4_Pravděpodobnost a pravděpodobnostní rozložení

Co je pravděpodobnost

- Způsob jak kvantifikovat nejistotu
- =pravděpodobnost/šance výskytu možných výsledků náhodného fenoménu (např.pokus/experiment/výběr) např. výhra v loterii, při hodu kostkou padne 6, bude pršet?….
- Možné výsledky jsou známy, ale je nejisté který nastane
- Příklady náhodného procesu: loterie, ruleta, hod kostkou/mincí, ale i výběr vzorku

Dlouhodobé chování náhodných jevů a „objektivní“ definice pravděpodobnosti

- Se 4 hody mincí nebude překvapivé dostat 4 pany, se 100 hodů velice překvapivé dostat 100 panen
- s narůstajícím počtem pokusů/observací nabývá proporce výskytu daného jevu očekávaných hodnot – tato proporce v dlouhodobém horizontu vytváří základ pro definici pravděpodobnosti…
  - …$p(A) = n(A) / n(S)$
  - …kdy pravděpodobnost (p) konkrétního výsledku odpovídá proporci (relativní četnosti) výskytu tohoto výsledku v dlouhodobém horizontu

Stanovení pravděpodobnosti

- 1. Definujeme základní množinu/prostor (S)
  - *= zjistíme sadu **všech možných výsledků***
  - ***Např. kostka = 1, 2, 3, 4, 5, 6***
  - ***Hod mincí 2 x = PP,PO,OP,OO***
  - ***Odpovědi na 3 otázky každá o 2 odpovědích (správně=A/chybně=N) = 8 možných výsledků (2*2*2)=AAA,AAN,ANA,ANN,NAA,NAN,NNA,NNN***

- 2. Definujeme podmnožinu základního prostoru = jev/jevy
  - *Sada/skupina výsledků – např. jev „lichá čísla“=1,3,5, nebo jev „studenti odpověděli alespoň 2 správně“=AAN, ANA, NAA, AAA*
  - *Každý výsledek i jev má určitou (p)*
  - *2 základní pravidla:*
    - *1. (p) každého jednotlivého výsledku je v rozmezí 0 až 1*
    - *2. Součet (p) všech jednotlivých výsledků (pozor ne však jevů, mezi kterými může být průnik(jeden výsledek obsažený ve více jevech)= 1*
  - *Pokud jsou všechny výsledky stejně pravděpodobné, pak p(jevu A) = počet výsledků obsahujících jev A / počet všech možných výsledků v prostoru*
    - *Např. p (lichá čísla)=3/6=1/2=0.5*
    - *P (studenti alespoň 2 správně) = 4/8 = 1/2 = 0.5*

Výpočet pravděpodobností z kontingenční tabulky

- Kont.t. ukazuje četnosti kombinací kategorií dvou kategorických proměnných

Příjem	Prošlo kontrolou	Neprošlo kontrolou	Celkem
Pod 200tis	1260	132147	133407
200tis-1mil	131	4311	4442
Více než 1mil	22	371	393
Celkem	1413	136829	138242

- Podmíněné proporce = 22/393 = 0,05
- Nepodmíněné/marginální proporce = 1413/138242= 0,01

Příjem	Prošlo kontrolou	Neprošlo kontrolou	Celkem
Pod 200tis	1260	132147	133407
200tis-1mil	131	4311	4442
Více než 1mil	22	371	393
Celkem	1413	136829	138242

- Základní prostor = všechny možné výsledky = 6
- Pravděpodobnost že plátce daně je kontrolován = 1413/138242 = 0,01
- Pravděpodobnost příjmu nad 1 mil. = 393/138242=0,003

Základní pravidla pro výpočet pravděpodobností dvou událostí

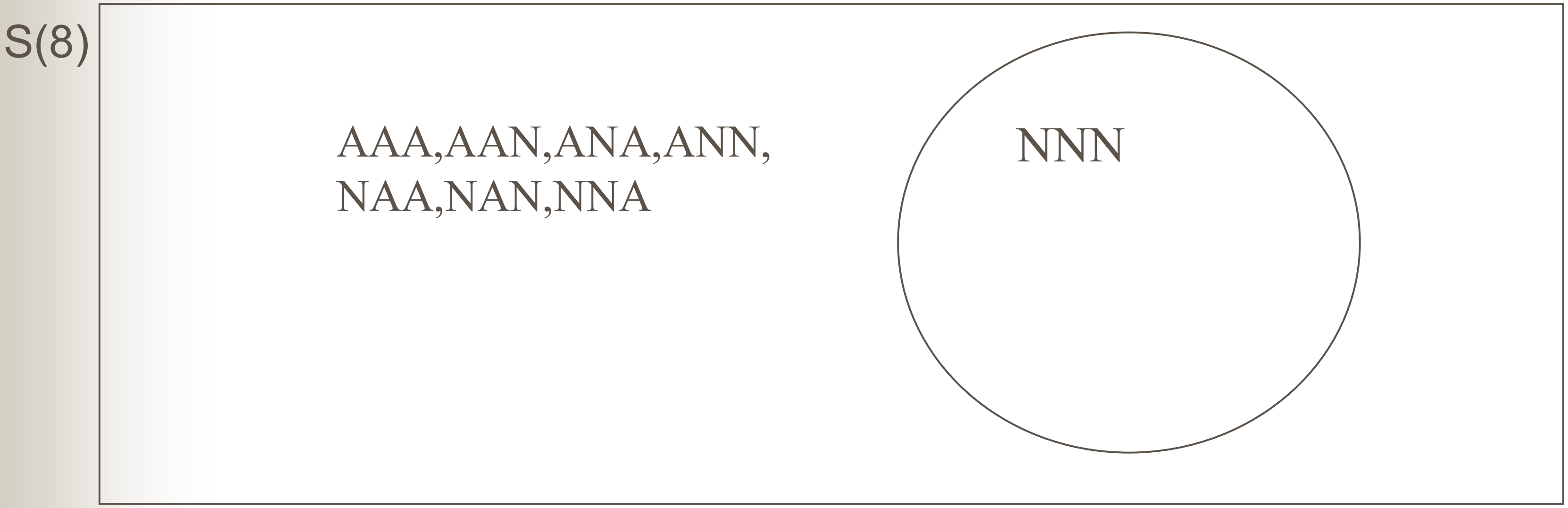
- Některé jevy jsou vyjádřeny jako výsledky které
  - a) nejsou obsaženy v jiných jevech = DOPLNĚK, VZÁJEMNĚ SE VYLUČUJÍCÍ JEVY
  - b) jsou v jednom jevu a zároveň i v druhém jevu = PRŮNIK
  - c) jsou v jednom jevu nebo v jiném = SJEDNOCENÍ

a) Doplněk

- Doplněk jevu A obsahuje všechny výsledky základního prostoru které nejsou v jevu A
- Součet pravděpodobností „A“ a „ne A“=1

  a proto p(~A) = 1 – p(A)
- Př. kontingenční tabulka: Jev A=„příjem 1mil a méně“ je doplňkem jevu B=„příjem nad 1mil“

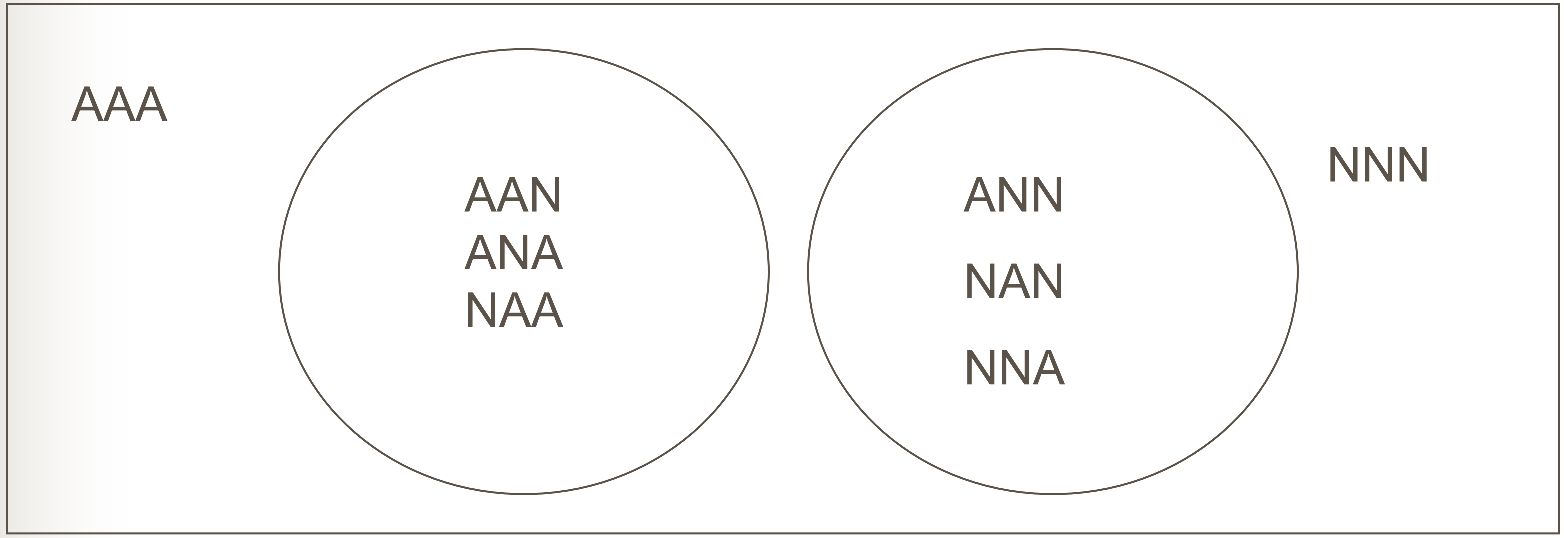
  P(A)=1 – p(B) = 1 - 0,003 = 0,997
- Př. studenti: P(nejméně jednu otázku správně) = 1 – p(žádná správně) = 1 – 1/8 = 7/8 = 0,875



a) Vzájemně se vylučující jevy

- = Jevy které nesdílejí žádný výsledek
- Např. Jev X =„právě 1 otázku správně“ a jev Y =„právě 2 otázky správně“ jsou vylučující se (oproti tomu žádný z jevů není vylučující se s jevem Z =„první otázka správně“, neboť tento jev má výsledky společné s X i Y

S=8



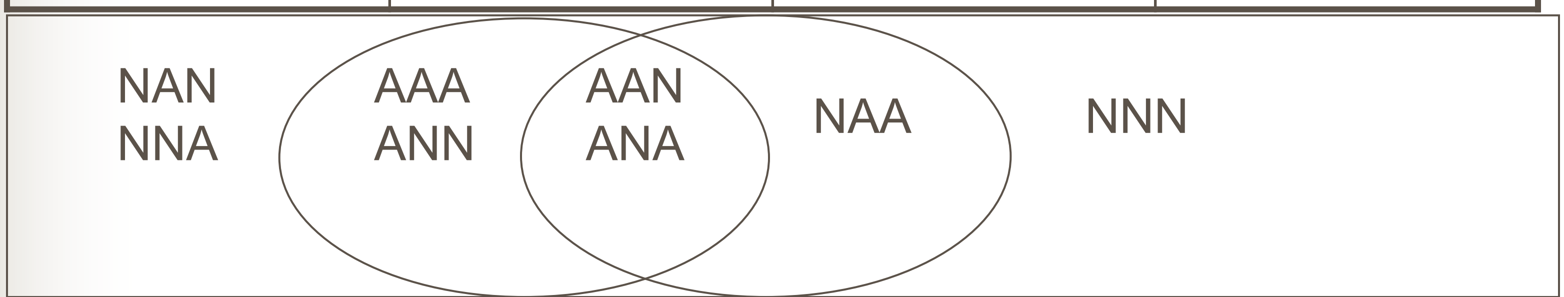
b) Průnik (A a B)

- = jev kdy nastane více jevů zároveň
- Je složen z výsledků které jsou zároveň v jevu A i B
- P(A a B) = p(A) * p(B) pokud jsou jevy nezávislé a p(B)*p(A|B) pokud jsou jevy závislé
- Př. studenti: průnikem jevu A = student odpoví první otázku dobře a jevu B=student odpoví 2 otázky dobře je jev C = AAN, ANA = 2/8 = 0,25
  - Užití vzorce: 3/8 * 2/3 = 2/8 =0,25
- Př. daně: průnik jevu „plátce je kontrolován“ a jevu „příjem nad mil.“ = 22/138242=0,0002
  - užití vzorce….vyzkoušejte si!☺

S=6

Příjem	Prošlo kontrolou	Neprošlo kontrolou	Celkem
Pod 200tis	1260	132147	133407
200tis-1mil	131	4311	4442
Více než 1mil	22	371	393
Celkem	1413	136829	138242

S=8



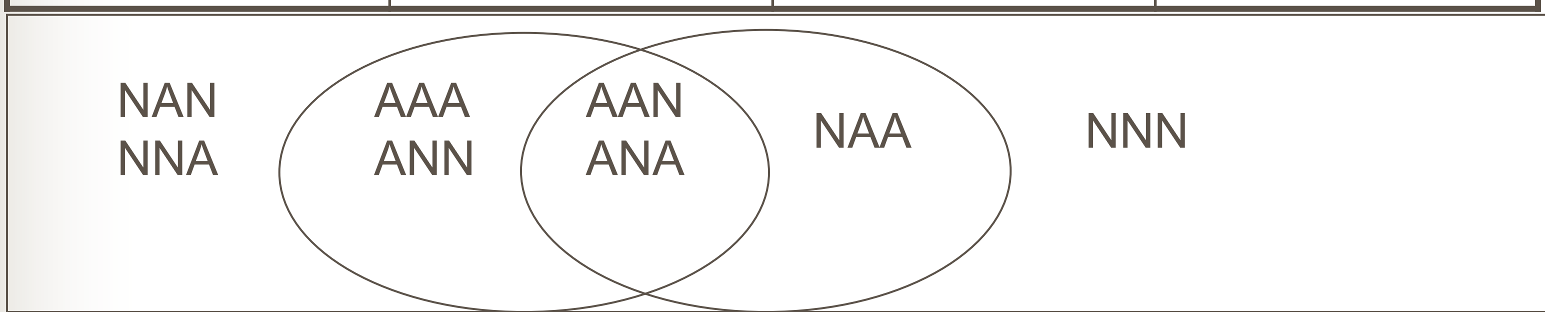
c) Sjednocení (A nebo B)

- = sjednocení A a B je složeno z výsledků které jsou v A nebo v B nebo v obou jevech
- P (A nebo B) = p(A) + p(B) – p(A a B)
- Př. daně: sjednocení jevu „plátce je kontrolován“ a jev „příjem nad mil.“ = (1260+131+22+371)/138242=0,013
  - užití vzorce: 1413/138242 + 393/138242 – 22/138242=0,013
- Př. studenti: sjednocením jevu A = student odpoví první otázku dobře a jevu B=student odpoví 2 otázky dobře je jev C = AAA, AAN, ANA, ANN, NAA = 5/8 = 0,625
  - Užití vzorce: 4/8 + 3/8 – 2/8 = 5/8 = 0,625

S=6

Příjem	Prošlo kontrolou	Neprošlo kontrolou	Celkem
Pod 200tis	1260	132147	133407
200tis-1mil	131	4311	4442
Více než 1mil	22	371	393
Celkem	1413	136829	138242

S=8



Závislé vs. nezávislé pokusy

- Pokusy jsou nezávislé pokud to co se stane v jednom pokusu neovlivňuje co nastane v jakémkoli jiném pokusu
  - Pak p(A a B) = p(A)*p(B)
  - Př. basket: hráč hází 2x na koš, p(koš)=0,8, jaká je (p) že dá oba koše?: p(A a B) = 0,8*0,8=0,64. KK=0,64, KO=0,8*0,2=0,16, OK=0,2*0,8=0,16 a OO=0,2*0,2=0,04. Bez ohledu na to zda první koš dal nebo ne, zůstává pro druhý koš pravděpodobnost stejná tedy 0,8.
- Naopak závislé pokusy jsou tehdy, pokud výsledek prvního ovlivňuje výsledek druhého
  - Př. studenti odpovídají na dvě otázky buďto správně (A) nebo špatně (N). Jev 1A=„první otázka dobře“ a jev 2A=„druhá otázka dobře“. Jev 1A=AA,AN=0,05+0,58=0,63. Jev 2A=AA,NA=0,58+0,11=0,69.
    - Pokud by tyto jevy byly nezávislé pak p(1A a 2A) = p(1A) * p(2A) = 0,63 * 0,69 = 0,43. Ve skutečnosti však p(1A a 2A)=0,58, jevy jsou tedy závislé.
    - Interpretace: Pokud student odpověděl první otázku správně, má vyšší pravděpodobnost, že odpověděl i druhou správně, než člověk který první neodpověděl správně.

	2A	2N	celkem
1A	0,58	0,05	0,63
1N	0,11	0,26	0,37
Celkem	0,69	0,31	1

Podmíněná pravděpodobnost

- = pravděpodobnost že nastane jev A když víme, že výsledek se nachází v nějaké konkrétní části základního prostoru
- Podmíněná pravděpodobnost jevu A pokud nastal jev B je rovna proporci výsledků v průniku A a B z celkového počtu výsledků v B, tedy P(A | B) = p (A a B) / p(B)
  - Např. pravděpodobnost kontroly daňového přiznání (A) když patřím do příjmové skupiny nad 1mil (B).
    - P(A | B) = p (A a B) / p(B) = 0,0002 / 0,0029 = 0,07
  - Pravděpodobnost správné odpovědi na druhou otázku když vím že jsem správně odpověděl první otázku
  - Pravděpodobnost koše při druhém hodu když jsem dal koš v prvním hodu

Příjem	*Prošlo kontrolou*	*Neprošlo kontrolou*	*Celkem*
Pod 200tis	0,0091 (1260)	0,9559 (132147)	0,9650 (133407)
200tis-1mil	0,0009 (131)	0,0312 (4311)	0,0321 (4442)
Více než 1mil	0,0002 (22)	0,0027 (371)	0,0029 (393)
Celkem	0,0102 (1413)	0,9898 (136829)	1 (138242)

Tabulka podmíněných pravděpodobností

Příjem	*Prošlo kontrolou*	*Neprošlo kontrolou*	*Celkem*
Pod 200tis	0,01 (1260)	0,99 (132147)	1 (133407)
200tis-1mil	0,03 (131)	0,97 (4311)	1 (4442)
Více než 1mil	0,07 (22)	0,93 (371)	1 (393)

Statistická nezávislost

- Jevy jsou statisticky nezávislé, pokud p(A|B) = P(A) tedy pokud pravděpodobnost že člověk prošel kontrolou je stejná jako pravděpodobnost že člověk prošel kontrolou pokud patří např. do nejvyšší příjmové skupiny
- Z předchozího snímku víme, že P(A | B) = 0,07, zatímco P(A) =0,01. Jevy tedy nejsou nezávislé – je mezi nimi souvislost/vztah
- Ze znalosti o průniku již víme, že jevy jsou také statisticky nezávislé pokud p(A a B) = p(A)*p(B). Z předchozího snímku víme, že p(A a B)=0,0002 a p(A)*p(B)=0,000029. Jevy tedy nejsou nezávislé.
- V praxi se častěji užívá třetí způsob ověření statistické nezávislosti pomocí podmíněných pravděpodobností: jevy jsou nezávislé pokud p(A|B1) = p(A|B2) , tedy když podmíněné pravděpodobnosti „kontroly“ se u jednotlivých příjmových podskupin neliší.
  - Z tabulky vidíme, že tomu tak není. Čím větší příjem člověk má, tím je větší pravděpodobnost, že bude kontrolován ((p) stoupá z 0,01 přes 0,03 na 0,07. P(„prošel kontrolou“ | „prijem nad 1mil“)=0,07. P(„prosel kontrolou“ | „prijem 200 až 1mil“)=0,05 a P(„prosel kontrolou“ | „prijem pod 200“)= 0,01)

Příjem	*Prošlo kontrolou*	*Neprošlo kontrolou*	*Celkem*
Pod 200tis	**0,01 (1260)**	0,99 (132147)	1 (133407)
200tis-1mil	**0,03 (131)**	0,97 (4311)	1 (4442)
Více než 1mil	**0,07 (22)**	0,93 (371)	1 (393)

Ukázka stat. nezávislosti 2

Příjem	*Prošlo kontrolou*	*Neprošlo kontrolou*	*Celkem*
Pod 200tis	0,01 (1364)	0,955 (132043)	0,965 (133407)
200tis-1mil	0,0003 (45)	0,032 (4397)	0,0323 (4442)
Více než 1mil	0,00003 (4)	0,003 (389)	0,00303 (393)
Celkem	0,01 (1413)	0,99 (136829)	1 (138242)

Příjem	*Prošlo kontrolou*	*Neprošlo kontrolou*	*Celkem*
Pod 200tis	**0,01 (1364)**	0,99 (132043)	1 (133407)
200tis-1mil	**0,01 (45)**	0,99 (4397)	1 (4442)
Více než 1mil	**0,01 (4)**	0,99 (389)	1 (393)

Ukázka stat.nezávislosti

- Znovu př. basket: hráč (a nebo např. 64 hráčů) hází 2x na koš…..

	2koš	*2mimo*	*celkem*
1koš	0,64 (41)	0,16 (10)	0,8 (51)
1mimo	0,16 (10)	0,04 (3)	0,2 (13)
celkem	0,8 (51)	0,2 (13)	1 (např. 64)

Vypočteme podmíněné pravděpodobnosti…

	2koš	*2mimo*	*celkem*
1koš	**0,8 (41)**	0,2 (10)	1 (51)
1mimo	**0,8 (10)**	0,2 (3)	1 (13)

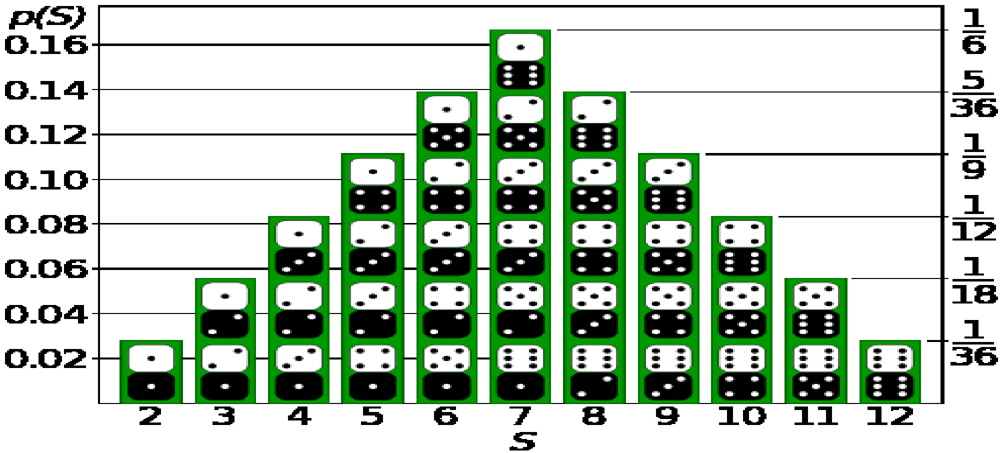
- Podmíněné pravděpodobnosti že hráč dá druhý koš když dal první p(2koš|1koš) = 0,64/0,8=0,8 a že dá druhý koš když první nedal p(2koš|1mimo) = 0,16/0,2=0,8 se rovnají - jevy „2koš“ a „1koš“ jsou tedy nezávislé. Jinými slovy, hráčova úspěšnost při druhém hodu není ovlivněna jeho úspěšností v hodu prvním.

Rozložení pravděpodobností

- Distribuce (p-tí) výskytů všech možných výsledků náhodného procesu (náhodný výběr, experiment) = (p-stní) rozložení náhodné proměnné
- (p-ti) všech možných výskytů se sčítají do 1 = součet (p-tí) v rozložení (p-tí)
- (p-stní) rozložení diskrétní vs. spojité proměnné

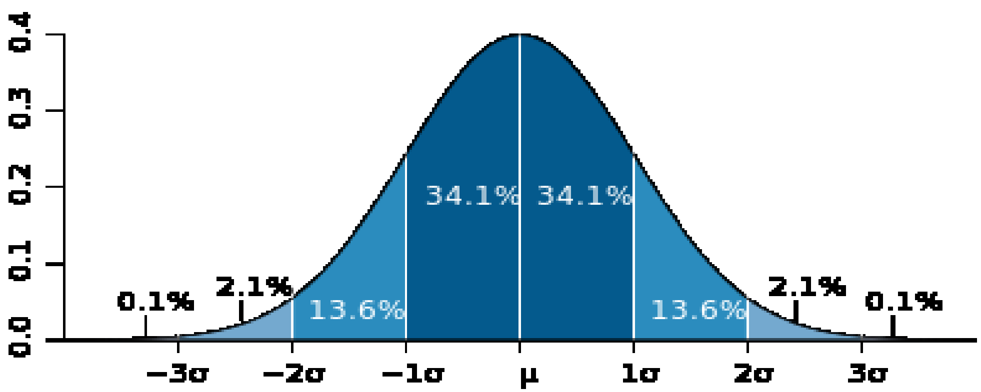
Rozložení (p-stí) diskrétní proměnné

- Diskrétní náhodná proměnná nabývá oddělených hodnot 0, 1, 2, 3….
- Pravděpodobnostní distribuce diskrétní proměnné přiděluje každé možné hodnotě pravděpodobnost
  - Pro každou takovou hodnotu (p) mezi 0 a 1
  - Suma p-stí pro všechny možné výsledky =1
  - Př. kostka x=1, 2, 3, 4, 5, 6
    - p(1)=p(2)=p(3)=p(4)=p(5)=p(6)=1/6
    - 6*1/6=1
  - Př. dvě kostky x=součet hodnot na obou kostkách
    - X = 2….12
    - P(2)=1/36 + p(3)=1/18 + p(4)=1/12 + p(5)=1/9 + p(6)=5/36 + p(7)=1/6 + p(8)=5/36 + p(9)=1/9 + p(10)=1/12 + p(11)=1/18 + p(12)=1/36
    - Součet =1



Rozložení (p-stí) spojité proměnné

- U spojitých proměnných sleduje p-ti intervalů hodnot spíše než konkrétních hodnot
  - Protože spojitá proměnná může nabývat jakýchkoli hodnot, je (p) konkrétní hodnoty blízká nule – proto hovoříme o hustotě pravděpodobnosti
  - Každý interval má p-nost mezi 0 a 1, pravděpodobnost konkrétního intervalu výsledků odpovídá velikosti oblasti pod křivkou nad daným intervalem
  - Interval obsahující všechny možné výsledky má p-nost 1, celková plocha pod křivkou = 1
  - př. normální rozložení



Parametry pravděpodobnostního rozložení

- Průměr = jakou hodnotu očekáváme když zprůměrujeme všechna pozorování v dlouhodobém horizontu = očekávaná hodnota
  - Př. hod kostkou: (1*1/6) + (2*1/6) + (3*1/6) + (4*1/6) + (5*1/6) + (6*1/6) = 3.5
  - př. počet homerunů během zápasu (0*0.3889) + (1*0.3148) + (2*0.2222)+(3*0.05556)+(4*0.0185)= 1
    - Vážený průměr neboť každý výsledek má jinou (p) výskytu

Binomické rozložení

- P-nostní rozdělení pro diskrétní proměnné
- Binární/dichotomická data/proměnná
  - Př. přijat vs. nepřijat, ano vs. ne, žena vs. nežena
- Zajímá nás počet/proporce případů, kdy nastane sledovaný výsledek
  - Př. kolik/jaká proporce šestek padne na kostce padne když hodím 6krát? Proměnná nabývá hodnot 0, 1, 2, 3, 4, 5, 6.
  - Př. Jaká je pravděpodobnost že nebude vybrána žádná žena do skupiny 10 zaměstnanců určených ke školení, pokud se jedná o náhodný výběr?

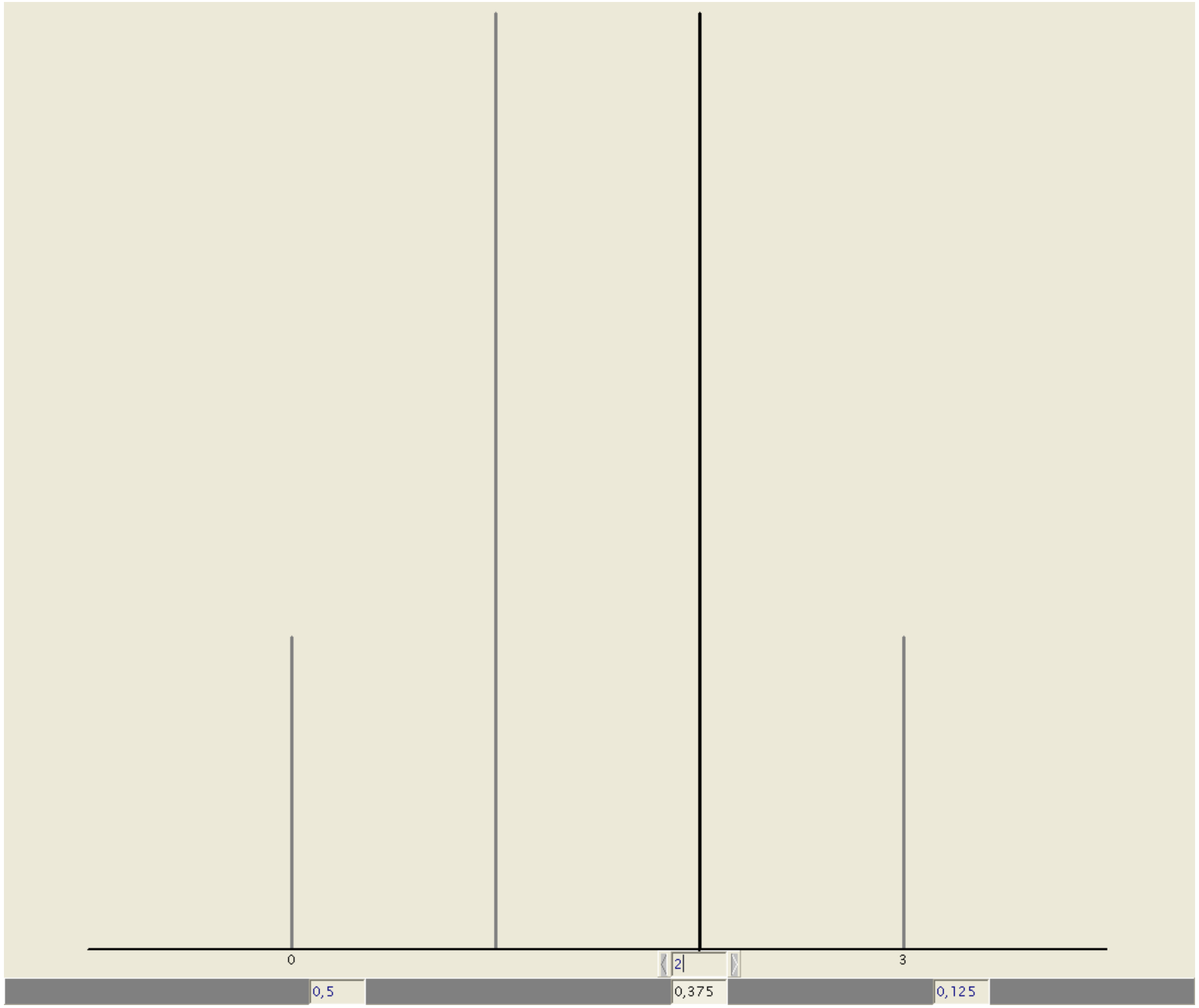
Binomické rozložení - definice

- Každý z pokusů/pozorování (tj.hody kostkou, velikost výběru) má dva možné výsledky: výsledek který sledujeme=úspěch (ano, šestka…) a ostatní výsledky=neúspěch (to ostatní)
- Pravděpodobnost úspěchu = p, pravděpodobnost neúspěchu tedy 1 – p (doplněk), pravděpodobnosti jsou stejné pro každý pokus
- Pokusy jsou nezávislé – výsledek prvního pokusu neovlivňuje výsledek druhého atd.

Pravděpodobnosti binomického rozložení

- Třikrát házíme mincí, zajímají nás počty pannen, např. jaká je p(dvakrát panna)
- Ze základního prostoru=(PPP,PPO,POP,OPP,POO,OPO,OOP,OOO) obsahují tři možnosti/sekvence výsledků 2xpanna a sice PPO,POP,OPP. Každý z výsledků má p=0.5*0.5*0.5=0.125, a proto p(2 panny)=3*(0.5*0.5*0.5)=0.375
- Když je počet pokusů velký používáme vzorec
- $n! / (n-x)!x! * p^x(1-p)^{n-x}$
  - 1. část vzorce tzv. binomický koeficient určuje počet sekvencí/výsledků s hledaným počtem úspěchů (x) ze všech pokusů (n) – zde 3! / (3-2)!2! = (3*2*1) / (2*1) *(1) = 3
  - 2. část pak pravděpodobnost každého takového kýženého výsledku - zde $(0.5)^2$*0.5=0,125

Binomická distribuce pro N=3,p=0.5, p(x=2)=0.375



Normální rozložení

- Speciální a nejpoužívanější druh pravděpodobnostního rozložení (rozdělení pravděpodobností) pro spojitá data (dalšími (p) rozděleními např. binomické, chi, poisson, F atd.)
- Jedná se o teoretické rozdělení, jemuž se rozložení v realitě více či méně blíží
- Intervalová/poměrová proměnná

Vlastnosti normálního rozložení

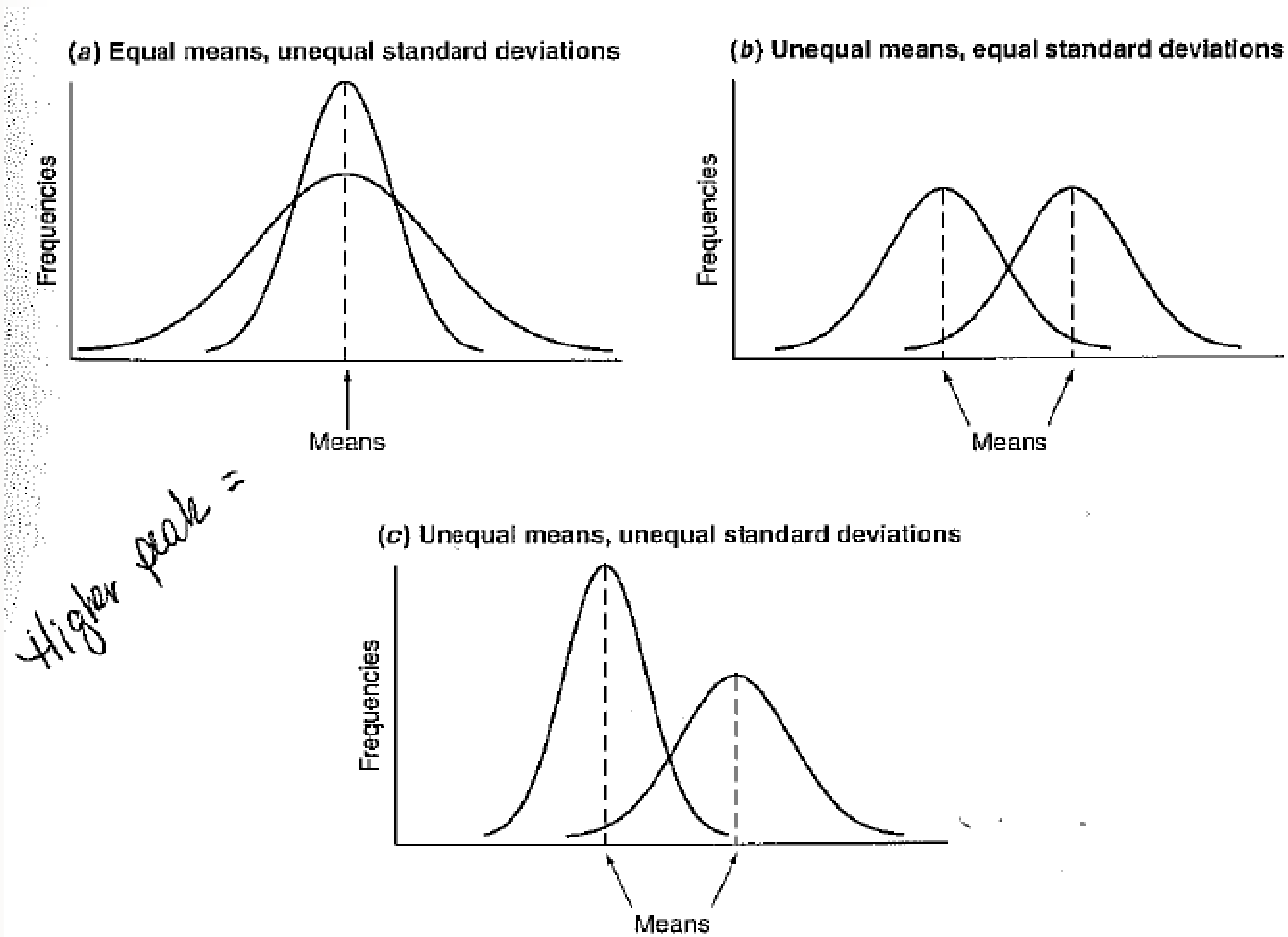
- Symetrická distribuce, zvonovitý tvar
- Šikmost, špičatost = 0
- Modus=median=průměr=nejvyšší bod
- Křivka se nedotýká osy x: extrémní hodnoty + proměnlivost populace
- Osa x rozdělená do 6 rovných jednotek = každá jednotka odpovídá 1 σ (1 SD)
- Pravděpodobnost mezi nějakými konkrétními směrodatnými odchylkami (např. mezi 0 a 1 SD=0.34) je stejná pro všechna normální rozložení

Proč je normální rozložení užitečné?

- Realita je často normálně rozložená
- Když není, lze ji často transformovat tak aby se normálnímu rozložení blížila, neboť některé statistické procedury předpokládají normální rozložení
- Při velkém počtu možných výsledků se normálnímu rozložení blíží i mnoho diskrétních rozložení
- Díky centrálnímu limitnímu teorému lze pomocí normálního rozložení za určitých podmínek analyzovat i v populaci nenormálně rozložená data = klíčová funkce při inferenční statistice

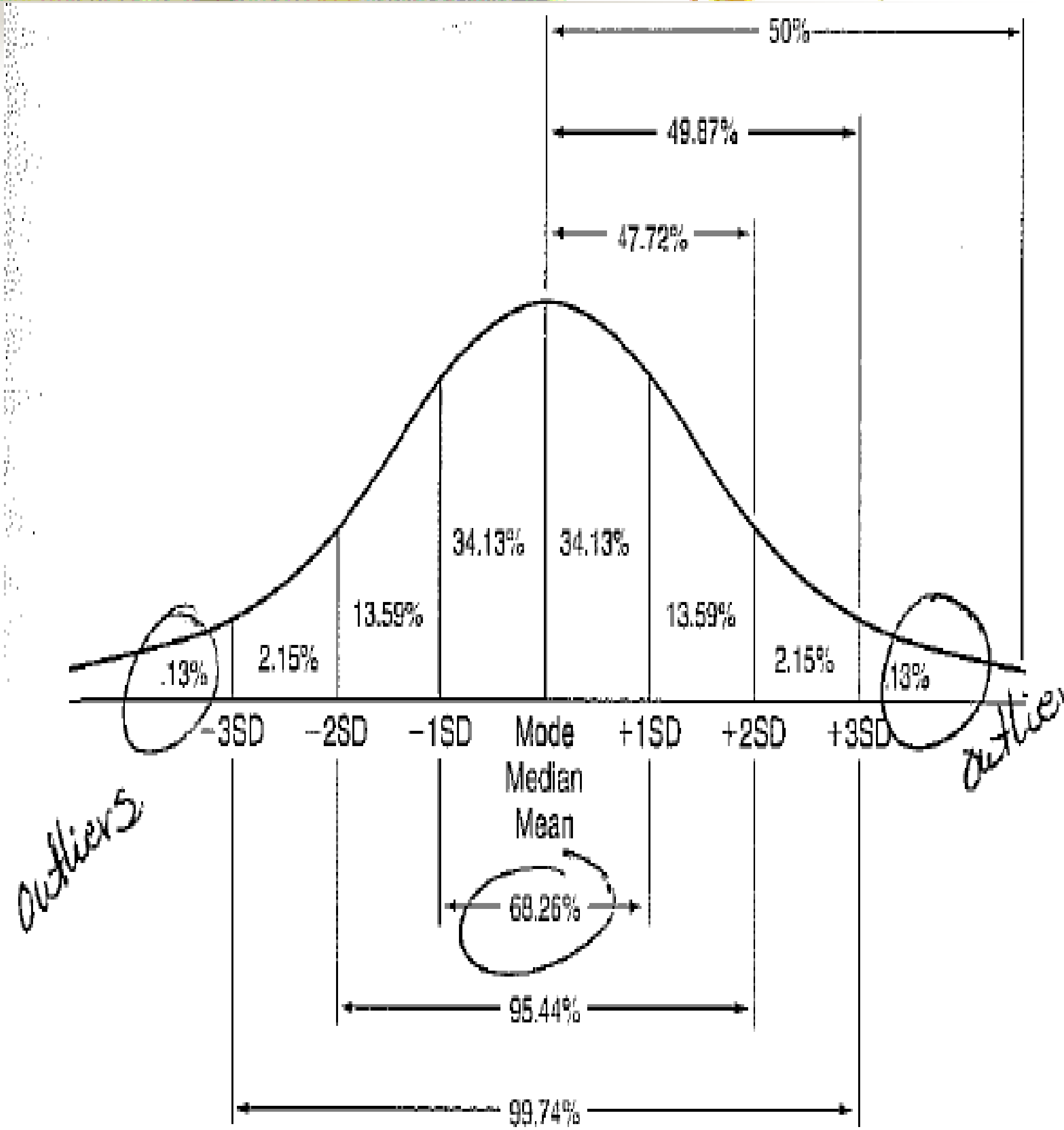
Různé tvary normálního rozložení

- Různé normální křivky - různé průměry a různé σ
- vyšší křivka = menší standardní odchylka



Rozložení pravděpodobností v normálním rozložení

- 50% oblasti pod křivkou (=naměřených hodnot proměnné) leží pod průměrem a 50% nad = symetrie
- 68% leží +/- 1 SD
- 95% leží +/- 2 SD
- 99 % leží +/- 3 SD
- Př. Pokud prům. výška=170cm a σ =5, pak 68% studentek sociální práce je vysokých 165 až 175 cm



Normální rozložení a z-skor

- V normálním rozložení lze tedy zjistit kumulativní pravděpodobnost výskytu jakéhokoli intervalu hodnot zkoumané proměnné

- A to prostřednictvím standardizace - převodu absolutní hodnoty proměnné na z-skor

Příklady kumulativní pravděpodobnosti normálního rozložení

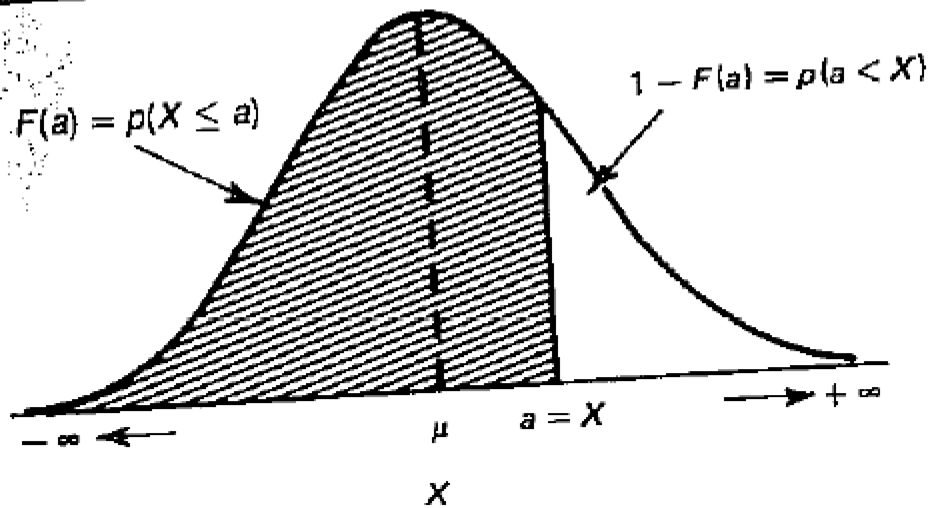


Figure 6.2.1

A cumulative probability.

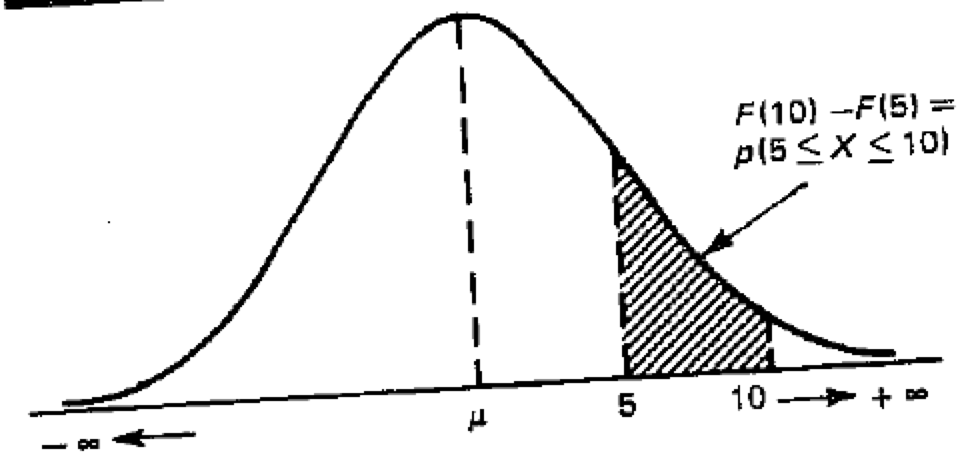


Figure 6.2.2

The probability of an interval in a norm
distribution.

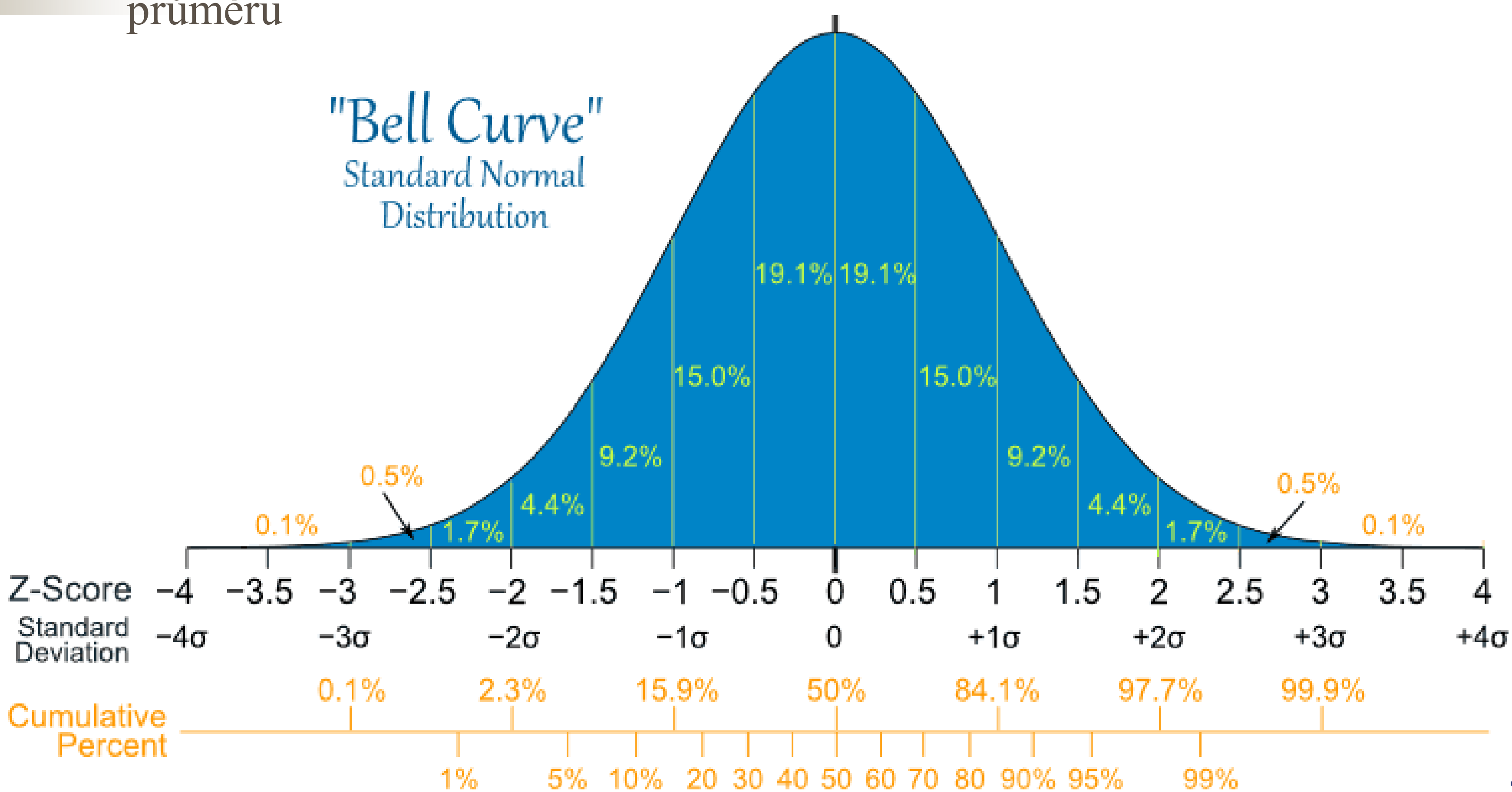
Z skóre - základ

- = z-skor pro hodnotu x náhodné proměnné představuje jak daleko (**kolik směrodatných odchylek**) od průměru se hodnota x nachází
- = rozdíl mezi individuální hodnotou (Xi) a průměrnou hodnotou (X prům) relativně k rozptylu distribuce (s)
  - Z = (Xi – X prům) / s
- Proto Z = 0 = průměr (μ)
- A také 1 z = 1 SD, 2 z = 2 SD atd.
  - Důkaz: z = (X – X prům) / s = (110-100)/10=s / s = 1
  - (120-100)/10=20/10=2 atd.

- standardizací původních hodnot distribuce vzniká Standardizované normální rozložení Z~N(0, 1)

Standardizované normální rozložení

- Mnoho statistických metod se vztahuje ke speciálnímu normálnímu rozložení zvanému **standardizované normální rozložení**
- Standardizované normální rozložení má průměr 0 a odchylku 1, Z~N(0, 1)
- Př. z = 2 se nachází 2 odchylky od průměru, z = -1.3 leží -1.3 odchylky od průměru



Z skóre - výklad

- Účel č.1: Zjištění relativní pozice individua k populaci

- Př. Výsledky IQ testu jsou aproximovány (blíží se) normálním rozložením o průměru $\mu = 100$ a $\sigma=16$. Bob skóroval 125. Jak „chytrý“ je Bob vzhledem k ostatním?
  - $Z = (X - \mu) / \sigma = (125-100) / 16 = 1.56$
  - Bob skóroval 1.56 standardní odchylky nad průměrem
  - Jaká část populace skórovala více (nebo méně)? Viz tabulka

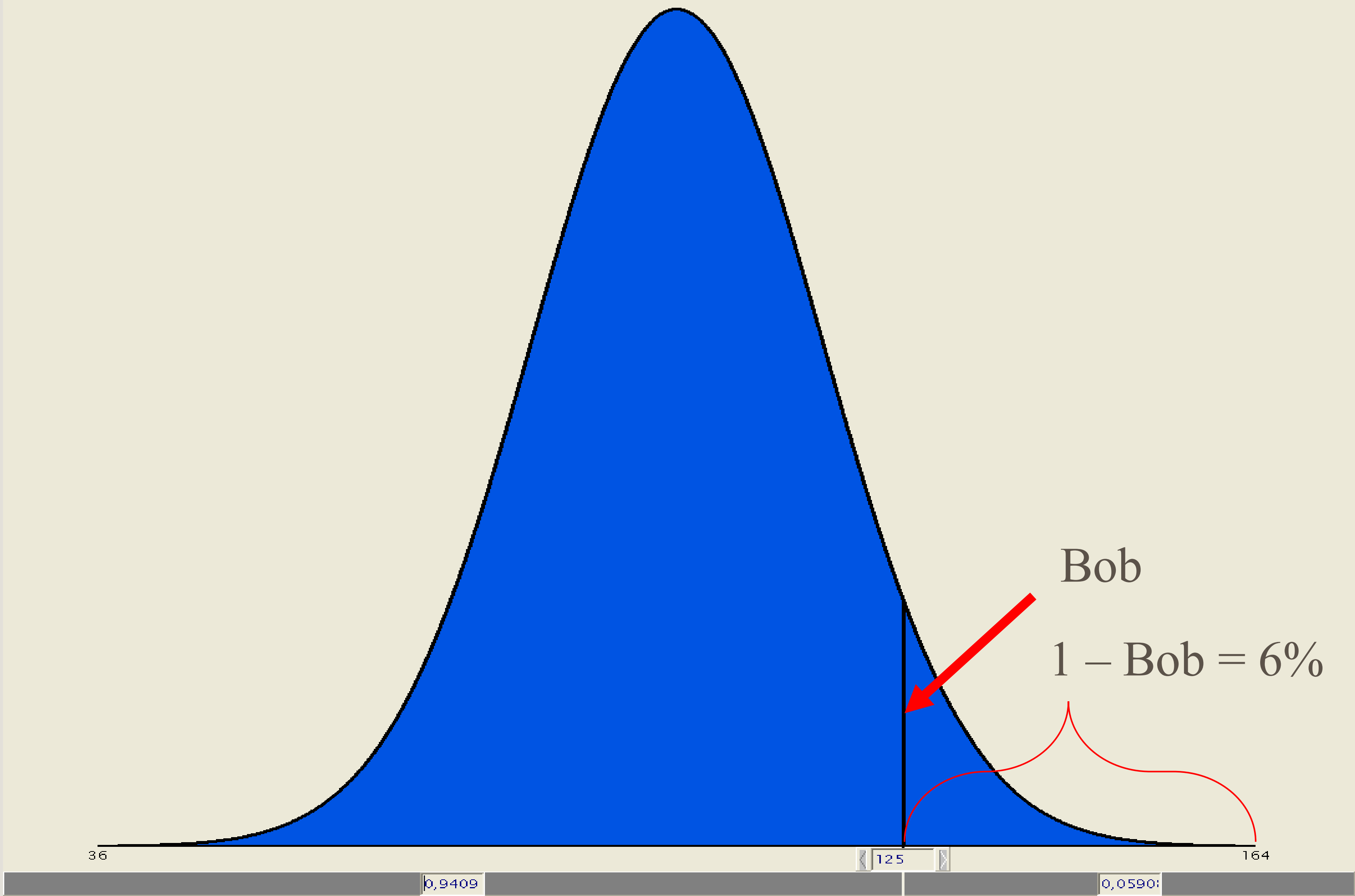
- Hodnota v tabulce odpovídá oblasti pod normální křivkou mezi průměrem a z-skórem
- Z=1.56 odpovídá hodnotě 44.06
- 50 + 44.06 = 94.06
- Interpretace:
  - a) bob leží na 94.06 percentilu, je chytřejší než 96% ostatních dětí v populaci
  - b) protože
    1 – 94 = 6, tak existuje 6% šance že člověk v populaci má vyšší skóre než Bob

TABLE

Area under the normal curve between mean and z score

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	00.00	00.40	00.80	01.20	01.60	01.99	02.39	02.79	03.19	03.59
0.1	03.98	04.38	04.78	05.17	05.57	05.96	06.36	06.75	07.14	07.53
0.2	07.93	08.32	08.71	09.10	09.48	09.87	10.26	10.64	11.03	11.41
0.3	11.79	12.17	12.55	12.93	13.31	13.68	14.06	14.43	14.80	15.17
0.4	15.54	15.91	16.28	16.64	17.00	17.36	17.72	18.08	18.44	18.79
0.5	19.15	19.50	19.85	20.19	20.54	20.88	21.23	21.57	21.90	22.24
0.6	22.57	22.91	23.24	23.57	23.89	24.22	24.54	24.86	25.17	25.49
0.7	25.80	26.11	26.42	26.73	27.04	27.34	27.64	27.94	28.23	28.52
0.8	28.81	29.10	29.39	29.67	29.95	30.23	30.51	30.78	31.06	31.33
0.9	31.59	31.86	32.12	32.38	32.64	32.90	33.15	33.40	33.65	33.89
1.0	34.13	34.38	34.61	34.85	35.08	35.31	35.54	35.77	35.99	36.21
1.1	36.43	36.65	36.86	37.08	37.29	37.49	37.70	37.90	38.10	38.30
1.2	38.49	38.69	38.88	39.07	39.25	39.44	39.62	39.80	39.97	40.15
1.3	40.32	40.49	40.66	40.82	40.99	41.15	41.31	41.47	41.62	41.77
1.4	41.92	42.07	42.22	42.36	42.51	42.65	42.79	42.92	43.06	43.19
1.5	43.32	43.45	43.57	43.70	43.83	43.94	44.06	44.18	44.29	44.41
1.6	44.52	44.63	44.74	44.84	44.95	45.05	45.15	45.25	45.35	45.45
1.7	45.54	45.64	45.73	45.82	45.91	45.99	46.08	46.16	46.25	46.33
1.8	46.41	46.49	46.56	46.64	46.71	46.78	46.86	46.93	46.99	47.06
1.9	47.13	47.19	47.26	47.32	47.38	47.44	47.50	47.56	47.61	47.67
2.0	47.72	47.78	47.83	47.88	47.93	47.98	48.03	48.08	48.12	48.17
2.1	48.21	48.26	48.30	48.34	48.38	48.42	48.46	48.50	48.54	48.57
2.2	48.61	48.64	48.68	48.71	48.75	48.78	48.81	48.84	48.87	48.90
2.3	48.93	48.96	48.98	49.01	49.04	49.06	49.09	49.11	49.13	49.16
2.4	49.18	49.20	49.22	49.25	49.27	49.29	49.31	49.32	49.34	49.36
2.5	49.38	49.40	49.41	49.43	49.45	49.46	49.48	49.49	49.51	49.52
2.6	49.53	49.55	49.56	49.57	49.59	49.60	49.61	49.62	49.63	49.64
2.7	49.65	49.66	49.67	49.68	49.69	49.70	49.71	49.72	49.73	49.74
2.8	49.74	49.75	49.76	49.77	49.77	49.78	49.79	49.79	49.80	49.81
2.9	49.81	49.82	49.82	49.83	49.84	49.84	49.85	49.85	49.86	49.86
3.0	49.87									
3.5	49.98									
4.0	49.997									
5.0	49.99997									

Kde leží Bob?



- Účel č.2: Porovnání relativních pozic dvou individuí z rozdílných vzorků (populací?)

- Př. Dvě kamarádky Rita a Miriam se účastnili jiných skupin kurzu praxe sociální práce, v ½ semestru složili zkoušku, Rita získala 21, Miriam 85 bodů, kdo byl lepší?
    - Srovnat maximální počet bodů v obou testech
        - Rita 21 z 25 = 84 %, Miriam 85 ze 100 = 85 %, je Miriam lepší? Co když je Miriaminých 85 % nejhorší výsledek ve skupině zatímco Ritiných 85 % nejlepší výsledek?
    - nebo srovnat jednotlivé výsledky s výsledky ostatních studentů pomocí z-skóru

- Př. Deborah pracuje jako sociální pracovnice ve studentském zdravotním centru a vede kurzy pro léčbu chronické úzkosti. Uvolnilo se jí místo ve skupině. Do skupiny se přijímá na základě testu „Škála úzkosti A“ (μ =70, σ =10) . Pouze studenti kteří dosáhnou min. 80 bodů na škále A mohou být přijati. Deborah se podívala do seznamu potenciálních klientů a zjistila že nejvyššího skóre 78 dosáhla Gina. Deborah však právě dostala doporučení o novém studentovi který trpí úzkostí a potřebuje pokračovat v léčbě. Doporučení také obsahovalo že student Tom dosáhl 66 bodů na jiné škále „Škále B“ (μ =50, σ =12).

- Co může Deborah udělat aby srovnala oba uchazeče a vybrala potřebnějšího?
  - A) Nechat Toma otestovat „Škálou A“
  - B) Zná-li průměr a směrodatnou odchylku obou škal, může porovnat Z-skóry.

- Řešení:
  - $Z_{Gina} = (78 – 70) / 10 = 0.8$
    - Tabulka Z 0.8 = 28.81 + 50 = 78.81 = 79th percentil
  - $Z_{Tom} = (66 – 50) / 12 = 1.33$
    - Tabulka Z 1.33 = 40.82 + 50 = 90.82 = 91st percentil
  - Tom byl vybrán jako potřebnější na základě relativně vyšší úrovně úzkosti

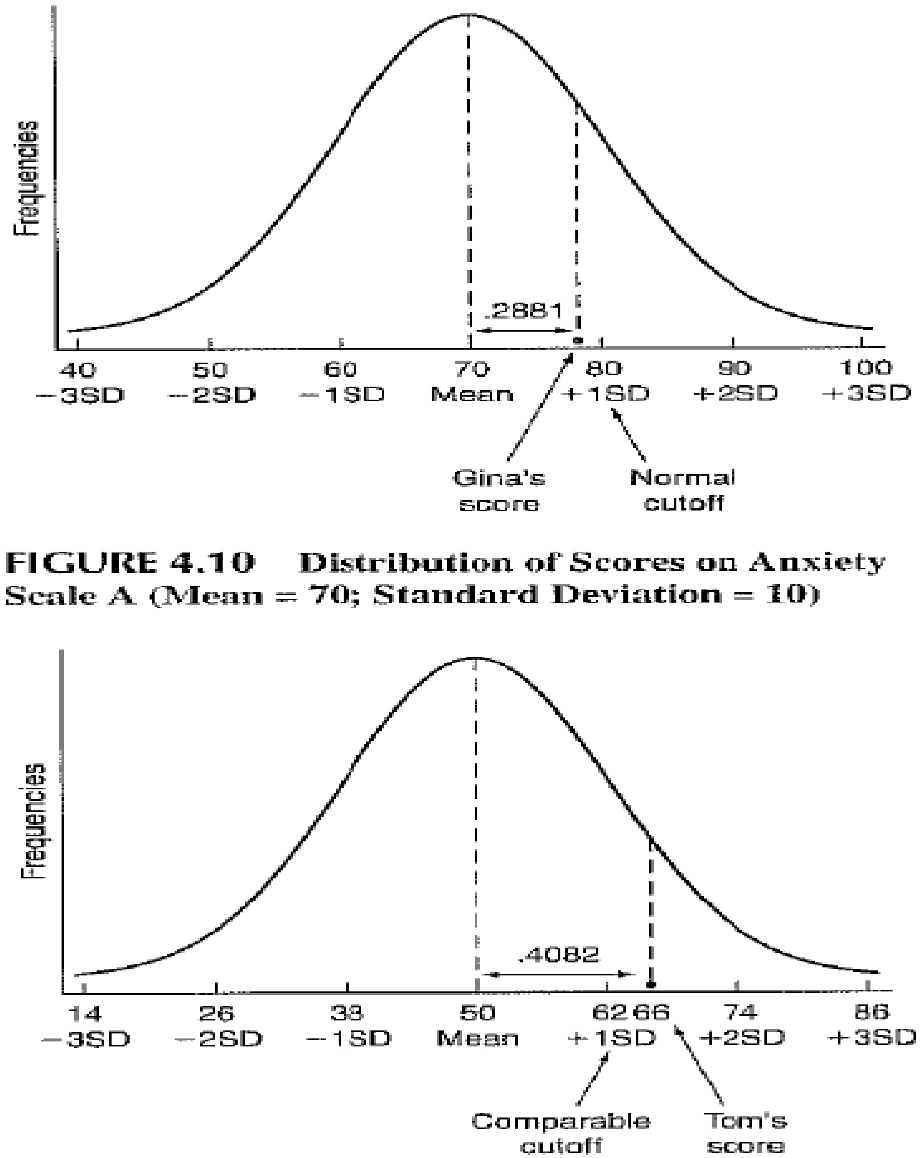


FIGURE 4.10 Distribution of Scores on Anxiety Scale A (Mean = 70; Standard Deviation = 10)

- Účel č. 3: Odvození syrového skóre z percentilu (z-skóru)
- Sociální pracovnice Lauren chce vytvořit skupinu pro léčbu studentů s vysokou úrovní úzkosti, na základě výsledků z testů na „Škále B“(μ =50, σ =12), přičemž chce přijmout jen horních 10 procent nejvážnějších případů.
- Řešení:
  - Lauren musí najít mezní bod (cut-off point) pro syrové skóre, který by nejlépe odpovídal 90th percentilu. Studenti nad toto skóre budou přijati, ostatní ne.
  - $X = \mu + z^* \sigma$
  - Postup: najít z-skor pro kumulativní pravděpodobnost 90 – 50 = 40
  - Jaké Z odpovídá hodnotě 40?: Z = 1.28
  - 1.28 (Z-skóre)= (x – 50) / 12

    (12*1.28) + 50 = x

    65.36 = x

Odpověď: Pro vstup do skupiny je třeba získat 66 bodů.

Příloha 1: Shrnutí základních pojmů

- Jednoduchý/náhodný pokus
  - *Akt vedoucí k jednomu výsledku - např. hod kostkou, zatočení ruletou, vytažení karty z balíčku, výběr osoby na ulici*
  - *Výsledkem je výskyt jednoduchého jevu/události*
- Jednoduchý výsledek
  - *člen základní množiny*
  - *výsledek jednoduchého pokusu - např. hodnota 1 na kostce, 0 na ruletě, sedmička srdcová, modrooká paní*
- Jev/třída jevů
  - *sada jednoduchých výsledků, podmnožina základního prostoru - např. lichá čísla, „srdce“, „piky“*
- Základní množina/prostor *(S)*
  - *sada všech jednoduchých jevů / všech možných výsledků*
- Spojené jevy *– nastávají když výsledek pokusu spadá pod jevy A(„srdce“) i B(„král“) např. „srdcový král“, popřípadě A nebo B např. „srdce“ nebo „král“*
  - Průnik (∩) *– např. průnik jevů A a B = A ∩ B nebo-li A a B*
    - *současné nastání dvou nebo více jevů*
  - Sjednocení (U) - *např. sjednocení jevů = A U B nebo-li A nebo B*
    - *sečtení dvou nebo více jednoduchých jevů bez průniku*
- Doplněk (~A)
  - *doplňkem jevu A je sada všech zbývajících jevů z S*
- Vzájemně vylučující se/neslučitelné jevy
  - *nemohou nastat současně, jejich ∩ = 0*
- Vyčerpávající jevy
  - *jevy vyplňují celý S, jejich U = S*
- Pravděpodobnost *(p)*
  - *míra jistoty nastání každého jevu ze základního prostoru - např. pravděpodobnost že padne 1 na kostce*
- Podmíněná pravděpodobnost (p (A|B))
  - *pravděpodobnost výskytu jevu A za předpokladu, že zároveň nastane jev B –*
  - *např. experiment: hod dvěma kostkami, událost: součet hodnot, otázka: jaká je pravděpodobnost výskytu události 4 když na jedné kostce padne 5?*
- Statistická nezávislost
  - *nepodmíněná pravděpodobnost jevu A a podmíněná pravděpodobnost jevu A stane-li se zároveň B jsou si rovny*
  - *tj. p(A) = p(A | B)*
  - *nebo když* p (A ∩ *B*) = p (A) * p (B)

Příloha 2: Pravidla pravděpodobnosti

- p(~A) = 1 – p(A) *(„doplňková pravděpodobnost")*
  - Př. Jaká je pravděpodobnost že vyberu „ne červenou" kuličku tj. jinou než „červenou"?
    - p(~červená) = 1 - p (červená) = .7

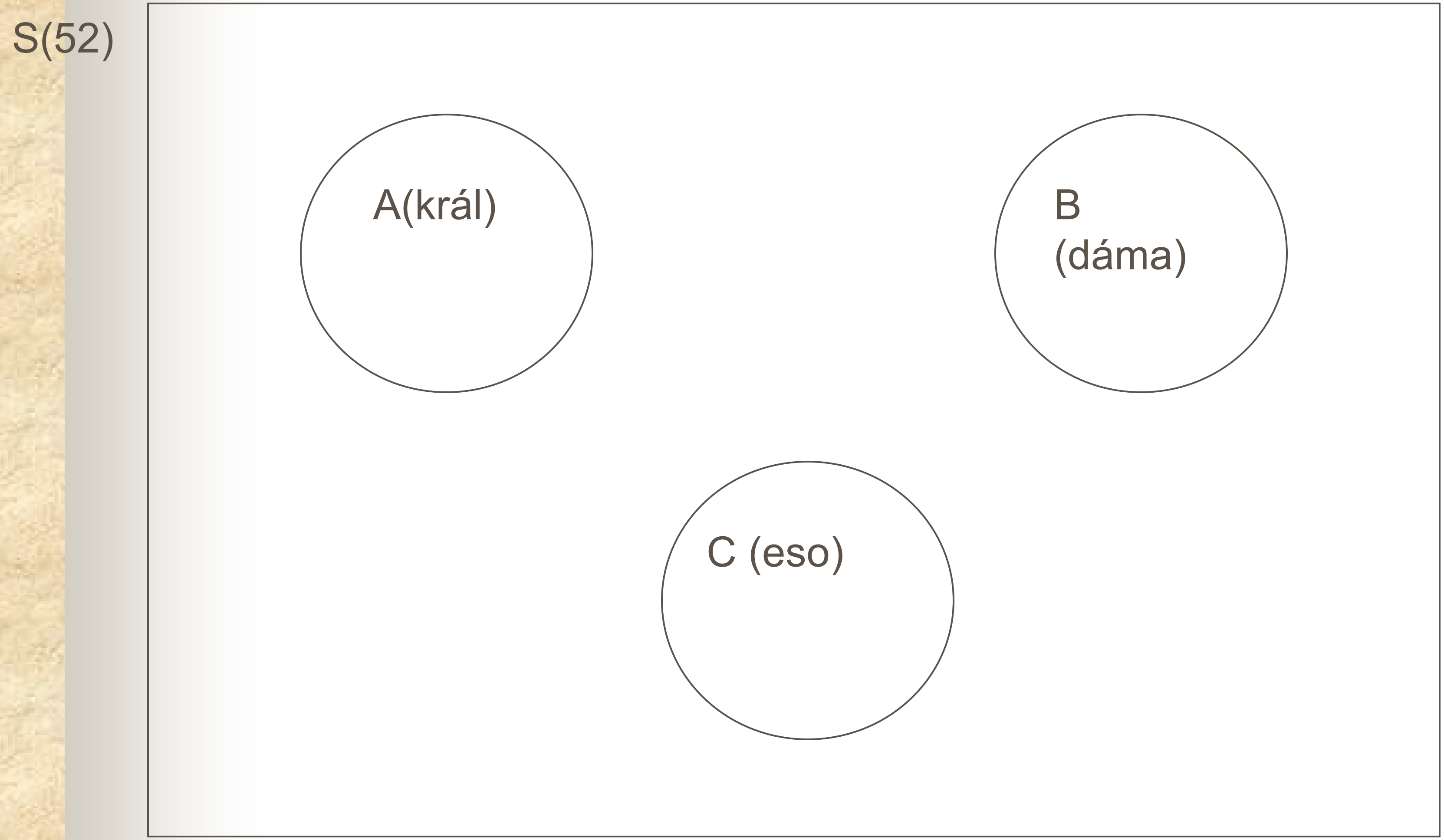
- $0 \leq p(A) \leq 1$ *(„rozsah pravděpodobnosti")*
- *(důkaz: pokud by nějaký jev měl p větší než 1 pak by podle pravidla 1 měl doplněk jevu p zápornou a to by odporovalo axiomu 1)*

- *p(Ø) = 0, pro jakékoli S („nemožný jev")*
  - Př. Jaká je pravděpodobnost že vyberu „bezbarvou" kuličku tj. jinou než „červenou" nebo „bílou" nebo „černou"?
    - p(bez barvy)=0

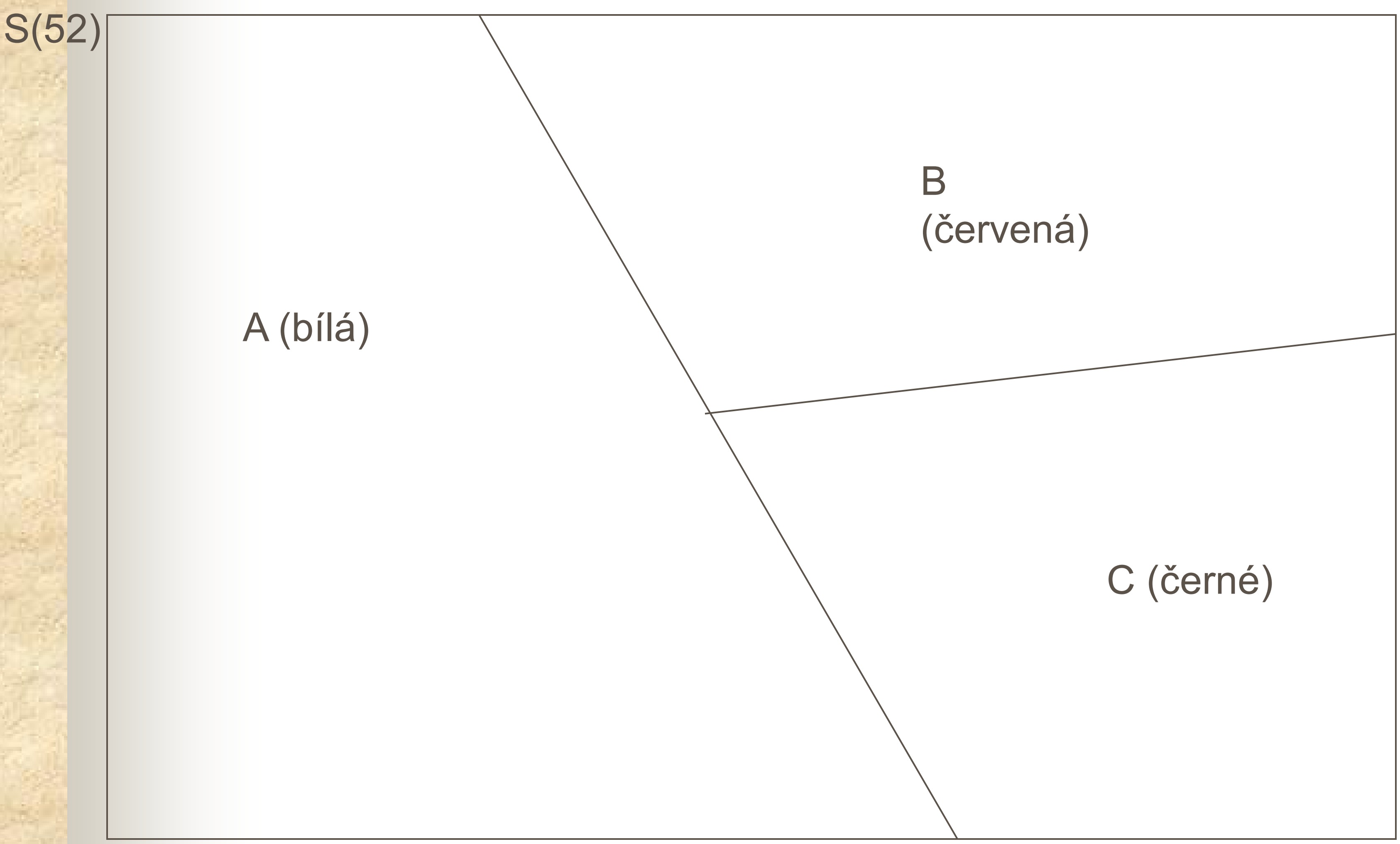
- *p (A U B) = p (A) + p (B) – p (A ∩ B) (tzv."nebo" pravidlo)*
  - *Př. Balíček 52 karet. Jaká je pravděpodobnost „krále" nebo „srdce"?*
    - *P (král) = 1/13, p(srdce)=1/4, p (král ∩ srdce)=1/52 (jeden z králů je srdcový)*
    - *p (král nebo srdce) = 1/13 + 1/4 – 1/52 = 16/52 = 4/13*
  - *Speciální případ: když jsou jevy vzájemně se vylučující, pak p (A ∩ B) =0*
  - *a proto p (A U B) = p (A) + p (B)*
    - *Př.* p(červená nebo bílá) = .30 + .50 = .80

- *Pokud A,...., L tvoří segmenty S, pak p (A* U ...... U L) = p (A) + …p (L) = 1
  - *Pokud jsou jevy A až L vylučující se a vyčerpávající, pak tvoří celý prostor S a součet jejich pravděpodobností musí být 1*
  - *Př.* p(červená nebo bílá nebo černá kulička) = .30 + .20 + .50 = 1.00

Příloha 3: A, B a C jsou vzájemně se vylučující jevy



Příloha 4: A, B a C jsou vzájemně se vylučující a vyčerpávající jevy



Příloha 5: užití podmíněné (p) při posuzování „přesnosti“ diagnostického testu

- Jedním ze způsobů určení „přesnosti testu“ je spočítat pravděpodobnosti dvou typů chyb, které chceme minimalizovat:
  - Falešná pozitivita = test říká, že nemoc je přítomná, ale ve skutečnosti přítomná není
  - Falešná negativita = test říká, že nemoc přítomna není, ale ve skutečnosti přítomna je
- Př. N=5282 žen nad 35 let, 48 Downových syndromů z 54 se podařilo tímto testem odhalit, zatímco 25 procent těhotenství bylo chybně identifikováno jako ohrožené Downovým syndromem
  - Falešná pozitivita = p(D „ano“| D (ne)) = 1307/5228 = 0,25
  - Falešná negativita = p(D „ne“ | D (ano)) = 6 / 54 = 0,11

	TEST		
	D(„ano“)	D(„ne“)	celkem
D(ano)	48	6	54
D(ne)	1307	3921	5228
Celkem	1355	3927	5282

- Užitečnost testu:
  - Na jednu stranu pouze 4% pozitivně diagnostikovaných žen ve skutečnosti disponují Downovým syndromem, neboť podmíněná pravděpodobnost že dítě má syndrom když test řekne ano = p(Dano|D“ano“) = 48/1355=0.035
  - Na druhou stranu snižuje riziko výskytu syndromu z P(Dano)=54/5282=0.01 na P(Dano | D“ne“) = 6/3927=0.0015