Name Fall, 2012

Applied Statistics Comprehensive Examination
Regression & Linear Models

1. (40 Points) A statistician studied the relationship between wins at home and wins on the road for the 16 National League baseball teams in 2012. Use the Minitab output below to help answer the following questions.

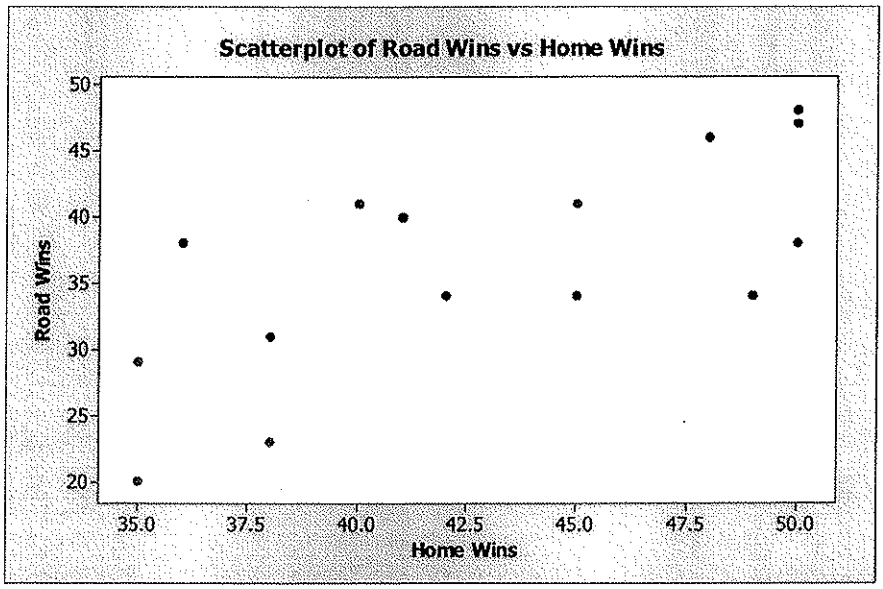
   (a) (5 Points) Comment on the appropriateness of doing a simple linear regression in this example.

   (b) (10 Points) Test for a positive relationship between wins at home and wins on the road.

   (c) (10 Points) Give the coefficient of determination, (R^2), and provide an interpretation for this value.

   (d) (10 Points) The 16 teams are split into 3 divisions of 5, 5, and 6 teams each. Create indicator variables for the divisions and write one regression model that would allow each division to have its own y-intercept and slope.

   (e) (5 Points) Based on your model in part d, briefly discuss how you would test for the equality of slopes between the three divisions.




```
The regression equation is
Road Wins = - 8.4 + 1.05 Home Wins


Predictor     Coef  SE Coef      T      P
Constant    -8.44    11.90  -0.71  0.490
Home Wins  1.0509   0.2738   3.84  0.002


Analysis of Variance

Source           DF       SS      MS      F      P
Regression        1   532.00  532.00  14.73  0.002
Residual Error   14   505.75   36.13
```

2. (15 Points) Briefly describe the following three stepwise procedures in terms of how they build a regression model: forward selection, backward elimination, and stepwise selection. Which method do you prefer and why?

3. (25 points) An experiment was run to study the effects of two factors, A and B, on a response variable. Factor A has two levels, while factor B has three levels. Assume that the data will be analyzed using a model $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$, where $i = 1, 2$, $j = 1, 2, 3$, $k = 1, 2$, α_i and β_j are fixed main effects, γ_{ij} is an interaction effect, and ϵ_{ijk} is an error term.

   (a) (15 Points) Write down two orthogonal contrasts in the interaction effects and determine whether they are estimable.

   (b) (10 Points) Write the design matrix using set-to-zero restrictions.

4. (20 Points) Consider a design with one treatment and one blocking factor in which each treatment appears once in each block. Answer the following questions.

   (a) (5 Points) Fill in the following ANOVA table.

Source of Variation	SS	df	MS
Treatments	12	?	?
Blocks	?	4	6
Error	36	?	?
Total	?	19	

   (b) (10 Points) Test at alpha=0.05 to determine whether there is a significant treatment effect.

   (c) (5 Points) Discuss the importance of the blocking factor in this example.