

# Homework 7: Sections 7.1 - 7.2

STA209-04: Applied Statistics

Assigned: 04/12/2019

Due: 04/19/2019

## From the Book:

**Questions:** 7.15, 7.21, 7.23, 7.41, 7.45, 7.53

**7.15** Between 2008 and 2011, the age distribution of users of social networking sites such as Facebook changed dramatically. In 2008, 70% of users were 35 years of or younger. In 2011, the age distribution was much more spread out. Table 7.10 shows the age distribution of 975 users of social networking sites from a survey reported in June 2011.

Table 7.10

Age	18-22	22-35	36-49	50-65	65+
Frequency	156	312	253	195	59

- Test an assumption that users are equally likely to be in each of the five age groups listed. Show all details of the test.
  - Which age group contributes the largest amount to the sum for the  $\chi^2$  test statistic? For this age group, is the observed count smaller or larger than the expected count?
- 7.21** Most medical school graduates in the US enter their residency programs at teaching hospitals in July. A recent study suggests that a spike in deaths due to medication errors coincides with this influx of new practitioners. The study indicates that the number of deaths is significantly higher than expected in July.
- What type of statistical analysis was probably done to arrive at this conclusion?
  - Is the  $\chi^2$  statistic likely to be relatively large or relatively small?
  - Is the p-value likely to be relatively large or relatively small?
  - What does the relevant categorical variable record?
  - What cell contributes the most to the  $\chi^2$  statistics?
  - In the cell referred to in part e), which is higher: the observed count or the expected count?

**7.23** Movies based on Ian Fleming’s novels starring British secret agent James Bond have become one of the longest running film series to date. As of 2016, six different actors have portrayed the secret agent. Which actor is the best James Bond? A sample of responses to this question is shown in Table 7.16.

Table 7.16

Actor	Frequency
Sean Connery	98
George Lazenby	5
Roger Moore	23
Timothy Dalton	9
Pierce Brosnan	25
Daniel Craig	51

- Does the sample provide evidence of a significant difference in popularity among the six actors, at a 5% significance level?
- Repeat the test from part a) if we ignore the results for George Lazenby, who only appeared in one Bond film. Do we find evidence of a significant difference in popularity among the remaining five actors?
- The message from Chapter 1 still holds true: Pay attention to where the data come from! These data come from a poll held on a James Bond fan site. Can we generalize the results of this poll to the movie-watching population?

**7.41** In Exercise 6.148 on page 445 we perform a test for the difference in the proportion of penguins who survive over a ten year period, between penguins tagged with metal tags and those tagged with electronic tags. We are interested in testing whether the type of tag has an effect on penguin survival rate, this time using a chi-square test. In the study, 10 out of the 50 metal-tagged penguins survived while 18 out of the 50 electronic-tagged penguins survived.

- Create a two-way table from the information given.
- State the null and alternative hypotheses.
- Give a table with the expected counts for each of the four categories.
- Calculate the chi-square test statistic
- Determine the p-value and state the conclusion using a 5% significance level.

**7.45** 478 middle school (grades 4 to 6) students from three school districts in Michigan were asked whether good grades, athletic ability, or popularity was most important to them. The results are shown below, broken down by gender.

	Grades	Sports	Popular
Boy	117	60	50
Girl	130	30	91

- Do these data provide evidence that grades, sports, and popularity are not all equally valued among middle school students in these school districts? State the null and alternative hypotheses, calculate a test statistic, find a p-value, and answer the question.
- Do middle school boys and girls have different priorities regarding grades, sports, and popularity? State the null and alternative hypotheses, calculate a test statistic, find a p-value, and answer the question.

**7.53** The study on genetics and fast-twitch muscles includes a sample of elite sprinters, a sample of elite endurance athletes, and a control group of non-athletes. Is there an association between genetic allele classification (R or X) and group (sprinter, endurance, control)? Computer output is shown for this chi-square test. In each cell, the top number is the observed count, the middle number is the expected count, and the bottom number is the contribution to the chi-square statistic.

	R	X	Total
Control	244 251.42 0.219	192 184.58 0.299	436
Sprint	77 61.70 3.792	30 45.30 5.166	107
Endurance	104 111.87 0.554	90 82.13 0.755	196
Total	425	312	737

Chi-Sq = 10.785, DF = 2, P-Value = 0.005

- How many endurance athletes were included in the study?
- What is the expected count for sprinters with the R allele? For this cell, what is the contribution to the chi-square statistic? Verify both values by computing them yourself.
- What are the degrees of freedom for the test? Verify this value by computing it yourself.
- What is the chi-square test statistic? What is the p-value? What is the conclusion of the test?
- Which cell contributes the most to the chi-square statistic? For this cell, is the observed count greater than or less than the expected count?
- Which allele is most over-represented in sprinters? Which allele is most over-represented in endurance athletes?