

Chapter 10

χ^2 tests for goodness of fit and independence

Prof. Tesler

Math 186
Winter 2018

Multinomial test

- Consider a k -sided die with faces $1, 2, \dots, k$.
- We want to simultaneously test that the probabilities p_1, p_2, \dots, p_k of rolling $1, 2, \dots, k$ are specified values.
- To test if a 6-sided die is fair,
$$H_0: (p_1, \dots, p_6) = (1/6, \dots, 1/6)$$
$$H_1: \text{At least one } p_i \neq 1/6$$
- Decision rule is based counting # 1's, 2's, etc. on n independent rolls of the die.
- For the fair coin problem, the exact distribution was binomial, and we approximated it with a normal distribution.
- For this problem, the exact distribution is multinomial.
We will combine the separate counts of $1, 2, \dots$ into a single test statistic whose distribution is approximately a χ^2 distribution.

10.3 Goodness of fit tests for Mendel's experiments

- In Mendel's pea plant experiments, yellow seeds (Y) are dominant and green (y) recessive; round seeds (R) are dominant and wrinkled (r) are recessive.
- Consider the phenotypes of the offspring in a "dihybrid cross" $YyRr \times YyRr$:

| Type | Expected fraction | Observed number |
|-------------------|-------------------|------------------|
| yellow & round | 9/16 | 315 |
| yellow & wrinkled | 3/16 | 101 |
| green & round | 3/16 | 108 |
| green & wrinkled | 1/16 | 32 |
| | | Total: $n = 556$ |

- Hypothesis test:

$$H_0: (p_1, p_2, p_3, p_4) = \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right)$$

H_1 : At least one p_i disagrees

Does the data fit the expected distribution?

| Type | Expected fraction | Observed number |
|-------------------|-------------------|-----------------|
| yellow & round | 9/16 | 315 |
| yellow & wrinkled | 3/16 | 101 |
| green & round | 3/16 | 108 |
| green & wrinkled | 1/16 | 32 |
| | Total: | $n = 556$ |

- The observed number of “yellow & round” plants is $O = 315$. (Don’t confuse the letter O with the number 0.)
- The expected number is
$$E = (9/16) \cdot 556 = 312.75.$$
- The goodness of fit test requires that we convert all the expected proportions into expected numbers.

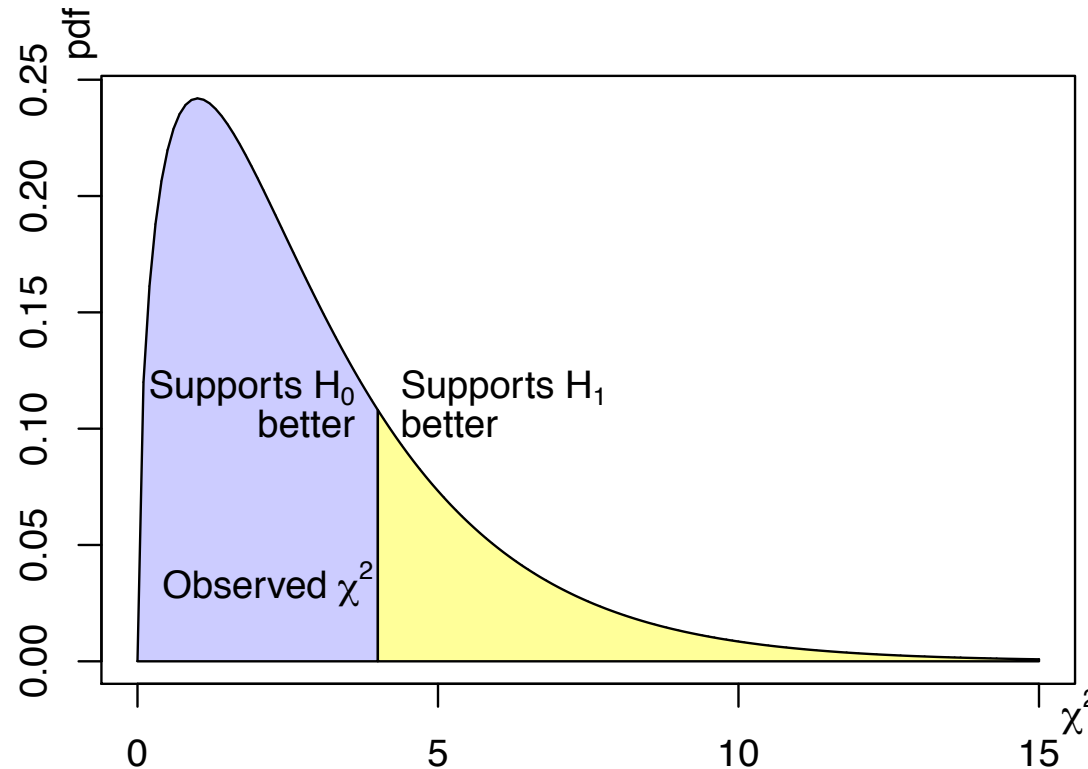
Goodness of fit test

| Type | Observed number O | Expected number E | $O - E$ | $(O - E)^2 / E$ |
|-------------------|------------------------|------------------------|----------|------------------|
| yellow & round | 315 | $(9/16)556 = 312.75$ | 2.25 | 0.0161871 |
| yellow & wrinkled | 101 | $(3/16)556 = 104.25$ | -3.25 | 0.1013189 |
| green & round | 108 | $(3/16)556 = 104.25$ | 3.75 | 0.1348921 |
| green & wrinkled | 32 | $(1/16)556 = 34.75$ | -2.75 | 0.2176259 |
| Total | 556 | 556 | 0 | 0.4700240 |

- $k = 4$ categories give $k - 1 = 3$ degrees of freedom.
(The O and E columns both total 556, so the $O - E$ column totals 0; thus, any 3 of the $(O - E)$'s dictate the fourth.)
- The test statistic is the total of the last column, $\chi_3^2 = 0.4700240$.
- The general formula is $\chi_{k-1}^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$.
- **Warning:** Technically, that formula only has an approximate chi-squared distribution. When $E \geq 5$ in all categories, the approximation is pretty good.

Goodness of fit test

- Smaller values of χ^2 indicate better agreement between the O and E values (so support H_0 better). Larger values support H_1 better.
It's a one-sided test.

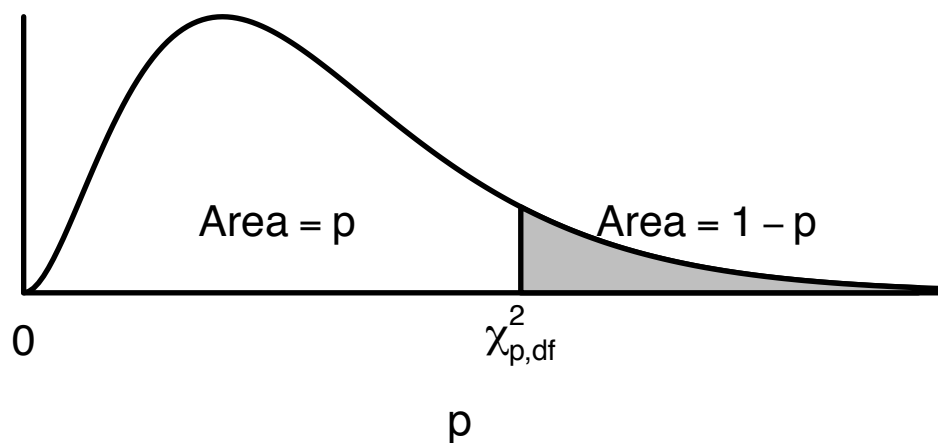


- In the χ^2 table, look at the row $df = 3$ to find 0.4700240; it's between $0.05 < p < 0.10$.
- Thus, $P(\chi_3^2 \leq 0.4700240)$ is between 0.05 and 0.10.
(With a computer, it's $P(\chi_3^2 \leq 0.4700240) = 0.0745741$.)

See χ^2 table in the back of the book (Table A.3)

Look up CDF of $\chi_3^2 = 0.4700240$; get $.05 < \text{CDF} < .10$.

χ^2 Distribution with df Degrees of Freedom



| df | 0.010 | 0.025 | 0.050 | 0.10 | 0.90 | 0.95 | 0.975 | 0.99 |
|----|----------|----------|--------------|--------------|--------|--------|--------|--------|
| 1 | 0.000157 | 0.000982 | 0.00393 | 0.015 | 2.705 | 3.841 | 5.023 | 6.634 |
| 2 | 0.020 | 0.050 | 0.102 | 0.210 | 4.605 | 5.991 | 7.377 | 9.210 |
| 3 | 0.114 | 0.215 | 0.351 | 0.584 | 6.251 | 7.814 | 9.348 | 11.344 |
| 4 | 0.297 | 0.484 | 0.710 | 1.063 | 7.779 | 9.487 | 11.143 | 13.276 |
| 5 | 0.554 | 0.831 | 1.145 | 1.610 | 9.236 | 11.070 | 12.832 | 15.086 |
| 6 | 0.872 | 1.237 | 1.635 | 2.204 | 10.644 | 12.591 | 14.449 | 16.811 |
| 7 | 1.239 | 1.689 | 2.167 | 2.833 | 12.017 | 14.067 | 16.012 | 18.475 |
| 8 | 1.646 | 2.179 | 2.732 | 3.489 | 13.361 | 15.507 | 17.534 | 20.090 |
| 9 | 2.087 | 2.700 | 3.325 | 4.168 | 14.683 | 16.918 | 19.022 | 21.665 |

Goodness of fit test

- $P(\chi_3^2 \leq 0.4700240) = 0.0745741$ is not too extreme.
It means that if H_0 is true and the experiment is repeated a lot, about 7.5% of the time, a χ_3^2 value supporting H_0 better (lower values of χ_3^2) will be obtained, and about 92.5% of the time, values supporting H_1 better (higher values of χ_3^2) will be obtained.
- *P*-value: The *P*-value is the probability, under H_0 , of a test statistic that supports H_1 as well as or better than the observed value:

$$P = P(\chi_3^2 \geq 0.4700240) = 1 - P(\chi_3^2 \leq 0.4700240) = .9254259$$

Connection to original χ^2 test

Technically, use of χ^2 for the “goodness of fit test” and “contingency tables” is just an approximation. The motivation:

- Recall $\chi_n^2 = Z_1^2 + \dots + Z_n^2$ if Z_i 's are i.i.d. standard normal.
- Our random variable is a count, O_i , the observed # of events.
- Approximate pdf of O_i by a Poisson distribution with

Mean $\lambda = E_i = \text{expected number of events}$

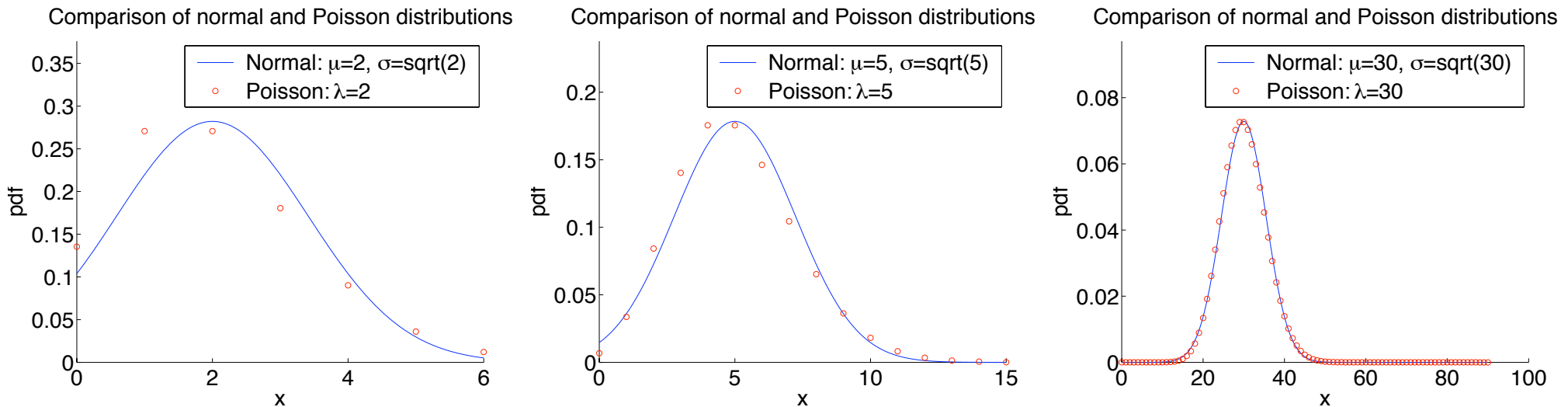
SD $\sigma = \sqrt{\lambda} = \sqrt{E_i}$

”z-score” $Z_i = \frac{O_i - E_i}{\sqrt{E_i}}$ (but it's not really a normal distribution)

- $Z_i^2 = (O_i - E_i)^2 / E_i$ in this notation.

Connection to original χ^2 test

- This approximates the normal distribution ($\mu = \lambda$, $\sigma = \sqrt{\lambda}$) pretty well for $\lambda \geq 5$ due to the Central Limit Theorem.



- The Z_i 's are not independent though, so we have $d.f. = n - 1$ in the goodness of fit test (and $d.f.$ reduced more in contingency tables).
- See Chapter 10 in book for a rigorous explanation.

Ronald Fisher (1890–1962)

- He made important contributions to both statistics and genetics.
- Connection: he invented statistical methods while working on genetics problems.
- Our way of using the normal, Student t , and χ^2 distributions in the same framework, is due to him.
- In genetics, he reconciled continuous variations (heights and weights) with Mendelian genetics (discrete traits), and developed much of population genetics.

Did Mendel fudge his data?

- For *independent experiments*, the values of χ^2 may be “pooled” by adding the χ^2 values and adding the degrees of freedom.
- Fisher pooled the data from Mendel’s experiments and got $\chi^2 = 41.6056$ with 84 degrees of freedom.
- Assuming Mendel’s laws are true, how often would we get χ^2_3 supporting H_0/H_1 better than this?

Support H_0 better:

$$P(\chi_{84}^2 \leq 41.6056) = 0.00002873$$

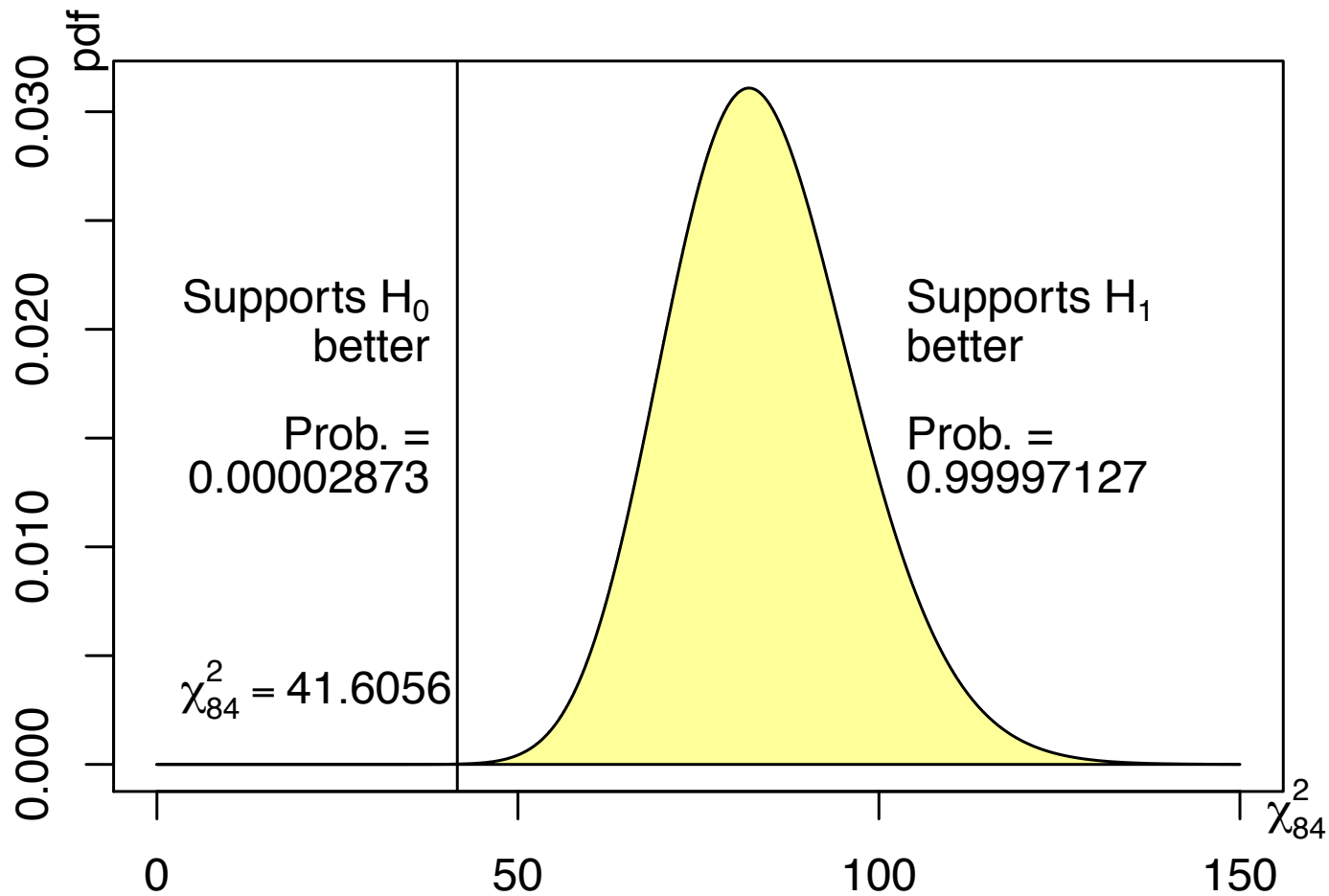
(on a computer; this is beyond what’s in the table in our book).

Support H_1 better:

$$P\text{-value } P = P(\chi_{84}^2 \geq 41.6056) = 1 - 0.00002873 = .99997127.$$

- So if Mendel’s laws hold and 1 million researchers independently conducted the same experiments as Mendel, about 29 of them would get data with as little or even less variation than Mendel had.

Did Mendel fudge his data?



Did Mendel fudge his data?

Based on this and similar tests, Fisher believed that something was fishy with Mendel's data:

- The values are “too good” in the sense that they are too close to what was expected.
- At the same time, they are “bad” in the sense that there is too little random variation.
- Some people have accused Mendel of faking data.
- Others speculate that he only reported his best data.
- Other people defend Mendel by speculating on biological explanations for why his results would be better than expected.
- All pro and con arguments have later been rebutted by someone else.

10.5 Tests of independence (“contingency tables”)

A study in 1899 examined 6800 German men to see if hair color and eye color are related.

| Observed counts O : | | Hair color | | | | Total |
|-----------------------|------------|------------|-------|------|-----|--------------|
| | | Brown | Black | Fair | Red | |
| Eye Color | Brown | 438 | 288 | 115 | 16 | 857 |
| | Gray/Green | 1387 | 746 | 946 | 53 | 3132 |
| | Blue | 807 | 189 | 1768 | 47 | 2811 |
| Total | | 2632 | 1223 | 2829 | 116 | 6800 |

Hypothesis test (at $\alpha = 0.05$)

H_0 : eye color and hair color are independent, vs.

H_1 : eye color and hair color are correlated

Meaning of independence

For all eye colors x and all hair colors y :

$$P(\text{eye color}=x \text{ and hair color}=y) = P(\text{eye color}=x) \cdot P(\text{hair color}=y)$$

Computing E table

Hypothesis test (at $\alpha = 0.05$)

H_0 : eye color and hair color are independent, vs.

H_1 : eye color and hair color are correlated

- The fraction of people with red hair is $116/6800$.
- The fraction with blue eyes is $2811/6800$.
- Use these as point estimates: $P(\text{hair color}=\text{red}) \approx 116/6800$ and $P(\text{eye color}=\text{blue}) \approx 2811/6800$.
- Under the null hypothesis, the fraction with red hair and blue eyes would be $\approx (116 \cdot 2811)/6800^2$.
- The expected number of people with red hair and blue eyes is $6800(116 \cdot 2811)/6800^2 = (116 \cdot 2811)/6800 = 47.95$.
(Row total times column total divided by grand total.)
- Compute E this way for all combinations of hair and eye color. As long as $E \geq 5$ in every cell (here it is) and the data is normally distributed (an assumption), the χ^2 test is valid.

Computing E and $O - E$ tables

| Expected counts E : | | Hair color | | | |
|-----------------------|------------|------------|--------|---------|-------|
| | | Brown | Black | Fair | Red |
| Eye | Brown | 331.71 | 154.13 | 356.54 | 14.62 |
| Color | Gray/Green | 1212.27 | 563.30 | 1303.00 | 53.43 |
| | Blue | 1088.02 | 505.57 | 1169.46 | 47.95 |

In each position, compute $O - E$.

For red hair and blue eyes, this is $O - E = 47 - 47.95 = -.95$:

| $O - E$: | | Hair color | | | |
|-----------|------------|------------|---------|---------|-------|
| | | Brown | Black | Fair | Red |
| Eye | Brown | 106.29 | 133.87 | -241.54 | 1.38 |
| Color | Gray/Green | 174.73 | 182.70 | -357.00 | -0.43 |
| | Blue | -281.02 | -316.57 | 598.54 | -0.95 |

Note all the row and column sums in the $O - E$ table are 0, so if we hid the last row and column, we could deduce what they are. Thus, this 3×4 table has $(3 - 1)(4 - 1) = 6$ degrees of freedom.

Computing test statistic χ^2

Compute $(O - E)^2/E$ in each position.

For red hair and blue eyes, this is $(-.95)^2/47.95 = 0.0189$.

(You could go directly to this computation after the E computation, without doing $O - E$ first.)

| $(O - E)^2/E$: | | Hair color | | | |
|-----------------|------------|------------|----------|----------|--------|
| | | Brown | Black | Fair | Red |
| Eye Color | Brown | 34.0590 | 116.2632 | 163.6301 | 0.1304 |
| | Gray/Green | 25.1852 | 59.2571 | 97.8139 | 0.0034 |
| | Blue | 72.5845 | 198.2220 | 306.3398 | 0.0189 |

Add all twelve of these to get

$$\chi^2 = 34.0590 + \dots + 0.0189 = 1073.5076$$

There are 6 degrees of freedom, so $\chi_6^2 = 1073.5076$.

Performing the test of independence

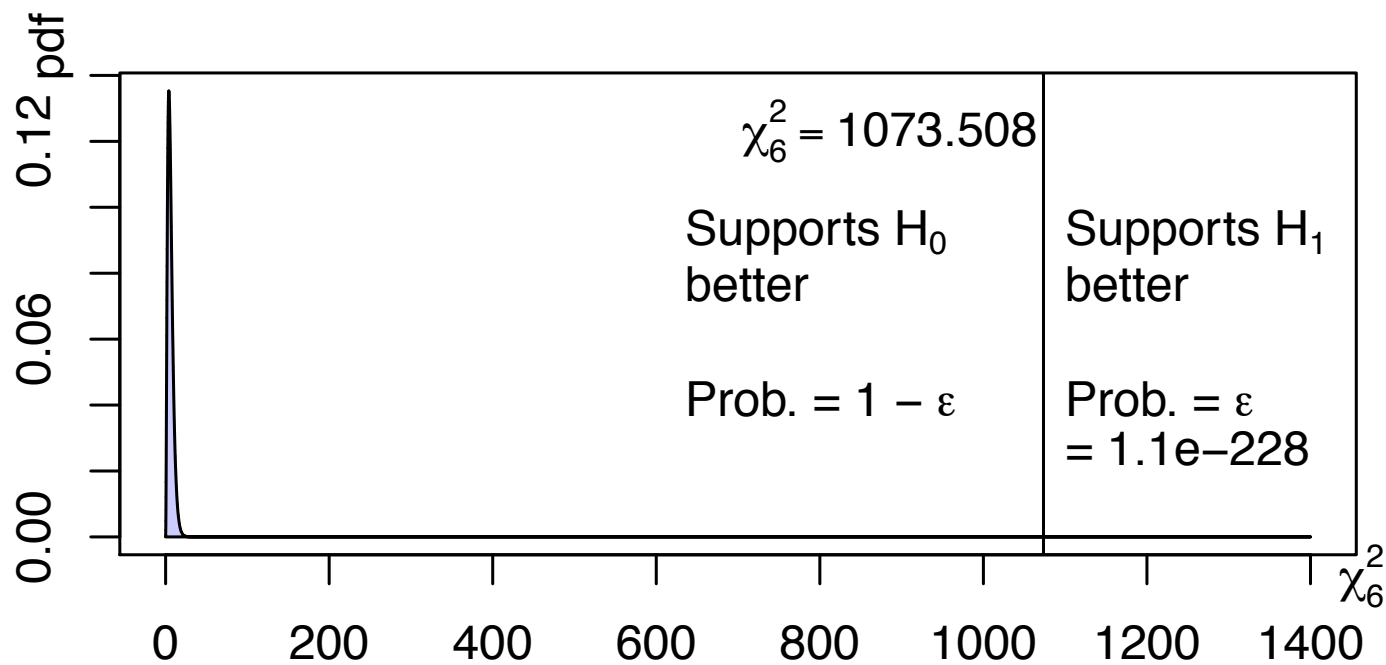
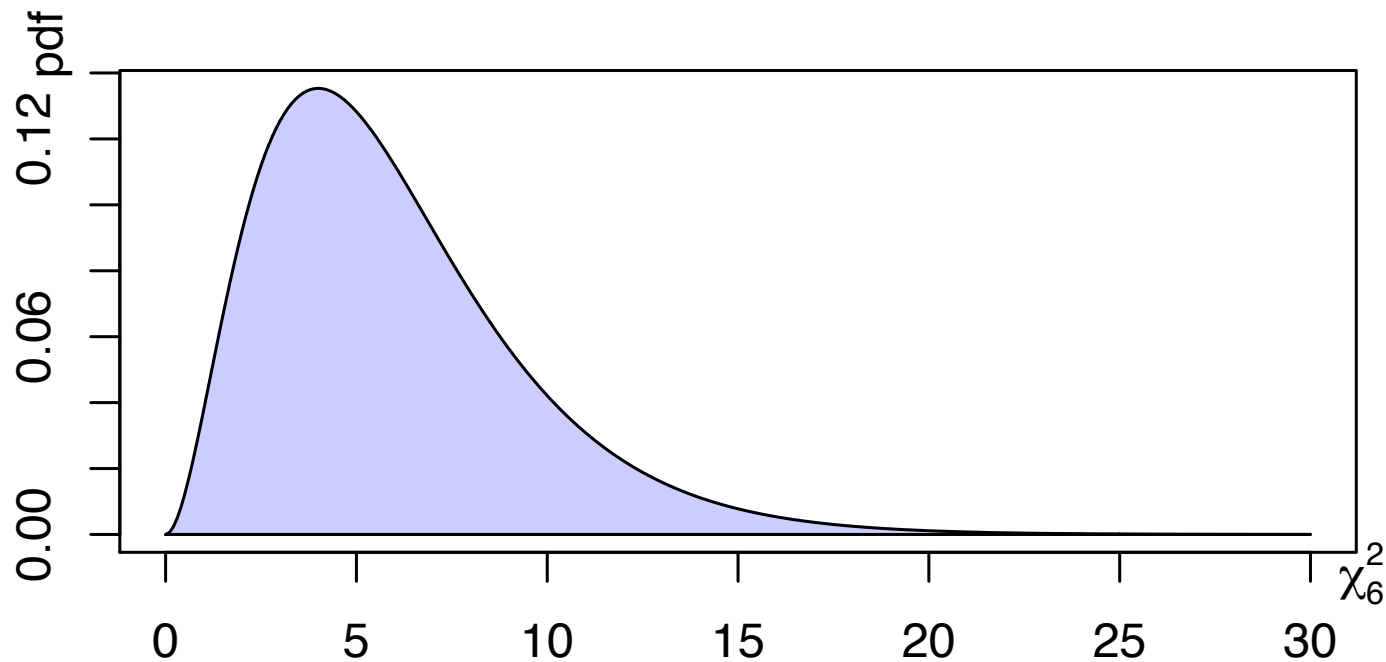
- χ^2 would be 0 if the traits were truly independent. Smaller values support H_0 better (traits independent). Larger values support H_1 better (traits correlated).
It's a one-sided test.

- At the 0.05 level of significance, we reject H_0 if

$$\chi_6^2 \geq \chi_{0.95,6}^2 = 12.592$$

- Indeed, $1073.5076 > 12.592$ so we reject H_0 and conclude that hair color and eye color are linked in this data.
- This doesn't prove that a particular hair color causes one to have a particular eye color, or vice-versa; it just says there's a correlation in this data.
- Using P -values: $P \approx 1.1 \cdot 10^{-228}$.
So $P \leq \alpha = 0.05$ and we reject H_0 .

Performing the test of independence



Another application

- Test the effects of a medical treatment.
- The rows could be different dosages of a medication (or different types of medication) and placebo.
- The columns could be “cured,” “improved,” “no effect,” “worsened,” “dead.”

Mendel's pea plants revisited: Are loci Y and R linked?

We will use the same data as in the goodness-of-fit test but for a different purpose. Consider the phenotypes of the offspring in a “dihybrid cross” $YyRr \times YyRr$:

| Observed counts O : | | Seed Shape | | Total |
|-----------------------|----------------|---------------|------------------|--------------|
| | | Round (R) | Wrinkled (r) | |
| Seed Color | Yellow (Y) | 315 | 101 | 416 |
| | Green (y) | 108 | 32 | 140 |
| Total | | 423 | 133 | 556 |

Hypothesis test (at $\alpha = 0.05$)

H_0 : Seed color and seed shape are independent, vs.

H_1 : Seed color and seed shape are correlated

Mendel's pea plants revisited: Are loci Y and R linked?

| | | Seed Color | | Seed Shape | | Total |
|---|----------------|---------------|------------------|----------------|---------------|--------------|
| | | Round (R) | Wrinkled (r) | Yellow (Y) | Green (y) | |
| O : (Observed #) | Yellow (Y) | 315 | 101 | | | 416 |
| | Green (y) | 108 | 32 | | | 140 |
| Total | | 423 | 133 | | | 556 |
| E : (Expected #) | Yellow (Y) | 316.4892 | 99.5108 | | | 416 |
| | Green (y) | 106.5108 | 33.4892 | | | 140 |
| Total | | 423 | 133 | | | 556 |
| $O - E$: (Deviation) | Yellow (Y) | -1.4892 | 1.4982 | | | 0 |
| | Green (y) | 1.4892 | -1.4892 | | | 0 |
| Total | | 0 | 0 | | | 0 |
| $(O - E)^2 / E$: (χ^2 contrib.) | Yellow (Y) | 0.0070 | 0.0223 | | | 0.0293 |
| | Green (y) | 0.0208 | 0.0662 | | | 0.0870 |
| Total | | 0.0278 | 0.0885 | | | 0.1163 |

Mendel's pea plants revisited: Are loci Y and R linked?

- Using χ^2 as the test statistic:

$$df = (2 - 1)(2 - 1) = 1$$

$$\chi_1^2 = .0070 + .0223 + .0208 + .0662 = 0.1163$$

$\chi_{0.95,1}^2 = 3.8415$ but $.1163 < 3.8415$ so it's not significant

- Using P -values:

$$P = P(\chi_1^2 \geq 0.1163) = 0.7330$$

$P \geq 0.05$ so Accept H_0 (genes not linked)

Comparison of the two tests

- At fertilization, if genes R and Y are not linked, then in an $RrYy \times RrYy$ cross, the expected proportions are

$$RY:Ry:rY:ry = 1:1:1:1.$$

If linked, it would be different.

- Some genotypes may not survive to the points at which the phenotype counts are made; e.g., hypothetically, 40% of individuals with Rr might not be born, might die before reproducing (affecting multigenerational experiments), etc.

This would change the ratio of

$$RR:Rr:rr \text{ from } 1:2:1 \text{ to } 1:1.2:1 = 5:6:5,$$

and round:wrinkled from 3:1 to 2.2:1 = 11:5.

Comparison of the two tests

- The goodness-of-fit test assumed all genotypes are equally viable.

Whether the genes are linked or not should be a separate matter.

If you know the yellow:green and round:wrinkled viability ratios, you can use the goodness-of-fit test on 4 phenotypes with 3 degrees of freedom by adjusting the proportions.

- If you don't know these viability ratios, you can estimate the ratios from data via contingency tables, at the cost of dropping to 1 degree of freedom.