# Week 3 Problem Based Learning and Practical Solutions

*Jeffrey O. Hanson*[1]

[1]*School of Biological Sciences, The University of Queensland, Brisbane, QLD, Australia*
*Correspondance should be addressed to [jeffrey.hanson@uqconnect.edu.au](mailto:jeffrey.hanson@uqconnect.edu.au)*

*20 March 2016*

## Contents

## Problem based learning workshop

### Exercise 1

In a genetics experiment on tomatoes, a dihybroid cross was made, with the frequencies of the progeny expected to be in the ratio 9:3:3:1.

The following table gives the observed frequencies

| Round/Yellow | Wrinkled/Yellow | Round/Green | Wrinked/Green |
|:---:|:---:|:---:|:---:|
| 56 | 19 | 17 | 8 |

1. *What are the expected proportions?*

- Remember that our expected ratio is 9:3:3:1. We can use this to calculate the expected proportions.

| Round/Yellow | Wrinkled/Yellow | Round/Green | Wrinked/Green |
|:---:|:---:|:---:|:---:|
| $\frac{9}{9+3+3+1} = 0.5625$ | $\frac{3}{9+3+3+1} = 0.1875$ | $\frac{3}{9+3+3+1} = 0.1875$ | $\frac{1}{9+3+3+1} = 0.0625$ |

2. *What are the expected frequencies?*

- To calculate the expected frequencies, we first need to calculate the number of individuals in the experiment.

- The total number of individuals in the experiment is $56 + 19 + 17 + 8 = 100$.

- We can then multiply the epxected proportions (which we claculated in the previous question) by the total number of observed individuals to yield the expected frequencies.

| Round/Yellow | Wrinkled/Yellow | Round/Green | Wrinked/Green |
|---|---|---|---|
| $0.5625 \times 100 = 56.25$ | $0.1875 \times 100 = 18.75$ | $0.1875 \times 100 = 18.75$ | $0.0625 \times 100 = 6.25$ |

3. *What is the $\chi^2$ value?*

   - We can use the following equation to calculate the $\chi^2$ value: $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$.
   - In this equation, $O_i$ is the observed frequency for the $i$'th combination and $E_i$ is the expected frequency for the $i$'th combination.
   - We can sub in our observed and expected frequencies into the equation, and calculate the $\chi^2$ value.

$$\chi^2 = \frac{(56 - 56.25)^2}{56.25} + \frac{(19 - 18.75)^2}{18.75} + \frac{(17 - 18.75)^2}{18.75} + \frac{(8 - 6.25)^2}{6.25}$$
$$\chi^2 = 0.001 + 0.003 + 0.163 + 0.49$$
$$\chi^2 = 0.657$$

4. *What are the degrees of freedom?*

   - To help understand how we calculate the degrees of freedom ($df$) let us first think about our data. We have individuals with combinations of different shape (round vs. wrinkled) and colors (green vs. yellow). We can think of shape and colors as different variables along which individuals can vary. We can use this equation to calculate the degrees of freedom.

$$df = (\text{number of groups in 1st variable} - 1) \times (\text{number of groups in 2nd variable} - 1)$$
$$df = (\text{number shapes} - 1) \times (\text{number colors} - 1)$$
$$df = (2 - 1) \times (2 - 1)$$
$$df = 1$$

   - **If this doesn't make sense, the example in second exercise is easier to understand.**

5. *What is the associated p-value?*

   - Given that we know the $\chi^2$ value is 0.657 and the degrees of freedom is 1, we can calculate the p-value using the following R code, where x is the $\chi^2$ and d is the degrees of freedom ($df$).

```
pchisq(x, df=d, lower.tail=FALSE)
```

- We can sub our values into this code, and calculate the p-value.

```
pchisq(0.657, df=1, lower.tail=FALSE)
```

- $\chi^2 = 0.4176211$

6. *Can you reject the null hypothesis?*

   No, given that the p-value is much greater than 0.05, we cannot reject the null hypothesis.

**Exercise 2**

Here we have a contingency table for case and control individuals genotyped at a diallelic marker locus.

| Affection status | AA | AB | BB |
|---|---|---|---|
| Case | 23 | 47 | 30 |
| Control | 12 | 40 | 48 |

1. *Calculate the expected values for each cell?*

   - First we need to propose a null hypothesis. Here, our null hypothesis will be that each outcome (eg. Case/AA) has an equal probability of occurring. Since there are six possible outcomes and our null hypothesis is that each outcome has an equal chance of occurring, each outcome has a probability of $\frac{1}{6}$.
   - We can express these expected probabilities using a contingency table.

| Affection status | AA | AB | BB |
|---|---|---|---|
| Case | $\frac{1}{6} = 0.1667$ | $\frac{1}{6} = 0.1667$ | $\frac{1}{6} = 0.1667$ |
| Control | $\frac{1}{6} = 0.1667$ | $\frac{1}{6} = 0.1667$ | $\frac{1}{6} = 0.1667$ |

   - We can then multiply these expected probabilities by the number of replicates in the experiment $(23 + 46 + 30 + 12 + 40 + 45 = 196)$ to calculate the expected frequencies

| Affection status | AA | AB | BB |
|---|---|---|---|
| Case | $0.167 \times 196 = 32.732$ | $0.167 \times 196 = 32.732$ | $0.167 \times 196 = 32.732$ |
| Control | $0.167 \times 196 = 32.732$ | $0.167 \times 196 = 32.732$ | $0.167 \times 196 = 32.732$ |

2. *Perform a $\chi^2$ test on those values.*

- Let's refresh our memory on what our observed frequencies are.

| Affection status | AA | AB | BB |
|---|---|---|---|
| Case | 23 | 47 | 30 |
| Control | 12 | 40 | 48 |

- Also, let's refresh our memory on what our expected frequencies are.

| Affection status | AA | AB | BB |
|---|---|---|---|
| Case | 32.732 | 32.732 | 32.732 |
| Control | 32.732 | 32.732 | 32.732 |

- We can calculate the $\chi^2$ value using this equation $\chi^2 = \sum_{i=1}^{n} \frac{O_i - E_i}{E_i}$.
- We can then calculate the $\chi^2$ value by subbing our values into the equation.

$$\chi^2 = \sum_{i=1}^{n} \frac{O_i - E_i}{E_i}$$

$$\chi^2 = \frac{(23 - 32.732)^2}{32.732} + \frac{(47 - 32.732)^2}{32.732} + \frac{(30 - 32.732)^2}{32.732} + \frac{(12 - 32.732)^2}{32.732} +$$
$$\frac{(40 - 32.732)^2}{32.732} + \frac{(48 - 32.732)^2}{32.732}$$

$$\chi^2 = 2.894 + 6.219 + 0.228 + 13.131 + 1.614 + 7.122$$

$$\chi^2 = 31.208$$

3. *Find the associated p-value.*

- First we need to calculate the degrees of freedom. Since our data is in a contingency table we can do this with **one simple trick**.

$$df = (\text{number of rows} - 1) \times (\text{number of columns} - 1)$$
$$df = (2 - 1) \times (3 - 1)$$
$$df = 2$$

- Now that we have the degrees of freedom ($df = 2$) and our $\chi^2$ statistic (31.208) which we calculated earlier, we can calculate the p-value using the following R code.

```
pchisq(31.208, df=2, lower.tail=FALSE)
```

$$\chi^2 = 1.6721256 \times 10^{-7}$$
$$\chi^2 < 0.001$$

4. Can you reject the null hypothesis at alpha $= 0.05$?

  - Yes. Yes, we can.

**Exercise 3**

1. Are extinction events, as observed in the fossil record, random in time, or do they have cluster (eg. mass extinctions)? In other words, do species go extinct at random intervals, or species tend to go extinct at the same time? To answer this question, use $\chi^2$ to measure the "fit" of the probability model to the data. Significant lack of fit implies rejection of the null hypothesis of "no departure from the model".

   Here we have a table that shows the number of time intervals between extinction events. The "number of extinctions" column contains the index (starting at zero) of the extinction (eg. the first extinction has a zero, the second has a one, etc). The "number of time intervals" describes the amount of time that has passed between conescutive extinctions (using a standardised unit of time).

| Number of extinctions | Number of Time Intervals |
|:---:|:---:|
| 0 | 0 |
| 1 | 13 |
| 2 | 15 |
| 3 | 16 |
| 4 | 7 |
| 5 | 10 |
| 6 | 4 |
| 7 | 2 |
| 8 | 1 |
| 9 | 2 |
| 10 | 1 |
| 11 | 1 |
| 12 | 0 |
| 13 | 0 |

| Number of extinctions | Number of Time Intervals |
|:---------------------:|:------------------------:|
| 14                    | 1                        |
| 15                    | 0                        |
| 16                    | 2                        |
| 17                    | 0                        |
| 18                    | 0                        |
| 19                    | 0                        |
| 20                    | 1                        |

We will use the Poisson distribution to express our null hypothesis. An expectation from the Poisson distribution can be calculated using

$Pr\{Y = k|\mu\} = \frac{e^{-u}\mu^k}{k!}$

## R practical session