Interval spolehlivosti – velký výběr

- Z centrální limitní věty vyplývá, že pro *n* konvergující k nekonečnu, distribuce náhodné proměnné *Z* (*t*-průměr)

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$$

konverguje ke standardnímu normálnímu rozdělení.
- Nechť $X_1,\ldots, X_n$ je náhodný výběr z distribuce *F* se střední hodnotou *μ.*
- Pokud *n* bude dostatečně velké, pak musí platit:

$$\mathrm{P}\left(-z_{\alpha/2} < \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} < z_{\alpha/2}\right) \approx 1 - \alpha$$

- Tedy pokud $x_1,\ldots, x_n$ jsou realizace náhodného výběru neznámě distribuce se střední hodnotou *μ* a *n* je dostatečně velké, pak 1 – *α* interval spolehlivosti pro *μ* bude:

$$\left(\bar{x}_n - z_{\alpha/2}\frac{s_n}{\sqrt{n}}, \bar{x}_n + z_{\alpha/2}\frac{s_n}{\sqrt{n}}\right)$$

Interval spolehlivosti – velký výběr

- Otázka nyní zní: Jak velké by mělo být *n*, aby šlo na náhodný výběr z libovolné distribuce aplikovat postup konstrukce intervalu spolehlivosti pro $N(0, 1)$?
- Nelze jednoznačně odpovědět. Hodně záleží na tom, jaká vlastně ta neznámá distribuce je.
- Př.: mějme 4 distribuce, dvě rozdílné velikosti náhodných výběrů z těchto distribucí a dvě rozdílné hladiny spolehlivosti.
- Provedeme simulaci podobně jako na str. 36 v před. 8 a dostaneme 10 000 intervalů spolehlivosti pomocí metody velkých výběrů.
- V tabulce pak máme simulované hladiny spolehlivosti (pravděpodobnostní míru, že mi simulovaný interval pokrývá střední hodnotu dané distribuce).
- Pareto distribuce je hodně asymetrická oproti exponenciální distribuci.

		γ	
Distribution	n	0.900	0.950
$Exp(1)$	20	0.851	0.899
$Exp(1)$	100	0.890	0.938
$Par(2.1)$	20	0.727	0.774
$Par(2.1)$	100	0.798	0.849

Interval spolehlivosti – Bin(n, p)

- Mějme statistický soubor = realizace náhodné proměnné *X* s *Bin*(*n*,p).
- Pomocí realizace *X* chceme stanovit parametr *p*.
- Př.: mějme volby, chceme stanovit jaký poměr *p* voličů volí kandidáta *G*.
- Uděláme náhodný výběr *n* prvků ze všech voličů. Nechť *x* voličů volí *G* a tedy poměr *x*/*n* je odhadem parametru *p*. Náhodná prom. *X* je modelována distribucí *Bin*(*n*,*p*).
- Hledáme interval spolehlivosti pro *p*. Hledáme takové odhadové funkce *L* a *U* pro výběrové parametry *l* a *u*, že musí platit: $\mathrm{P}(L < p < U) = 1 - \alpha$

Interval spolehlivosti – Bin(n, p)

- Obecně neexistuje řešení tohoto problému.
- Pokud je *n* dostatečně velké, pak lze pomocí tzv. „Wilsonovy“ metody najít interval spolehlivosti s hladinou spolehlivosti přibližně (1 - α).
- Pozn.: jak moc je hladina spolehlivosti blízko (1 - α) závisí na velikosti hledaného *p*. Ideálně by *p* mělo být „daleko“ od 0 nebo 1.
- Z centrální limitní věty pro dostatečně velké *n* plyne, že *X* lze aproximovat standardním normálním rozdělením: $\dfrac{X - np}{\sqrt{np(1-p)}}$

Interval spolehlivosti – Bin(n, p)

- Jmenovatel i čitatel vydělíme *n*: $$\frac{\frac{X}{n}-p}{\sqrt{\frac{p(1-p)}{n}}}.$$
- Pro velké *n* musí platit: $$\mathrm{P}\left(-z_{\alpha/2}<\frac{\frac{X}{n}-p}{\sqrt{\frac{p(1-p)}{n}}}<z_{\alpha/2}\right)\approx 1-\alpha.$$
- Pravděpodobnostní jev je ekvivalentní s:

$$\left(\frac{\frac{X}{n}-p}{\sqrt{\frac{p(1-p)}{n}}}\right)^2<\left(z_{\alpha/2}\right)^2$$

- Což lze upravit na:

$$\left(\frac{X}{n}-p\right)^2-\left(z_{\alpha/2}\right)^2\,\frac{p(1-p)}{n}<0.$$

Interval spolehlivosti – Bin(n, p)

- Teď už stačí nerovnost upravit tak, aby odpovídala $L < p < U$ a získáme tak interval spolehlivosti pro p.
- Konkrétní příklad: bylo náhodně vybráno $n = 125$ voličů z toho $x = 78$ volilo kandidáta G. Jaký je 95% interval spolehlivosti pro p?
- Víme, že $z_{\alpha/2} = z_{0,025} = 1,96$. Dosadíme do nerovnosti:

$$\left(\frac{78}{125} - p\right)^2 - \frac{(1.96)^2}{125}\,p(1-p) < 0$$

- Úpravou získáme:

$$1.0307\,p^2 - 1.2787\,p + 0.3894 < 0$$

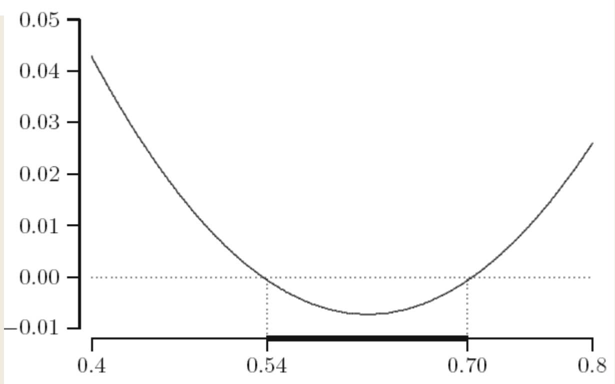
Interval spolehlivosti – Bin(n, p)

- Dostali jsme rovnici paraboly. Hledáme pro jaké *p* bude parabola záporná. Řešením kvadratické rovnice získáme interval, kdy je parabola záporná -> interval spolehlivosti pro *p*.
- Pro tento konkrétní případ:

$$p_{1,2} = \frac{-(-1.2787) \pm \sqrt{(-1.2787)^2 - 4 \cdot 1.0307 \cdot 0.3894}}{2 \cdot 1.0307} = 0.6203 \pm 0.0835;$$

- A tedy interval spolehlivosti:

  (0,54, 0,70).



Interval spolehlivosti – obecná metoda

- Známe postup jak stanovit interval spolehlivosti pro odhad $\bar{X}_n$ střední hodnoty μ a parametru p binomického rozdělení.
- V obecném případě máme statistický výběr $X_1,\ldots, X_n$ z nějaké distribuce. Zajímá nás v jakém intervalu spolehlivosti leží hodnota nějakého parametru θ této distribuce.
- Odhadová funkce $T = f(X_1,\ldots, X_n)$ a funkční hodnota t je odhadem parametru θ.
- Obecně distribuce T musí být funkcí jak $X_1,\ldots, X_n$ tak θ. Pak lze nalézt takové funkce g(θ) a h(θ), že platí:

$$\mathrm{P}(g(\theta) < T < h(\theta)) = 1 - \alpha$$

Interval spolehlivosti – obecná metoda

- Jestliže funkce *g* a *h* jsou striktně rostoucí, pak lze nerovnost přepsat na: $h^{-1}(T) < \theta < g^{-1}(T)$
- Pak pro každé θ platí:

$$\mathrm{P}\left(h^{-1}(T) < \theta < g^{-1}(T)\right) = 1 - \alpha$$

- Poslední rovnice implikuje, že intervalem spolehlivosti 1 – α pro parametr θ je: $(h^{-1}(t), g^{-1}(t))$

Jednostranný interval spolehlivosti

- Interval spolehlivosti říkal, s jakou pravděpodobností hodnota parametru leží v nějakém rozpětí hodnot.
- Teď nás zajímá: s jakou pravděpodobností γ bude hodnota parametru větší jak nějaká hraniční předdefinovaná hodnota.
- Př.: hledáme 95% jednostranný interval spolehlivosti, že hodnota parametru překročí limit 31,00.
- Podobně jako jsme hledali interval spolehlivosti střední hodnoty, tak nás bude zajímat jen spodní mez nerovnosti definující interval spolehlivosti. Tedy máme:

$$\mathrm{P}\left(\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} < t_{n-1,\alpha}\right) = 1 - \alpha,$$

Jednostranný interval spolehlivosti

- což lze upravit na: $$\mathrm{P}\left(\bar{X}_n - t_{n-1,\alpha}\frac{S_n}{\sqrt{n}} < \mu\right) = 1 - \alpha.$$
- Pro případ, že neznáme rozptyl statistického souboru použijeme *t* rozdělení.
- A jednostranný (1 - α) interval spolehlivosti pro střední hodnotu μ je: $$\left(\bar{x}_n - t_{n-1,\alpha}\frac{s_n}{\sqrt{n}},\, \infty\right)$$
- Nechť α = 0,05, n = 22, $\langle x_n \rangle$ = 31,012 a s_n = 0,1294, pak: $$\left(31.012 - 1.721\frac{0.1294}{\sqrt{22}},\, \infty\right) = (30.964, \infty).$$

Jednostranný interval spolehlivosti

- Vidíme, že 95% jednostranný interval spolehlivosti je menší jak 31,00. Musíme tedy provést více měření, abychom se dostali na požadovanou hodnotu 31,00 se spolehlivostí aspoň 95%.
- Obecná definice jednostranného intervalu spolehlivosti pak je: $\mathrm{P}(L_n < \theta) = \gamma$ a interval bude: (l_n, ∞)
- Číslo l_n je pak γ dolní spolehlivostní hranice pro parametr θ.
- Číslo u_n je pak γ horní spolehlivostní hranice parametru θ s jednostranným intervalem spolehlivosti: $(-\infty, u_n)$

Testování hypotéz

Testování hypotéz

- Známe statistické metody, které umožňují odhadnout určité vlastnosti modelových distribucí odpovídající hledaným parametrům.
- Tyto odhady jsou ve formě bodových odhadů nebo intervalů spolehlivosti.
- Často nastává situace, kdy výstupem má být nějaké tvrzení (ne číslo!!!) a my chceme ověřit jeho pravdivost na základě získaných naměřených dat.
- Říkáme, že provádíme **testování hypotéz**, které se vzájemně vylučují.
- Př.: technologové chtějí vědět zda je suché vrtání rychlejší jak chlazené vrtání; chci vědět, zdali životnost ložisek odpovídá životnosti garantované výrobcem; je-li účinnost jedné metody vyšší než druhé metody; atp.
- Tedy vybíráme mezi dvěma možnostmi a hledáme postup, jak určit, která z hypotéz je pravdivá.

Nulová hypotéza

- Vraťme se k příkladu s hledáním počtu vyrobených automobilů ze znalosti náhodně určených sériových čísel – viz přednáška 7 str. 19.
- Mějme nyní náhodný výběr 5 sériových čísel: 61, 19, 56, 24, 16.
- Z „neověřených“ zdrojů máme informaci, že zkoumaná automobilka vyrábí měsíčně 350 aut.
- Podle získaných náhodných sériových čísel jsou získané „informace“ o 350 autech měsíčně více než nadhodnocené.
- Úkolem je vybrat mezi dvěma tvrzeními to správné:
  – měsíční výroba je 350 aut
  – měsíční výroba je menší jak 350 aut
- Dvě protichůdné tvrzení nazýváme jako **nulová (testovaná) hypotéze** H_0 a jako **alternativní hypotéza** H_1.

Nulová hypotéza

- Je to koncepčně stejná situace jako posuzování soudu o něčí vině nebo nevině.
- H_0 odpovídá faktu, že obžalovaný je považován za nevinného, pokud není přesvědčivými důkazy prokázána jeho vina.
- H_1 odpovídá obžalobě vznesené proti žalovanému.
- K rozhodnutí o tom, zdali je H_0 nepravdivá užijeme statistický model.
- 5 náhodně vybraných čísel bez vracení jsou realizace náhodné proměnné $X_1,\ldots, X_5$ z čísel 1, 2, 3,…, N, kde N představuje celkový počet vyrobených aut v měsíci.
- Dvě hypotézy jsou: $H_0 : N = 350$ a $H_1 : N < 350$

Nulová hypotéza

- Odmítnutím nulové hypotézy akceptujeme alternativní hypotézu (odmítáme H_0 ve prospěch H_1).
- Tedy výběr tvrzení H_1 (alternativní hypotézy) je velmi důležitý!!!
- Výrok H_1 představuje něco čemu budeme věřit, že platí, když odmítneme H_0. Tedy H_1 musí být velmi dobře formulováno, aby se to vztahovalo na náš problém.
- Dále musíme vybrat **kritérium** na základě náhodného výběru $X_1, \ldots, X_5$, které poskytne **indikátor** o tom, zdali je H_0 pravdivé či nikoli.
- Kritérium určíme pomocí **statistického testu (testu významnosti)**.

Statistický test

TEST STATISTIC. Suppose the dataset is modeled as the realization of random variables $X_1, X_2, \ldots, X_n$. A *test statistic* is any sample statistic $T = h(X_1, X_2, \ldots, X_n)$, whose numerical value is used to decide whether we reject H_0.

- V našem modelovém příkladu statistickým testem bude odhadová funkce *T*, jejíž hodnota *t* bude kritériem rozhodnutí:

$$T = \max\{X_1, X_2, \ldots, X_5\}$$

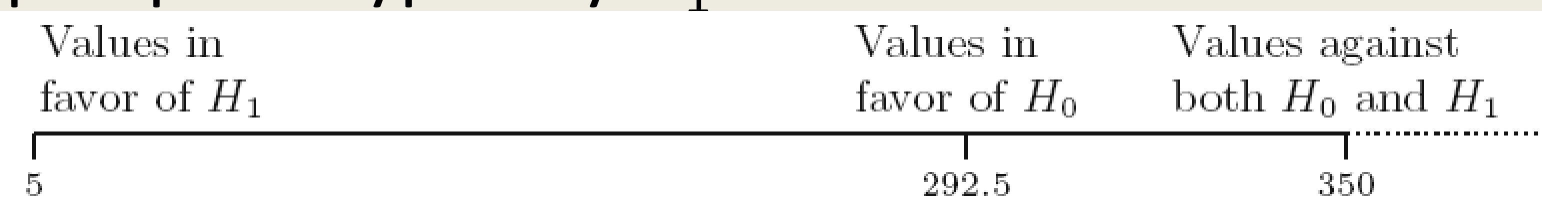
- Budeme zkoumat, jakých hodnot *t* může nabývat. Z možných hodnot *t* vytvoříme **stupnici věrohodnosti**.
- Musíme rozhodnout, které z hodnot *t* poskytují důkaz ve prospěch H_0 a které ve prospěch H_1.
- Pokud *t* bude nabývat hodnot větších jak 350, tak hned víme, že tvrzení obě H_0 a H_1 jsou nepravdivé.

Stupnice věrohodnosti

- Možné, pro nás zajímavé hodnoty odhadové funkce *T* jsou celá čísla ležící v intervalu (5, 350).
- Z přednášky 7, str. 25 víme, že pokud $n = 5$, pak střední hodnota *T* je: E[T] = 5(N+1)/6.
- Tedy distribuce *T* musí být centrovaná kolem hodnoty E[T].
- Pokud je H_0 pravdivé, pak typické hodnoty odhadové funkce *T* vyjádřené čísly *t* musí ležet v blízkém okolí hodnoty E[T] = 5*351/6 = 292,5.
- Hodnoty *t* hodně vzdálené od 292,5 jsou důkazem proti platnosti hypotézy H_0.
- Hodnoty hodně větší jak 292,5 jsou proti H_0 ale zároveň poskytují silnější důkaz proti H_1. Pak neodmítáme hypotézu H_0 ve prospěch hypotézy H_1 !!!

Stupnice věrohodnosti

- Hodnoty *t*, jež jsou trošku menší jak 292,5 jsou důvodem neodmítnutí nulové hypotézy H_0, protože jsme přesvědčeni věřit ve správnost H_0.
- Na druhou stranu hodnoty *t* blízko malých čísel kolem 5 nás přesvědčují o tom, že máme silné důkazy o nutnosti odmítnout hypotézu H_0 ve prospěch hypotézy H_1.



- Pro náš konkrétní příklad realizace náhodného výběru hodnota odhadové funkce *T* je:

$$t = \max\{61, 19, 56, 24, 16\} = 61$$

Stupnice věrohodnosti

- Jak se rozhodneme?
- Ze stupnice věrohodnosti plyne, že hodnota t leží hodně nalevo od E[T] a tak tedy **hypotézu H_0 odmítáme ve prospěch H_1: N < 350**.
- **Otázka zní**: můžeme odmítnout H_0 nebo můžeme připustit fakt, že náš náhodný výběr generující $T = 61$ byla jen náhoda nemající nic společného s realitou?
- Neboli: můžeme H_0 odmítnout nade vší pochybnost?
- Abychom to rozhodli, budeme zkoumat, jak pravděpodobné je, že by se pozorovaly takové hodnoty T, které by ještě silněji svědčily proti pravdivosti H_0, než je hodnota $T = 61$.
- Pokud to bude velmi nepravděpodobné, pak $T = 61$ už samo o sobě je silný důkaz proti platnosti hypotézy H_0.

Okrajová pravděpodobnost

- Hodnoty $T < 61$ poskytují silnější důkaz proti platnosti H_0. Proto spočítejme pravděpodobnost $P(T \leq 61)$ – **okrajová pravděpodobnost**.
- Pokud $N = 350$, odhadová funkce je maximum z 5 náhodně vybraných čísel bez vracení z řady 1, 2, 3, …, 350 pak platí:

$$
\begin{aligned}
\mathrm{P}(T \leq 61) &= \mathrm{P}(\max\{X_1, X_2, \ldots, X_5\} \leq 61) \\
&= \frac{61}{350} \cdot \frac{60}{349} \cdots \frac{57}{346} = 0.00014.
\end{aligned}
$$

- Vidíme, že pravděpodobnost je extrémně malá a tedy hodnota $T = 61$ je dostatečně silný důkaz pro odmítnutí hypotézy H_0.
- **Závěr**: $T = 61$ je mimořádně malé na to, aby H_0 bylo pravdivé.

Okrajová pravděpodobnost

- Nyní můžeme udělat dvě možná rozhodnutí:
  - buď věříme tomu, že H_0 je pravdivé a stalo se něco neočekávatelného při náhodném výběru $X_1,\ldots, X_5$
  - nebo věříme, že události s tak malou pravděpodobností (0,00014) v reálném životě nenastávají a tak tedy hodnoty $T \leq 61$ mohly nastat jen díky tomu, že hypotéza H_0 je nepravdivá.
- Věříme druhé možnosti a odmítáme H_0: $N = 350$ ve prospěch H_1: $N < 350$. Hypotéza H_0 neplatí nade vší pochybnosti.

P-hodnoty

- V předchozím příkladě jsme počítali tzv. levou okrajovou pravděpodobnost P($T \leq 61$).
- Může nastat situace, že silnější důkaz proti H_0 budou hodnoty T pozorované napravo od E[T].
- Budeme tedy hledat pravou okrajovou pravděpodobnost P($T \geq 330$).
- Takovéto pravděpodobnosti nazýváme jako ***p*–hodnoty**.
- Tedy velikost *p*–hodnoty odráží, jak moc důkazů pozorovaná hodnota T poskytuje proti platnosti H_0.
- Čím **menší** je *p*–hodnota, tím **silnější** důkaz je hodnota T proti platnosti H_0.

Chyby I a II druhu

- Předpokládejme, že hodnota odhadové funkce T vyjde $T = 200$ místo 61.
- Je nyní hodnota $T = 200$ dostatečně vlevo, abychom odmítli hypotézu H_0?
- Spočítejme si p-hodnotu:

$$\begin{aligned}\mathrm{P}(T \leq 200) &= \mathrm{P}(\max\{X_1, X_2, \ldots, X_5\} \leq 200) \\ &= \frac{200}{350} \cdot \frac{199}{349} \cdots \frac{196}{346} = 0.0596.\end{aligned}$$

- Je tedy pravděpodobnost 0,0596 dostatečně malá na to, abychom odmítli H_0?
- Rozlišujeme teď dvě situace:
  - jak je to ve skutečnosti: H_0 nebo H_1 je pravdivé
  - jaký je náš názor: odmítáme H_0 ve prospěch H_1 nebo neodmítáme H_0 na základě získaných náhodných dat

Chyby I a II druhu

		True state of nature	
		H_0 is true	H_1 is true
Our decision on the basis of the data	Reject H_0	*Type I error*	Correct decision
	Not reject H_0	Correct decision	*Type II error*

- Nastávají dvě situace, kdy rozhodnutí na základě náhodných dat jsou nesprávná.

TYPE I AND II ERRORS. A *type I error* occurs if we falsely reject H_0. A *type II error* occurs if we falsely do not reject H_0.

- Chyba I druhu: odsoudíme nevinného obžalovaného
- Chyba II druhu: zprostíme viny kriminálníka

Chyby I a II druhu

- V našem příkladě jestliže H_0: $N = 350$ je pravdivá, potom rozhodnutí odmítnou H_0 je chybou I druhu.
- Nikdy nebudeme vědět zdali jsme udělali chybu I druhu!!!
- Pokud budeme mít **rozhodovací pravidlo**, pak můžeme stanovit pravděpodobnost, že jsme spáchali chybu I druhu.
- Rozhodovací pravidlo: „odmítáme H_0: $N = 350$, pokaždé když $T \leq 200$“.
- Pak pravděpodobnost spáchání chyby I druhu je $P(T \leq 200) = 0{,}0596$.
- Jak malá musí být *p*-hodnota, abychom mohli udělat rozhodnutí nade vší pochybnosti?

Chyby I a II druhu

- Praktická zásada velikosti *p*-hodnoty rovné 0,05 se užívá jako úroveň, kdy začínají pochybnosti.
- Cokoliv nastane s pravděpodobností menší jak 0,05 je považováno za jako příliš výjimečné.
- Ale není to obecné pravidlo, jak malá musí být *p*-hodnota, abychom museli odmítnout H_0. Může to být libovolně jiné číslo mezi 0 a 1.
- Nejlepší je zvolit *p*-hodnotu na základě daného statistického souboru. Je to objektivní a je v tom obsaženo nejvíce informací.
- Kdo pak dělá rozhodnutí o nulové hypotéze si může podle uvážení zvolit vlastní rozhodovací hladinu.

Hladina významnosti

- V předchozím příkladu jsme rozhodovali o platnosti nulové hypotézy na základě statistického souboru a spočítání *p*-hodnoty. Tím jsme stanovili rozhodovací kritérium.
- Často nastává **opačná situace**, kdy rozhodovací hladina je pevně dána a hledáme jakou hodnotu T by musel mít statistický test, abychom odmítli H_0.
- Př.: rychlostní limit na silnici je 120 km/h. Rychlost vozidel změřená radarem je modelována jako náhodný výběr X_1, X_2, X_3 a zařízení hned spočítá $\bar{X}_3$. Na základě velikosti $\bar{X}_3$ je udělena nebo neudělena pokuta.
- Jakých hodnot by mělo $\bar{X}_3$ nabývat k pokutování řidičů, jestliže připustíme, že 5% řidičů bude pokutováno nespravedlivě?

Hladina významnosti

- Předpokládáme, že: měření rychlosti = skutečná rychlost vozidla + chyba měření.
- Chyba měření: náhodná proměnná (typicky s normální distribucí) s nulovou střední hodnotou a rozptylem třeba $\sigma^2 = 4$.
- Pak X_1, X_2, X_3 jsou náhodný výběr z $N(\mu,4)$.
- μ = skutečná rychlost vozidla
- Testované hypotézy jsou: H_0: $\mu = 120$ a H_1: $\mu > 120$.
- Statistický test: $T = (X_1+X_2+X_3)/3 = \bar{X}_3$.
- Z vlastnosti normálního rozdělení plyne, že $\bar{X}_3$ musí mít distribuci $N(\mu,4/3)$.
- Hodnoty T blízké 120 -> H_0 přijímáme.

Hladina významnosti

- Hodnoty *T* daleko od 120 považujeme jako silný důkaz pro odmítnutí H_0.
- Hodnoty *T* větší jak 120 ukazují na fakt, že μ > 120 a tedy odmítáme H_0 ve prospěch H_1.
- Hodnoty *T* menší jak 120 -> odmítáme H_0, ale zároveň máme ještě silnější důkaz proti H_1.
- Odmítnutí H_0 ve prospěch H_1 odpovídá pokutování řidiče.
- Nespravedlivě pokutovaný řidič odpovídá chybnému odmítnutí H_0 -> spáchali jsme chybu I druhu.
- Požadujeme, aby pravděpodobnost spáchání chyby I druhu byla *a priori* 5%.
- Pro jaké hodnoty *T* bychom měli odmítnout H_0?
- **Rozhodovací pravidlo** pro odmítnutí H_0 musí být takové, že odpovídající **pravděpodobnost** spáchání **chyby** I druhu je 0,05.

Hladina významnosti

- Hodnotu 0,05 nazýváme jako **hladinu významnosti**.

> SIGNIFICANCE LEVEL. The *significance level* is the largest acceptable probability of committing a type I error and is denoted by α, where $0 < \alpha < 1$.

- V našem případě testujeme H_0: $\mu = 120$ proti H_1: $\mu > 120$ na hladině významnosti $\alpha = 0{,}05$.
- Stanovme pro jaké hodnoty $T \equiv \bar{X}_3$ odmítáme H_0 na hladině významnosti $\alpha = 0{,}05$ ve prospěch H_1.
- Pokutujme každého řidiče, jehož $\bar{X}_3 \geq 121$, tedy odmítáme H_0 pokaždé když $T \geq 121$.
- Jak velká je pravděpodobnost spáchání chyby I druhu $P(T \geq 121)$?

Hladina významnosti

- Náhodná proměnná $T = \bar{X}_3$ má distribuci $N(120,4/3)$.
- Změnou proměnné na Z provedeme transformaci na $N(0,1)$: $Z = \frac{T-120}{2/\sqrt{3}}$
- Potom: $\mathrm{P}(T \geq 121) = \mathrm{P}\left(\frac{T-120}{2/\sqrt{3}} \geq \frac{121-120}{2/\sqrt{3}}\right) = \mathrm{P}(Z \geq 0.87)$
- Snadno zjistíme, že $\mathrm{P}(Z \geq 0.87) = 0{,}1922$.
- Tedy pravděpodobnost spáchání chyby I druhu je větší jak zvolená hladina významnosti $\alpha = 0{,}05$.
- Tedy v tomto případě neodmítáme H_0.
- Musíme tedy zvolit nové kritérium pro odmítnutí H_0.

Hladina významnosti

- Pokutujme každého řidiče, jehož $\bar{X}_3 \geq 122$, tedy odmítáme H_0 pokaždé když $T \geq 122$.
- Potom pravděpodobnost spáchání chyby I druhu bude $P(T \geq 122) = 0.0416$.
- Vidíme, že pravděpodobnost je menší jak zvolená hladina významnosti $\alpha = 0,05$ a proto hypotézu H_0 odmítáme.
- Hraničním případem je taková hodnota T, že platí: $P(T \geq c) = 0,05$. K nalezení c řešíme rovnici:

$$\mathrm{P}\left(Z \geq \frac{c-120}{2/\sqrt{3}}\right) = 0.05.$$

- Z tabelovaných hodnot Φ plyne, že $z_{0,05} = 1,645$.

Hladina významnosti

- Potom máme: $$\frac{c-120}{2/\sqrt{3}} = 1.645, c = 120 + 1.645 \cdot \frac{2}{\sqrt{3}} = 121.9.$$
- Pokud hladina významnosti je $\alpha = 0{,}05$, potom odmítáme hypotézu H_0 ve prospěch H_1 pokaždé, když $T = \bar{X}_3 \geq 121{,}9$.
- **Závěr**: jestliže průměrná zaznamenaná rychlost vozidla bude větší nebo rovna 121,9 km/h, pak pokutujeme řidiče.
- Toto rozhodovací pravidlo zaručuje, že nejvíce 5% řidičů bude pokutováno nespravedlivě (aniž by to tušili).
- **Hladina významnosti** je taková úroveň, pod kterou je ***p*-hodnota** dostatečně malá k odmítnutí H_0. Tedy pro pozorované hodnoty $T \geq 121{,}9$ *p*-hodnota je max. 0,05.

Kritický obor

- V předchozím příkladě množina K = [121,9, ∞) obsahující hodnoty T, pro které odmítáme H_0, se nazývá **kritický obor**. Hraniční hodnota 121,9 mezi odmítnutím a přijetím H_0 se nazývá **kritická hodnota**.

CRITICAL REGION AND CRITICAL VALUES. Suppose we test H_0 against H_1 at significance level α by means of a test statistic T. The set $K \subset \mathbb{R}$ that corresponds to all values of T for which we reject H_0 in favor of H_1 is called the *critical region*. Values on the boundary of the critical region are called *critical values*.

- Kritický obor je určen použitou hodnotou hladiny významnosti α a statistického testu T.
- Vždy musí platit $\mathrm{P}(T \in K) \leq \alpha$ v případě, že H_0 je pravdivé.
- Pravděpodobnost lze zapsat při využití notace s podmíněnou pravděpodobnosti jako: $\mathrm{P}(T \in K \mid H_0)$

P-hodnoty vs. kritický obor

- Nechť $T = \bar{X}_3 = 124$ -> naměřená hodnota spadá do kritického regionu -> odmítáme H_0 ve prospěch H_1.
- Spočítejme p-hodnotu pro $T = 124$:

$$\mathrm{P}(T \geq 124 \mid H_0) = \mathrm{P}\left(\frac{T-120}{2/\sqrt{3}} \geq \frac{124-120}{2/\sqrt{3}}\right) = \mathrm{P}(Z \geq 3.46) = 0.0003$$

- Tedy spočítaná p-hodnota je mnohem menší jak hladina významnosti $\alpha = 0{,}05$.
- Nyní můžeme vyslovit tvrzení:

$$t \in K \quad \Leftrightarrow \quad \text{the } p\text{-value corresponding to } t \text{ is less than or equal to } \alpha.$$

- Z výše uvedeného snadno nahlédneme, že pro **hladinu významnosti** musí platit: $\alpha = \mathrm{P}(T \geq c_\alpha | H_0)$, kde c_α je taková hodnota T, pro kterou je p-hodnota rovna α.

P-hodnoty vs. kritický obor

- Analogicky pak *p*-hodnota = $P(T \geq t|H_0)$.
- Rozhodnutí o odmítnutí H_0 můžeme udělat na základě porovnání c_α s t nebo *p*-hodnoty s α.
- *P*-hodnotu pak nazýváme jako **pozorovanou hladinou významnosti**.
- Koncept kritické hodnoty a *p*-hodnoty má svůj význam.
- Kritický obor s odpovídajícími kritickými hodnotami přesně specifikuje jaké hodnoty T mi vedou k odmítnutí H_0 na dané hladině významnosti α.
- Toto mohu udělat bez znalosti statistického souboru a počítání hodnoty t pomocí testu významnosti.
- Na druhou stranu:
  - *p*-hodnota mi představuje „sílu důkazů“, že pozorovaná hodnota t svědčí proti H_0
  - ale *p*-hodnota mi nespecifikuje všechny hodnoty T, které odmítají H_0

Diskrétní kritický obor

- Kritický obor K mohu konstruovat takový, že $P(T \in K \mid H_0)$ se rovná přímo α – viz předchozí příklad.
- Pokud je distribuce náhodné proměnné T diskrétní, pak to nemohu vždy udělat.
- Ukažme si to na jednoduchém příkladě.
- Chceme otestovat minci 1 EURO jestli je poctivá.
- Provedeme 10 hodů a X mi zaznamená kolikrát padl orel. X musí mít Bin(10,p) distribuci. Chceme zjisti, zdali se p liší od ½.
- Testujeme hypotézy H_0: $p = ½$ proti H_1: $p \neq ½$.
- Hodnota X bude test významnosti a $\alpha = 0{,}05$.
- Pro jaké hodnoty X odmítneme H_0 ve prospěch H_1?

Diskrétní kritický obor

- Pokud je H_0 pravdivá, pak E[X] = 10· ½ = 5. Tedy X blízko 5 jsou ve prospěch H_0.
- X blízko 10 -> p > ½ nebo X blízko 0 -> p < ½ -> odmítám H_0 ve prospěch H_1.
- Odmítneme H_0 ve prospěch H_1, pokud $X \leq c_l$ a $X \geq c_u$. Kritický obor pak je: $K = \{0, 1, \ldots, c_l\} \cup \{c_u, \ldots, 9, 10\}$
- c_l a c_u nazýváme jako levá a pravá kritická hodnota.
- Kritický region musí být co možná největší, ale zároveň splňovat podmínku:

$$\mathrm{P}(X \in K \mid H_0) = \mathrm{P}\left(X \leq c_l \mid p = \tfrac{1}{2}\right) + \mathrm{P}\left(X \geq c_u \mid p = \tfrac{1}{2}\right) \leq 0.05.$$

- Nepreferujeme žádnou část K, pak:

$$\mathrm{P}\left(X \leq c_l \mid p = \tfrac{1}{2}\right) \leq 0.025 \quad \text{and} \quad \mathrm{P}\left(X \geq c_u \mid p = \tfrac{1}{2}\right) \leq 0.025.$$

Diskrétní kritický obor

- Tabulka s hodnotami *F* pro Bin (10, ½).

k	$\mathrm{P}(X \leq k)$	k	$\mathrm{P}(X \leq k)$
0	0.00098	6	0.82813
1	0.01074	7	0.94531
2	0.05469	8	0.98926
3	0.17188	9	0.99902
4	0.37696	10	1.00000
5	0.62305		

- Hned vidíme, že $c_l = 1$
- Pak lze najít $c_u = 9$, protože

$$\mathrm{P}(X \geq 9) = 1 - \mathrm{P}(X \leq 8) = 1 - 0.98926 = 0.01074,$$
$$\mathrm{P}(X \geq 8) = 1 - \mathrm{P}(X \leq 7) = 1 - 0.94531 = 0.05469.$$

- Kritický obor tedy je: $K = \{0, 1, 9, 10\}$.
- Odpovídající chyba I druhu je:

$$\mathrm{P}(X \in K) = \mathrm{P}(X \leq 1) + \mathrm{P}(X \geq 9) = 0.01074 + 0.01074 = 0.02148$$

- Je navíc menší jak hladina významnosti.

Chyba II druhu

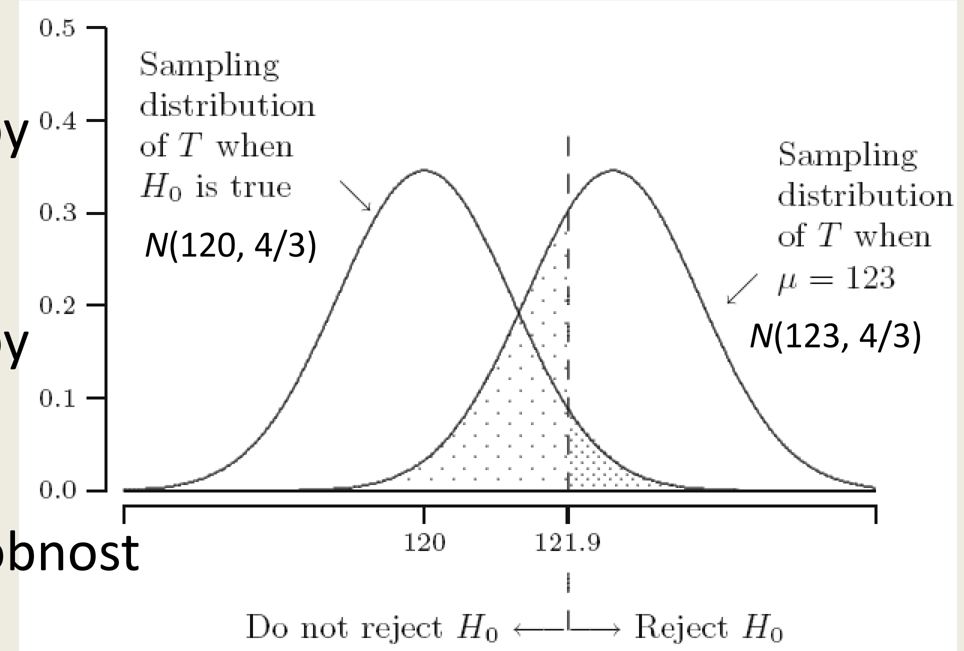
- Už víme, že nastavením hladiny významnosti α kontrolujeme pravděpodobnost, že spácháme chybu I druhu.
- V příkladu s měřením rychlosti – viz str. 37, rozhodovací kritérium odpovídá T = 121,9 km/h a α = 0,05.
- Jaká je pravděpodobnost spáchání chyby II druhu?
- Odpovídá to procentu řidičů, jejichž reálná rychlost je vyšší jak 120 km/h, ale nejsou pokutováni, protože naměřená $\bar{X}_3$ < 121,9 km/h.
- Např.: nechť reálná rychlost auta je μ = 125 km/h. $T = \bar{X}_3$ má distribuci $N(125, 4/3)$
- Chyba II druhu nastane pokud T < 121,9 km/h a pravděpodobnost :

$$\mathrm{P}(T < 121.9 \mid \mu = 125) = \mathrm{P}\left(\frac{T-125}{2/\sqrt{3}} < \frac{121.9-125}{2/\sqrt{3}}\right) = \Phi(-2.68) = 0.0036.$$

- Pokud μ = 123 km/h, pak pravděpodobnost:

$$\mathrm{P}(T < 121.9 \mid \mu = 123) = \mathrm{P}\left(\frac{T-123}{2/\sqrt{3}} < \frac{121.9-123}{2/\sqrt{3}}\right) = \Phi(-0.95) = 0.1711.$$

Chyba II druhu



- Plocha vpravo od 121,9 je pravděpodobnost spáchání chyby I druhu 5%.
- Plocha vlevo od 121,9 je pravděpodobnost spáchání chyby II druhu 17,11%.
- Posouváním distribuce *T* vpravo nebo vlevo měníme pravděpodobnost spáchání chyby II druhu.
- Pravděpodobnost spáchání chyby II druhu závisí na velikosti μ v alternativní hypotéze H_1: $\mu > 120$.
- Pravděpodobnost chyby I druhu je vždy maximálně α na rozdíl od pravděpodobnosti chyby II druhu, jejíž velikost může být libovolně blízko k $1 - \alpha$.

Intervaly spolehlivosti vs kritické hodnoty

- V příkladu na str. 35 jsme kritické hodnoty získali pomocí vzorce: $c_{0.05} = 120 + 1.645 \cdot \frac{2}{\sqrt{3}}$.
- Z přednášky 8 str. 41 víme, že pokud náhodná proměnná $\bar{X}_3$ má distribuci $N(\mu, 4/3)$, pak spodní hranice 95% jednostranného **intervalu spolehlivosti** pro μ je dána vztahem: $l_n = \bar{x}_3 - 1.645 \cdot \frac{2}{\sqrt{3}}$.
- Na první pohled to vypadá, že statistické testování hypotéz a konstruování intervalů spolehlivosti jsou dvě naprosto odlišné procedury, ve skutečnosti jsou úzce spjaty.
- Pro př. ze str. 30 H_0: $\mu = 120$ odmítáme ve prospěch H_1: $\mu > 120$ na hladině významnosti 0,05 právě tehdy když:

Intervaly spolehlivosti vs kritické hodnoty

$$\bar{x}_3 \geq 120 + 1.645 \cdot \frac{2}{\sqrt{3}}$$

neboli

$$\bar{x}_3 - 1.645 \cdot \frac{2}{\sqrt{3}} \geq 120$$

- Nebo také, když 120 není v 95% jednostranném intervalu spolehlivosti pro μ.
- Není to náhoda. Můžeme to zobecnit.
- Nechť pro nějaký parametr θ testujeme nulovou hypotézu H_0: $\theta = \theta_0$.
- Potom odmítáme H_0: $\theta = \theta_0$ ve prospěch H_1: $\theta > \theta_0$ na hladině významnosti α, tehdy a jen tehdy, pokud θ_0 není v jednostranném $(1 - \alpha)$ intervalu spolehlivosti pro θ.
- Stejná definice platí i pro H_1: $\theta < \theta_0$.
- Podobné definice platí i pro H_1: $\theta \neq \theta_0$ a dvoustranný interval spolehlivosti.

Intervaly spolehlivosti vs kritické hodnoty

- Tedy odmítáme H_0: $\theta = \theta_0$ ve prospěch H_1: $\theta \neq \theta_0$ na hladině významnosti α, tehdy a jen tehdy, pokud θ_0 není ve dvoustranném $(1 - \alpha)$ oboru spolehlivosti pro θ.
- **Def.**: $(1 - \alpha)$ **obor spolehlivosti** pro parametr θ je množina hodnot θ_0, pro které nulová hypotéza H_0: $\theta = \theta_0$ není odmítnuta na hladině α.
- Tyto vztahy platí jen v případě, že **náhodná proměnná**, ze které se konstruuje *interval spolehlivosti*, odpovídá **náhodné proměnné**, která tvoří *statistický test*.
- Např. není možné konstruovat interval spolehlivosti pro μ z $\bar{X}_n$ a zároveň jako statistický test nulové hypotézy použít odhadovou funkci pro μ ve tvaru výběrového mediánu Med_n.