χ^2 hypothesis test

Let $p_1,\ldots,p_k$ be a probability distribution (so $p_1+\ldots+p_k=1$)

Let $N_1,\ldots,N_k$ be an observation from a multinomial distribution with parameters $(p_1,\ldots,p_k)$ and n trials. Define

$$X^2 := \sum_{i=1}^{k} \frac{(\text{observed}_i - \text{expected}_i)^2}{\text{expected}_i}$$

$$= \frac{(N_1 - np_1)^2}{np_1} + \cdots + \frac{(N_k - np_k)^2}{np_k}$$

Theorem: As $n\to\infty$, X^2 converges to the χ^2 distribution with (k-1) degrees of freedom.

X^2 hypothesis test

This is the basis of a test of the null hypothesis, that the data ($n_1,\ldots,n_k$) were drawn from the multinomial distribution with parameters ($p_1,\ldots,p_k$).

1. Compute the X^2 statistic from the data.

2. Compute the probability that X^2 with k-1 degrees of freedom is bigger than this value. This is the p-value.

3. If you're working with a table, it will simply give you the value such that X^2 has a fixed probability of being bigger than it. For instance, if you're testing at the .05 level, you look in the corresponding column, and reject the null if the X^2 you compute from the data is bigger than the number on the table.

Even under the null hypothesis, the X^2 we compute from the data should only have <u>approximately</u> χ^2 distribution.

How big should n be for the distribution to be "close enough"?

Rule of thumb: The expected value in each class should be significantly bigger than 1.

Example: Did the atomic bomb cause leukemia in the survivors?

Null hypothesis: No effect. In other words, leukemia cases are randomly distributed through the city.

If we break up the city into regions, n leukemia cases should be distributed among the regions like n picks from a multinomial distribution, with p=proportion of the population

Alternative hypothesis: People living closer to the bomb site should be more likely to develop leukemia.

Table 1.—*Leukemia Incidence in Exposed Residents of Hiroshima*

Hiroshima City	Distance from Hypocenter in Meters				Total Exposed
	Under 1,000	1,000-1,499	1,500-1,999	2,000-10,000	
No. Cases 1950-1957 *	15	32	9	14	69
Population **	1282	10557	17654	60999	90492
Estimated Annual Incidence ***	1460	380	57	29	95

* Resident of Hiroshima City at Time of Onset.

** 1953 Census Figures.

*** Per Million Per Year. Eight Years at Risk Assumed.

Probabilities	Expected numbers	Observed numbers
$p_1=1282/90492=.014$	$E_1=69p_1=.98$	$N_1=15$
$p_2=10557/90492=.117$	$E_2=8.0$	$N_2=32$
$p_3=17654/90492=.195$	$E_3=13$	$N_3=9$
$p_4=60999/90492=.674$	$E_4=46$	$N_4=14$

Probabilities	Expected numbers	Observed numbers
$p_1 = 1282/90492 = .014$	$E_1 = 69p_1 = .98$	$N_1 = 15$
$p_2 = 10557/90492 = .117$	$E_2 = 8.0$	$N_2 = 32$
$p_3 = 17654/90492 = .195$	$E_3 = 13$	$N_3 = 9$
$p_4 = 60999/90492 = .674$	$E_4 = 46$	$N_4 = 14$

Class 1 is kind of small, so we lump it together with class 2.

Probabilities	Expected numbers	Observed numbers
$p_{1,2} = 11839/90492 = .131$	$E_{1,2} = 9.0$	$N_{1,2} = 47$
$p_3 = 17654/90492 = .195$	$E_3 = 13$	$N_3 = 9$
$p_4 = 60999/90492 = .674$	$E_4 = 46$	$N_4 = 14$

Probabilities	Expected numbers	Observed numbers
$p_{1,2}=11839/90492=.131$	$E_{1,2}=9.0$	$N_{1,2}=47$
$p_3=17654/90492=.195$	$E_3=13$	$N_3=9$
$p_4=60999/90492=.674$	$E_4=46$	$N_4=14$

$$X^2 = \frac{(47-9.0)^2}{9.0} + \frac{(9-13)^2}{13} + \frac{(14-46)^2}{46} = 184$$

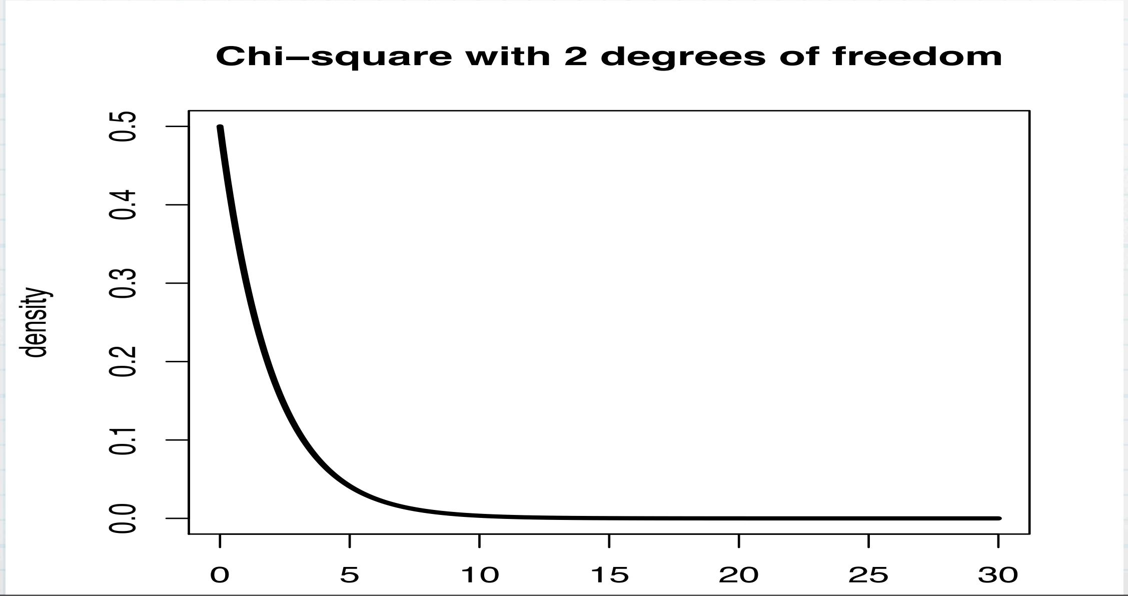
Testing at .01 significance level:
Rejection threshold 9.2

Probabilities	Expected numbers	Observed numbers
$p_{1,2}$=11839/90492=.131	$E_{1,2}$=9.0	$N_{1,2}$=47
p_3=17654/90492=.195	E_3=13	N_3=9
p_4=60999/90492=.674	E_4=46	N_4=14

$$X^2 = \frac{(47-9.0)^2}{9.0} + \frac{(9-13)^2}{13} + \frac{(14-46)^2}{46} = 184$$

Testing at .01 significance level:

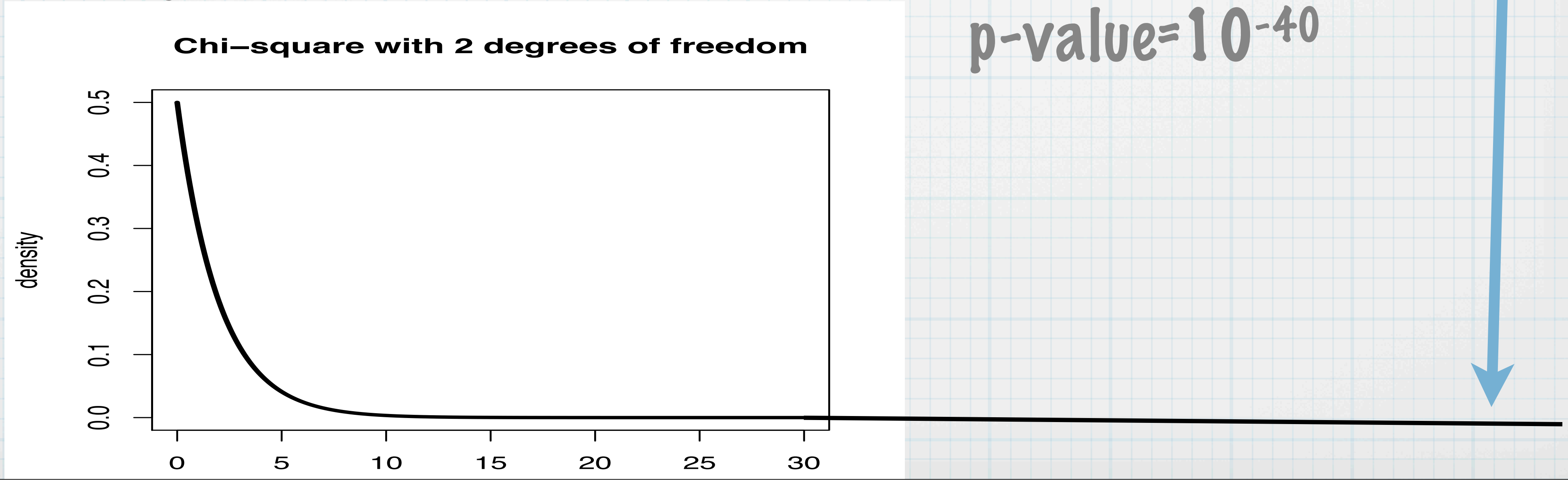
Rejection threshold 9.2



Probabilities	Expected numbers	Observed numbers
$p_{1,2}=11839/90492=.131$	$E_{1,2}=9.0$	$N_{1,2}=47$
$p_3=17654/90492=.195$	$E_3=13$	$N_3=9$
$p_4=60999/90492=.674$	$E_4=46$	$N_4=14$

$$X^2 = \frac{(47-9.0)^2}{9.0} + \frac{(9-13)^2}{13} + \frac{(14-46)^2}{46} = 184$$

Testing at .01 significance level:
Rejection threshold 9.2



p-value=10^{-40}

200 random digits

3 1415926535 8979323846 2643383279 5028841971 6939937510
5820974944 5923078164 0628620899 8628034825 3421170679
8214808651 3282306647 0938446095 5058223172 5359408128
4811174502 8410270193 8521105559 6446229489 549303819

Could these have come from a uniform distribution?

Digit	0	1	2	3	4	5	6	7	8	9
# occurrences	19	20	24	20	22	20	15	12	25	23

Could these have come from a uniform distribution?

Digit	0	1	2	3	4	5	6	7	8	9
# occurrences	19	20	24	20	22	20	15	12	25	23

For χ^2 with 9 degrees of freedom, testing at .05 significance level, the rejection threshold is 16.9. For 7.2 the p-value is 0.62.

Could these have come from a uniform distribution?

Digit	0	1	2	3	4	5	6	7	8	9
# occurrences	19	20	24	20	22	20	15	12	25	23

$$X^2 = \frac{(19-20)^2}{20} + \frac{(20-20)^2}{20} + \frac{(24-20)^2}{20} + \frac{(20-20)^2}{20}$$

$$+ \frac{(22-20)^2}{20} + \frac{(20-20)^2}{20} + \frac{(15-20)^2}{20}$$

$$+ \frac{(12-20)^2}{20} + \frac{(25-20)^2}{20} + \frac{(23-20)^2}{20} = 7.2$$

For χ^2 with 9 degrees of freedom, testing at .05 significance level, the rejection threshold is 16.9. For 7.2 the p-value is 0.62.

Mendel's peas

smooth is dominant over wrinkled, yellow dominant over green.
Theory predicts 3/4 dominant, and characters are independent.

Mendel's data for 556 peas:

Proportions aren't exactly right.
Could the deviations be due to chance?

	smooth	wrinkled
yellow	315	108
green	102	31

Expected

	smooth	wrinkled
yellow	556*9/16 =312.75	556*3/16 =104.25
green	556*3/16 =104.25	556*1/16 =34.75

observed

	smooth	wrinkled
yellow	315	108
green	102	31

$$X^2 = \frac{2.25^2}{312.75} + \frac{3.75^2}{104.25} + \frac{2.25^2}{104.25} + \frac{3.75^2}{34.75} = 0.6$$

For χ^2 with 3 degrees of freedom, testing at .05 significance level, the rejection threshold is 7.8. For 0.6 the p-value is 0.9.

That is, you would expect to get a fit this good only one time in 10.

V2 Rocket problem

537 bomb hits in London. Were they targeted?

Idea: Divide the city up into 576 equal-sized squares. If no targeting, the rockets should be equally likely to fall in any square.

Then the number of hits in a square should be Poisson distributed, all with the same distribution.

	0	1	2	3	4	5	6	7	8+
# squares	229	211	93	35	7	0	0	1	0

Null hypothesis: Number of hits in each square is Poisson distributed with the same parameter λ. In other words, the probability of getting k hits is $\frac{e^{-\lambda}\lambda^k}{k!}$

If we put each grid square into a box for the number of hits it got, the probability of box k is $p_k(\lambda) = e^{-\lambda}\lambda^k/k!$.

The counts of squares in each box should be like a draw from a multinomial with parameters $(p_0(\lambda), p_1(\lambda), p_2(\lambda), p_3(\lambda), p_{4+}(\lambda))$.

Problem: Which λ?

Find the best fit (method of moments or MLE):

$$\hat{\lambda} = \text{average \# hits} = .9323$$

New null hypothesis: The counts in the boxes are like 576 draws from a multinomial with parameters $p_k(\hat{\lambda})$.

	0	1	2	3	4+
# squares	229	211	93	35	8
Poisson prob.	.395	.367	.171	.0529	.0145
Expected (Poisson)	226.7	211.4	98.5	30.6	8.7

	0	1	2	3	4+
# squares	229	211	93	35	8
Poisson prob.	.395	.367	.171	.0529	.0145
Expected (Poisson)	226.7	211.4	98.5	30.6	8.7

$$X^2 = \frac{2.3^2}{226.7} + \frac{0.4^2}{211.4} + \frac{5.5^2}{98.5} + \frac{4.4^2}{30.6} + \frac{0.7^2}{8.7} = 1.02$$

Threshold for χ^2 with 4 degrees of freedom at the .05 significance level is 9.5. So, we don't reject. p-value=0.9.

But wait... We cheated. We said we were testing whether these came from any Poisson distribution. But we only computed χ^2 for the one Poisson distribution with the best fit -- that is, the lowest χ^2. Would that tend to make our p-value too high or too low?

But wait... We cheated. We said we were testing whether these came from any Poisson distribution. But we only computed χ^2 for the one Poisson distribution with the best fit -- that is, the lowest χ^2. Would that tend to make our p-value too high or too low?

Parameter-fitting correction: If you chose which distribution to compare to your observations by fitting j parameters, the χ^2 statistic you compute should have χ^2 distribution with k-j-1 degrees of freedom. This is a theorem (kind of hard...)

But wait... We cheated. We said we were testing whether these came from any Poisson distribution. But we only computed χ^2 for the one Poisson distribution with the best fit -- that is, the lowest χ^2. Would that tend to make our p-value too high or too low?

Parameter-fitting correction: If you chose which distribution to compare to your observations by fitting j parameters, the χ^2 statistic you compute should have χ^2 distribution with k-j-1 degrees of freedom. This is a theorem (kind of hard...)

For the V2 data: 3 degrees of freedom.
Threshold at .05 significance level is now 7.8.
p-value is now 0.8.

But wait... We cheated. We said we were testing whether these came from any Poisson distribution. But we only computed χ^2 for the one Poisson distribution with the best fit -- that is, the lowest χ^2. Would that tend to make our p-value too high or too low?

Parameter-fitting correction: If you chose which distribution to compare to your observations by fitting j parameters, the χ^2 statistic you compute should have χ^2 distribution with k-j-1 degrees of freedom. This is a theorem (kind of hard...)

For the V2 data: 3 degrees of freedom.
Threshold at .05 significance level is now 7.8.
p-value is now 0.8.

That is, without the correction we overestimated how good the fit was.