo rozdělení. Protože ale máme dostatečně velký vzorek, můžeme jej aproximovat normálním rozdělením.

Řešení 1: Interval spolehlivosti. Nepokryje-li 95% interval spolehlivosti I hodnotu μ_0, zamítneme H_0:

$$
\begin{aligned}
I &= \bar{x} \pm \frac{s_x}{\sqrt{n}}\Phi^{-1}(1-\frac{\alpha}{2}) = \\
&= 34.46 \pm \frac{5.835}{\sqrt{216}} 1.96 \text{ g/l} \\
&= 34.46 \pm 0.78 \text{ g/l} \\
I &= (33.68, 35.24) \text{ g/l}
\end{aligned}
$$

Předpokládaná hodnota $\mu_0 = 33.5$ g/l nepatří do tohoto intervalu, což je výsledek, který bychom pozorovali v méně než 5 % případů, kdyby skutečně platilo $\mu = 33.5$ g/l. Proto zamítáme H_0 na hladině významnosti 5 %.



Příklad: Oboustranný test střední hodnoty

Zadání: 216 pacientům jsme změřili koncentraci bílkovin v krevním séru: $\bar{x} = 34.46$ g/l, $s_x = 5.835$ g/l. Ověřte hypotézu, že střední hodnota koncentrace bílkovin u pacientů tohoto typu je $\mu_0 = 33.5$ g/l, proti možnosti, že se od této hodnoty liší.

Poznámka: Protože neznáme skutečný rozptyl σ^2, měli bychom použít Studentovo rozdělení. Protože ale máme dostatečně velký vzorek, můžeme jej aproximovat normálním rozdělením.

Řešení 2: Oboustranný test hypotézy.

Formulujeme hypotézy:

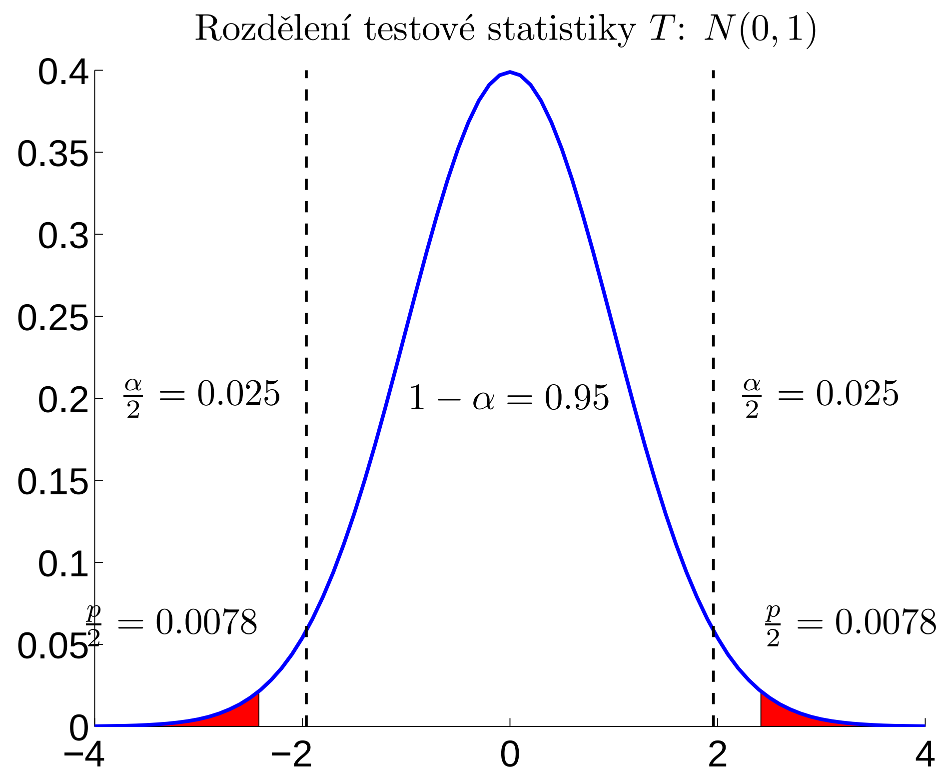
$$H_0: \mu = 33.5 \text{ g/l} \quad \text{a} \quad H_A: \mu \neq 33.5 \text{ g/l}$$

Realizace testové statistiky

$$t = \frac{\bar{x} - \mu}{s_x}\sqrt{n} = \frac{34.46 - 33.5}{5.835}\sqrt{216} = 2.418$$

T má přibližně rozdělení $\Phi = N(0,1)$. Dosažená hladina významnosti:

$$p = 2(1 - \Phi(|t|)) = 0.0156$$



Pravděpodobnost, že bychom pozorovali hodnotu $t = 2.418$ nebo větší, kdyby platila H_0, je pouze 1.56%. Na hladině významnosti 5 % bychom H_0 zamítli. Na hladině významnosti 1 % bychom H_0 zamítnout nemohli.

Recept: Test rozptylu $N(\mu, \sigma^2)$

Realizaci testové statistiky

$$t = \frac{(n-1)s_x^2}{c}$$

porovnáme s kvantily χ^2 *rozdělení s* $n-1$ *stupni volnosti*:

H_0	H_A	H_0 zamítáme, když	dosažená významnost P
$\sigma^2 \leq c$	$\sigma^2 > c$	$t > q_{\chi^2(n-1)}(1-\alpha)$	$1 - F_{\chi^2(n-1)}(t)$
$\sigma^2 \geq c$	$\sigma^2 < c$	$t < q_{\chi^2(n-1)}(\alpha)$	$F_{\chi^2(n-1)}(t)$
$\sigma^2 = c$	$\sigma^2 \neq c$	$t > q_{\chi^2(n-1)}(1-\frac{\alpha}{2})$ nebo $t < q_{\chi^2(n-1)}(\frac{\alpha}{2})$	$2\min(F_{\chi^2(n-1)}(t), 1 - F_{\chi^2(n-1)}(t))$

Neparametrické testy

Neparametrické?

Neparametrické testy nejsou založeny na nějaké parametrizované rodině rozdělení, tj.

- jsou použitelné bez ohledu na typ rozdělení, ale
- jsou slabší (ke stejnému závěru potřebujeme více dat než u parametrických testů, jsou-li aplikovatelné oba druhy).
- Narozdíl od parametrických testů jsou často použitelné i na kvalitativní data, tj. pro ordinální či nominální škálu.

Znaménkový test

Jak otestovat hypotézu o poloze rozdělení, když nemůžeme použít test střední hodnoty (který vyžaduje rozdělení, u něhož střední hodnota existuje)?

- Otestujme medián: $H_0 : q_X(0.5) = c$
- Platí-li H_0, pak jsou kladné i záporné odchylky od c stejně pravděpodobné.

Znaménkový test

Jak otestovat hypotézu o poloze rozdělení, když nemůžeme použít test střední hodnoty (který vyžaduje rozdělení, u něhož střední hodnota existuje)?

- Otestujme medián: $H_0: q_X(0.5) = c$
- Platí-li H_0, pak jsou kladné i záporné odchylky od c stejně pravděpodobné.

Testovací statistikou T je počet kladných odchylek, které testujeme na rozdělení $\text{Bi}(n, 0.5)$. (Z výběru jsme předem vyloučili nulové odchylky.)

H_0	H_A	H_0 zamítáme, když	dosažená významnost P
$q_X(0.5) \leq c$	$q_X(0.5) > c$	$t > q_{\text{Bi}(n,\frac{1}{2})}(1-\alpha)$	$1 - F_{\text{Bi}(n,\frac{1}{2})}(t)$
$q_X(0.5) \geq c$	$q_X(0.5) < c$	$t < q_{\text{Bi}(n,\frac{1}{2})}(\alpha)$	$F_{\text{Bi}(n,\frac{1}{2})}(t)$
$q_X(0.5) = c$	$q_X(0.5) \neq c$	$t > q_{\text{Bi}(n,\frac{1}{2})}(1-\frac{\alpha}{2})$ nebo $t < q_{\text{Bi}(n,\frac{1}{2})}(\frac{\alpha}{2})$	$2\min(F_{\text{Bi}(n,\frac{1}{2})}(t), 1 - F_{\text{Bi}(n,\frac{1}{2})}(t))$

Znaménkový test

Jak otestovat hypotézu o poloze rozdělení, když nemůžeme použít test střední hodnoty (který vyžaduje rozdělení, u něhož střední hodnota existuje)?

- Otestujme medián: $H_0\colon q_X(0.5) = c$
- Platí-li H_0, pak jsou kladné i záporné odchylky od c stejně pravděpodobné.

Testovací statistikou T je počet kladných odchylek, které testujeme na rozdělení $\mathrm{Bi}(n, 0.5)$. (Z výběru jsme předem vyloučili nulové odchylky.)

H_0	H_A	H_0 zamítáme, když	dosažená významnost P
$q_X(0.5) \le c$	$q_X(0.5) > c$	$t > q_{\mathrm{Bi}(n,\frac{1}{2})}(1-\alpha)$	$1 - F_{\mathrm{Bi}(n,\frac{1}{2})}(t)$
$q_X(0.5) \ge c$	$q_X(0.5) < c$	$t < q_{\mathrm{Bi}(n,\frac{1}{2})}(\alpha)$	$F_{\mathrm{Bi}(n,\frac{1}{2})}(t)$
$q_X(0.5) = c$	$q_X(0.5) \ne c$	$t > q_{\mathrm{Bi}(n,\frac{1}{2})}(1-\frac{\alpha}{2})$ nebo $t < q_{\mathrm{Bi}(n,\frac{1}{2})}(\frac{\alpha}{2})$	$2\min(F_{\mathrm{Bi}(n,\frac{1}{2})}(t), 1 - F_{\mathrm{Bi}(n,\frac{1}{2})}(t))$

Pro velká n používáme CLV a testujeme

$$T_0 = \frac{2T - n}{\sqrt{n}}$$

na rozdělení $\mathrm{N}(0, 1)$.

Jednovýběrový Wilcoxonův test

Testuje H_0: X má rozdělení symetrické kolem hodnoty c.

- Má-li X rozdělení symetrické kolem c, pak je c mediánem i střední hodnotou.
- Často se používá jako neparametrická alternativa testu střední hodnoty, je silnější než znaménkový test.
- Z realizace $\boldsymbol{x} = (x_1, \ldots, x_n)$ vypočteme posloupnost $(z_1, \ldots, z_n)$, kde $z_j = x_j - c$.
- Seřadíme ji vzestupně podle $|z_j|$, čímž j-tému prvku přiřadíme pořadí r_j. Je-li více stejných rozdílů, přiřadíme jim stejné pořadí rovné aritmetickému průměru.
- Testovou statistikou je

$$T_1 = \sum_{j:z_j>0} r_j$$

nebo

$$T_2 = \min\left(\sum_{j:z_j>0} r_j, \sum_{j:z_j<0} r_j\right).$$

- Porovnáváme s tabulkou kritických hodnot pro tento test.

Testování hypotéz a jeho metodika

Kuchařka pro testování hypotéz

Neparametrické testy
- Neparametrické?
- Znaménkový test
- Jednovýběrový Wilcoxonův test

Jednovýběrový Wilcoxonův test

Testuje H_0: X má rozdělení symetrické kolem hodnoty c.

- Má-li X rozdělení symetrické kolem c, pak je c mediánem i střední hodnotou.
- Často se používá jako neparametrická alternativa testu střední hodnoty, je silnější než znaménkový test.
- Z realizace $\boldsymbol{x} = (x_1, \ldots, x_n)$ vypočteme posloupnost $(z_1, \ldots, z_n)$, kde $z_j = x_j - c$.
- Seřadíme ji vzestupně podle $|z_j|$, čímž j-tému prvku přiřadíme pořadí r_j. Je-li více stejných rozdílů, přiřadíme jim stejné pořadí rovné aritmetickému průměru.
- Testovou statistikou je

$$T_1 = \sum_{j: z_j > 0} r_j$$

nebo

$$T_2 = \min\left(\sum_{j: z_j > 0} r_j, \sum_{j: z_j < 0} r_j \right).$$

- Porovnáváme s tabulkou kritických hodnot pro tento test.

V dalších přednáškách uvidíte příklady dalších neparametrických testů.