Math 263: Solutions to Excel Assignment 8

.

1. This can be done with either a z-score or with chi-squared; we use z.

There are two populations: Population 1 is the male salmon swimming upstream; Population 2 is female salmon swimming upstream. The variable is categorical: having ichthyophonus.

Null hypothesis $H_0: p_1 = p_2$ and alternate hypothesis $H_0: p_1 < p_2$.

The sample proportions are $\hat{p}_1 = 235/961 = 0.245$ and are $\hat{p}_2 = 212/670 = 0.316$. Then the pooled proportion is

$$\hat{p} = \frac{235 + 212}{961 + 670} = 0.274.$$

$$z = \frac{0.245 - 0.316}{\sqrt{0.274(1 - 0.274)\left(\frac{1}{961} + \frac{1}{670}\right)}} = -3.16.$$

The p-value is $p = 0.0008$.

Since the p-value is so small (0.08%), the difference in proportions of infections between male and female fishes is very unlikely to have occurred by chance if the two genders were really infected at the same rate. Thus, we either conclude that the infection rates in the two populations are different.

2. This hypothesis test uses chi-squared because there are more two categories.

We have a two-way table with infected and uninfected counts for each of the six site. Since the data is categorical and we have more than two samples, we use a χ^2test with 5 degrees of freedom.

The null hypothesis is that there is no association between site and infection; the alternate hypothesis is that there is an association. In other words, the null hypothesis is that the proportion of females infected varies significantly between the sites.

The data reports the observed numbers of females sampled and infected at each site, so we calculate the number uninfected at each site. We calculate the expected number infected and uninfected under the null hypothesis. The observed and expected data is shown in the following two tables.

Females: Original

Site	Emmonak	Galena	Rapids (1)	Circle	Border (2)	Whitehorse	Total
#infected_female	54	7	68	15	50	18	212
#ininfected_female	145	11	126	20	57	99	458
#sample_female	199	18	194	35	107	117	670

Females: Estimated

Site	Emmonak	Galena	Rapids (1)	Circle	Border (2)	Whitehorse	All
#infected_female	62.967	5.696	61.385	11.075	33.857	37.021	212
#uninfected_female	136.033	12.304	132.615	23.925	73.143	79.979	458
#sample_female	199	18	194	35	107	117	670

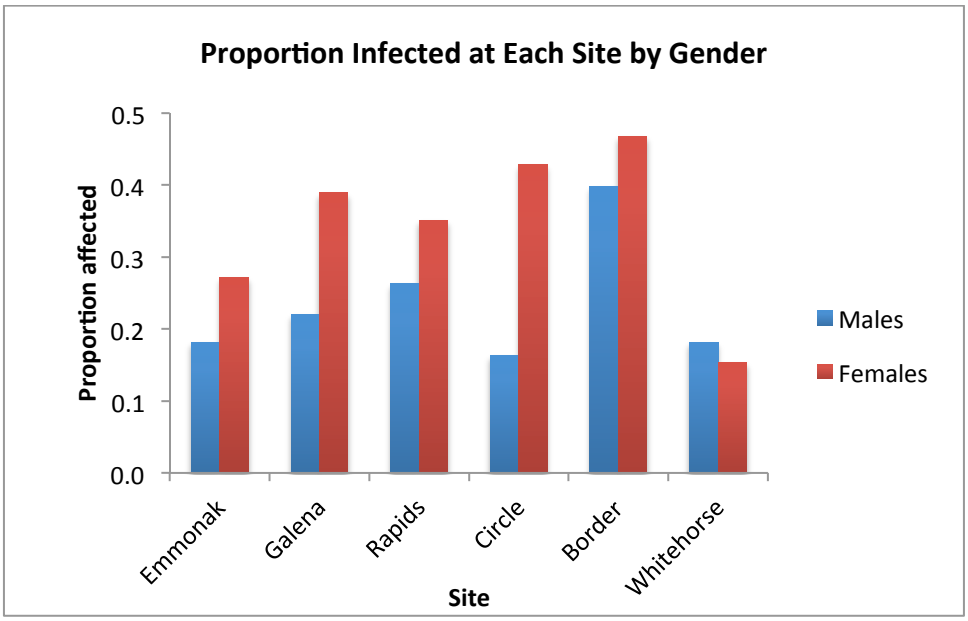
Using the data in these tables, we calculate that

$$\chi^2 = \frac{(54-62.967)^2}{62.967} + \frac{(7-5.696)^2}{5.696} + \cdots + \frac{(99-79.979)^2}{79.979} = 30.9$$

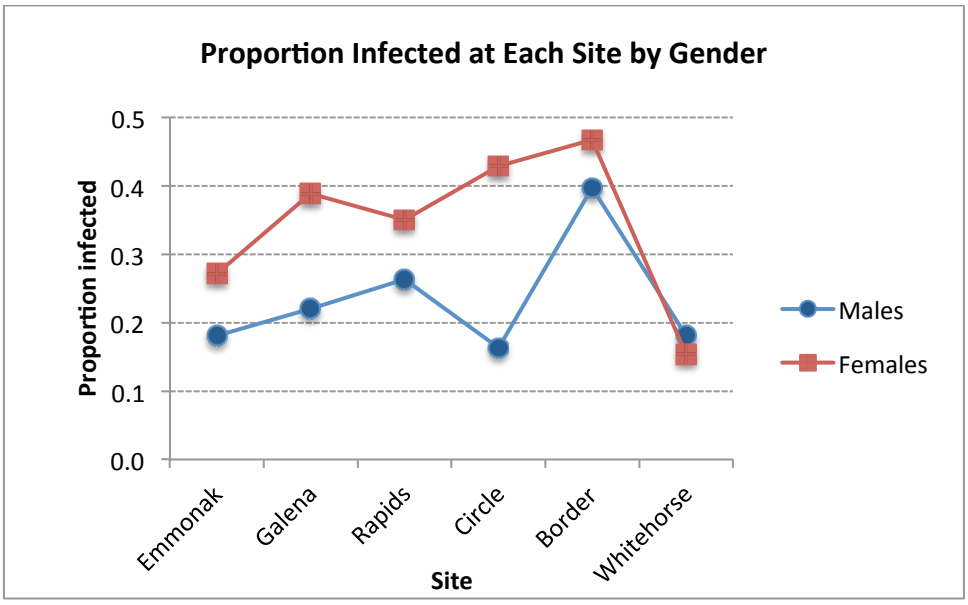
There are 5 degrees of freedom, with $p = 9 \cdot 10^{-6} = 0.0000096$. Since this p-value is so small, we conclude there is an interaction between infection rate and site.

3. The sites are listed in order in the Excel file. From left to right, the sites are listed in order as we move upstream.

(a) With the sites in order from left to right as we move up stream:
(i) The column graph follows:



(ii) The line graph follows:



(b) Both genders start with about 20-25% infection rate at the mouth of the river; this rate mostly increases as the fish move upstream. At the last site (Whitehorse), however, there is a significant drop-off in the infection rate for both genders. Males, but not females, have a significant drop off at one intermediate site

(Circle). The infection rate for females is higher than for males throughout the river except at the last site, where males have a higher infection rate.

(c) The graphs fit the explanation well. It starts at approximately the right level of infection with the females higher, then increases and decreases as some of the fish die. The features of the graphs that do not fit the explanation are the drop in male infection rate at Circle and the drop in female infection rates below the males at Whitehorse.