CHAPTER 9

Hypothesis testing about the variance

Is the die loaded?

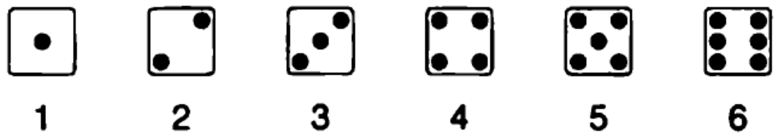
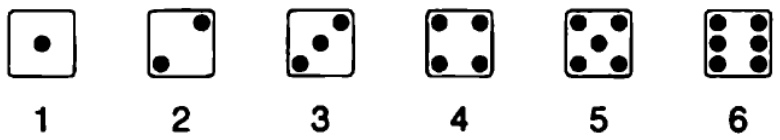


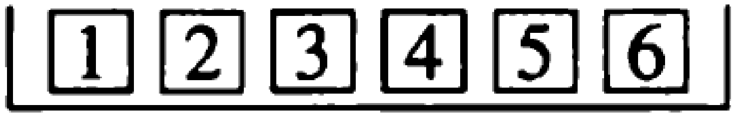
Table 1. Sixty rolls of a die, which may be loaded.

4	3	3	1	2	3	4	6	5	6
2	4	1	3	3	5	3	4	3	4
3	3	4	5	4	5	6	4	5	1
6	4	4	2	3	3	2	4	4	5
6	3	6	2	4	6	4	6	3	2
5	4	6	3	3	3	5	3	1	4

Categorical Random Variable



The Box Model



The random variable is not Numerical
The random variable is "Categorical"

There are 6 mutually exclusive categories
Each observation falls in one and only one category

We don't calculate the average number of spots of the 60 observations
We tabulate the frequency with which each category occurs

Observed frequencies

Roll the die $n = 60$ times.

The sample average is 3.75;
but we are interested in the total distribution

Table 1. Sixty rolls of a die, which may be loaded.

4	3	3	1	2	3	4	6	5	6
2	4	1	3	3	5	3	4	3	4
3	3	4	5	4	5	6	4	5	1
6	4	4	2	3	3	2	4	4	5
6	3	6	2	4	6	4	6	3	2
5	4	6	3	3	3	5	3	1	4

Value	*Observed frequency*
1	4
2	6
3	17
4	16
5	8
6	9
sum	60

Incorrect: *z*-test

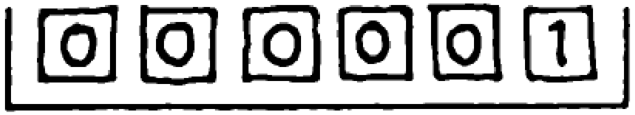
3 spots: high frequency

n = 60
Count = Sum of 1s = 17
Expected value = 60 x 1/6 = 10

Average of Box = 1/6
SD of Box = $\sqrt{1/6 \times 5/6} \approx 0.37$
SE = $\sqrt{60} \times$ SD of Box ≈ 2.9

z = (17 - 10) / 2.9 = 2.4
p = 1%

Box Model



Value	*Observed frequency*
1	4
2	6
3	17
4	16
5	8
6	9
sum	60

"Data Snooping"

Incorrect Null Hypothesis:
"The chance of getting 3 spots is 1/6"

This Null Hypothesis is formulated
after the fact (!).

With multiple categories,
one of them is likely
to have a large *z*-value

Must formulate Null Hypothesis
before you run the trial

Correct Null hypothesis:
"The die is fair"

Value	*Observed frequency*	*Expected frequency*
1	4	10
2	6	10
3	17	10
4	16	10
5	8	10
6	9	10
sum	60	60

Pearson's Chi-squared statistic

$$\chi^2 = \text{sum of } \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}} \qquad = 14.2$$

Value	*Observed frequency*	*Expected frequency*
1	4	10
2	6	10
3	17	10
4	16	10
5	8	10
6	9	10
sum	60	60

$$\frac{(4-10)^2}{10} + \frac{(6-10)^2}{10} + \frac{(17-10)^2}{10} + \frac{(16-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(9-10)^2}{10}$$

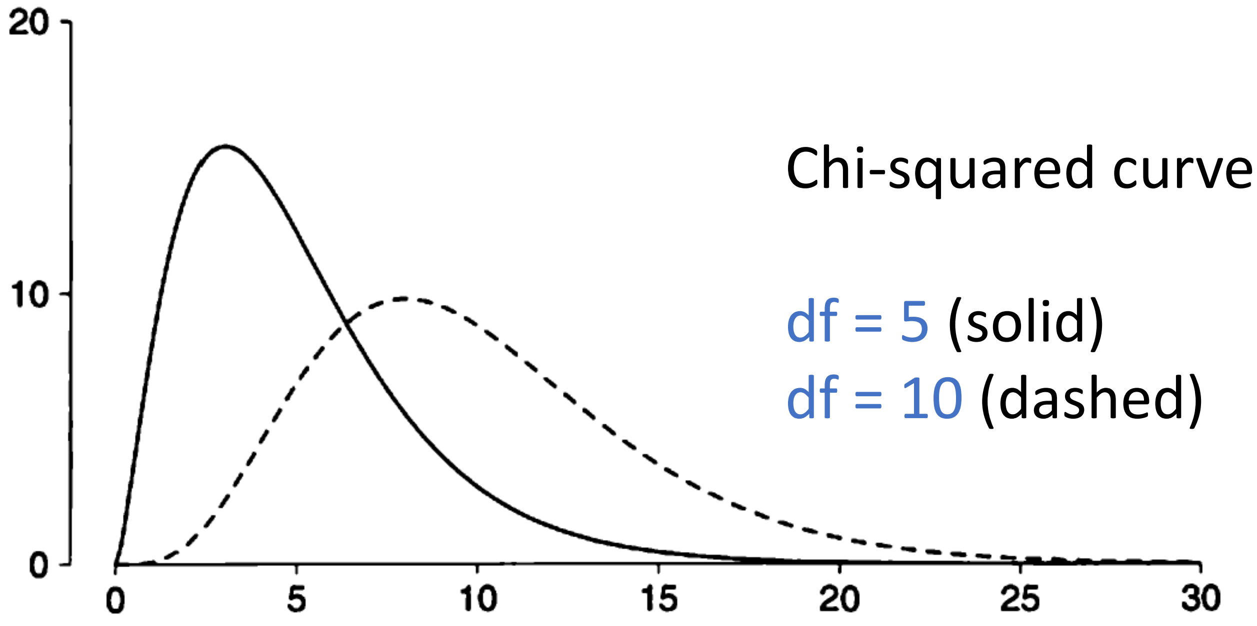
$$= \frac{142}{10} = 14.2$$

Chi-squared distribution: *P*-value

Chi-squared = 14.2
Number of categories: k = 6
Degrees of freedom: k – 1 = 5

P = 1 - CHISQ.DIST(14.2, 5, TRUE) = 1.4%

P = CHISQ.TEST(observed range, expected range)

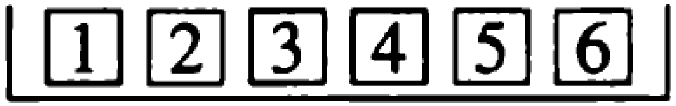


Chi-squared curves

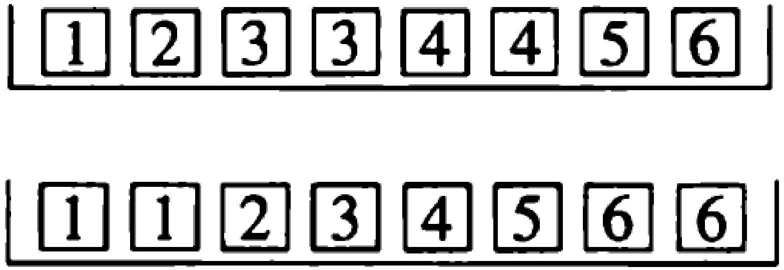
df = 5 (solid)
df = 10 (dashed)

Chi-square versus *z*-test

- *z*-test: to compare the average of a sample with an expected average
- Chi-squared test: to compare the entire distribution of the sample with an expected distribution



Chi-squared:
Hypothesis is about distribution of categories box



z-test:
Hypothesis is only about average of numbers in box

Input for the Chi-squared test

- Observed frequencies for all categories
- Null hypothesis: Expected frequencies for all categories

P = CHISQ.TEST(observed range, expected range)

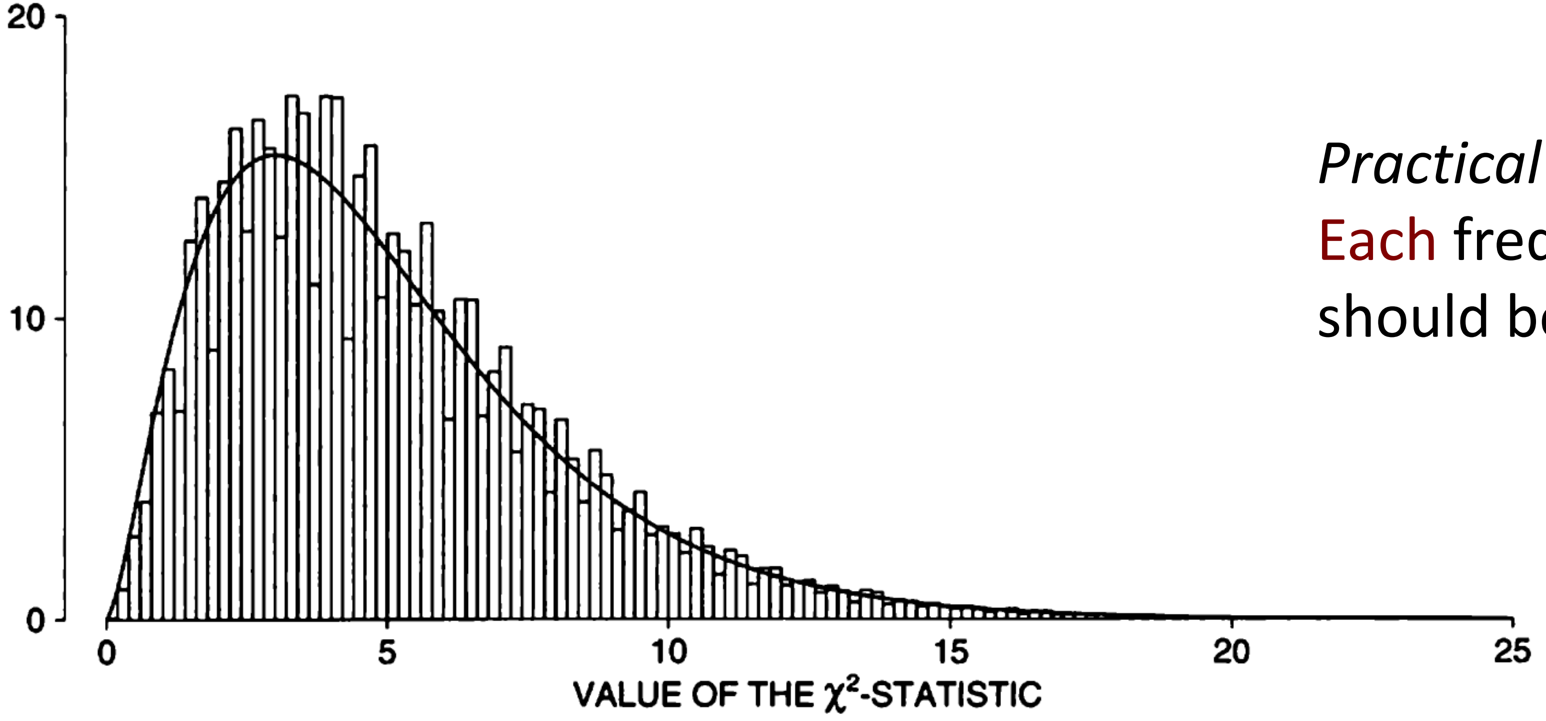
n = number of “draws” from the box

Plays no role in the calculation

Degrees of freedom = k – 1 (with k = number of categories)

Chi-squared curve: Approximate for large *n*

The real distribution of the chi-squared statistic for 60 rolls of a die
is more jagged than the theoretical chi-squared curve



Practical requirement:
Each frequency in the table
should be at least 5

Chi-squared test for independence

One "Box"

Two Categorical Random Variables

Use Chi-squared test to see whether the two variables are independent?

Handedness and gender

	Men	*Women*
Right-handed	934	1,070
Left-handed	113	92
Ambidextrous	20	8

	Men	*Women*
Right-handed	87.5%	91.5%
Left-handed	10.6%	7.9%
Ambidextrous	1.9%	0.7%

Calculating Expected Values

Null hypothesis: Gender and Handedness are independent

P(man and left-handed) = P(man) x P(left-handed)

	Men	Women	Total
Right-handed	934	1,070	2,004
Left-handed	113	92	205
Ambidextrous	20	8	28
Total	1,067	1,170	2,237

P(man) = 1,067 / 2,237 = 47.7%

P(left-handed) = 205/2,237 = 9.16%

Expected

P(left-handed men) = 4.37%

0.0437 x 2,237 = 97.7

Observed and Expected Frequencies

	Observed		*Expected*	
	Men	*Women*	*Men*	*Women*
Right-handed	934	1,070	956	1,048
Left-handed	113	92	98	107
Ambidextrous	20	8	13	15

Differences

−22	22
15	−15
7	−7

Chi-squared = 12

$$
\begin{aligned}
\chi^2 &= \text{sum of } \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}} \\
&= \frac{(934-956)^2}{956} + \frac{(1{,}070-1{,}048)^2}{1{,}048} \\
&\quad + \frac{(113-98)^2}{98} + \frac{(92-107)^2}{107} \\
&\quad + \frac{(20-13)^2}{13} + \frac{(8-15)^2}{15} \\
&\approx 12
\end{aligned}
$$

Degrees of freedom = 2

Degrees of freedom for m x n table = (m - 1) x (n – 1)

Degrees of freedom in 3 x 2 table = 2 x 1 = 2

−22	22
15	−15
7	−7

Conclusion

Chi-squared = 12
Degrees of freedom = 2
P = 1 – CHISQ.TEST(12, 2, TRUE) = 2.5%

Small P-value:

The Null hypothesis is rejected:
Based on the HANES data, handedness and gender are very likely not independent

Null hypothesis:
"Handedness and gender are independent"

	Men	*Women*
Right-handed	87.5%	91.5%
Left-handed	10.6%	7.9%
Ambidextrous	1.9%	0.7%

Chi-squared statistics can be pooled

Two or more independent experiments:

- Can add the separate chi-squared statistics
- Can add up the degrees of freedom

(Do not need to know the sample sizes)

$$\chi^2 = 5.8 \text{ with 5 degrees of freedom}$$
$$\chi^2 = 3.1 \text{ with 2 degrees of freedom}$$

Total Chi-squared = 5.8 + 3.1 = 8.9

Total degrees of freedom = 5 + 2 = 7

P = 1 – CHISQ.DIST(8.9, 7, TRUE) = 26%

When *P* is very large

Fisher's use of chi-squared test on Mendel's data

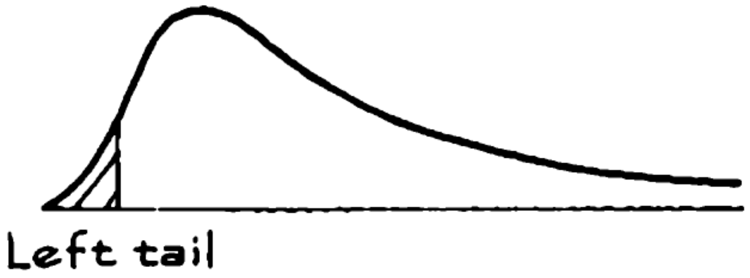
Type of pea	*Observed number*	*Expected number*
Smooth yellow	315	313
Wrinkled yellow	101	104
Smooth green	108	104
Wrinkled green	32	35

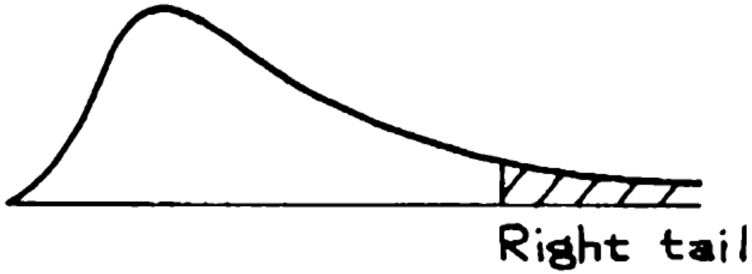
Chi-squared = 0.51,
degrees of freedom = 3

CHISQ.DIST(0.51, 3, TRUE) = 8.3%

CHISQ.TEST(..., ...) = 91.7%

P = 91.7%





Fisher and Mendel

Fisher pooled chi-squared stats for many of Mendel's genetic experiments.

Null hypothesis: *"Mendel's genetic model is correct"*

The pooled P-value was very large.
We can not reject the Null hypothesis.

But: The probability of getting experimental data
that are this close to the expected outcome is extremely small

Most likely: Mendel's experimental data were manipulated.