

13

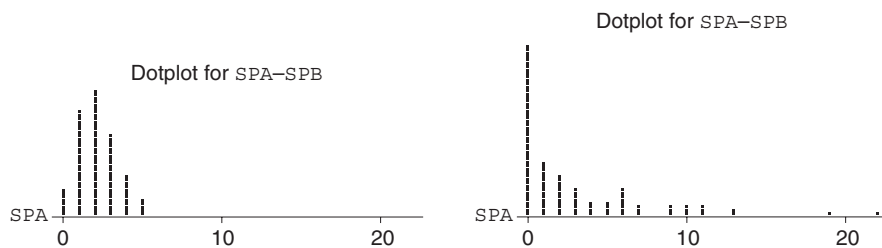
Categorical data

13.2 The Poisson distribution

The number of parasites per host are plotted here on the same axes. Combining the appearance of these histograms with descriptive data (means and variances), it is thought that Species A may follow a Poisson distribution, and Species B an aggregated distribution. This hypothesis is tested in the main text.

MINITAB COMMANDS FOR FIG. 13.2 Two plots of the parasite data

Commands	<pre>%dotplot 'SPA' 'SPB'; onepage.</pre>
Menu route	Graph > Dotplot SPA SPB → Variables ⊙ Each column constitutes a group



Minitab output for Fig. 13.2 Two dotplots of the parasite data

We can also obtain descriptive statistics for these distributions by the normal route:

MINITAB COMMANDS FOR BOX 13.1 Descriptive statistics of the parasite data

Commands	<pre>describe SPA SPB.</pre>
Menu route	Stat > Basic Statistics > Display Descriptive Statistics SPA SPB → Variables

MINITAB OUTPUT FOR BOX 13.1 Descriptive statistics of the parasite data

Descriptive Statistics: SPA, SPB

Variable	N	Mean	Median	TrMean	StDev	SE Mean
SPA	119	2.126	2.000	2.093	1.232	0.113
SPB	119	2.739	1.000	2.215	4.045	0.371

Variable	Minimum	Maximum	Q1	Q3
SPA	0.000	5.000	1.000	3.000
SPB	0.000	22.000	0.000	4.000

This information can then be used to perform a dispersion test, as described in the main text. The first step is to calculate the chi-squared value, and the second to obtain a p -value. This can be done for species A using the following commands:

MINITAB COMMANDS FOR BOX 13.2 A dispersion test for the helminth parasites in bird species A

Commands

```
let K1 = 118*(1.232**2)/2.126.

cdf K1;
chisquare 118.
```

Menu route

```
Calc > Calculator
K1 → Store result in variable
118*(1.232**2)/2.126 → Expression

Calc > Probability Distributions > Chi-square
⊙ Cumulative Probability
118 → Degree of freedom
⊙ Input constant
K1 → Input constant
```

MINITAB OUTPUT FOR BOX 13.2 A dispersion test for the helminth parasites in bird species A

Cumulative Distribution Function

Chi-Square with 118 DF

X	P(X ≤ x)
84.2442	0.00081

Step 1

The chi-squared value is calculated from the variance: mean ratio and degrees of freedom for species A, and is stored in K1. Under the null hypothesis these data

are from a Poisson distribution, so κ_1 should come from the chi-squared distribution with 118 degrees of freedom.

Step 2

The second and third lines of command input ask for the probability of getting the value κ_1 , or anything less than this, if it comes from a chi-squared distribution with 118 degrees of freedom. In other words, it asks for the area of the probability distribution of χ_{118}^2 to the left of κ_1 . In reply, Minitab gives the probability of 0.0081.

Step 3

Is this the final probability for our test? No—the dispersion test is applied here as a two-tailed test, because we are interested in whether the variance is greater or less than the mean. To convert this p -value into the appropriate p -value for our test, it should be multiplied by 2. So the final answer for the dispersion test is a p -value of 0.0162. In this case the variance is significantly less than the mean, and we conclude that the distribution is underdispersed (more uniform than expected).

This may then be repeated for Species B using the same commands.

13.3 The chi-square test in contingency tables

Chi-square tests can be conducted in Minitab in two different ways. When the table is entered into the worksheet as you would lay it out on the page (in other words, as below),

	Hermaphrodite	Monoecious	Dioecious
Herbs	379	102	19
Shrubs	88	30	31
Trees	56	40	12

then the chi-square test command can be used as follows:

MINITAB COMMANDS FOR THE CHI-SQUARE TEST	
Commands	ChiSquare Hermaphrodite Monoecious Dioecious.
Menu route	Stat > Tables > Chi-square test Hermaphrodite Monoecious Dioecious → Columns containing the table

MINITAB OUTPUT FOR THE CHI-SQUARE TEST

Chi-Square Test: Hermaphrodite, Monoecious, Dioecious

Expected counts are printed below observed counts

	Hermaphr	Monoecio	Dioeciou	Total
1	379 345.44	102 113.61	19 40.95	500
2	88 102.94	30 33.85	31 12.20	149
3	56 74.62	40 24.54	12 8.85	108
Total	523	172	62	757

Chi-Sq = 3.260 + 1.186 + 11.767 +
 2.169 + 0.439 + 28.952 +
 4.644 + 9.741 + 1.125 = 63.282

DF = 4, P-Value = 0.000

However, if the data are entered in the usual subscripts form (i.e. mating strategy coded as 1, 2 or 3 and trees size as 1, 2 or 3), then a chi-square test is done via the cross-tabulation command. This is more flexible, in that it allows you to request the standardised residual for each cell. This can be helpful in the interpretation of significance patterns (see main text).

MINITAB COMMANDS FOR BOX 13.4 **Chi-squared analysis of Darwin's hypothesis**

Commands table STRATEGY PLSIZE;
 frequencies NOSP;
 chisquare 3.

Menu route Stat > Tables > Cross Tabulation
 STRATEGY PLSIZE → Classification variables
 Chi-square analysis
 Above and std. residual
 Frequencies are in:
 NOSP → Frequencies are in

MINITAB OUTPUT FOR BOX 13.4 Chi-squared analysis of Darwin's hypothesis

Tabulated Statistics: STRATEGY, PLSIZE

Rows: STRATEGY

Columns: PLSIZE

	1	2	3	All
1	379	88	56	523
	345.44	102.94	74.62	523.00
	1.81	-1.47	-2.16	--
2	102	30	40	172
	113.61	33.85	24.54	172.00
	-1.09	-0.66	3.12	--
3	19	31	12	62
	40.95	12.20	8.85	62.00
	-3.43	5.38	1.06	--
All	500	149	108	757
	500.00	149.00	108.00	757.00
	--	--	--	--

Chi-Square = 63.282, DF = 4, P-Value = 0.000

Cell Contents --

Count
Exp Freq
St Resid

No new commands are introduced for the rest of this chapter, so we move onto the exercises.

13.6 Exercises

Soya beans revisited

MINITAB COMMANDS FOR BOX 13.14 Soya beans revisited

```
Commands      name C3 = 'SQRTDAM'
              let SQRTDAM = SQRT(DAMAGE)

              glm SQRTDAM = WDKLR;
              brief 3.
```

Menu route Calc > Calculator
 SQRTDAM → Store result in variable
 SQRT (DAMAGE) → Expression

 Stat > ANOVA > General Linear Model
 SQRTDAM → Response
 WDKLR → Model

Results...

⊙ In addition, coefficients for all terms

MINITAB OUTPUT FOR BOX 13.14 **Soya beans revisited**

General Linear Model: SQRTDAM versus WDKLR

Factor	Type	Levels	Values
WDKLR	fixed	3	1 2 3

Analysis of Variance for SQRTDAM, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
WDKLR	2	46.180	46.180	23.090	83.18	0.000
Error	21	5.829	5.829	0.278		
Total	23	52.010				

Term	Coef	SE Coef	T	P
Constant	5.3268	0.1075	49.53	0.000
WDKLR				
1	-1.6860	0.1521	-11.09	0.000
2	-0.0256	0.1521	-0.17	0.868

Fig trees in Costa Rica

The Minitab output for this exercise may be found in the answers to exercises in Chapter 14.