Stat571 Fall 2004 Final Exam – Brief Solutions

1. (a) Since there are 5 treatments, there are 4df for treatment. With 35 total observations, there are 34 df total. Df(error) = 5*(7-1)=30. The ANOVA table is:

source	df	SS	MS
trt	4	8098.17	2024.54
error	30	24960.57	832.02
total	34	33058.74	——

$H_0 : \mu_A = \mu_B = \mu_C = \mu_D = \mu_E$; H_A: not H_0 F=MSTrt/MSError=2.43 on 4,30 df. $.05 < p < .10$ and there is weak evidence against the null that all the means are the same.

(b) i. The null hypothesis states that the mean of treatment A is equal to the average of the means of treatments B, C, D, and E. Thus the average of the means of the treatments with the 'new' formulation is equal to the mean of the standard formulation.

ii. The null hypothesis states that there is no linear relationship between the level of the additive and the mean plant weight for the 4 treatments with the new formulation.

(c) The se for each contrast can be written: $s_p\sqrt{(1/n) * \sum(\lambda_i)^2}$. Substituting 1,2,3,4,5 for A,B,C,D,E, the lambdas for the 2 contrasts are: (i) 1,-1/4,-1/4,-1/4,-1/4; and (ii) 0,-3,-1,1,3. For (i), t=-37.04/12.19=-3.04 on 30df. The comp-wise p-val is: $.002 < p < .01$. Using Bonferroni with 2 contrasts (it is the number of contrasts, not the number of treatments that is important!!!), the comp-wise p-val needs to be doubled. Thus $.004 < p < .02$ which is less than .05; thus, there is solid evidence against the equality of A and the average of the new formulation treatments. For (ii), t=33.55/48.76=0.69. Using Bonferroni, $p > .40$ and there is no meaningful evidence against the null of no linear trend for additive.

2. (a) $\hat{b_1} = -22.93/196.90 = -.116$. This means that for each increase in temperature of one degree C, the rate of oxygen consumption decreases by .116 ml/g/hr.

(b) $SSReg = \hat{b_1} * (-22.93) = 2.67$. Thus, the ANOVA table for regression can be written:

source	df	SS	MS
regr	1	2.67	2.67
error	8	.41	.051
total	9	3.08	——

The stated test is $H_0 : b_1 = 0$ vs $H_A : b_1 \neq 0$. F=2.67/.051= 52.1 on (1,8) df. Thus, $p < .001$ and the rate depends strongly on the temperature.

(c) Here we focus on $\hat{Y}$ as an estimator. $se(\hat{Y}_e) = s\sqrt{(1/n) + (5-\bar{x})^2/(\sum(x_i-\bar{x})^2)} = .079$. Now, $t_{8,.025} = 2.306$ and $\hat{Y} = 3.326$. Thus, a 95% confidence interval for $\hat{Y}_e$ is (3.144, 3.508). Thus, at 95 % confidence, a plausible range for the mean rate of oxygen consumption is from 3.144 to 3.508 ml/g/hr. Since the value of "3" does not fall within this range, it is not a plausible value with 95% confidence; thus the hypothesis that the true rate is 3 is rejected at $\alpha = 0.05$.

3. There are two equivalent approaches. Let $m = 4n$ be the total number of trials. Due to independence, we can solve directly for m and choose the smallest integer equal to or larger than m that is divisible by 4. $P(atleast 1 success) = 1 - P(0 successes) = 1 - .98^m$. We wish to choose m so that $.98^m = 0.20$. This can be done by trial and error or, directly, using logs. $m * ln(.98) = ln(.20)$. This results in $m = 79.66$. Rounding up to 80 (and dividing by 4) results in $n = 20$. Equivalently, let Y be the number of successes out of 4 trials. Then, $P(Y = 0) = .98^4 = .92237$. Then, with n as the number of 'sets' of trials, $.92237^n = 0.20$ using similar reasoning to the above. Trial and error – or logs – again results in $n = 20$.

4. (a) FALSE — The only way one can obtain a negative value of F is if one has made a mistake. Even if all data values are negative, F can *never* be less than 0.

(b) FALSE — Since you are in 'planning' mode, the underlying variance should be viewed as known. We know $Var(\hat{b_1}) = \sigma^2_{y.x}/\sum(x_i - \bar{x})^2$. With the 10 data points, the sum in the denominator is 30; for the 6 data points situation, the sum is 70. Thus, $Var(\hat{b_1})$ will be smaller for the 6 data point case. (Even if one were concerned with CI width by considering the difference in t-values for 4df and 8df, the conclusion would not change.)

5. The random variable of interest is D, the difference in glucose levels between B and A. The stated condition requires the distribution of D under the given alternative. In this case, $D \sim N(.5, 1.8)$. The given condition is $P(\bar{D} \geq 0.3|\mu_D = .5) = .90$. We know $P(Z \geq -1.28) = .90$. Thus, $-1.28 = (.3 - .5)/\sqrt{(1.8/n)}$. Solving results in $\sqrt{n} = 8.59$ and, rounding up, $n = 74$.

Grade Distribution

```
90-99:20
80-89:19
70-79:37        median =  74
60-69:23
50-59:24
below: 5
```