

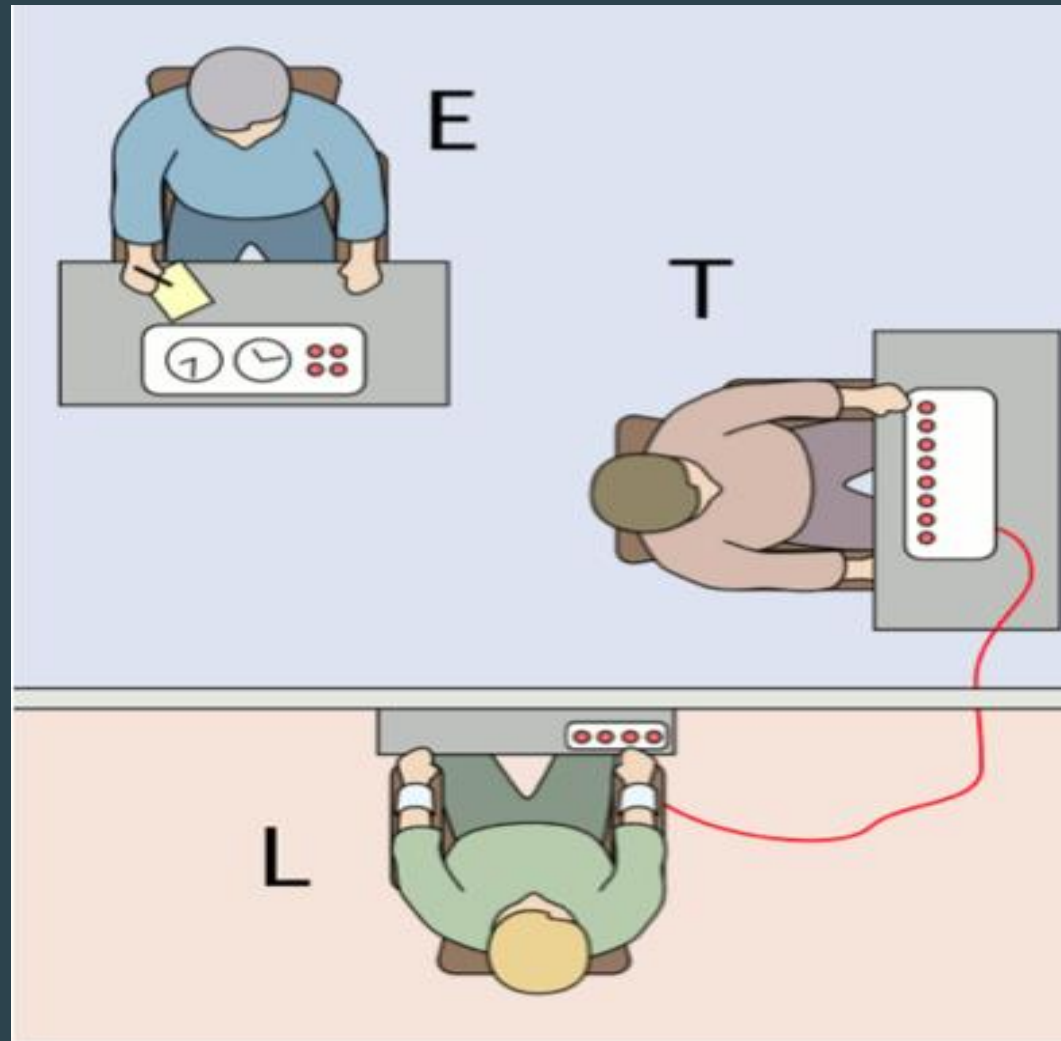


Analýza dát

Prednáška 4

Geometrické a binomické rozdelenie

Milgramov experiment



Milgramov experiment

- Stanley Milgram, psychológ z Univerzity v Yale, uskutočnil sériu experimentov zameraných na poslušnosť voči autorite (1963). Experimentátor (E) nariaďuje učiteľovi (T), ktorý je predmetom experimentu, aby žiakovi (L) spôsobil vážne elektrické šoky, zakaždým, keď študent nesprávne odpovie na otázku. Študent je v skutočnosti herec a elektrické šoky nie sú skutočné, ale pri každom zásahu učiteľa elektrickým prúdom sa prehrá vopred nahraný zvuk.
- [Odkaz na bližšie informácie](#)

Milgramov experiment

- Tieto experimenty merali ochotu účastníkov štúdie poslúchať autoritu, ktorá im dala pokyny, aby konali činy, ktoré sú v rozpore s ich osobným svedomím.
- Milgram zistil, že asi **65% ľudí** by sa podriadilo autorite a spôsobilo takéto šoky.
- V priebehu rokov ďalší výskum naznačil, že tento počet je približne konzistentný vo všetkých komunitách a čase

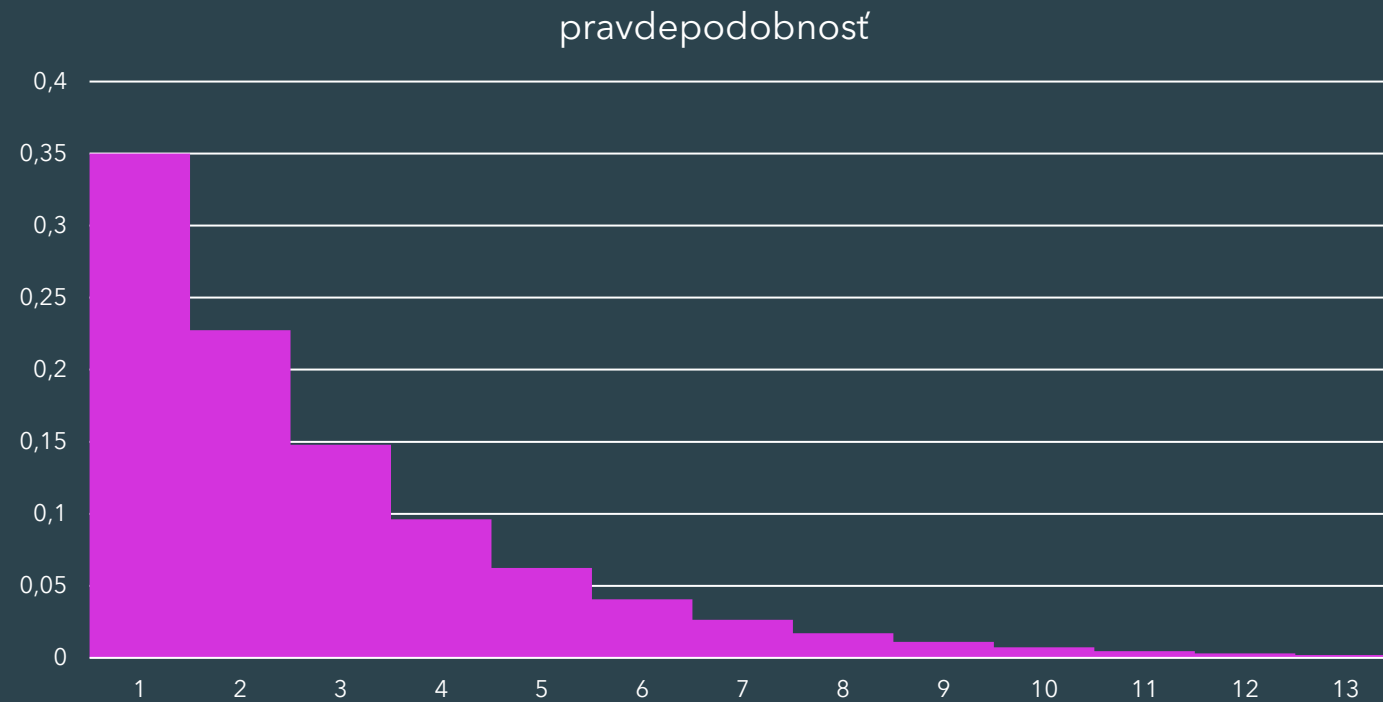
Bernoulliho náhodná premenná

- Každá osoba v Milgramovom experimente môže byť považovaná za pokus.
- Osoba je označená **za úspešnú**, ak odmietne aplikovať závažný šok, **a neúspešnú**, ak tento šok vykoná.
- Pretože iba 35% ľudí odmietlo vykonať šok, pravdepodobnosť úspechu je $p = 0,35$.
- Ak má individuálny pokus iba dva možné výsledky, potom sa príslušná náhodná premenná nazýva **Bernoulliho náhodná premenná**.

Geometrické rozdelenie - príklad

- Doktorka Smithová chce opakovať Milgramove experimenty, ale chce ich opakovať dovtedy, kým nenájde osobu, ktorá odmietne dať ťažký šok. Aká je pravdepodobnosť, že sa zastaví po prvej, druhej, tretej, desiatej osobe?

Histogram geometrického rozdelenia pri $p = 0,35$



Geometrické rozdelenie

- **Geometrické rozdelenie** popisuje „čakaciu dobu“ nezávislých a identicky distribuovaných Bernoulliho náhodných premenných, kým nenastane úspech
- **nezávislosť**: výsledky pokusov sa navzájom neovplyvňujú
- **identickosť**: pravdepodobnosť úspechu je rovnaká pre každý pokus

Geometrické rozdelenie

Ak:

- **p** = predstavuje pravdepodobnosť úspechu,
 - **(1 - p)** = predstavuje pravdepodobnosť zlyhania
 - **n** predstavuje počet nezávislých pokusov
-
- **P(úspech nastane pri n-tom pokuse) =**
 $(1 - p)^{n-1} \cdot p$

Príklad

- Môžeme pomocou geometrického rozdelenia vypočítať pravdepodobnosť, že pri hádzaní kockou padne šestka po prvýkrát pri šiestom pokuse?
- Riešenie: áno, za predpokladu, že za úspech budeme považovať hod šestky a za neúspech budeme považovať, že padlo čokoľvek iné ako šestka. Tieto javy môžeme považovať za nezávislé a identicky distribuované.
- $P(\text{šestka padne v 6. pokuse}) =$

Príklad

- Koľko ľudí má doktorka Smithová otestovať kým nájde prvú osobu, ktorá odmietne podať šok? Čiže aká je očakávaná priemerná hodnota μ a jej smerodajná odchýlka σ ?
- Očakávaná priemerná hodnota μ bude predstavovať priemerne koľko pokusov treba uskutočniť, aby sme dosiahli úspešný pokus.

Očakávaná hodnota (priemer) a smerodajná odchýlka geometrického rozdelenia

- Nech p je pravdepodobnosť úspechu, potom

očakávaná priemerná hodnota μ geometrického rozdelenia je:

$$\mu = \frac{1}{p}$$

Smerodajná odchýlka σ geometrického rozdelenia je:

$$\sigma = \sqrt{\frac{1-p}{p^2}}$$

Príklad - riešenie

- $P = 0,35$

$$\mu = \frac{1}{p} = \frac{1}{0,35} = 2,86$$

Doktorka Smithová by mala uskutočniť pokus 2,86 krát

$$\sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-0,35}{0,35^2}} = 2,3$$

Binomické rozdelenie

Binomické rozdelenie pracuje s **Bernoulliho náhodnou premennou** (t.j. ak má individuálny pokus iba dva možné výsledky), ale zaujíma nás počet úspechov v stanovenom počte pokusov, na rozdiel od geometrického rozdelenia, pri ktorom nás zaujíma pravdepodobnosť, že úspech nastane po prvýkrát v stanovenom pokuse.

Binomické rozdelenie -príklad

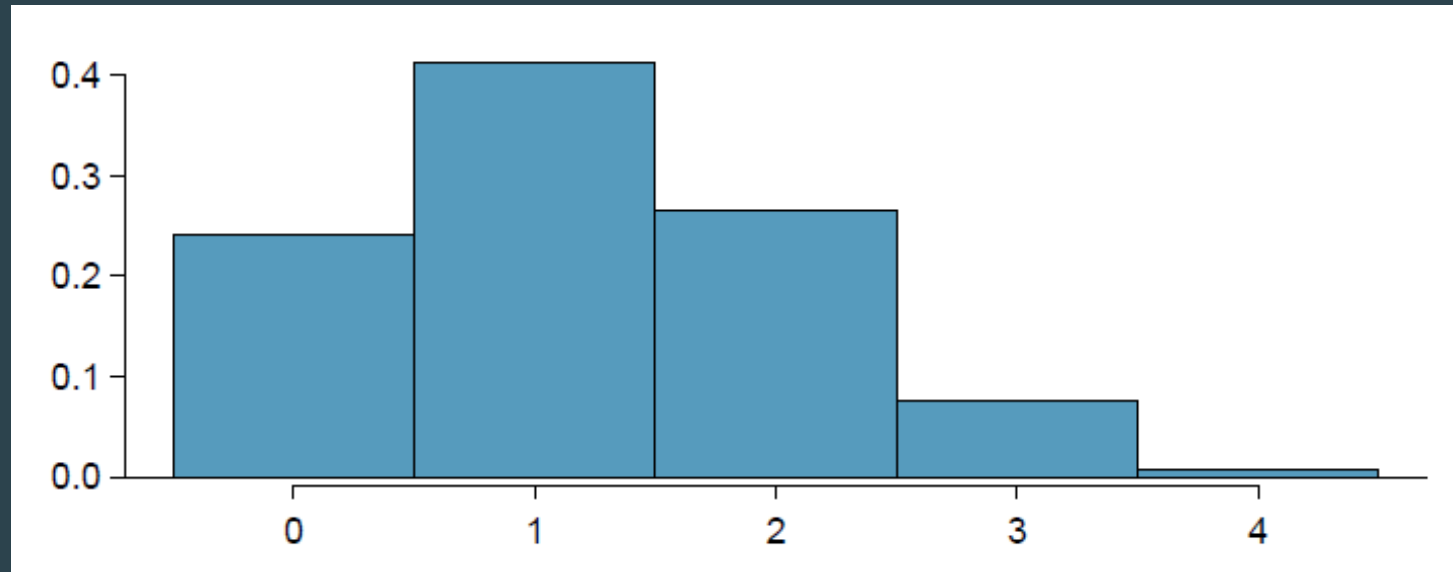
Z prieskumu svetovej zdravotníckej organizácie sa zistilo, že 30% všetkých fajčiarov počas svojho života ochorie na závažnú pľúcnu chorobu. Budeme náhodne vyberať 4 fajčiarov. Aké je rozdelenie diskkrétnej náhodnej premennej = počet fajčiarov so závažnou chorobou pľúc medzi vybranými 4 fajčiarmi?

Tabuľka binomického rozdelenia

Tabuľka binomického rozdelenia

Xi	P(xi)
0	$P(0 \text{ zo } 4) = \binom{4}{0} 0,7^4 = 0,24$
1	$P(1 \text{ zo } 4) = \binom{4}{1} 0,3 \cdot 0,7^3 = 0,41$
2	$P(2 \text{ zo } 4) = \binom{4}{2} 0,3^2 \cdot 0,7^2 = 0,26$
3	$P(3 \text{ zo } 4) = \binom{4}{3} 0,3^3 \cdot 0,7 = 0,08$
4	$P(4 \text{ zo } 4) = \binom{4}{4} 0,3^4 = 0,008$

Histogram binomického rozdelenia



Očakávaná priemerná hodnota a smerodajná odchýlka binomického rozdelenia

- Pre binomické rozdelenie s parametrami n a p , kde n je počet nezávislých pokusov a p je pravdepodobnosť úspešného pokusu platí:

$$\mu = np$$

$$\sigma = \sqrt{np(1 - p)}$$

Príklad

Pravdepodobnosť, že náhodne vybraný fajčiar počas svojho života ochorie na závažnú pľúcnu chorobu je 0,3. Koľko očakávame ľudí chorých na závažnú pľúcnu chorobu medzi 40 náhodne vybranými fajčiarmi? Aká je smerodajná odchýlka?

Normálna aproximácia binomického rozdelenia

- V prípade, že pri binomickom rozdelení máme n väčšie číslo, výpočet pravdepodobnosti pomocou binomického pravidla je pomerne náročný.
- Napríklad: Vieme, že 20% populácie sú fajčiari. Aká je pravdepodobnosť, že medzi 400 ľuďmi bude maximálne 60 fajčiarov?

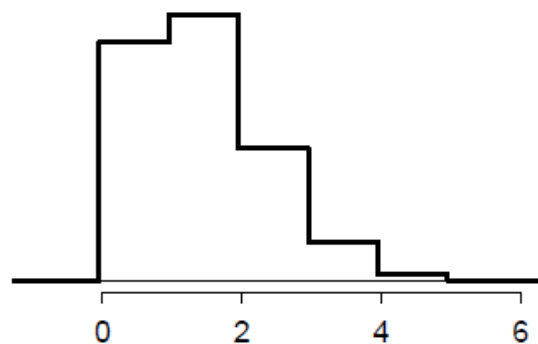
Ako by mohlo vyzerat' riešenie

Náhodná premenná X =počet fajčiarov medzi 400 vybranými ľuďmi má binomické rozdelenie s parametrami: $p = 0,2; n = 400$.

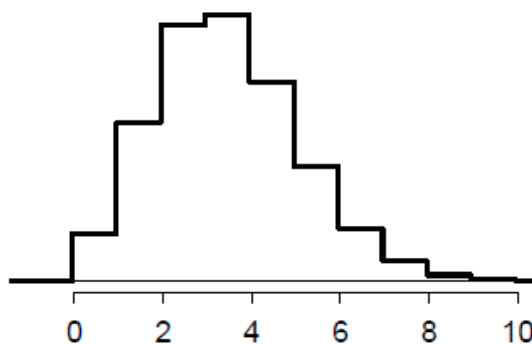
Medzi 400 vybranými ľuďmi nemá byť viac ako 60 fajčiarov, t.z. ani jeden alebo jeden alebo dvaja alebo... 60: preto $k \in \{0,1, \dots, 60\}$

Ako by mohlo vyzerat' riešenie

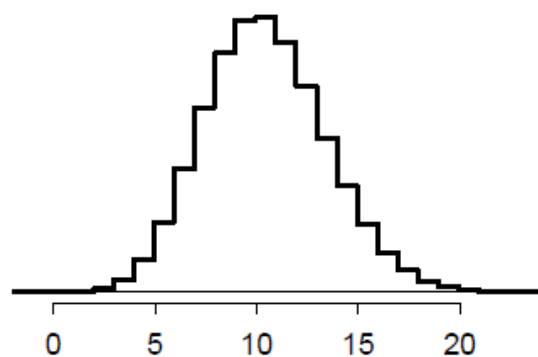
Príklady histogramov binomického rozdelenia pre rôzne hodnoty n a $p = 0,1$



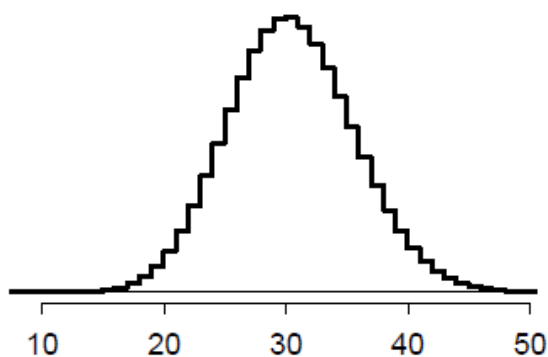
$n = 10$



$n = 30$



$n = 100$



$n = 300$

Normálna aproximácia binomického rozdelenia

Binomické rozdelenie s pravdepodobnosťou úspechu p môžeme považovať za **normálne**, ak pre hodnoty np a $n(1 - p)$ platí:

$$np \geq 10$$

$$n(1 - p) \geq 10$$

Normálna aproximácia binomického rozdelenia

Ak sú dodržané predchádzajúce podmienky, tak normálna aproximácia binomického rozdelenia má parametre normálneho rozdelenia, t.j.

$$\mu = np$$

$$\sigma = \sqrt{np(1 - p)}$$

Príklad opäť, ale inak

Vieme, že 20% populácie sú fajčiari. Aká je pravdepodobnosť, že medzi 400 ľuďmi bude maximálne 60 fajčiarov?

Príklad - riešenie

Príklad - riešenie