Pst – přednáška 10: uvedení do úsudkové statistiky

Poslední tři týdny předmětu se týkají třetího oboru, který máme v tomto předmětu probrat, a sice úsudkové statistiky (inferential statistics) či statistického usuzování (statistical inference).

Objevují se zde obtížnost a algoritmy, které nelze vyvodit ze předchozích dvou oborů (popisné statistiky a pravděpodobnostních modelů), ikdyž pojmy ze předchozích dvou oborů jsou používány; je tedy důležitá zvýšená pozornost ze strany studenta. Situace je o to těžší, že odvození či zdůvodnění většiny postupů je natolik matematicky náročné, že v běžných učebnicích ani nebývá uvedeno, a na studentovi se žádá, aby postupy používal, aniž jim bez chyby rozumí. Pokusím se ovšem o výklad co možná nejsrozumitelnější a napojený na dosud užívané pojmy.

A. Chyby 1. druhu a 2. druhu ve statistickém rozhodování (skripta BM43-stare.pdf, str. 176–178)

Představte si rozhodování na jednoduchém soudu, kde existují jen dvě možnosti reality (realita 1: obžalovaný je vinen, realita 0: obžalovaný je nevinen) a dva verdikty soudu (verdikt 1: obžalovaný je vinen; verdikt 0: obžalovaný je nevinen). Kombinací zde dostaneme čtyři možnosti verdiktu vzhledem k realitě:

- všimněte si, že číslo 0 je spojeno se záporkou ne- u realit i u verdiktů ... to není náhoda, ale pomůcka k zapamatování!!

	realita 0: nevinen	realita 1: vinen
verdikt 0: nevinen	O.K.	chyba 2. druhu
verdikt 1: vinen	chyba 1. druhu	O.K.

Dvě ze těchto možností jsou dobré/správné, ale druhé dvě možnosti jsou chybné, protože realita neodpovídá verdiktu (verdikt neodpovídá realitě)

- většina lidí se shoduje na tom, že závažnější chyba je odsouzení nevinného, proto tato chyba dostala označení CHYBA PRVNÍHO DRUHU; dopouští se jí soudce, který je hodně přísný a odsoudí člověka při sebemenším podezření
- chyby 2. druhu (osvobození darebáka) se dopouští soudce benevolentní, který není pohnut k verdiktu VINEN ani vážnými důkazy

O velmi podobné rozhodování se bude jednat v úsudkové statistice ohledně chování náhodné veličiny X. Bude volit mezi dvěma hypotézami, H_0 (nějaká hodnota či veličina nezávisí na jiné veličině)
H_1 (— " — závisí na jiné veličině);
vzhledem k verdiktům mohou nastat čtyři možnosti:

	realita H_0: ...nezávisí na...	realita H_1: ...závisí na...
verdikt 0: H_0 nezamítáme	O.K.	chyba 2. druhu
verdikt 1: H_0 zamítáme ve prospěch platnosti H_1	chyba 1. druhu	O.K.

tedy chybou 1. druhu je: H_0 zamítáme, třebaže platí!
chybou 2. druhu je: H_0 nezamítáme, třebaže neplatí.

Protože závažnější je chyba 1. druhu, statistické testy a intervaly spolehlivosti pracují zejména s ní a označují α = pst výskytu chyby 1. druhu ve statistickém usuzování.

B. Statistický test střední hodnoty binomického rozdělení psti → ve skriptech není dobře popsán

Celou problematiku budu vysvětlovat na příkladech, které čísluji od začátku, tj. začneme

Příkladem 1. Expert podniku tvrdí, že o nový model výrobku bude mít zájem 20% zákazníků. Při průzkumu u 600 zákazníků jich o nový model projevilo zájem 135. Na hladině významnosti $\alpha = 0{,}05$ testujte hypotézu H_0, že zájem o výrobek je ZHRUBA stejný, jak expert předpokládal.

řešení: (hladina významnosti α je právě pst výskytu chyby 1. druhu – tuto pst volíme na začátku, před provedením úsudku – je to vlastně nastavení jakési přesnosti/tolerance našeho úsudku; α se obvykle volí $\leq 0{,}05$)

krok 1: budeme rozhodovat mezi hypotézami (X = počet zájemců ze 600 lidí)

H_0 : zájem nezávisí na novém typu výrobku a je stále 20% : $EX = 120$

H_1 : zájem závisí na novém typu a změnil se od dřívějška : $EX \neq 120$

(je vidět, že v anketě $X = 135$, tj. zájem je „trochu větší než" 20 procent ze 600 lidí, otázkou ovšem kritická mez: můžeme pořád tvrdit, že zájem je ZHRUBA u 20% lidí?
To „ZHRUBA" zde znamená střední hodnotu EX : jaký je „průměrný" zájem?

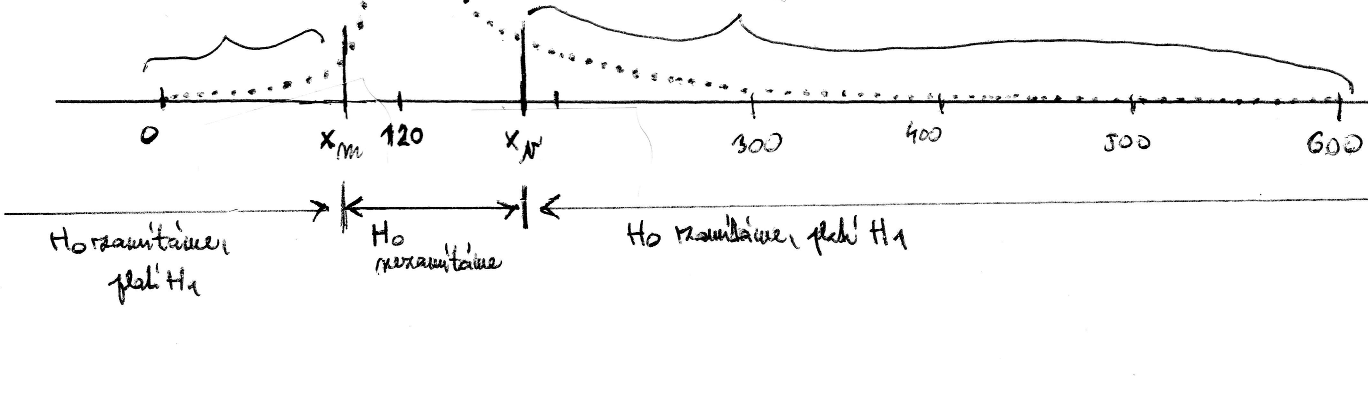
krok 2: Kritériem našeho rozhodování je veličina X = počet zájemců o nový výrobek ze 600 lidí.
V našem jediném měření $X = 135$, ale bude tomu tak vždy? Jaký je průměrný zájem?

krok 3: Popišme chování veličiny X za předpokladu, že platí H_0; to vlastně znamená, že každý člověk má 20%-ní šanci, že ho nový výrobek upoutá, nezávisle na ostatních lidech.
Čili
$$X \overset{H_0}{\sim} Bi(N = 600,\ p = 0{,}2)$$
X tedy může nabýt hodnot $0, 1, 2, \ldots, 599, 600$ s pstí
$$p(k) = \binom{600}{k} \cdot 0{,}2^{k} \cdot 0{,}8^{600-k}$$

krok 4: pro předem zvolené $\alpha = 0{,}05$ najdeme kritické meze našeho rozhodování: když počet zájemců bude podstatně menší než 120 nebo podstatně větší než 120, zamítneme H_0 a prohlásíme, že platí H_1. Vyjdeme ze grafu pstní funkce $p(k)$:
máme zde 601 pstí, jejichž součet je roven 1.
„USEKNEME" na obou stranách/koncích tolik hodnot $\underline{k}$, aby:

součet $\sum_{0}^{x_m} p(k) \doteq \frac{\alpha}{2} = 0{,}025$

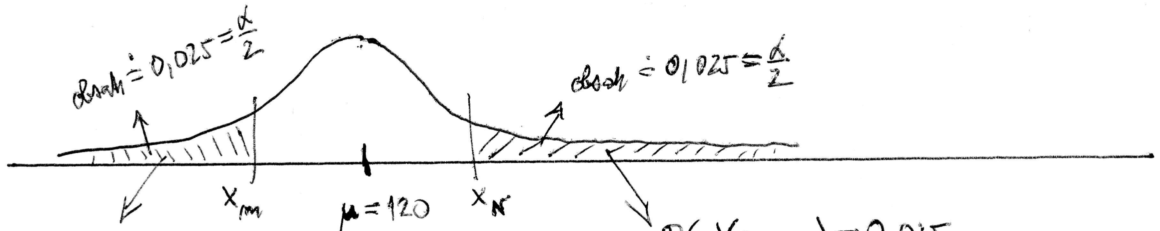
součet $\sum_{x_N}^{600} p(k) \doteq \frac{\alpha}{2} = 0{,}025$



Nyní máme dvě možnosti: a) pracovat s binomickým rozdělením, ovšem při hledání x_v musíme sečítat sumu se 400 nebo více členy, to je hodně pracné

b) využít faktu z předpokladu 9, že Bi lze dobře aproximovat = přibližně popsat pomocí normálního rozdělení s korekcí – pak bychom hledanou mez x_m, x_v našli pomocí jednoho výpočtu – PROVEDEME NYNÍ:

$$Bi(N=600, p=0{,}20) \sim No(\mu = N\cdot p = 120, \sigma^2 = N\cdot p(1-p) = 96)$$



$P(X < x_m) = 0{,}025$

$P(X \in (-\infty; x_m)) = 0{,}025$

$F(x_m) - \underbrace{F(-\infty)}_{=0} = 0{,}025$

0 … v minus nekonečnu ještě není nakumulováno nic

dle předpokladu 9: výpočet převedeme na $No(\mu=0, \sigma^2=1)$:

$$\Phi\Big(\underbrace{\frac{x_m - 120}{\sqrt{96}}}_{u_m}\Big) = 0{,}025$$ … hledáme v tabulce, pro které u_m platí $\Phi(u_m) = 0{,}025$

$P(X > x_v) = 0{,}025$

$P(X \in (x_v; \infty)) = 0{,}025$

$\underbrace{F(\infty)}_{=1} - F(x_v) = 0{,}025$

1 … v plus nekonečnu už budeme mít nakumulováno všechno, co lze naměřit

$$0{,}975 = F(x_v) = \Phi\Big(\underbrace{\frac{x_v - 120}{\sqrt{96}}}_{u_v}\Big)$$ … hledáme v tabulce, pro které u_v platí $\Phi(u_v) = 0{,}975$ … nacházíme $u_v \doteq 1{,}96$

$u_m = -1{,}96$ protože hustota $No(0;1)$ je sudou funkcí, hodnoty u_m, u_v se liší jen o znaménko

tedy $\dfrac{x_m - 120}{\sqrt{96}} = -1{,}96$

$x_m = 120 - 1{,}96\cdot\sqrt{96} \doteq \underline{100{,}796}$

tedy $\dfrac{x_v - 120}{\sqrt{96}} = 1{,}96$

$x_v = 120 + 1{,}96\cdot\sqrt{96} \doteq \underline{139{,}204}$

<u>krok 5</u>: rozhodnutí statistického testu:

$x = 135 \in \langle x_m; x_v\rangle = \langle 100{,}796; 139{,}204\rangle$, tedy H_0 nezamítáme.

$\boxed{H_0: \text{Zájem o nový výrobek je zhruba u 20\% zákazníků}}$

(neprokázalo se, že by zájem o nový výrobek byl statisticky významně jiný než u 20% zákazníků, tj. zájem o nový výrobek není statisticky významně nižší nebo vyšší než 20%).

Pozn.:
H_0 bychom zamítli až v případě, kdyby ze 600 lidí mělo zájem 100 a méně lidí nebo 140 a více lidí.

C. Statistický test střední hodnoty průměru ze normálního rozdělení

Příklad 2 (test typu μ = konstanta)

Je známo, že počet bodů na testu z matematiky v pololetí maturitního ročníku má u náhodně vybraného studenta přibližně normální rozdělení pro $\mu = 500$ bodů, $\sigma^2 = 10\,000$ (tj. $\sigma = 100$ bodů).

Firma KAPPA sestavila program INTEL, aby zlepšila znalosti matematiky NAVÍC k výuce ve škole, a náhodně jej zaslala **25** studentům, kteří po půlroce jeho užívání dosáhli v testu výsledku $\bar{X} = 540$ bodů (průměr 25 bodových výsledků). Stačí toto zlepšení bodového hodnocení na závěr, že půl roku používání programu NAVÍC K VÝUCE statisticky významně zlepšuje výsledek testu?

Řešení: provedeme podobný test jako v příkladu 1 s tím rozdílem, že hypotéza H_1 bude v tzv. jednostranné variantě, tj. nikoli $\mu \neq const$, ale $\mu > const$ ... můžeme totiž podle způsobu výuky předpokládat, že procvičující program je NAVÍC K VÝUCE, tj. výsledek písemky může jenom zlepšit.

krok 1: budeme rozhodovat mezi hypotézami ($\bar{X} = \frac{1}{25}\sum_1^{25} X_i$, kde výsledek X_i každého studenta má rozdělení $No(\mu = 500, \sigma^2 = 10\,000)$)

H_0: výsledek testu nezávisí na použití programu, tj. $\boxed{E\bar{X} = 500 \text{ bodů}}$ i při využití programu

H_1: výsledek testu závisí na užití programu $\boxed{E\bar{X} > 500 \text{ bodů}}$

krok 2: Naším rozhodovacím kritériem je veličina $\bar{X} := \frac{1}{25}\cdot\sum_1^{25} X_i$ ... průměrné bodové hodnocení testu u 25 studentů.

krok 3: popíšeme chování veličiny $\bar{X}$ za předpokladu, že platí H_0:

$$X_i \overset{H_0}{\sim} No(\mu = 500 \text{ bodů}, \sigma^2 = 10\,000)$$

$$\bar{X} \overset{H_0}{\sim} No\Big(\mu_{\bar{X}} = E\bar{X} = E\Big(\frac{1}{25}\cdot\sum_1^{25} X_i\Big) = \frac{1}{25}\cdot\sum_1^{25} \underbrace{EX_i}_{\mu} = \frac{1}{25}\cdot\sum \mu = \frac{25\mu}{25} = \mu = 500,$$

$$\sigma^2_{\bar{X}} = D\Big(\frac{1}{25}\cdot\sum_1^{25} X_i\Big) = \frac{1}{625}\cdot\sum_1^{25} DX_i = \frac{1}{625}\cdot\sum_1^{25} 10\,000 = \frac{25\cdot 10\,000}{625} = \frac{10\,000}{25} = 400 = \frac{\sigma^2}{N}$$

při vytýkání konstanty z argumentu rozptylu ji musíme umocnit na druhou, protože rozptyl vyjadřuje kvadratickou odchylku od průměru

vysvětlení vzorce $\boxed{\sigma^2_{\bar{X}} = \frac{\sigma^2}{N}}$ → jedno měření má rozptyl $\sigma^2 = 10\,000$, ovšem průměr $25 = N$ hodnot bude mít menší rozptyl než jedna hodnota, tento rozptyl průměru je menší a menší pro rostoucí počet měření, ze kterých průměr počítáme: $\sigma^2_{\bar{X}} = \frac{\sigma^2}{N}$

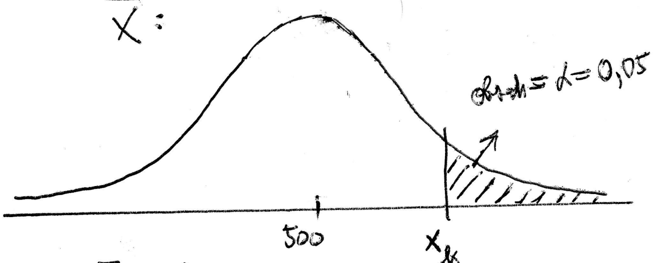
sigma na druhou s indexem $\bar{X}$ = rozptyl veličiny $\bar{X}$

Tedy $\bar{X} \overset{H_0}{\sim} No(\mu_{\bar{X}} = 500, \sigma^2_{\bar{X}} = 400$, tj. $\sigma_{\bar{X}} = \sqrt{400} = 20)$

$\mu_{\bar{X}} = \mu$ $\qquad$ $\sigma^2_{\bar{X}} = \frac{\sigma^2}{N}$

krok 4: pro předem zvolené $\alpha = 0{,}05$ najdeme kritické meze našeho rozhodování:

$\overline{X}$:



obsah $= \alpha = 0{,}05$ u jednostranného testu „usekáváme" pouze na té straně od konstanty 500, na kterou „směřuje" nerovnost: $H_1: E\overline{X} > 500$

$P(\overline{X} > x_k) = 0{,}05$

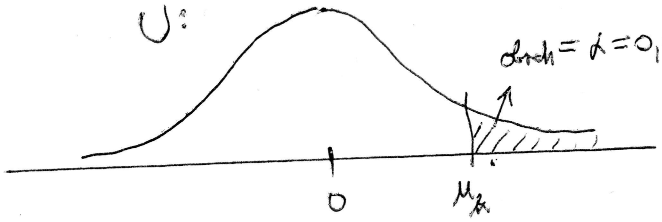
$P(\overline{X} \in (x_k; \infty)) = 0{,}05$

$\underbrace{F(\infty)}_{=1} - F(x_k) = 0{,}05$

1 … v nekonečnu už máme nakumulovány všechny hodnoty veličiny, které lze naměřit

$0{,}95 = F(x_k) = \Phi\left(\underbrace{\frac{x_k - 500}{20}}_{= u_k}\right)$

U:



obsah $= \alpha = 0{,}05$

$0{,}95 = \Phi(u_k)$

$1{,}64 \doteq u_k$

$x_k = 500 + 1{,}64 \cdot 20 \doteq 532{,}8$

krok 5: rozhodnutí: $\overline{X} = 540 \notin (-\infty; 532{,}8) \Rightarrow H_0$ zamítáme, prokázalo se, že program statisticky významně zlepšuje výsledek.

Poznámka: $u_k = 1{,}64$ je vlastně 0,95-kvantil našeho rozdělení U.
$x_k = 532{,}8$ je vlastně 0,95-kvantil našeho rozdělení $\overline{X}$ } zkrátka a dobře, stejně jako u popisné statistiky jsme k výpočtu kvantilu užili kumulativní relativní četnosti,

nyní u náhodných veličin na VŠ úrovni, když máme k dispozici distribuční funkci, užíváme této distribuční fce

x_k je vlastně $x_{0,95}$ neboli 0,95 kvantil: taková hodnota veličiny $\overline{X}$, že $P(\overline{X} < x_{0,95}) = 0{,}95$

Příklad 3. Konkurenční firma DELTA vyrábí program KILL využívající počítače při výuce matematiky. Ředitel jedné ZŠ povolal jako experta jednoho absolventa pedagogické fakulty MU pro matematiku na 2. stupni, aby zjistil, který program je lepší, zda INTEL (z příkladu 2) nebo KILL.

Expert rozeslal INTEL náhodně vybraným $N_1 = 25$ studentům, jejich výsledek na testu … $\overline{X}_1 = 600$ bodů
KILL náhodně vybraným $N_2 = 30$ studentům, jejich výsledek na testu … $\overline{X}_2 = 575$ bodů

Lze říci na hladině významnosti $\alpha = 0{,}05$, že INTEL je statisticky významně lepší než KILL?

Řešení: musíme použít oboustranný stat. test, $H_1: \mu_1 \neq \mu_2$, protože je možné, že INTEL je náhodou s lepším výsledkem a mohl by vést i k horším výsledkům.

krok 1: budeme rozhodovat mezi hypotézami $\left(\bar X_1 = \frac{1}{25}\cdot\sum_1^{25} x_i\,,\ \bar X_2 = \frac{1}{30}\cdot\sum_1^{30} x_i\right)$

H_0: výsledek testu nezávisí na programu: $\mu_{\bar X_1} = \mu_{\bar X_2}$

H_1: výsledek testu závisí na programu: $\mu_{\bar X_1} \neq \mu_{\bar X_2}$

krok 2: kritériem našeho statistického testu bude rozdíl průměrů $\bar X_1 - \bar X_2$.

krok 3: popíšeme chování veličiny $\bar X_1 - \bar X_2$ za předpokladu, že platí H_0:

$$\bar X_1 - \bar X_2 \overset{H_0}{\sim} No\left(\mu_{\bar X_1 - \bar X_2} = \mu_{\bar X_1} - \mu_{\bar X_2} \overset{H_0}{=} \underline{\underline{0}}\right.$$

$$\sigma^2_{\bar X_1 - \bar X_2} = D(\bar X_1 - \bar X_2) = D\bar X_1 + D\bar X_2 = D\left(\frac{1}{25}\cdot\sum_1^{25} x_i\right) + D\left(\frac{1}{30}\cdot\sum_1^{30} x_i\right) =$$

hodnota 10 000 je rozptyl jednoho měření x_i, tato informace je vzata z př. 2 a platí i zde v př. 3

rozptyl veličiny, která vznikne rozdílem dvou jiných veličin, se nikdy neodčítá, ale sčítá – plyne to i z faktu, že rozptyl funguje jako kvadratická odchylka, tj. (−1) se při vytknutí umocňuje na druhou

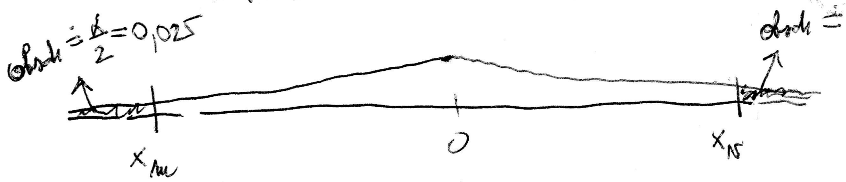
$$= \frac{10\,000}{25} + \frac{10\,000}{30} = 733{,}3\bar 3 \implies \sigma_{\bar X_1 - \bar X_2} = \sqrt{733{,}\bar 3} = \underline{\underline{27{,}08}}$$

využijeme vzorce

$$D(\bar X) = \frac{\sigma^2}{N}$$

$$\left(\text{tedy } \bar X_1 - \bar X_2 \overset{H_0}{\sim} No\left(\mu_{\bar X_1 - \bar X_2} = 0\,,\ \sigma^2_{\bar X_1 - \bar X_2} = 733{,}33\right)\right)$$

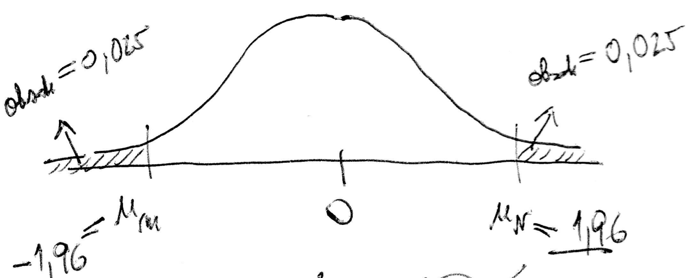
krok 4: pro předem zvolené $\alpha = 0{,}05$ najdeme kritické meze našeho rozhodování: $\bar X_1 - \bar X_2$:



$$\frac{x_m - 0}{27{,}08} = -1{,}96 \qquad \frac{x_v - 0}{27{,}08} = +1{,}96$$

$$x_m \doteq \underline{\underline{-53{,}08}} \qquad x_v \doteq \underline{\underline{53{,}08}}$$

U:



hodnoty získáme stejně jako v příkladě 1

krok 5: rozhodnutí: $\bar X_1 - \bar X_2 = 600 - 575 = 25 \in (-53{,}08; 53{,}08)$, a tedy H_0 nezamítáme, oba programy mají průměrně stejný dopad, rozdíl mezi průměrnými výsledky není statisticky významný!