Applied Statistics Comprehensive Examination

Statistical Methods I & II

1. (30 pts) The data below are the number of home runs hit by National League home runs leaders in the years 1990 – 2009.

$$40 \quad 38 \quad 35 \quad 46 \quad 43 \quad 40 \quad 47 \quad 49 \quad 70 \quad 65$$
$$50 \quad 73 \quad 49 \quad 47 \quad 48 \quad 51 \quad 58 \quad 50 \quad 48 \quad 47$$

Note: $\sum y = 994$, $\sum y^2 = 51,290$ and $\sum (y - \bar{y})^2 = 1888.2$.

   (a) Construct a stem and leaf plot with two lines per stem for these data and comment on the distribution with respect to symmetry, normality and outliers.

   (b) Compute the mean, median, standard deviation, IQR and range for these data.

   (c) Compute the proportion of observations that are within two standard deviations of the mean and compare this proportion to what is expected by the empirical rule.

2. (20 pts) In a biology laboratory, the mating of two red-eyed fruit flies yielded $n = 432$ offspring, among which 254 were red-eyed, 69 were brown-eyed, 87 were scarlet-eyed and 22 were white-eyed. Use these data to test the hypothesis that the true frequencies of the offspring follow the ratio 9:3:3:1 by calculating the appropriate test statistic and estimating and interpreting its P–value.

3. (30 pts) Five measurements of the tar content of a certain kind of cigarette yielded 14.5, 14.2, 14.4, 14.3 and 14.6 mg/cigarette. Assume that the data are a random sample from a normal population.

   (a) Use a 5% level of significance to show that the null hypothesis $\mu = 14.0$ should be rejected in favor of the alternative hypothesis $\mu \neq 14.0$.

   (b) Suppose that the first measurement is incorrectly recorded as 16.0 instead of 14.5. Show that the null hypothesis $\mu = 14.0$ should not be rejected in favor of the alternative hypothesis $\mu \neq 14.0$, using a 5% level of significance.

   (c) Explain the apparent paradox that even though the difference between the sample mean and μ_0 has increased, the result is no longer significant.

4. (20 pts) An apiarist (bee keeper) uses 16 bee hives to conduct an experiment to compare four different types of materials for insulating hives for the winter survival of bee colonies. Four hives were assigned to each of the four insulating materials (A,B,C,D), four hives were exposed to each direction of the compass (N,S,E,W) and four hives were assigned to each of four randomly selected colonies of bees (1,2,3,4) in such a way that each combination of (direction and colony), (direction and material) and (colony and material) was assigned exactly once in the experiment. The specific layout of the experimental design appears below:

| Direction | Colony | | | |
	1	2	3	4
N	A	B	C	D
S	C	D	A	B
E	D	A	B	C
W	B	C	D	A

(a) Identify the experimental design and state the appropriate mathematical model. Be sure to state all assumptions and indicate which factors are fixed and which factors are random.

(b) Suppose the sum of squares for materials, directions, colonies and total are 165, 90, 210 and 525 respectively. Complete the ANOVA table and make all appropriate inferences; that is, test all fixed effects using $\alpha = 0.05$ and provide a point estimate for all variance components.