The experiment was conducted by constructing 36 Web sites consisting of all possible combinations of the four factors described above. Each potential customer who visited the company's Web site during the trial period was randomly redirected to one of the 36 configurations. The number of visitors n_{ijkl} to the $ijkl$th configuration and the number that signed up x_{ijkl} was logged. x_{ijkl} is then binomially distributed

$$B(x_{ijkl}, n_{ijkl}, p_{ijkl}) = \left(\begin{array}{c} n_{ijkl} \\ x_{ijkl} \end{array} \right) p_{ijkl}^{x_{ijkl}} (1 - p_{ijkl})^{(n_{ijkl} - x_{ijkl})} \tag{3.9}$$

where n_{ijkl} is the number of visitors to the $ijkl$th configured Web page during the testing period.

Below is the R code to open the raw data, and print the first six lines of the data frame.

```
> library(daewr)
> data(web)
> head(web)
  A B C D visitors signup
1 1 1 1 1     1016     22
2 1 1 1 2     1145     16
3 1 1 2 1     1145     17
4 1 1 2 2     1082     19
5 1 2 1 1     1121     28
6 1 2 1 2     1103     28
```

The correct procedure must be utilized to analyze the data, determine if any of the factor effects are significant, and to predict the optimal Web page configuration. Since the responses for the individual visitors to each configuration of the Web site are Bernoulli, the aggregate response data is binomial with large sample sizes (i.e., number of visitors to each possible Web configuration).

To incorporate the binomial distribution the method of maximum likelihood can be used to fit the model (3.8). This can be done fitting a general linear model using the R `glm` function. The unequal variances for the responses (i.e. the variance of a binomial response, p, is pq/n) is similar to having an unequal number of replicates for each combination of factor levels. Therefore, the type III sums of squares of the form $(\boldsymbol{L}\hat{\boldsymbol{\beta}})'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{L}\hat{\boldsymbol{\beta}})$ should be calculated using the `Anova` function in the `car` package as shown in section 3.5.3. The type III option in the `Anova` function will not work with a saturated model. Therefore, the saturated model was fit with sequential sums of squares as shown in the R code on the web page. The final term ABCD, that is adjusted for all others, was insignificant ($p > .73$) and dropped. The model was refit using the type III option as shown in the code on the next page. The sums of squares will be approximately distributed as chi-squares under the null hypothesis. The approximation will be accurate due to the large number of visitor in each factor combination.

```
> library(car)
> modb <- glm(cbind(signup,visitors-signup)~A+B+C+D+A:B+
          A:C+A:D+B:C+B:D+C:D+A:B:C+A:B:D+A:C:D+B:C:D,
          data = web, family = binomial,
          contrasts=(list(A=contr.sum, B=contr.sum,
          C=contr.sum,D=contr.sum)))
> Anova(modb, type="III")
```

`signup/visitors` is the observed proportion signing up. The option `family = binomial` declares the response to be binomially distributed, and the option `type="III"` in the call to the `Anova` function requests a table of the type III adjusted chi-square tests. The results are shown below. The command `summary(modb)` prints a table of the parameter estimates produced by the maximum likelihood method, which is similar to the summary of an object created by the function `lm`, and it is not shown here.

```
Analysis of Deviance Table (Type III tests)

Response: cbind(signup, visitors - signup)
      LR Chisq Df Pr(>Chisq)
A      10.1116  2   0.006372 **
B       5.7081  2   0.057611 .
C       2.4453  1   0.117879
D       4.7582  1   0.029159 *
A:B     6.5032  4   0.164591
A:C     0.8247  2   0.662101
A:D     0.0295  2   0.985335
B:C     4.1839  2   0.123443
B:D     3.8263  2   0.147615
C:D     0.1492  1   0.699346
A:B:C   2.1136  4   0.714874
A:B:D   5.8815  4   0.208175
A:C:D   6.1536  2   0.046106 *
B:C:D   0.0219  2   0.989104
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

In this output we can see that (at the $\alpha = 0.05$ level of significance) factors A (background style), factor D (sign-up button or link) were significant along with the three-way interaction ACD, where factor C represents the text color.