**Question 1**

A number of websites have appeared that purport to help consumers make product choices. Some researchers are interested in how many clicks are required for consumers to make a choice at different types of websites. In this study there are two types of web sites: EBA (elimination-by-aspects)—at which consumers click on minimum acceptable levels for product attributes (e.g., my new car must get at least 25 mpg) and RATE (ratings of attributes)—at which consumers give numerical ratings of how important each attribute is to them (e.g., I rate gas mileage as being very important, a 4 on a 5-point scale). In this study, each participant made a choice at both types of websites and the order of websites was counterbalanced across consumers. For each consumer the correlation among product attributes were, for both websites, either POSITIVE (i.e., good things go together) or NEGATIVE (i.e., to get a good level on one attribute one must give up something on another attribute—there are tradeoffs). The variables below are available for each consumer.

| | |
|---|---|
| EXPER | User's self-rated experience with using the web. |
| EBAClicks | Number of mouse clicks to make a choice at a website organized according to an "elimination-by-aspects" scheme. |
| RATEClicks | Number of mouse clicks to make a choice at a website organized according to an attribute rating scheme. |
| CORR | whether attribute correlations at both websites were POSITIVE or NEGATIVE |

1. Both measures of the number of clicks ranged from 5 to 120, with many more scores near the bottom of the distribution rather than the top. In light of this, would you consider transforming them prior to conducting the analyses? Why and what transformation might you try?

2. Ignoring the type of website and the attribute correlation, do the number of clicks decrease with increased web experience? Specify the MODEL C/A comparison that would answer this question. (For this and the following model questions, use the simplest models possible. And remember to define any new variables you construct.)

3. Still ignoring the type of website and also ignoring experience, are the number of clicks greater for websites with negative correlations (where presumably the choices are more difficult) than for those with positive correlations?

4. Ignoring the attribute correlation and user experience, do EBA websites require more clicks than RATE websites for making a choice?

5. Does the answer to the previous question depend on the attribute correlation at the website?

6. Does the dependence of the website difference on the attribute correlation (i.e., the previous question) decrease for users with more experience?

**Question 2**

The 1982 General Social Survey asked respondents whether or not they voted for Ronald Reagan for President (versus Jimmy Carter and others). Respondents also rated themselves on a 7-point scale of political views with a score of 1 representing "extremely liberal" and a score of 7 representing "extremely conservative." Respondents were also categorized as White or Nonwhite. In the following logistic regression, WHITE is a dummy code with 1 representing Whites and 0 representing Nonwhites. The variable POLVD is a mean-deviated version of the political views ratings (the mean was 4.19). The response variable was coded 1 if the respondent voted for Reagan and 0 otherwise. Interpret ***each*** of the parameters (including the intercept) in the following logistic model. Below each parameter are the odds ratio estimates.

$$\hat{logit} = -2.87 + 2.9White + 0.37PolvD + 0.13White * PolvD$$

$$Odds \quad 0.057 \quad 18.23 \quad 1.44 \quad 1.12$$

Only the first two parameters (the intercept and the parameter for White) are significantly different from zero. Given that, write a brief summary of the results.

**Question 3**

The bark of some tropical trees seems to offer protection from termites. If this protection could be harnessed, the trees that produce the resin might become a valuable resource. Experimenters investigated the effects of these tree resins on termites. The resin was dissolved in a solvent and placed on filter paper in different doses (0 mg, 5mg, and 10mg). For each dosage level, five dishes are set up with 25 termites in each dish. The termites are fed the dosed filter paper and a daily count is made of the number of termites surviving.

The following data indicate for each dish the number of termites surviving after 4, 8, and 12 days.

| dish | dose | day4 | day8 | day12 |
|---|---|---|---|---|
| 1 | 0 | 22 | 21 | 19 |
| 2 | 0 | 25 | 21 | 20 |
| 3 | 0 | 24 | 22 | 21 |
| 4 | 0 | 25 | 23 | 23 |
| 5 | 0 | 22 | 22 | 20 |
| 6 | 5 | 23 | 15 | 14 |
| 7 | 5 | 24 | 19 | 11 |
| 8 | 5 | 23 | 14 | 12 |
| 9 | 5 | 23 | 18 | 17 |
| 10 | 5 | 23 | 18 | 16 |
| 11 | 10 | 12 | 8 | 4 |
| 12 | 10 | 14 | 6 | 2 |
| 13 | 10 | 16 | 5 | 0 |
| 14 | 10 | 15 | 1 | 0 |
| 15 | 10 | 16 | 7 | 3 |

Below are the means from each day for each dosage level:

| | day4 | day8 | day12 |
|---|---|---|---|
| 0 mg | 23.6 | 21.8 | 20.6 |
| 5 mg | 23.2 | 16.8 | 14.0 |
| 10 mg | 14.6 | 5.4 | 1.8 |

The following variables are computed on these data for purposes of analysis:

```
d1      =   -2*(dose=0)+1*(dose=5)+1*(dose=10);
d2      =   0*(dose=0)-1*(dose=5)+1*(dose=10);
aveday  =   (day4+day8+day12)/3;
linday  =   day4-day12;
quaday  =   2*day8-day4-day12;
```

On the following pages are a series of proc reg results, regressing aveday, linday, quaday on d1 and d2. In light of these results and the above cell means, answer the following questions.

1. On average across the three days, is there evidence that the resin is toxic to the termites? (Present PRE, F*, N-PA and PA-PC, and interpret the parameter of interest.)

2. On average across the three days, is there evidence of a dosage effect? (i.e., is the resin more toxic at higher dosage levels?) (Present PRE, F*, N-PA and PA-PC, and interpret the parameter of interest.)

3. What model comparison would you make to test if the resin makes a difference on termite survival even after only 4 days? (Indicate Models A and C.)

4. In the model where LINDAY is the dependent variable, all three parameter estimates are significant. Provide substantive interpretations of each of the three.

5. Interpret the significant intercept in the model where QUADAY is the dependent variable.

6. These data consist of counts of surviving termites, with 25 starting at day 1, and decreasing over time. In light of this, would you expect their distribution to be skewed? If so, what sort of transformation might you consider to deal with the problem?

---

Model: MODEL1
Dependent Variable: AVEDAY

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | 2 | 580.45926 | 290.22963 | 199.903 | 0.0001 |
| Error | 12 | 17.42222 | 1.45185 | | |
| C Total | 14 | 597.88148 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 1.20493 | R-square | 0.9709 |
| Dep Mean | 15.75556 | Adj R-sq | 0.9660 |
| C.V. | 7.64764 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob > \|T\| |
|---|---|---|---|---|---|
| INTERCEP | 1 | 15.755556 | 0.31111111 | 50.643 | 0.0001 |
| D1 | 1 | -3.122222 | 0.21998878 | -14.193 | 0.0001 |
| D2 | 1 | -5.366667 | 0.38103174 | -14.085 | 0.0001 |

Model: MODEL2
Dependent Variable: LINDAY

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | 2 | 245.73333 | 122.86667 | 19.000 | 0.0002 |
| Error | 12 | 77.60000 | 6.46667 | | |
| C Total | 14 | 323.33333 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 2.54296 | R-square | 0.7600 |
| Dep Mean | 8.33333 | Adj R-sq | 0.7200 |
| C.V. | 30.51557 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob > \|T\| |
|---|---|---|---|---|---|
| INTERCEP | 1 | 8.333333 | 0.65659052 | 12.692 | 0.0001 |
| D1 | 1 | 2.666667 | 0.46427961 | 5.744 | 0.0001 |
| D2 | 1 | 1.800000 | 0.80415587 | 2.238 | 0.0449 |

Model: MODEL3
Dependent Variable: QUADAY

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | 2 | 63.33333 | 31.66667 | 2.189 | 0.1547 |
| Error | 12 | 173.60000 | 14.46667 | | |
| C Total | 14 | 236.93333 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 3.80351 | R-square | 0.2673 |
| Dep Mean | -3.26667 | Adj R-sq | 0.1452 |
| C.V. | -116.43389 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob > \|T\| |
|---|---|---|---|---|---|
| INTERCEP | 1 | -3.266667 | 0.98206132 | -3.326 | 0.0060 |
| D1 | 1 | -1.333333 | 0.69442222 | -1.920 | 0.0789 |
| D2 | 1 | -1.000000 | 1.20277457 | -0.831 | 0.4220 |