



Gregor Mendel v kontextu historie statistiky (200 let narození zakladatele genetiky)

Jan Kalina

Ústav informatiky AV ČR, v.v.i.

Mendelův životopis

- 1822 narozen v Hynčicích
- studia v Hynčicích, Lipníku nad Bečvou, Opavě, Olomouci
- 1843 vstoupil k augustiniánům
- 1847 ordinován jako mnich
- 1851–1853 studia ve Vídni
 - Christian Doppler
 - Andreas von Ettingshausenkniha o kombinatorice, mj. zavedl $\binom{n}{k}$
- 1856–1863 experimenty s hrachem
- Současně učil fyziku a botaniku na gymnáziu
- 1865 dvě přednášky v Brně
- 1866 **článek**
- 1867 zvolen opatem
- 1884 † v Brně



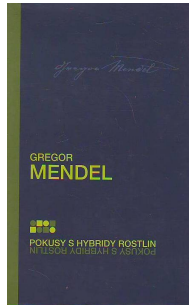
Jaké jsou 3 hlavní cíle této přednášky?

Gregor Mendel (1822-1884)



Gregor Mendel v kontextu historie statistiky

- Úvod
- Čím mě zaujal Mendelův článek
- Příliš dobré výsledky
- Jak Mendel ovlivnil historii statistiky



- 1 Mendel G. (1866): Versuche über Pflanzen-Hybriden. *Verhandlungen des Naturforschenden Vereines in Brünn* 4, 3–47.

Ukázka z Mendelova článku (1866)

				in Verhältniss gestellt:
Generation	<i>A</i>	<i>Aa</i>	<i>a</i>	<i>A</i> : <i>Aa</i> : <i>a</i>
1	1	2	1	1 : 2 : 1
2	6	4	6	3 : 2 : 3
3	28	8	28	7 : 2 : 7
4	120	16	120	15 : 2 : 15
5	496	32	496	31 : 2 : 31
<i>n</i>				$2^n - 1$: 2 : $2^n - 1$

In der 10. Generation z. B. ist $2^n - 1 = 1023$. Es gibt somit unter je 2048 Pflanzen, welche aus dieser Generation hervorgehen, 1023 mit dem constanten dominirenden, 1023 mit dem recessiven Merkmale und nur 2 Hybriden.

$$\begin{aligned} & \frac{AB}{AB} + \frac{AB}{Ab} + \frac{AB}{aB} + \frac{AB}{ab} + \frac{Ab}{AB} + \frac{Ab}{Ab} + \frac{Ab}{aB} + \frac{Ab}{ab} + \frac{aB}{AB} + \\ & \frac{aB}{Ab} + \frac{aB}{aB} + \frac{aB}{ab} + \frac{ab}{AB} + \frac{ab}{Ab} + \frac{ab}{aB} + \frac{ab}{ab}, \text{ oder} \\ & AB + ABb + AaB + AaBb + ABb + Ab + AaBb + Aab + AaB + \\ & AaBb + aB + aBb + AaBb + Aab + aBb + ab + AB + Ab + aB + \\ & ab + 2 ABb + 2 aBb + 2 AaB + 2 Aab + 4 AaBb. \end{aligned}$$

In ganz ähnlicher Weise erklärt sich die Entwicklungsreihe der Hybriden, wenn in denselben dreierlei differirende Merkmale

Mendelův článek

- Svěží čtení, nadšení z pokusů i výsledků
- Důsledně definuje pojmy, které používá
- Věnuje se svému konkrétnímu experimentu, včetně praktických potíží, od jednoduchého ke složitému
- Když čtenář pochopí značení, jsou postupy výborně popsány
- Věta (*der Satz*), zákony
- Je dobrý v aritmetice
- Dvojková soustava: dominantní vs. recesivní

- ① Nešetřil J., Nešetřilová H. (2017): Remarks on Mendel's mathematics. *Folia Mendeliana* **53**, 5–22.

Mendelův článek, náhoda a pravděpodobnost

- Návrh experimentu
 - Faktoriální (bifaktoriální, trifaktoriální)
- Formulace genetických zákonů
 - Binomické rozdělení
 - Nezávislost znaků
 - *“Je zcela ponecháno náhodě, který z obou druhů pylu se spojí s kterou jednotlivou vaječnou buňkou. Přitom na základě principu pravděpodobnosti v průměru mnoha případů musí se vždy ...”*
- Vyhodnocení výsledků
 - Chápe variabilitu
 - Chápe centrální limitní větu
 - Empirické četnosti (*“přibližné hodnoty”*) považuje za dostatečně blízké očekávaným, což bere za důkaz (*der Beweis*) vyslovených vět
 - Přesvědčivé

Gregor Mendel v kontextu historie statistiky

- Úvod
- Čím mě zaujal Mendelův článek
- **Příliš dobré výsledky**
- Jak Mendel ovlivnil historii statistiky

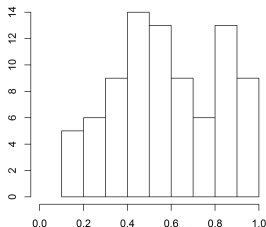
Co znamená shoda pozorovaných dat s očekávanými hodnotami?

- Nejjednodušší experimenty pro jednotlivých 7 znaků
- Teoretický poměr vždy 3:1

Experiment	Dominantní	Recesivní	Celkem	Poměr	<i>p</i> -hodnota
1	Tvar semen 5474	1850	7324	2,96:1	0,608
2	Zabarvení dělohy 6022	2001	8023	3,01:1	0,903
3	Barva květu 705	224	929	3,15:1	0,532
4	Tvar lusku 882	299	1181	2,91:1	0,801
5	Barva lusku 428	152	580	2,82:1	0,502
6	Umístění květu a lusku na stonku 651	207	858	3,14:1	0,554
7	Velikost stonku 787	277	1064	2,84:1	0,436

- H_0 : poměr 3:1 (χ^2 test)
- Pro první dva experimenty dává Mendel i mezivýsledky

p -hodnoty pro Mendelova data (χ^2 -test)



- Histogram spočten z Mendelových 84 experimentů
- Celková p -hodnota 0,999 97
- Fisher (1936): výsledky jsou až příliš dobré
- Možnost sekvenčního návrhu, který tehdy ovšem nebyl dostatečně prostudován (Prášková, 1991):
 - ❶ Havránek T. (1986): G. Mendel a experimentální data. *Vesmír* **65**, 331–333.
- Z nových výpočtů se zdá, že výsledky odpovídají dvoustupňovému sekvenčnímu postupu.

Statistické argumenty k tématu “příliš dobrých” Mendelových výsledků

- 1986: Edwards A.W.F. Are Mendel's results really too close? *Biological Reviews* **61**, 295–312.
- 1991: Prášková Z. Upravoval Mendel výsledky svých pokusů? *Informační bulletin České statistické společnosti* **2** (5), 19–24.
- 1994: Nissani M. Psychological, historical, and ethical reflections on the Mendelian paradox. *Perspectives in Biology and Medicine* **37**, 182–196.
- 2001: Fairbanks D.J., Rytting B. Mendelian controversies: A botanical and historical review. *American Journal of Botany* **88**, 737–752.
- 2004: Novitski C.E. Revision of Fisher's analysis of Mendel's garden pea experiments. *Genetics* **166**, 1139–1140.
- 2008: Franklin A., Edwards A.W.F., Fairbanks D.J., Hartl D.L., Seidenfeld T. *Ending the Mendel–Fisher Controversy*. University of Pittsburgh Press, Pittsburgh.
- 2010: Pires A.M., Branco J.A. A statistical model to explain the Mendel-Fisher controversy. *Statistical Science* **25**, 545–565.
- 2014: Kalina J. Gregor Mendel, his experiments and their statistical evaluation. *Acta Musei Moraviae, Scientiae biologicae* **99**(1), 87–99.
- 2016: Kalina J. Gregor Mendel's genetic experiments: A statistical analysis after 150 years. *European Journal for Biomedical Informatics* **12**(2), 20–26.

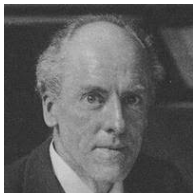
Gregor Mendel v kontextu historie statistiky

- Úvod
 - Čím mě zaujal Mendelův článek
 - Příliš dobré výsledky
 - **Jak Mendel ovlivnil historii statistiky**
-
- Kalina J. (2012): Ronald Fisher, otec biostatistiky. *Pokroky matematiky, fyziky a astronomie* **57**, 186–190.
 - Kalina J. (2011): 75. výročí úmrtí Karla Pearsona, zakladatele biometrie. *Biologie, chemie, zeměpis* **20**, 123–127.

Interpretace Mendelova objevu po 1900

- Škola biometriků: Karl Pearson (1857–1936)
 - Odmítali zákony dědičnosti
- Škola Mendelistů: William Bateson (1861–1926)
 - Pojem genetika (1906)
 - Statistiku kompletně odmítali
- Ronald A. Fisher (1890–1962)
 - Moderní (evoluční) syntéza
 - Obdivoval Mendela a jeho práci
 - Otec statistiky
 - Faktoriální design (Mendel: *die Faktoren*)

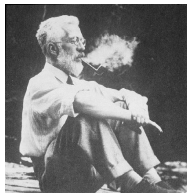
Pearson:



Bateson:



Fisher:



Moderní teorie navrhování experimentů

- Fisher (1935): *The Design of Experiments*.
- Cílem je studovat vliv různých faktorů na odezvu (např. vliv ošetření na terapii)
- Návrh vždy závisí na cíli studie
- Randomizace & replikace
- Možnost interakce mezi faktory
- Optimální postupy
- Kolik pozorování?
- Testování hypotéz

Mendel praotcem **Mendelovské randomizace**:

- Problém: praktická omezení při randomizovaných studiích
- Některé faktory nelze randomizovat (genetická informace, kouření, alkoholismus)
- Využití genetických a statistických principů v genetické epidemiologii, konkrétně využití 2. Mendelova zákona (o náhodné segregaci genů) při návrhu experimentu
- Statistická korekce systematické chyby (*confounding*), metoda instrumentálních proměnných

Gregor Mendel v kontextu historie statistiky: závěry

- Úvod
 - 200 let narození Gregora Mendela (* 20.7. 1822)
 - Čím mě zaujal Mendelův článek
 - Příliš dobré výsledky
 - Jak Mendel ovlivnil historii statistiky
 - S geny pracoval jako s černou skříňkou
 - Objev dědičné bioinformace
 - Předchůdce bioinformatiky
- 1 Sekerák J. (2008): Mendel v černé skříňce. Moravské zemské muzeum, Brno.