

GENOME 560, Spring 2012

Problem Set #4

(Due May 31th 9:00am)

1. [20 points] ANOVA Table

Complete the ANOVA table above from the values given. The sample size is $N = 20$. (Hint: See slide 21 of lecture note 7.)

Source of Variation	Sum of Squares	df	Mean Square	F
Group	56.7			
Error		14	13.5	
Total				

2. [80 points] Chi-Square Test

An expensive private school also asks for donations. Here are (actual) data (we have anonymized the school) on how many of the parents in each of its graduating classes have donated to the school's fund drive. The school is K-12. These are from the same year (2009) so the 2009 class means current 12-th graders, 2010 means current 11th graders, and so on.

Year	Donated	Total Parents
2009	35	51
2010	42	56
2011	39	70
2012	37	60
2013	38	53
2014	35	54
2015	32	53
2016	19	27
2017	22	31
2018	20	30
2019	17	32
2020	34	34
2021	28	32

- (a) Do a Chi-square analysis (be careful to set up the table correctly) to find out whether there is any sign of parent burnout – are donations equally likely in all grades?
- (b) Is this to be done as one-tailed or two-tailed? Why? (Does a low Chi-square mean a departure from the expected proportions?)
- (c) Think of some way to lump parts of the table to make the test focus more on the question at issue, and not waste effort on detecting whether there are differences that do not represent a long term trend. Carry it out and describe the results. Note that you can sum column 2 of rows 1 to 8 of a table by the R command `sum(a[1:8,2])`
- (d) What is the effect on the Chi-square test, on average, if some of the parents have two (or more) children in more than one grade, and thus are listed as donating (or not) in both of those grades on the basis of the same donation or the same non-donation?