

Worksheet 5 - Analysis of frequencies

Frequency analysis references

- Fowler *et al.* (1998) - Chpt 13
- Holmes *et al.* (2006) - Chpt 5
- Quinn & Keough (2002) - Chpt 14

Question 1 - Goodness of fit test

A fictitious plant ecologist sampled 90 shrubs of a dioecious plant in a forest, and each plant was classified as being either male or female. The ecologist was interested in the sex ratio and whether it differed from 50:50. The observed counts and the predicted (expected) counts based on a theoretical 50:50 sex ratio follow.

Format of fictitious plant sex ratios - note, not a file

Expected and Observed data (50:50 sex ratio).

	Female	Male	Total
Observed	40	50	90
Expected	45	45	90



Note, it is not necessary to open or create a data file for this question.

Q1-1. First, what is the **appropriate test** to examine the sex ratio of these plants?

Q1-2. What null hypothesis is being tested by this test?

Q1-3. What are the degrees of freedom associated with this data for this test?

Q1-4. Perform a **Goodness-of-fit test** to test the null hypothesis that these data came from a population with a 50:50 sex ratio. Identify the following:

a. X^2 statistic

b. df

c. P value

Q1-5. What are your conclusions (statistical and biological)?

Lets now extend this fictitious endeavor. Recent studies on a related species of shrub have suggested a 30:70 female:male sex ratio. Knowing that our plant ecologist had similar research interests, the authors contacted her to inquire whether her data contradicted their findings.

Q1-6. Using the same observed data, **test the null hypothesis** that these data came from a population with a 30:70 sex ratio. From a 30:70 female:male sex ratio, what are the expected frequency counts of females and males from 90 individuals and what is the X^2 statistic?.

a. Expected number of females

b. Expected number of males

c. X^2 statistic


Q1-7. Do the plant ecologist's data dispute the findings of the other studies? (y or n)

Question 2 - Contingency tables

Here is a modified example from Quinn and Keough (2002). Following fire, French and Westoby (1996) cross-classified plant species by two variables: whether they regenerated by seed only or vegetatively and whether they were dispersed by ant or vertebrate vector. The two variables could not be distinguished as response or predictor since regeneration mechanisms could just as conceivably affect dispersal mode as vice versa.

Format of french.csv data files

REGENERATION MODE	DISPERSAL MODE	
	ANT	Vertebrate
SeedOnly	25	6
Vegetative	61	21



Note, it is not necessary to open or create a data file for this question.

Q2-1. What null hypothesis is being tested by this test?

Q2-2. Using a **2 x 2 (two way) contingency table**, test this null hypothesis and identify the following:

a. χ^2 statistic

b. df

c. P value

Q2-3. What are your conclusions (statistical and biological)?

Question 3 - Contingency tables

Arrington et al. (2002) examined the frequency with which African, Neotropical and North American fishes have empty stomachs and found that the mean percentage of empty stomachs was around 16.2%. As part of the investigation they were interested in whether the frequency of empty stomachs was related to dietary items. The data were separated into four major trophic classifications (detritivores, omnivores, invertivores, and piscivores) and whether the fish species had greater or less than 16.2% of individuals with empty stomachs. The number of fish species in each category combination was calculated and a subset of that (just the diurnal fish) is provided.

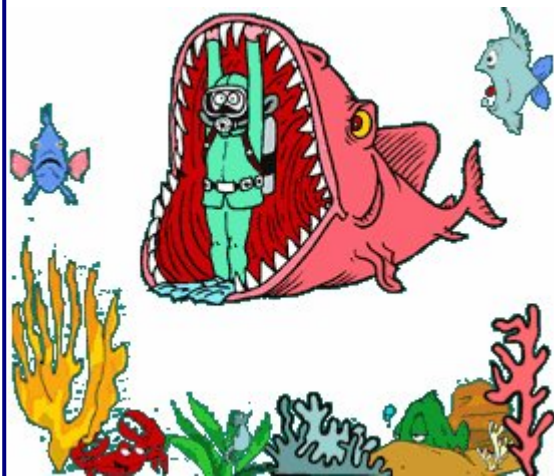
Format of arrington.csv data file

STOMACH	TROPIC
< 16.2	DET
..	..
< 16.2	OMN
..	..
< 16.2	PISC
..	..
< 16.2	INV
..	..

STOMACH Categorical listing of the proportion of individuals in the species with empty stomachs (< 16.2% or > 16.2%).

TROPIC Categorical listing of the trophic classification (DET = detritovore, OMN = omnivore, INV = invertivore, PISC = piscivore).

Trophic classification	% Stomachs empty	
	< 16.2	> 16.2
DET	18	4
OMN	45	8
INV	58	15
PISC	16	34



[Open](#) the arrington data file.

Q3-1. Using a **two-way contingency table**, test the null hypothesis that the percentage of empty stomachs was independent of trophic classification. What would you conclude from the analysis? Write the results out as though you were writing a research paper/thesis. For example (select the phrase that applies and fill in gaps with your results):
The percentage of empty stomachs was (choose the correct option)

(choose correct option ☐ trophic classification. ($\chi^2 =$, $df =$, $P =$).

Q3-2. Calculate the residuals associated with the above contingency test and complete the following table of standardized residuals.


	< 16.2%	> 16.2%
DET	<input type="text"/>	<input type="text"/>
OMN	<input type="text"/>	<input type="text"/>
INV	<input type="text"/>	<input type="text"/>
PISC	<input type="text"/>	<input type="text"/>

Q3-3. What further conclusions would you draw from the standardized residuals?

Question 4 - Contingency tables

Here is an example (13.5) from Fowler, Cohen and Parvis (1998). A field biologist collected leaf litter from a 1 m² quadrats randomly located on the ground at night in two locations - one was on clay soil the other on chalk soil. The number of woodlice of two different species (*Oniscus* and *Armadilidium*) were collected and it is assumed that all woodlice undertake their nocturnal activities independently. The number of woodlice are in the following contingency table.

Format of Woodlice data set		
SOIL TYPE	WOODLICE SPECIES	
	Oniscus	Armadilidium
Clay	14	6
Chalk	22	46



Note, it is not necessary to open or create a data file for this question.

Q4-1. What null hypothesis is being tested by this test?

Q4-2. Using a **2 x 2 (two way) contingency table**, test this null hypothesis. What is the P value?

Q4-3. If the assumption is OK, test this null hypothesis (HINT) and identify the following.

a. X^2 statistic

b. df

c. P value

Q4-4. Generate the residuals (HINT) associated with the above contingency test and complete the following table of standardized residuals.

	oniscus	armadilidium
CLAY	<input type="text"/>	<input type="text"/>
CHALK	<input type="text"/>	<input type="text"/>


Q4-5. What are your conclusions (statistical and biological)?

Question 5 - Contingency tables

An invertebrate biologist investigating the frequency of banding and color patterns in coastal and hedgerow populations of the terrestrial snail *Cepea nemoralis* classified captured snails as being of one of four color patterns (banded yellow, non-banded yellow, banded pink and non-banded pink). The biologist was primarily interested in whether the frequency of color patterns was associated (related) to habitat. Her compiled findings are presented in the following table.

Format of snails data sett

Habitat	Shell color and pattern			
	Banded yellow	Non-banded yellow	Banded pink	Non-banded pink
Coastal	10	19	5	16
Hedgerow	17	8	19	11



Note, it is not necessary to open or create a data file for this question.

Q5-1. Using a **two-way contingency table**, test the null hypothesis that the frequency of color patterns was independent of the location from which the snails were collected. What would you

conclude form the analysis?
Write the results out as though you were writing a research paper/thesis. For example (select the phrase that applies and fill in gaps with your results):
The frequency of color patterns was (choose the correct option)
(choose correct option ☐ habitat. ($\chi^2 =$, $df =$, $P =$).

Q5-2. Calculate the residuals associated with the above contingency test and complete the following table of standardized residuals.

	Banded yellow	Non-banded yellow	Banded pink	Non-banded pink
Coastal	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Hedgerow	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Q5-3. What further conclusions would you draw from the standardized residuals?

Welcome to the end of Worksheet 5