Lecture 34 Chi-square Tests For Categorical Data

BIO210 Biostatistics

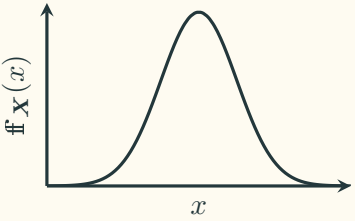
Xi Chen

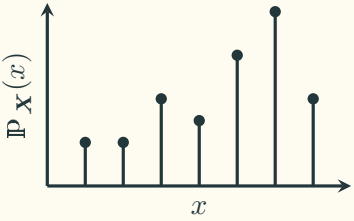
Spring, 2024

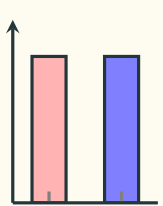
School of Life Sciences
Southern University of Science and Technology

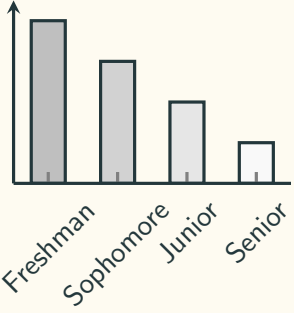
南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

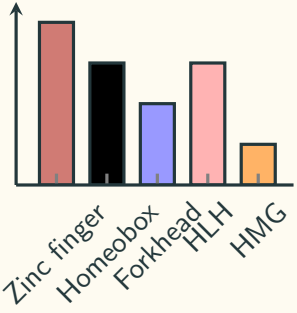
Categorical Distribution





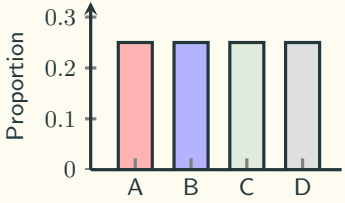




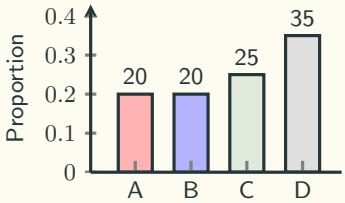


Goodness of Fit

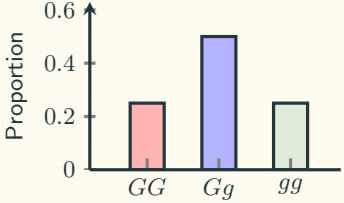
Population: Correct answers in all multiple choices (expectation)



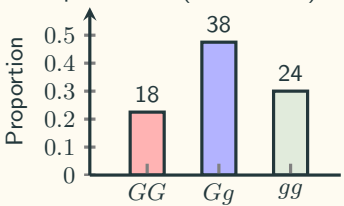
Sample: Correct answers in a sample $n = 100$ (observation)



Population: Theoretical genotype frequencies (expectation)



Sample:Genotype frequencies in a sample $n = 80$ (observation)



Observation vs Expectation

Does the observation agree with the expectation?

How good does the observation fit the expectation?

Chi-square Test For Goodness-of-fit

Example: Correct answers in multiple choices

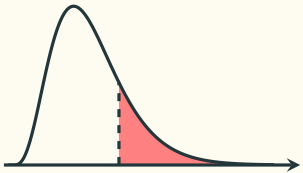
H_0 : equal distribution

H_1 : unequal distribution

$$\Updownarrow$$

$H_0 : \chi^2 = 0$

$H_1 : \chi^2 \neq 0 \Rightarrow \chi^2 > 0$



Observed:

Total	A	B	C	D
100	20	20	25	35

Expected (if H_0 were true):

Total	A	B	C	D
100	$100 \times 0.25 = 25$	$100 \times 0.25 = 25$	$100 \times 0.25 = 25$	$100 \times 0.25 = 25$

The test statistic:

$$\chi^2 = \sum_{\text{cells}} \frac{(O_i - E_i)^2}{E_i}, \nu = 3$$

$$= \frac{(20-25)^2}{25} + \frac{(20-25)^2}{25} + \frac{(20-25)^2}{25} + \frac{(20-25)^2}{25} = 6$$

Chi-square of The Multiple Choices Example

All multiple choices questions, equal distribution of correct answers

↓

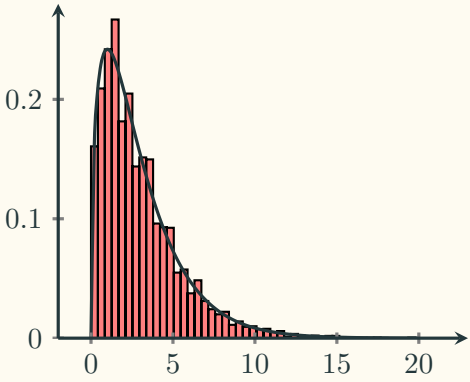
Sample $n = 100$

↓

Compute and record the test statistic:

$$\sum \frac{(O_i - E_i)^2}{E_i}$$

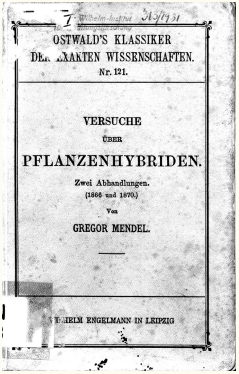
10,000 simulations



Mendel's Pea Plant Experiments

- Conducted between 1856 and 1863.
- Published in 1866: Versuche über Pflanzenhybriden (**Experiments on Plant Hybridization**).
- A classical piece in teaching biology.

1. The difference in the form of the ripe seeds.
2. The difference in the colour of the seed albumen.
3. The difference in the colour of the seed coat.
4. The difference in the form of the ripe pod.
5. The difference in the colour of the unripe pod.
6. The difference in the placement of the flowers.
7. The difference in the length of the stem.


OSTWALD'S KLASSIKER
DER EXAKTEN WISSENSCHAFTEN.
Nr. 121.

VERSUCHE
ÜBER
PFLANZENHYBRIDEN.
Zwei Abhandlungen.
(1866 und 1870.)
Von
GREGOR MENDEL.

WILHELM ENGELMANN IN LEIPZIG

Mendel's Pea Plant Experiments

Was die Gestalt der Hybriden in diesem Falle anbelangt, zeigten die Versuche übereinstimmend, dass dieselbe stets jener der beiden Stammpflanzen näher steht, welche die grössere Anzahl von dominirenden Merkmalen besitzt. Hat z. B. die Samenpflanze eine kurze Axe, endständige weisse Blüthen und einfach gewölbte Hülsen; die Pollenpflanze hingegen eine lange Axe, axenständige violett-rothe Blüthen und eingeschnürte Hülsen, so erinnert die Hybride nur durch die Hülsenform an die Samenpflanze, in den übrigen Merkmalen stimmt sie mit
[19] der Pollenpflanze überein. Besitzt eine der beiden Stammarten nur dominirende Merkmale, dann ist die Hybride von derselben kaum oder gar nicht zu unterscheiden.

Mit einer grösseren Anzahl Pflanzen wurden zwei Versuche durchgeführt. Bei dem ersten Versuche waren die Stammpflanzen in der Gestalt der Samen und in der Färbung des Albumens verschieden; bei dem zweiten in der Gestalt der Samen, in der Färbung des Albumens und in der Farbe der Samenschale. Versuche mit Samenmerkmalen führen am einfachsten und sichersten zum Ziele.

Um eine leichtere Uebersicht zu gewinnen, werden bei diesen Versuchen die differirenden Merkmale der Samenpflanze mit *A*, *B*, *C*, jene der Pollenpflanze mit *a*, *b*, *c* und die Hybridformen dieser Merkmale mit *Aa*, *Bb*, *Cc* bezeichnet.

Erster Versuch: *AB* Samenpflanze, *ab* Pollenpflanze,
A Gestalt rund, *a* Gestalt kantig,
B Albumen gelb, *b* Albumen grün.

Die befruchteten Samen erschienen rund und gelb, jenen der Samenpflanze ähnlich. Die daraus gezogenen Pflanzen gaben Samen von vierlei Art, welche oft gemeinschaftlich in einer Hülse lagen. Im Ganzen wurden von 15 Pflanzen 556 Samen erhalten, von diesen waren:

315 rund und gelb,
101 kantig und gelb,
108 rund und grün,
32 kantig und grün.

Alle wurden im nächsten Jahre angebaut. Von den runden gelben Samen gingen 11 nicht auf und 3 Pflanzen kamen nicht zur Fruchtbildung. Unter den übrigen Pflanzen hatten:

38 runde gelbe Samen . . . . . *AB*
65 runde gelbe und grüne Samen . . . . *ABb*

http://www.biolib.de/mendel/index.html

in ascertaining whether the law of development discovered in these applied to each pair of differentiating characters when several diverse characters are united in the hybrid by crossing. As regards the form of the hybrids in these cases, the experiments showed throughout that this invariably more nearly approaches to that one of the two parental plants which possesses the greater number of dominant characters. If, for instance, the seed plant has a short stem, terminal white flowers, and simply inflated pods; the pollen plant, on the other hand, a long stem, violet-red flowers distributed along the stem, and constricted pods; the hybrid resembles the seed parent only in the form of the pod; in the other characters it agrees with the pollen parent. Should one of the two parental types possess only dominant characters, then the hybrid is scarcely or not at all distinguishable from it.

Two experiments were made with a considerable number of plants. In the first experiment the parental plants differed in the form of the seed and in the colour of the albumen; in the second in the form of the seed, in the colour of the albumen, and in the colour of the seed-coats. Experiments with seed characters give the result in the simplest and most certain way.

In order to facilitate study of the data in these experiments, the different characters of the seed plant will be indicated by *A*, *B*, *C*, those of the pollen plant by *a*, *b*, *c*, and the hybrid forms of the characters by *Aa*, *Bb*, and *Cc*.

Expt. 1. — *AB*, seed parents; *ab*, pollen parents;
A, form round; *a*, form wrinkled;
B, albumen yellow. *b*, albumen green.

The fertilised seeds appeared round and yellow like those of the seed parents. The plants raised therefrom yielded seeds of four sorts, which frequently presented themselves in one pod. In all, 556 seeds were yielded by 15 plants, and of these there were:

315 round and yellow,
101 wrinkled and yellow,
108 round and green,
32 wrinkled and green.

https://www.biodiversitylibrary.org/item/124139#page/8/mode/1up

Mendel's Pea Plant Experiments

In order to facilitate study of the data in these experiments, the different characters of the seed plant will be indicated by A, B, C, those of the pollen plant by a, b, c, and the hybrid forms of the characters by Aa, Bb, and Cc.

Expt. 1. — AB, seed parents; ab, pollen parents;
A, form round; a, form wrinkled;
B, albumen yellow. b, albumen green.

The fertilised seeds appeared round and yellow like those of the seed parents. The plants raised therefrom yielded seeds of four sorts, which frequently presented themselves in one pod. In all, 556 seeds were yielded by 15 plants, and of these there were:

315 round and yellow,
101 wrinkled and yellow,
108 round and green,
32 wrinkled and green.

Parents: Seed $(AABB)$ $\times$ Pollen $(aabb)$

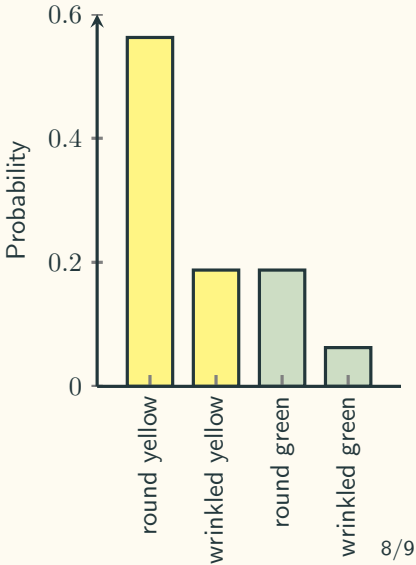
F1: round, yellow $(AaBb)$

F2:

Phenotype	observed
round, yellow	315
wrinkled, yellow	101
roiund, green	108
wrinkled, green	32

Mendel's Pea Plant Experiments - Analysis

		Pollen $(AaBb)$			
		AB	Ab	aB	ab
Seed $(AaBb)$	AB	$AABB$	$AABb$	$AaBB$	$AaBb$
	Ab	$AABb$	$AAbb$	$AaBb$	$Aabb$
	aB	$AaBB$	$AaBb$	$aaBB$	$aaBb$
	ab	$AaBb$	$Aabb$	$aaBb$	$aabb$



Mendel's Pea Plant Experiments - Analysis

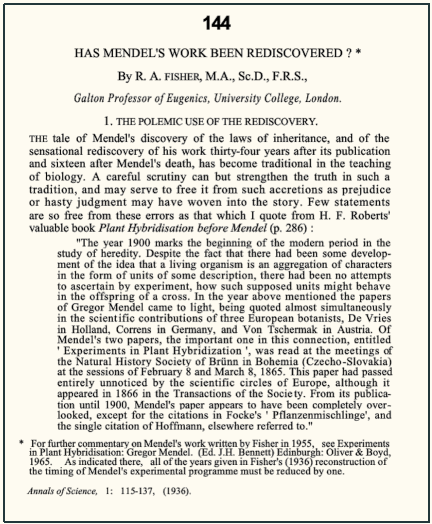
H_0 : ratio is 9:3:3:1

H_1 : ratio is not 9:3:3:1

Phenotype	Observed	Expected
round, yellow	315	312.75
wrinkled, yellow	101	104.25
round, green	108	104.25
wrinkled, green	32	34.75
Total	556	556

$$\chi^2 = \sum_{\text{cells}} \frac{(O_i - E_i)^2}{E_i} = 0.47\,,$$

$$p = \mathbb{P}\left(\chi^2_3 \geqslant 0.47\right) = 0.925$$



144

HAS MENDEL'S WORK BEEN REDISCOVERED ? *

By R. A. FISHER, M.A., Sc.D., F.R.S.,

Galton Professor of Eugenics, University College, London.

1. THE POLEMIC USE OF THE REDISCOVERY.

THE tale of Mendel's discovery of the laws of inheritance, and of the sensational rediscovery of his work thirty-four years after its publication and sixteen after Mendel's death, has become traditional in the teaching of biology. A careful scrutiny can but strengthen the truth in such a tradition, and may serve to free it from such accretions as prejudice or hasty judgment may have woven into the story. Few statements are so free from these errors as that which I quote from H. F. Roberts' valuable book *Plant Hybridisation before Mendel* (p. 286) :

"The year 1900 marks the beginning of the modern period in the study of heredity. Despite the fact that there had been some development of the idea that a living organism is an aggregation of characters in the form of units of some description, there had been no attempts to ascertain by experiment, how such supposed units might behave in the offspring of a cross. In the year above mentioned the papers of Gregor Mendel came to light, being quoted almost simultaneously in the scientific contributions of three European botanists, De Vries in Holland, Correns in Germany, and Von Tschermak in Austria. Of Mendel's two papers, the important one in this connection, entitled ' Experiments in Plant Hybridization ', was read at the meetings of the Natural History Society of Brünn in Bohemia (Czecho-Slovakia) at the sessions of February 8 and March 8, 1865. This paper had passed entirely unnoticed by the scientific circles of Europe, although it appeared in 1866 in the Transactions of the Society. From its publication until 1900, Mendel's paper appears to have been completely overlooked, except for the citations in Focke's ' Pflanzenmischlinge', and the single citation of Hoffmann, elsewhere referred to."

* For further commentary on Mendel's work written by Fisher in 1955, see Experiments in Plant Hybridisation: Gregor Mendel. (Ed. J.H. Bennett) Edinburgh: Oliver & Boyd, 1965. As indicated there, all of the years given in Fisher's (1936) reconstruction of the timing of Mendel's experimental programme must be reduced by one.

Annals of Science, 1: 115-137, (1936).