Chapter 22

Logistic Regression - Advanced Topics

Contents

22.1 Introduction ................................................................. 1493
22.2 Sacrificial pseudo-replication .............................................. 1493
22.3 Example: Fox-proofing mice colonies - dealing with sacrificial pseudo replication .................................................. 1494
  22.3.1 Using the simple proportions as data ................................ 1495
  22.3.2 Logistic regression using overdispersion ......................... 1496
  22.3.3 GLIMM modeling the random effect of colony .................. 1499
22.4 Example: Over-dispersed Seed Germination Data ..................... 1499
  22.4.1 Using the simple proportions as data ............................ 1502
  22.4.2 Logistic regression using overdispersion ....................... 1504
  22.4.3 GLIMM modeling the random effect of pots ...................... 1506
22.5 Example: Are mosquitos choosy? A preference experiment .............. 1507
  22.5.1 Using the simple proportions as data ............................ 1511
  22.5.2 Using the empirical logits as data ............................... 1513
  22.5.3 Logistic regression using overdispersion ....................... 1514
  22.5.4 GLIMM modeling the random effect of replicates ................ 1515
  22.5.5 Recommendation ................................................... 1516
22.6 Example: Reprise: Are mosquitos choosy? A preference experiment with complete blocks ............................................. 1516
  22.6.1 Using the simple proportions as data ............................ 1521
  22.6.2 Using the empirical logits as data ............................... 1527
  22.6.3 Logistic regression using overdispersion ....................... 1530
  22.6.4 GLIMM modeling ................................................... 1533
  22.6.5 Recommendation ................................................... 1534
22.7 Example: Reprise: Are mosquitos choosy? A preference experiment with INCOMPLETE blocks ............................................. 1534
  22.7.1 Using the simple proportions as data ............................ 1537
  22.7.2 Using the empirical logits as data ............................... 1538
  22.7.3 Logistic regression using overdispersion ....................... 1540
  22.7.4 GLIMM modeling ................................................... 1542
  22.7.5 Recommendation ................................................... 1542

The suggested citation for this chapter of notes is:
22.1 Introduction

The previous chapters on chi-square tests, logistic regression, and logistic ANOVA only considered the simplest of experiment designs where the data were collected under a completely randomized design, i.e. every observation is independent of every other observation with complete randomization over experimental units and treatments.

It is possible to extend logistic regression and logistic ANOVA to more complex experimental designs. My course notes in a graduate course Stat-805 [http://www.stat.sfu.ca/~cschwarz/Stat-805](http://www.stat.sfu.ca/~cschwarz/Stat-805) have some details on these more advanced topics.

It is only recently that software has become readily available to analyze these types of experiments. In this chapter some variations from the simple CRD will be discussed.

22.2 Sacrificial pseudo-replication

In many experiments, the experimental unit is a collection of individuals, but measurements take place on the individual.

Hurlbert (1984) cites the example of an experiment to investigate the effect of fox predation upon the sex ratio of mice. Four colonies of mice are established. Two of the colonies are randomly chosen and a fox-proof fence is erected around the plots. The other two colonies serve as controls with out any fencing.

Here are the data (Table 6 of Hurlbert (1984)):

<table>
<thead>
<tr>
<th>Colony</th>
<th>% Males</th>
<th>Number males</th>
<th>Number females</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foxes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A₁</td>
<td>63%</td>
<td>22</td>
<td>13</td>
</tr>
<tr>
<td>A₂</td>
<td>56%</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>No foxes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B₁</td>
<td>60%</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>B₂</td>
<td>43%</td>
<td>97</td>
<td>130</td>
</tr>
</tbody>
</table>

This data has the characteristics of a chi-square test or logistic ANOVA. The factor (type of fencing) is categorical. The response, the sex of the mouse, is also categorical. Many researchers would simply pool over the replicates to give the pooled table:

<table>
<thead>
<tr>
<th>Colony</th>
<th>% Males</th>
<th>Number males</th>
<th>Number females</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foxes</td>
<td>61%</td>
<td>31</td>
<td>20</td>
</tr>
<tr>
<td>No foxes</td>
<td>44%</td>
<td>112</td>
<td>140</td>
</tr>
</tbody>
</table>

If a $\chi^2$ test is applied to the pooled data, the $p$-value is less than 5% indicating there is evidence that...
the sex ratio is not independent of the presence of foxes.

This “pooled analysis” is INCORRECT. According to Hurlbert (1984), the major problem is that individual units (the mice) are treated as independent objects, when in fact, there are not. Experimenters often pool experimental units from disparate sets of observations in order to do simple chi-square tests or logistic ANOVA. He specifically labels this pooling as sacrificial pseudo-replication.

Hurlbert (1984) identifies at least 4 reasons why the pooling is not valid:

- **non-independence of observation.** The 35 mice caught in $A_1$ can be regarded as 35 observations all subject to a common cause, as can the 16 mice in $A_2$, as each group were subject to a common influence in the patches. Consequently, the pooled mice are NOT independent; they represent two sets of interdependent or correlated observations. The pooled data set violates the fundamental assumption of independent observations.

- **throws away some information.** The pooling throws out the information on the variability among replicate plots. Without such information there is no proper way to assess the significance of the differences between treatments. Note that in previous cases of ordinary pseudo-replication (e.g. multiple fish within a tank), this information is also discarded but is not needed - what is needed is the variation among tanks, not among fish. In the latter case, averaging over the pseudo-replicates causes no problems.

- **confusion of experimental and observational units.** If one carries out a test on the pooled data, one is implicitly redefining the experimental unit to be individual mice and not the field plots. The enclosures (treatments) are applied at the plot level and not the mouse level. This is similar to the problem of multiple fish within a tank that is subject to a treatment.

- **unequal weighting.** Pooling weights the replicate plots differentially. For example, suppose that one enclosure had 1000 mice with 90% being male; and a second enclosure has 10 mice with 10% being male. The pooled data would have 1000 + 10 mice with 900 + 1 being male for an overall male ratio of 90%. Had the two enclosures been given equal weight, the average male percentage would be (90%+10%)/2=50%. In the above example, the number of mice captured in the plots varies from 16 to over 200; the plot with over 200 mice essentially drives the results.

There are multiple ways to analyze this data that avoid the problem that render the pooled analysis invalid.

### 22.3 Example: Fox-proofing mice colonies - dealing with sacrificial pseudo replication

Hurlbert (1984) cites the example of an experiment to investigate the effect of fox predation upon the sex ratio of mice. Four colonies of mice are established. Two of the colonies are randomly chosen and a fox-proof fence is erected around the plots. The other two colonies serve as controls with out any fencing.

Here are the data (Table 6 of Hurlbert (1984)):
CHAPTER 22. LOGISTIC REGRESSION - ADVANCED TOPICS

<table>
<thead>
<tr>
<th>Colony</th>
<th>% Males</th>
<th>Number males</th>
<th>Number females</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foxes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_1$</td>
<td>63%</td>
<td>22</td>
<td>13</td>
</tr>
<tr>
<td>$A_2$</td>
<td>56%</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>No foxes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B_1$</td>
<td>60%</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>$B_2$</td>
<td>43%</td>
<td>97</td>
<td>130</td>
</tr>
</tbody>
</table>

The data are stored in a JMP data table in the usual way:

Don’t forget to specify that colony, treatment and sex are all nominally scaled variables.

22.3.1 Using the simple proportions as data

Hurlbert (1984) suggests the proper way to analyze the above experiment is to essentially compute a single number for each plot and then do a two-sample t-test on the percentages. [This is equivalent to the ordinary averaging process that takes place in ordinary pseudo-replication or sub-sampling.]

We convert from the long to the wide format and compute the proportion of males in each colony:

Then a simple $t$-test is done on the observed proportion of males:
The estimated difference in the sex ratio between colonies that are subject to fox predation and colonies not subject to fox predation is .082 (SE .092) with p-values of .47 (pooled t-test) and .51 (unpooled t-test) respectively. As the p-values are quite large, there is NO evidence of a predation effect.

With only two replicates (the colonies), this experiment is likely to have very poor power to detect anything but gross differences.

The above analysis is not entirely satisfactory. The proportion of males have different variabilities because they are based on different number of total mice. As well, there may be over dispersion among colonies under the same treatment, i.e. the variation in the proportion of males may be larger among the two colonies under the same treatment than expected.

### 22.3.2 Logistic regression using overdispersion

Another “approximate” method to deal with the potential overdispersion among the colonies within the same treatment group (the colony effect) is to use a standard logistic regression but use the goodness-of-fit test to estimate an overdispersion effect. This overdispersion is then used to adjust the standard errors of estimates and the test statistics for hypothesis tests. Please consult the chapter on Logistic Regression for more details.

In order to get the correct estimate of over dispersion, the data table needs to be split and summarized to the colony level, similar to what was done for the simple analysis in the previous section:

---

1Not computed by *JMP.*
Because data has been summarized to the colony level, we can now use, and it is highly recommended that the second way to specify the model be used where the $Y$ box as the number of “success” first, and then the total number of trials second.

The Analyze->Fit Model platform can be used to fit a generalized linear model adjusting for overdispersion:

The first part of the output estimates the over dispersion using the goodness of fit test and divides the goodness-of-fit statistic by the degrees of freedom:

There are two ways to compute the lack-of-fit – using a Pearson chi-square test statistic and using the deviance test statistic. Both measures should give similar measures of goodness-of-fit and similar estimates of over dispersion. If these two measures differ greatly, it usually is because of sparse data. Generally speaking, the deviance test statistic is preferred for computing an over dispersion factor – unfortunately, JMP does not have the option of changing the computation.

Here JMP estimates the over dispersion factor a 1.4. The test for no treatment effect is adjusted for over dispersion

CAUTION: As always the coefficient presented in the estimates section for categorical vari-
ables is NOT TO BE TRUSTED! In this case it estimates 1/2 of the treatment effect. To get estimates of the treatment effect, use the *Contrast* feature under the upper red triangle to perform the estimation of the effect:

The estimated difference of .66 is on the log-odds scale, and implies that the odds-ratio of the proportion of males between colonies with foxes and without foxes is $\exp(.66) = 1.94x$ but the 95% confidence interval for the odds ratio is from $\exp(-.0791) = .92$ to $\exp(1.40) = 4.05$ which includes the value of 1 (indicating no difference in the odds of males between treatments.).
The use a simple overdispersion factor is not completely satisfactory. It assumes a single correction factor for all of the estimates and again estimates the different amount of mice in each colony.

### 22.3.3 GLIMM modeling the random effect of colony

A more "refined" analysis is now available using Generalized Linear Mixed Models (GLIMM) which have been implemented in \textit{SAS} and \textit{R}.

GLIMM allow the specification of random effects in much the same way as in advanced ANOVA models. This is a very general treatment and now allows us to analyze data from very complex experimental designs.

In this model, the model would be specified as:

\[
\text{logit}(p_{males}) = \text{Treatment Colony(Treatment)}(R)
\]

where the \text{Colony(Treatment)} would be the random effect of the experimental units (the colonies). Again, it is good practice to use unique labels for nested effects to make the model syntax a little easier to digest:

\[
\text{logit}(p_{males}) = \text{Treatment Colony(R)}
\]

A logistic type model is used.

There are many ways to fit these GLIMM and a discussion of the pros and cons of the various methods is beyond the scope of this course. The key problem is that an integration over the random effects must be performed to evaluate the likelihood. In standard linear mixed models based on normal theory, this can be done relatively painlessly because of the normality assumption. This is not true in GLIMMs and is an active area of research and so the results that are presented here may change over time.

\textit{JMP} does not have this capability (V.11). Ignore the rest of this section.

\textit{JMP} is unable to fit a GLIMM. Please refer to the \textit{SAS} or \textit{R} version of these notes.

In this case, given the limited sample sizes, I would have likely gone with a Bayesian analysis where the random effects can be modeled directly and there is no need to try and approximate the likelihood.

### 22.4 Example: Over-dispersed Seed Germination Data

This data is from the \textit{SAS} manual.

In a seed germination test, seeds of two cultivars were planted in pots of two soil conditions. The following data contains the observed proportion of seeds that germinated for various combinations of cultivar and soil condition. Variable \textit{n} represents the number of seeds planted in a pot, and \textit{r} represents the number germinated. \textit{CULT} and \textit{SOIL} are indicator variables, representing the cultivar and soil condition, respectively.
The data is available in the file `germination.csv` in the Sample Program Library at: [http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms](http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms) It is read in the usual ways in most computer packages.

Notice that the experimental unit is the pot (i.e. soil and cult were applied to the pot level), but the observational unit (what is actually measured) is the individual seed. The response variable for each individual seed is the either yes or no depending if it germinated or not. The fact that the data have been summarized to the pot level doesn’t change the observational variable.

Consequently, this is an example of pseudo-replication (experiment units not equal to the observational unit) and any analysis must account for this mismatch. As was seen in the previous section, there are four ways to analyze this data:

- Compute a response variable at the pot level. This is often the “average” response when the response variable is continuous. In this case, the “average” will be the empirical proportion that germinated in each pot. You would now have 20 values to be analyzed. The analysis would be a two-factor (cultivar and soil) completely randomized design (cultivar and soil were randomized to the pots) with the empirical proportion that hatched as the response variable.

- Perform a generalized-linear model but adjust the results for over dispersion. This is known as a quasi-binomial or quasi-logit analysis. The goodness-of-fit statistics can be used to estimate the over dispersion factor which is then used to adjust the standard errors, test statistics, and p-values.

- Perform a generalized linear mixed model analysis where the ordinary logistic regression is augmented with random effects. The key problem here is the need to integrate over the random effects which can be computationally challenging.
• Perform a Bayesian analysis. This analysis uses MCMC sampling to do the numerical integration over the likelihood. This analysis is beyond the scope of these notes.

Is there evidence of over dispersion? This would mean that the variation in pot-to-pot results is larger than expected under simple binomial variation. For example, here is plot of the 95% confidence interval for the proportion hatched compared to the average proportion hatched for each combination of soil and cultivar.

Notice that several of the confidence intervals from the individual pots do not cover the average germination proportion for each treatment. This indicates that there is some (random) pot effect that is influencing all of the seeds within the pot simultaneously (e.g. pot location). One way to estimate this would be to compare the variation of \( \hat{p} \) among pots within the same soil-cultivar combination with the theoretical variation based on binomial sampling within each pot. In order to account for the differing sample sizes in each pot, we will compute a “standardized normal” variable for pot \( i \) within soil-cultivar combination \( j \) as:

\[
z_{ij} = \frac{\hat{p}_{ij} - p_j}{\sqrt{p_j(1-p_j)/n_{ij}}}
\]

where \( p_j = \frac{\sum_{i} r_{ij}}{n_{ij}} \) is the average germination rate for the soil-cultivar combination \( j \).

If the additional pot-to-pot random variation was negligible, then \( Z \) should have an approximate standard normal distribution with a variance of 1. The actual variance of \( Z \) was found to be 4.5 indicating that the pot-to-pot variation in \( \hat{p} \) was about \( 4 \times \) larger than expected from a simple binomial variation.

Because of this extra-binomial variation, it is not proper to simply “ignore” the pot and pool over the five pots for each cultivar-soil combination. This would be an example of sacrificial pseudo-replication as outlined by Hurlbert (1984). As you will below, the pot-to-pot variation in the proportion that germinate is more than can be explained by the simple binomial variation, i.e. there is a large random effect of pots. 

---

\[2\] This plot was prepared by \( R \).
that must be incorporated.

### 22.4.1 Using the simple proportions as data

As suggested by Hurlbert (1984), a simple analysis could proceed by finding the proportion of seeds that germinated in each pot (e.g. for pot 1, \( \hat{p} = 8/16 = 0.50 \)) and then doing a two-factor CRD analysis on these proportions using the model:

\[
\hat{p} = \text{Soil} \times \text{Cult}
\]

This is not completely satisfactory because the number of seeds in each pot (\( n \)) varies considerably from pot-to-pot, and hence the variance of \( \hat{p} \) also varies\(^3\) A weighted analysis could be performed which would partially solve this problem.

The derived variable \( phat \) is computed using a formula and then is used in the Analyze->Fit Model platform:

Note that we weight the responses by the number of seeds in each pot.

This give an estimate of the residual standard deviation of:

\[
\text{Summary of Fit}
\]

Here some care must be taken in the interpretation depending if the analysis was a weighted or unweighted analysis. If the analysis was an unweighted analysis, then the residual variance is a combination of pot-to-pot variance and the variability of the \( \hat{p} \) in each pot. As a rough guess, the average germination rate is around 0.5 with an average sample size of around 45. This would give a binomial variance of \( .5(1 - .5)/45 = .005 \). Hence the pot-to-pot variance is about \( .026 - .005 = .020 \) which is about \( 4 \times \) that of the binomial variance which we saw earlier. If the analysis is a weighted analysis (as

\(^3\)The binomial variance of each \( \hat{p} \) for a pot would be found as \( \sqrt{p(1-p)/n} \)
was done), the residual variance is the individual seed variation plus the pot-to-pot variation (standardized to the individual seeds). As a rough guess, the average germination rate is around 0.5. This would give a binomial variance for individual seeds of \(0.5(1 - 0.5) = 0.25\). Hence the pot-to-pot variance is about \(1.00 - 0.25 = 0.75\) which is about \(3 \times\) that of the individual binomial variance.

The following results for the tests for no main effects and interactions:

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
\text{Source} & \text{Nparam} & \text{DF} & \text{Sum of Squares} & \text{F Ratio} & \text{Prob > F} \\
\hline
\text{cult} & 1 & 1 & 1.575667 & 1.5667 & 0.2287 \\
\text{soil} & 1 & 1 & 10.917590 & 10.8556 & 0.0046^* \\
\text{cult*soil} & 1 & 1 & 0.055253 & 0.0549 & 0.8177 \\
\hline
\end{array}
\]

Hence the naive analysis find no evidence of an interaction effect of soil and cultivar, no evidence of a main effect of cultivar, but strong evidence of a main effect of soil upon the germination rate.

The marginal estimates and differences can be found in the usual ways. In SAS, the \textit{lsmeans} statement provides estimates of marginal means and differences in marginal means. In \textit{JMP}, look at the leverage plots and effect tests areas of the output. Here are the estimates of the marginal means and difference in the germination rates for the effects of soil. Similar tables can be produced for the effects of the other factor and their interaction, but are not shown.
The estimated marginal mean germination rates are relatively precise. The standard errors are not equal because of the differing sample sizes in the pots in the various soil-cultivar combinations.

### 22.4.2 Logistic regression using overdispersion

The pots serve as “cluster” in this experiment, so the *ad hoc* methods of correcting for overdispersion caused by “cluster” effects can also be done, i.e. estimating the overdispersion factor ($\hat{c}$) and multiplying the standard errors by the $\sqrt{\hat{c}}$. This is known as a quasi-binomial or quasi-logit analysis.

Because data has been summarized to the pot level, we can use, *and it is highly recommended* that the second way to specify the model be used where the $Y$ box as the number of “success” first, and then the total number of trials second. Then a generalized linear model is fit is used in the *Analyze->Fit Model* platform:
Note that both the number of germinated seeds and the number trials are specified in the $Y$ box. Also don’t forget to check the box to adjust for over dispersion. **CAUTION:** If you stack the data and use the `freq` button for the individual counts of the responses, your estimate of over dispersion will be too small as **JMP** then does not aggregate up to the pot level to compute the over dispersion factor.

This will use the deviance to degrees of freedom to estimate the overdispersion factor. This gives the following estimate for the overdispersion factor:

<table>
<thead>
<tr>
<th>Goodness Of Fit Statistic</th>
<th>ChiSquare</th>
<th>DF</th>
<th>Prob&gt;ChiSq</th>
<th>Overdispersion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>66.7619</td>
<td>16</td>
<td>&lt;.0001*</td>
<td>4.1726</td>
</tr>
<tr>
<td>Deviance</td>
<td>68.3465</td>
<td>16</td>
<td>&lt;.0001*</td>
<td></td>
</tr>
</tbody>
</table>

We see that the overall variation is over $4 \times$ larger than expected under the binomial model. It is not possible to obtain an explicit estimate of the actual pot-to-pot variance.

The tests for no effects of soil or cultivar or their interaction on the germination rate are:

<table>
<thead>
<tr>
<th>Effect Tests Source</th>
<th>DF</th>
<th>ChiSquare</th>
<th>L-R</th>
<th>Prob&gt;ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>cultiv</td>
<td>1</td>
<td>1.5822698</td>
<td>0.2084</td>
<td></td>
</tr>
<tr>
<td>soil</td>
<td>1</td>
<td>10.84101</td>
<td>0.0011*</td>
<td></td>
</tr>
<tr>
<td>cult*soil</td>
<td>1</td>
<td>0.0472887</td>
<td>0.8279</td>
<td></td>
</tr>
</tbody>
</table>

The results are similar to the naive analysis seen earlier.

The estimated marginal “means” for the soil effect are now on the logit scale:

Unfortunately, **JMP** does not produce tables of marginal means and so you must estimate the marginal means and any contrasts using the `Estimate` and `Contrast` items from the red triangle in the output giving:
These can be converted back to the regular scale using the inverse transformation:

\[ \hat{p}_{soil=0} = \expit(-.5266) = .37 \]

which is similar to the previous results. Note that the \( se \) must be converted using the delta-method and not simply using the \( \expit \) transformation and is found to be

\[ se(\hat{p}_{soil=0}) = se(\logit(\hat{p}_{soil=0}))(\hat{p}_{soil=0})(1 - \hat{p}_{soil=0}) = .2063(.37)(1 - .37) = .047 \]

which is again pretty close to results from the simple analysis. The comparison in germination rates can also be done. The raw estimates are on the log-odds scale and need to be converted back to the odds-scale as illustrated elsewhere in these notes.

### 22.4.3 GLIMM modeling the random effect of pots

Finally, a generalized linear mixed model logistic ANOVA can be done that explicitly models the random effects of pots directly. Unfortunately this cannot be done in \textit{JMP} (Version 11). Ignore the rest of this section.

This corresponds to the model in the shorthand notation:

\[ \logit(p) = cult \ soil \ cult \times \ soil \ pots(cult \times soil) - R \]
or the generalized linear model:

\[ r_{ij} \sim Binomial(n_{ij}, hp_{ij}) \]

\[ \theta_{ij} = \logit(p_{ij}) \]

\[ \theta_{ij} = cult \ soil \ cult * soil \ pots(cult * soil) - R \]

Note that pots are nested within each cultivar-soil combination.

The estimated pot-to-pot variance (on the logit scale) is found as:

There is no easy way to convert this to an estimate of the pot-to-pot variation on the regular scale.

Tests for no main effects and interactions are:

These results are similar to those seen in the previous analyses.

Similarly, estimates of marginal effects (on the logit scale) of soil are found to be:

These are again comparable to the previous results and can be converted back to the original scale in the usual way. For example, the estimated difference (on the logit scale) between the two marginal means of the soil combinations is \(-.83\) (SE .30). This can be converted to an odds ratio using the methods seen earlier, i.e.

\[ OR_{soil \ 0:1} = exp(-.8257) = .44 \]

which implies that the odds of germination in soil=0 is only 44% of the odds of germination in soil=1. The 95% confidence interval for the odds-ratio is from (.23 \(\rightarrow\) .83).

One of the the advantage of the using generalized linear mixed model procedures is the availability of model diagnostic plots. For example, the residual plot panel:

indicates the potential presence of at least one outlier with a germination rate (on the logit scale) is well below that predicted. The normal-probability plot (on the logit scale) also identifies this one potential outlier.

In this simple experimental design, there is no obvious advantage to using logistic ANOVA with random effects; however, in more complex designs such as split-plot designs, this is the only way to proceed.

### 22.5 Example: Are mosquitos choosy? A preference experiment.

This (fictitious) data is based on experiment conducted by Dan Peach in the Department of Biological Sciences, Simon Fraser University.

Both sexes of mosquitoes consume sugar from flowers and from honeydew (no, not the fruit\(^4\)). Volatile chemicals emitted by flowers may be used by foraging mosquitoes to locate a source of sugar. In this experiment, certain oils from a plant were extracted and applied to delta traps within an enclosure (see below) with one trap having the extracted oils (the treatment) and the other delta trap serving as a control.

\(^4\)Visit [https://en.wikipedia.org/wiki/Honeydew_(secretion)] if you are curious.
Then approximately 60 mosquitos were released in the enclosure. After 2 hours, the traps were removed and the mosquitos in each trap counted. Not all mosquitos choose one of the traps. These mosquitos can be ignored as they are non-informative about preferences. Each experiment took one day, i.e. each replicate is done on a separate day over a three-week period.

The data is available in the file mosquitos.csv in the Sample Program Library at: http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms. Here is the raw (fictitious) data
CHAPTER 22. LOGISTIC REGRESSION - ADVANCED TOPICS

<table>
<thead>
<tr>
<th>Replicate</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>28</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>28</td>
<td>21</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>23</td>
</tr>
<tr>
<td>8</td>
<td>28</td>
<td>35</td>
</tr>
<tr>
<td>9</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>11</td>
<td>23</td>
<td>17</td>
</tr>
<tr>
<td>12</td>
<td>28</td>
<td>22</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>43</td>
<td>44</td>
</tr>
<tr>
<td>15</td>
<td>34</td>
<td>23</td>
</tr>
</tbody>
</table>

The data are imported into a *JMP* datatable and the derived variables for the total number of mosquitoes that chose, and the empirical proportion that chose the treatment (*phat*) and the empirical logit of the proportion that chose the treatment (*logit*) are computed in the usual way:

<table>
<thead>
<tr>
<th>Replicate</th>
<th>Treatment</th>
<th>Control</th>
<th>Total</th>
<th>phat</th>
<th>logit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32</td>
<td>23</td>
<td>55</td>
<td>0.582</td>
<td>0.330</td>
</tr>
<tr>
<td>2</td>
<td>28</td>
<td>7</td>
<td>35</td>
<td>0.800</td>
<td>1.386</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>16</td>
<td>42</td>
<td>0.619</td>
<td>0.486</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
<td>5</td>
<td>23</td>
<td>0.783</td>
<td>1.281</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>12</td>
<td>30</td>
<td>0.600</td>
<td>0.405</td>
</tr>
<tr>
<td>6</td>
<td>28</td>
<td>21</td>
<td>49</td>
<td>0.571</td>
<td>0.288</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>23</td>
<td>53</td>
<td>0.566</td>
<td>0.266</td>
</tr>
<tr>
<td>8</td>
<td>28</td>
<td>35</td>
<td>63</td>
<td>0.444</td>
<td>-0.223</td>
</tr>
</tbody>
</table>

Notice that the experimental unit is the trap, but the observational unit (what is actually measured) is the individual mosquito. The response variable for each individual mosquito is the either *Treatment* or *Control* trap. The fact that the data have been summarized to the trap level doesn’t change the observational unit. It is very tempting to think that the counts in each trap are the response variable, but that is incorrect – these are simply a summary of the data.

Consequently, this is yet another example of pseudo-replication (experiment units not equal to the observational unit) and any analysis must account for this mismatch. As was seen in the previous section, there are several ways to analyze this data:

- Compute a response variable at the trap level. This is often the “average” response when the response variable is continuous. In this case, the “average” will be the empirical proportion that choose the *Treatment* trap in each replicate. You would now have 15 values to be analyzed. The analysis would be a simple *t*-test on this empirical proportion to see if there is evidence that the proportion that choose the Treatment is different from 0.5.
• Similarly, compute the logit of the proportion that chose the treatment, and do a one-sample t-test on the empirical logits to see if there is evidence that the mean logit is different from $0 = \text{logit}(0.5)$.

• Fit a generalized-linear model but adjust the results for over dispersion. This is known as a quasi-binomial or quasi-logit analysis. The goodness-of-fit statistics can be used to estimate the over dispersion factor which is then used to adjust the standard errors, test statistics, and p-values.

• Perform a generalized linear mixed model analysis where the ordinary logistic regression is augmented with random effects. The key problem here is the need to integrate over the random effects when computing the likelihood and this can be computationally challenging.

• Perform a Bayesian analysis. This analysis uses MCMC sampling to do the numerical integration over the likelihood. This analysis is beyond the scope of these notes.

On the surface, the data looks like a paired design with two counts (corresponding to the two treatments) in each replicate. Many students would like to analyze the difference in the counts, i.e. in replicate 1, the difference is $9 = 32 - 23$. This is not a good choice of analysis for several reasons:

• The difference ignores the size of the base counts. For example a difference of $4 = 8 - 4$ has the same impact as a difference of $4 = 80 - 76$. Clearly a difference of 4 when the base counts are around 6 is more “interesting” than a difference of 4 when the base counts are around 78. This is why the proportion choosing one of the treatments is a better choice because now $0.67 = 8/12$ is more interesting than $.51 = 80/176$.

• Many effects operate multiplicatively on counts. For example, consider counts of $(8,4)$ and $(80,40)$. In both bases the proportion choosing the treatment cell (the first cell) is 0.67 despite the large difference in the base counts. A simple difference of 4 or 40 will hide the consistent effects. Consequently it makes sense then to use the raw proportions $0.67 = 8/12 = 80/120$ or the \( \log(RATIO) \) of the two counts $\log(2.0) = \log(8/4) = \log(80/40)$ as both give consistent effects. The \( \log() \) of the ratio is used to make the analysis symmetrical. For example, it should not matter if you analyze the proportion that chose the treatment or the proportion that chose control as one is simply the complement of the other. Similarly, $\log(8/4) = - \log(4/8)$ so the analysis will again be symmetrical depending on how you compute the ratio. Note that the ratio of the counts is actually the logit in disguise, i.e.

$$2.0 = \log(8/4) = \log((8/12)/(4/12)) = \log(p/(1-p)) = \text{logit}(p)$$

It is also tempting to do a simple chi-square test based on the pooled data, i.e. find the total number of mosquitos that choose each type of trap (371 for Treatment; 271 for control) and test if the pooled proportion is 0.5 (indicating no preference) using a one sample \( z \)-test for proportions. For example, here is the output from R:

```
*** It is tempting (but wrong) to try a one-sample z-test on the total proportions ***

1-sample proportions test with continuity correction
data: sum(mos$Treatment) out of sum(mos$total.mos), null probability 0.5X-squared = 15.266, df = 1, p-value = 9.336e-05alternative hypothesis: true p is not equal to 0.595 percent confidence interval:0.5385411 0.6162769sample estimates:
```
The (incorrect) p-value is .00009 with the estimated proportion choosing the treatment trap of 0.57 (\(se\) 0.019) and (incorrect) 95% confidence interval of 0.53 → 0.61. As you will see later, the reported standard error is too small and the reported confidence interval is too narrow. The key problem is that ignoring pseudo-replication ignores extra experimental unit variation – this is called overdispersion in the logistic world.

Is there evidence of overdispersion? This would mean that the variation in results is larger than expected under simple binomial variation. For example, here is plot of the 95% confidence intervals for the proportion that chose the treatment trap compared to the overall proportion that chose the treatment trap.

Notice that several of the confidence intervals from the individual enclosures do not cover the overall proportion that chose each treatment. This indicates that there is some (random) day and enclosure effect that is influencing all of the mosquitoes simultaneously (e.g. daily humidity; daily temperature, etc.).

Because of this extra-binomial variation, it is not proper to simply “ignore” the days and pool over all of the days and do a simple chi-square test or simple logistic regression. This would be an example of sacrificial pseudo-replication as outlined by Hurlbert (1984).

### 22.5.1 Using the simple proportions as data

As suggested by Hurlbert (1984), a simple analysis could proceed by finding the proportion of mosquitoes that choose the treatment trap in each replicate (e.g. for replicate 1, \(\hat{p} = 32/55 = 0.58\)) and then doing a one-sample t-test analysis on these proportions. This is not completely satisfactory because the number of mosquitoes in each replicate \((n)\) varies considerably from replicate-to-replicated, and hence the variance of \(\hat{p}\) also varies. A weighted analysis could be performed which would partially solve this problem.

---

5This plot was prepared by R.

6The binomial variance of each \(\hat{p}\) for a replicate would be found as \(\sqrt{p(1-p)/n}\)
If the mosquitos were choosing the traps simply at random, we would expect that \( p = 0.5 \) - this is the hypothesized value in the \( t \)-test:

\[
H_0 : \mu_p = 0.5 \\
H_A : \mu_p \neq 0.5
\]

A plot of the empirical probability of selecting the treatment trap shows that most (but not all of the proportion exceed 0.5. The approximate confidence interval does not appear to include the hypothesized value of 0.5.

The derived variable \( \text{phat} \) is computed using a formula and then is used in the Analyze->Distribution platform to do a test that the mean proportion is 0.5:
The \( p \)-value for no preference is .00078. The estimated proportion that choose the treatment trap is 0.60 (se 0.024) with a 95% ci of 0.55 → 0.66. This analysis is not completely satisfactory for two reasons:

- The distribution of the empirical proportions may not have a nice “normal” shape, especially if the estimates are close to 0 or 1. As well, the reported confidence bounds may fall outside of 0 and 1.
- Each replicate is given equal weight even though some replicates had close to 60 mosquitos making a choice and some replicates only had 10 mosquitos that made a choice.

The first item can be solved using an analysis on the empirical logits. The second item can be resolved using a weighted analysis, or as shown below a logistic analysis (adjusted for overdispersion).

### 22.5.2 Using the empirical logits as data

We proceed by finding the empirical logit of the proportion of mosquitos that choose the treatment trap in each replicate (e.g. for replicate 1, \( \text{logit} = \logit(32/55) = \logit(0.58) = \log(0.58/0.42) = 0.33 \)) and then doing a one-sample \( t \)-test analysis on these logits. This is not completely satisfactory because the number of mosquitos in each replicate (\( n \)) varies considerably from replicate-to-replicate, and hence the variance of the logits also varies.

If the mosquitos were choosing the traps simply at random, we would expect that \( p = 0.5 \) or the \( \text{logit} = 0.0 \). Hence our hypothesis is now

\[
H_0 : \mu_{\text{logit}} = 0.0 \]

\[
H_A : \mu_{\text{logit}} \neq 0.0
\]

The derived variable \( \text{logit} \) is computed using a formula and then is used in the Analyze->Distribution platform:
Note that we weight the responses by the number of mosquitos in each replicate.

The \( p \)-value for no preference is .0013. The estimated logit of the proportion that chose the treatment is 0.45 (\( se \) 0.11) which can be back-transformed to estimate the proportion that choose the treatment trap as 0.61 (\( se \) 0.027) with a 95% ci of 0.55 → 0.66. These results are very similar to the the previous analysis.

### 22.5.3 Logistic regression using overdispersion

The enclosures serve as “cluster” in this experiment, so the \textit{ad hoc} methods of correcting for overdispersion caused by “cluster” effects can also be done, i.e. estimating the overdispersion factor (\( \hat{\phi} \)) and multiplying the standard errors by the \( \sqrt{\hat{\phi}} \). This is known as a quasi-binomial or quasi-logit analysis.

Because data has been summarized to the replicate level, we can use, \textit{and it is highly recommended} that the second way to specify the model be used where the \( Y \) box as the number of “success” first, and then the total number of trials second. Then a generalized linear model is fit is used in the \textit{Analyze}–\textit{Fit Model} platform:

Note that both the number of mosquitos that chose the treatment trap and the number mosquitos that...
chose either trap are specified in the Y box. There are NO terms in the Effects box to give the intercept only model. Also don’t forget to check the box to adjust for over dispersion. **CAUTION:** If you stack the data and use the freq button for the individual counts of the responses, your estimate of over dispersion will be too small as *JMP* then does not aggregate up to the replicate level to compute the over dispersion factor.

This gives:

![Image of a statistical output](image)

The *p*-value for no preference is .0008. Note that different packages may give slight different values for the *p*-value depending on the way the overdispersion factor is computed (e.g. using the deviance or the Pearson chi-square) and the way the *p*-value is computed (e.g. using quasilikelihood or a Wald statistic). All are asymptotically equivalent but can differ in small sample sizes – contact me for more details. The estimated logit of the preference for the treatment is 0.31 (se .094) which can be converted back to the estimate of the proportion choosing the treatment of 0.57 (se 0.23) with a 95% confidence interval from 0.53 → 0.62.

Again, the results are similar to what was seen previously.

The estimated overdispersion factor is 1.39 (see above) indicating that standard errors from a simple logistic fit must be adjusted by a factor of $\sqrt{1.39}$. This was automatically done above.

### 22.5.4 GLIMM modeling the random effect of replicates

Finally, a generalized linear mixed model logistic ANOVA can be done that explicitly models the random effects of replicates directly. **Unfortunately this cannot be done in *JMP* (Version 11). Ignore the rest of this section.**

This corresponds to the model in the shorthand notation:

$$\text{logit}(p) = 1 + \text{replicates}(R)$$
or the generalized linear model:

\[
\text{treatment}_i \sim \text{Binomial}(n_i, p_i) \\
\theta_i = \logit(p_i) \\
\theta_i = 1 + \text{replicates}(R)
\]

The estimated replicate variance (on the logit scale) is found as:

The results of the fit are:

The \(p\)-value is The \(p\)-value may differ slightly between packages depending a likelihood ratio or Wald test is used (contact me for details). The estimated logit of the preference for the treatment is 0.34 (\(se\) .098) which can be converted back to the estimate of the proportion choosing the treatment of 0.58 (\(se\) 0.24) with a 95% confidence interval from 0.54 → 0.64.

It is again tempting to fit a simple generalized linear model with FIXED block effects rather than random block effects. This gives similar results as seen in the previous analysis but is again not completely satisfactory. The key problem is that the analysis with fixed block effects assumes that if the experiment was replicated, that the replicate effects would be identical the next time around. Given the the replicate effect are “day” effects and represent the effects of external factors that vary by day (e.g. temperature, humidity, handling effects), this is unlikely to be true. An example of the analysis using fixed block effects is found in the \(R\) code. As expected, the estimates are similar, but the reported standard errors are too small and the confidence interval too narrow because that extra source of variation if new replicates were done is ignored.

22.5.5 Recommendation

As long as the number of mosquitos that select one of the traps is roughly the same among all of the enclosures, the analysis of the empirical logits using a \(t\)-test analysis is simple and relatively straightforward to do with modern statistical software.

If the number of mosquitos that makes a choice is highly variable and/or the empirical probabilities of selection are close to 0 or 1, then the more advanced methods will be preferred. Note that if some of the cages has 100% or 0% of mosquitos choosing one of the traps, then many of the methods will fail because the empirical logit are \(±\infty\) respectively. You can try using the adjustment to the empirical logits of adding \(1/2\) to the number of failures and the number of successes, or Bayesian methods often work quite well.

22.6 Example: Reprise: Are mosquitos choosy? A preference experiment with complete blocks.

This (fictitious) data is based on experiment conducted by Dan Peach in the Department of Biological Sciences, Simon Fraser University.

Both sexes of mosquitoes consume sugar from flowers and from honeydew (no, not the fruit\[7\]. Volatile chemicals emitted by flowers may be used by foraging mosquitoes to locate a source of sugar. In this experiment, three different oils \((T1, T2, T3)\) from a plant were extracted and applied to delta traps

within one of three enclosures (see previous section for a picture) on each day with one trap having the
extracted oils (the treatment) and the other delta trap serving as a control.

Then approximately 60 mosquitos were released in the enclosure. After 2 hours, the traps were
removed and the mosquitos in each trap counted. Not all mosquitos choose one of the traps. These
mosquitos can be ignored as they are non-informative about preferences. Each experiment took one day,
i.e. all three oils were tested on a single day in 3 enclosures and were randomly assigned to the three
enclosures in a day similar to:

<table>
<thead>
<tr>
<th>Day</th>
<th>Chamber 1</th>
<th>Chamber 2</th>
<th>Chamber 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T1 vs. C</td>
<td>T2 vs. C</td>
<td>T3 vs. C</td>
</tr>
<tr>
<td>2</td>
<td>T2 vs. C</td>
<td>T3 vs. C</td>
<td>T1 vs. C</td>
</tr>
<tr>
<td>3</td>
<td>T2 vs. C</td>
<td>T1 vs. C</td>
<td>T3 vs. C</td>
</tr>
<tr>
<td>etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that an alternative design where all three oils and the control traps are placed in the same cage
(i.e. 4 traps per enclosure) is much more difficult to analyze. They key problem with this revised design
is the constraint that mosquitos must choose one of the traps and so the sum of the counts in the trap must
be less than or equal to the number released. This introduces a negative dependence among the traps, i.e.
if one trap catches many mosquitos, the other traps must capture fewer. This negative dependence also
exists in the two traps/enclosure designs, but by measuring just the proportion of mosquitos selecting
one of the traps or by using the logit of the capture probability, this negative dependency is translated
into just a single variable. A similar analysis pathway occurs if there were 4 trap types per enclosure –
the 4 counts are translated into 3 response variables, but there are now many ways in which this can be
done. Please contact me for details on the analysis of such experiments.

There are now TWO different questions of interest

- Is there evidence for each oil (T1, T2, T3) that mosquitos choose it with a probability different
  than 0.5, i.e. is there evidence of a an attractiveness or repulsion for each oil

- Can we compare the efficacy of the three treatments

The first questions above is similar to what was done in the the previous section when only 1 treatment
is compared to a control trap.

The data is available in the file *mosquitos2.csv* in the Sample Program Library at: [http://www.stat.sfu.ca/~cshwarz/Stat-650/Notes/MyPrograms](http://www.stat.sfu.ca/~cshwarz/Stat-650/Notes/MyPrograms) The full dataset is not given here
but snippets are shown below.

The data are imported into a *JMP* datatable and the derived variables for the total number of mosquitos
that chose, and the empirical proportion that chose the treatment (*p_treatment*) and the empirical logit
of the proportion that chose the treatment (*logit_treatment*) are computed in the usual way:
CHAPTER 22. LOGISTIC REGRESSION - ADVANCED TOPICS

Notice that the experimental unit is the trap, but the observational unit (what is actually measured) is the individual mosquito. The response variable for each individual mosquito is the either Treatment or Control trap. The fact that the data have been summarized to the replicate level doesn’t change the observational unit. It is very tempting to think that the counts in each trap are the response variable, but that is incorrect – these are simply a summary of the data.

Consequently, this is yet another example of pseudo-replication (experiment units not equal to the observational unit) and any analysis must account for this mismatch. As was seen in the previous sections, there are several ways to analyze this data:

- Compute a response variable at the trap (experimental unit) level. This is often the “average” response when the response variable is continuous. In this case, the “average” will be the empirical proportion that choose the Treatment trap in each replicate. The analysis would be a simple t-test on this empirical proportion to see if there is evidence that the proportion that choose the Treatment is different from 0.5 for each of the individual oils. A combined analysis that analyzes all the oils simultaneously can also be used which has some advantages as will be seen. A simple RCB analysis can be used to compare the efficacy of the oils.

- Similarly, compute the logit of the proportion that chose the treatment, and do a one-sample t-test on the empirical logits to see if there is evidence that the mean logit is different from \( 0 = \text{logit}(0.5) \). A combined analysis that analyzes all the oils simultaneously can also be used which has some advantages as will be seen. A simple RCB analysis can be used to compare the efficacy of the oils among themselves.

- Fit a generalized-linear model but adjust the results for over dispersion. This is known as a quasi-binomial or quasi-logit analysis. The goodness-of-fit statistics can be used to estimate the over dispersion factor which is then used to adjust the standard errors, test statistics, and p-values.

- Perform a generalized linear mixed model analysis where the ordinary logistic regression is augmented with random effects. The key problem here is the need to integrate over the random effects when computing the likelihood and this can be computationally challenging.

- Perform a Bayesian analysis. This analysis uses MCMC sampling to do the numerical integration over the likelihood. This analysis is beyond the scope of these notes.

On the surface, the data looks like a paired design with two counts (corresponding to the two treatments) in each enclosure. Many students would like to analyze the difference in the counts from each enclosure. This is not a good choice of analysis for several reasons:
• The difference ignores the size of the base counts. For example a difference of $4 = 8 - 4$ has the same impact as a difference of $4 = 80 - 76$. Clearly a difference of 4 when the base counts are around 6 is more “interesting” than a difference of 4 when the base counts are around 78. This is why the proportion choosing one of the treatments is a better choice because now $0.67 = 8/12$ is more interesting than $.51 = 80/176$.

• Many effects operate multiplicatively on counts. For example, consider counts of $(8, 4)$ and $(80, 40)$. In both bases the proportion choosing the treatment cell (the first cell) is 0.67 despite the large difference in the base counts. A simple difference of 4 or 40 will hide the consistent effects. Consequently it makes sense then to use the raw proportions $0.67 = 8/12 = 80/120$ or the log($RATIO$) of the two counts $\log(2.0) = \log(8/4) = \log(80/40)$ as both give consistent effects. The $\log()$ of the ratio is used to make the analysis symmetrical. For example, it should not matter if you analyze the proportion that chose the treatment or the proportion that chose control as one is simply the complement of the other. Similarly, $\log(8/4) = -\log(4/8)$ so the analysis will again be symmetrical depending on how you compute the ratio. Note that the ratio of the counts is actually the logit in disguise, i.e.

$$2.0 = \log(8/4) = \log((8/12)/(4/12)) = \log(p/(1-p)) = \text{logit}(p)$$

It is also tempting to do a simple chi-square test based on the pooled data, i.e. find the total number of mosquitos that choose each type of trap and test if the pooled proportion is 0.5 (indicating no preference) using a one sample z-test for proportions for each oil. For example, here is the output from $R$ for the three oils:

```r
*** It is tempting (but wrong) to try a one-sample z-test on the total proportions for the pooled data. The estimated overall proportion isn’t badly estimated, but reported SE are too small. 

1-sample proportions test with continuity correction
data: sum(x$Treatment) out of sum(x$total.mos), null probability 0.5
X-squared = 28.136, df = 1, p-value = 1.131e-07
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
0.5618727 0.6338312
sample estimates:
p 0.5983718

1-sample proportions test with continuity correction
data: sum(x$Treatment) out of sum(x$total.mos), null probability 0.5
X-squared = 107.22, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
0.6576330 0.7259704
sample estimates:
p 0.6928375

1-sample proportions test with continuity correction
data: sum(x$Treatment) out of sum(x$total.mos), null probability 0.5
X-squared = 421.8, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
0.8069666 0.8848268
sample estimates:
p 0.8458469
```
X-squared = 6.7246, df = 1, p-value = 0.009509
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
0.5119093 0.5864170
sample estimates:
p
0.549435

<table>
<thead>
<tr>
<th>trt</th>
<th>sum.treatment</th>
<th>sum.control</th>
<th>statistic</th>
<th>df</th>
<th>p.value</th>
<th>phat</th>
<th>se</th>
<th>lcl</th>
<th>ucl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T1</td>
<td>441</td>
<td>296</td>
<td>28.135685</td>
<td>1.131011e-07</td>
<td>0.5983718</td>
<td>0.01805776</td>
<td>0.5618727</td>
<td>0.6338312</td>
</tr>
<tr>
<td>2</td>
<td>T2</td>
<td>503</td>
<td>223</td>
<td>107.219008</td>
<td>3.985953e-25</td>
<td>0.6928375</td>
<td>0.01712109</td>
<td>0.6576330</td>
<td>0.7259704</td>
</tr>
<tr>
<td>3</td>
<td>T3</td>
<td>389</td>
<td>319</td>
<td>6.724576</td>
<td>9.509332e-03</td>
<td>0.5494350</td>
<td>0.01869908</td>
<td>0.5119093</td>
<td>0.5864170</td>
</tr>
</tbody>
</table>

The estimated proportions are about correct, but the reported standard errors and reported \( p \)-values will be TOO SMALL. The key problem is that ignoring pseudo-replication ignores extra experimental unit variation – this is called overdispersion in the logistic world.

Similarly, it is tempting to pool over the data and do a chi-square test to compare the three oils, i.e. create a two-way table of oil by choosing the treatment or control trap. For example, here is the output from \( R \) for the pooled chi-square test on the counts for the three oils:

```
*** It is tempting (but wrong) to try a pooled-chisquare to compare treatments ***
The reported p-value is too small
<table>
<thead>
<tr>
<th>trt</th>
<th>total.Treatment</th>
<th>total.control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T1</td>
<td>441</td>
</tr>
<tr>
<td>2</td>
<td>T2</td>
<td>503</td>
</tr>
<tr>
<td>3</td>
<td>T3</td>
<td>389</td>
</tr>
</tbody>
</table>

Pearson’s Chi-squared test

Data:  bad.table[, 2:3]
X-squared = 32.252, df = 2, p-value = 9.923e-08
```

The Pearson chi-square test is inappropriate to compare the efficacy because of the pseudo-replication and overdispersion in the experiment.

Is there evidence of over dispersion? This would mean that the variation in results is larger than expected under simple binomial variation. For example, here is plot of the 95% confidence intervals for the proportion that chose the treatment trap compared to the overall proportion that chose the treatment trap[†]

[†]This plot was prepared by \( R \). A similar plot was also prepared using SAS but is not shown here.
Notice that several of the confidence intervals from the individual cages do not cover the overall proportion that chose each treatment for each oil. This indicates that there is some (random) day and enclosure effect that is influencing all of the mosquitoes simultaneously (e.g. replicate humidity; replicate temperature, etc.).

Because of this extra-binomial variation, it is not proper to simply “ignore” the replicates and pool over all of the replicates and do a simple chi-square test or simple logistic regression. This would be an example of sacrificial pseudo-replication as outlined by Hurlbert (1984).

Before doing any analysis, it is always a good idea to make sure there are no missing data, and if a blocked design, to see if the design is complete:

<table>
<thead>
<tr>
<th>Day</th>
<th>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>T1 T1 T1 T1 T1 T1 T1 T1 T1 T1 T1 T1 T1 T1 T1</td>
</tr>
<tr>
<td>T1</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td>
</tr>
<tr>
<td>T2</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td>
</tr>
<tr>
<td>T3</td>
<td>1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td>
</tr>
</tbody>
</table>

where a 1 indicates that the treatment (oil) was tested on that day. We see that every oil is tested on every day and every day has every oil tested. The design is complete. This makes the analysis somewhat simpler.

22.6.1 Using the simple proportions as data

As suggested by Hurlbert (1984), a simple analysis could proceed by finding the proportion of mosquitoes that choose the treatment trap in each replicate and then doing a one-sample $t$-test analysis on these proportions to see if proportion is different from 0.5 This is not completely satisfactory because the number of mosquitoes in each replicate ($n$) varies considerably from replicate-to-replicated, and hence the variance of $\hat{p}$ also varies. A weighted analysis could be performed which would partially solve this problem.

If the mosquitoes were choosing the traps simply at random, we would expect that $p = 0.5$ - this is

\[ \text{Binomial variance of } \hat{p} \text{ for a replicate would be found as } \sqrt{p(1-p)/n} \]
the hypothesized value in the $t$-test for each oil:

$$H_0 : \mu_p = 0.5$$

$$H_A : \mu_p \neq 0.5$$

A plot of the empirical probability of selecting the treatment trap

The approximate confidence intervals (notches on the boxplots) can be used to (approximately) test the hypothesized value of 0.5.

The derived variable $p_{\text{treatment}}$ is computed using a formula and then is used in the Analyze->Distribution platform to do a $t$-test that the mean proportion is 0.5 for each individual treatment:
The estimated proportion that choose the treatment trap for each oil are shown along with the \( p \)-value for each oil’s test and the 95% confidence interval for the proportion that choose the oil. This analysis is not completely satisfactory for several reasons:

- The distribution of the empirical proportions may not have a nice “normal” shape, especially if the estimates are close to 0 or 1. As well, the reported confidence bounds may fall outside of 0 and 1.
- Each replicate is given equal weight even not all trails had the same number of mosquitos that made a choice.
- Each oil was analyzed separately, even though they were all tested on the same day. Consequently a separate estimate of variance was computed for each oil’s analysis which is inefficient and leads to larger reported standard errors and a loss of power to detect effects.
The first item can be solved using an analysis on the empirical logits. The second item can be resolved using a weighted analysis, or as shown below a logistic analysis (adjusted for overdispersion). The third item can be resolved by fitting an RCB on the empirical proportions and then testing if the marginal “means” have the value of 0.5.

In order to make the analysis more efficient, we fit an RCB model with RANDOM blocks to fully capture the variability in the data. If an RCB model with fixed blocks was fit, it would assume that the exact same set of days and enclosures would be used in the future. While we may have control over the enclosures, we have no control over environmental variables such as humidity that may change from day to day. The Analyze-＞Fit Model platform is used to fit the RCB model with random blocks:

This gives the marginal means and confidence limits, but does not let you test if the individual marginal proportions are equal to 0.5 except through a custom test (not shown, but see me for details). You could simply see if the confidence interval for the marginal proportion for each treatment contains the value 0.5.

Notice that now all of the marginal “means” all have the same standard error. This is because the design was complete (i.e. every block had every treatment), and so a pooled estimate of the residual variation was used to estimate the standard error for each estimate. On average, these standard errors will be slightly smaller than those computed previously and the tests of the hypothesis will be slightly more powerful. The effect of analyzing each oil separately vs. a combined analysis will not be that great unless there only a few blocks (days) of data.

Once the RCB analysis is fit, it is then straight forward to compare the efficacy of the oils in the usual way, i.e. using a multiple comparison procedure following the overall test for now differences in efficacy: The test for the equality of treatment effect and the estimated difference between the margin values and the compact letter display are shown in the usual way:
The compact letter display (*cld*) is interpreted in the usual fashion.

A plot of the results can be make in the usual fashion as shown using *R*. 

<table>
<thead>
<tr>
<th>Source</th>
<th>Nparm</th>
<th>DF</th>
<th>DFDen</th>
<th>F Ratio</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>2</td>
<td>2</td>
<td>28</td>
<td>9.6828</td>
<td>0.0006*</td>
</tr>
</tbody>
</table>

**LSMeans Differences Tukey HSD**

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean[j]-Mean[j]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std Err Dif</td>
<td>0</td>
<td>0.0949</td>
<td>0.05112</td>
</tr>
<tr>
<td>Lower CL Dif</td>
<td>0</td>
<td>0.03368</td>
<td>0.03368</td>
</tr>
<tr>
<td>Upper CL Dif</td>
<td>0</td>
<td>-0.1763</td>
<td>-0.0322</td>
</tr>
<tr>
<td>T1</td>
<td>-0.0949</td>
<td>0</td>
<td>0.05112</td>
</tr>
<tr>
<td>T2</td>
<td>0.09493</td>
<td>0</td>
<td>0.14605</td>
</tr>
<tr>
<td>T3</td>
<td>-0.0511</td>
<td>-0.1461</td>
<td>0</td>
</tr>
</tbody>
</table>

**Least Squares Means (LsMeans)**

<table>
<thead>
<tr>
<th>Level</th>
<th>Sq Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2</td>
<td>0.69233580</td>
</tr>
<tr>
<td>T1</td>
<td>0.59740430</td>
</tr>
<tr>
<td>T3</td>
<td>0.54628159</td>
</tr>
</tbody>
</table>

Levels not connected by same letter are significantly different.
CHAPTER 22. LOGISTIC REGRESSION - ADVANCED TOPICS

22.6.2 Using the empirical logits as data

In an analogous fashion, we proceed by finding the empirical logit of the proportion of mosquitos that choose the treatment trap in each replicate and then doing a one-sample $t$-test analysis on these logits and/or an RCB analysis on the logits.

If the mosquitos were choosing the traps simply at random, we would expect that $p = 0.5$ or the logit $= 0.0$ for each marginal logit. Hence our hypothesis is now:

$H_0 : \mu_{\text{logit}} = 0.0$

$H_A : \mu_{\text{logit}} \neq 0.0$

for each marginal logit.

The derived variable logit_treatment is computed using a formula and then is used in the Analyze-Distribution platform to do a $t$-test that the mean logit is 0.0 for each individual treatment.
The estimated logit of the proportion can be back-transformed to estimate the proportion that choose the treatment trap. These results are very similar to the previous analysis.

We again fit the RCB model to the logits to do the comparison of the marginal logits for all three oils simultaneously and to compare the efficacy of the three oils: The Analyze->Fit Model platform is used to fit the RCB model with random blocks in a similar fashion as before. This gives the marginal means and confidence limits, but does not let you test if the individual marginal logits are equal to 0.0 except through a custom test (not shown, but see me for details). You could simply see if the confidence interval for the marginal logit for each treatment contains the value 0.0.
Notice that now all of the marginal “logits” all have the same standard error for the same reasons as when we analyzed the empirical proportions directly.

Once the RCB analysis is fit, it is then straight forward to compare the efficacy of the oils in the usual way, i.e. using a multiple comparison procedure following the overall test for now differences in efficacy: The test for the equality of treatment effect and the estimated difference between the margin values and the compact letter display are shown in the usual way:

The compact letter display (cld) is interpreted in the usual fashion. Notice that when we back transform the difference in logits, it is equivalent to the odds ratio (OR) for choosing the treated trap compared to the control trap and an $OR = 1$ indicates no evidence of a difference in efficacy between the oils.

A plot of the results can be make in the usual fashion as shown using $R$. 
CHAPTER 22. LOGISTIC REGRESSION - ADVANCED TOPICS

22.6.3 Logistic regression using overdispersion

The enclosures serve as “cluster” in this experiment, so the \textit{ad hoc} methods of correcting for overdispersion caused by “cluster” effects can also be done, i.e. estimating the overdispersion factor (\( \hat{c} \)) and multiplying the standard errors by \( \sqrt{\hat{c}} \). This is known as a quasi-binomial or quasi-logit analysis.

Because data has been summarized to the replicate level, we can use, \textit{and it is highly recommended} that the second way to specify the model be used where the \( Y \) box as the number of “success” first, and then the total number of trials second. Then a generalized linear model is fit is used in the \textit{Analyze} -> \textit{Fit Model} platform:
Note that both the number of mosquitos that chose the treatment trap and the number mosquitos that chose either trap are specified in the $Y$ box. There are NO terms in the Effects box to give the intercept only model. Also don’t forget to check the box to adjust for over dispersion. **CAUTION:** If you stack the data and use the **Freq** button for the individual counts of the responses, your estimate of over dispersion will be too small as *JMP* then does not aggregate up to the replicate level to compute the over dispersion factor.

This gives:

![Generalized Linear Model Fit](image)

This analysis is not completely satisfactory because *Day* is treated as a fixed effect and so the estimated marginal logits will tend to have reported standard errors that are too small. See me for additional details.

Note that different packages may give slight different values for the $p$-values for the marginal logits depending on the way the overdispersion factor is computed (e.g. using the deviance or the Pearson chi-square) and the way the $p$-value is computed (e.g. using quasilikelihood or a Wald statistic). All are asymptotically equivalent but can differ in small sample sizes – contact me for more details.

Once the RCB analysis is fit, it is then straight forward to compare the efficacy of the oils in the usual way, i.e. using a multiple comparison procedure after the test for no differences in efficacy: The output includes the effect tests:
but does not automatically give the differences in marginal logits nor the multiple comparison (groan).
We need to do an individual contrast for each pair of treatments and remember to adjust for multiple
testing (e.g. using a Bonferonni correction). For example, here is the contrast between the first two
treatments:

It is a pity that JMP doesn’t make this more convenient.

The compact letter display \((cld)\) is interpreted in the usual fashion. Notice that when we back
transform the difference in logits, it is equivalent to the odds ratio (OR) for choosing the treated trap compared
to the control trap and an \(OR = 1\) indicates no evidence of a difference in efficacy between the oils.
A plot of the results can be made in the usual fashion as well (but is not shown) in these notes.

The estimated overdispersion factor is:

<table>
<thead>
<tr>
<th>Goodness Of Fit Statistic</th>
<th>ChiSquare</th>
<th>DF</th>
<th>Prob&gt;ChiSq</th>
<th>Overdispersion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>50.9485</td>
<td>28</td>
<td>0.0051*</td>
<td>1.8196</td>
</tr>
<tr>
<td>Deviance</td>
<td>51.0747</td>
<td>28</td>
<td>0.0049*</td>
<td></td>
</tr>
</tbody>
</table>

indicating that standard errors from a simple logistic fit must be adjusted by the square root of the factor. This was automatically done above.

This analysis is not completely satisfactory because Day is treated as a fixed effect and so the estimated marginal logits will tend to have reported standard errors that are too small. See me for additional details.

22.6.4 GLIMM modeling

Finally, a generalized linear mixed model logistic ANOVA can be done that explicitly models the random effects of days and enclosures (the combination of day and cage) directly. Unfortunately this cannot be done in JMP (Version 11). Ignore the rest of this section.

This corresponds to the model in the shorthand notation:

\[
\text{logit}(p) = \text{trt} + \text{Day}(R) + \text{Experiment}(R)
\]

or the generalized linear model:

\[
treatment_i \sim \text{Binomial}(n_i, p_i) \\
\theta_i = \text{logit}(p_i) \\
\theta_i = \text{trt} + \text{Day}(R) + \text{Experiment}(R)
\]

where the Experiment term represent the combination of day and cage. You should not use Cage as the random effect because cage 1 on day 1 is not necessarily the same “cage” on day 2, despite it having the same physical location etc. Perhaps something happened to the cage on day 2 that made it subtly different from the same cage on the previous day.

The usual comparison among treatments is made after the test for no treatment effect.

The \( p \)-values may differ slightly between packages depending a likelihood ratio or Wald test is used (contact me for details).

The difference of logits corresponds to the \( \log(OddsRation) \) of the proportion choosing the treatment trap vs the control trap between two different oils. A back transformation can give estimates of the odds ratio (and confidence limits) directly. Notice that you do NOT take the anti-log of the \( se \) of the \( \log(\text{odds ratio}) \) to get the \( se \) of the odds-ratio, but must do a delta-method transformation. Contact me for details.

The estimated variance components (on the logit scale) are found as:
22.6.5 Recommendation

As long as the number of mosquitoes that select one of the traps is roughly the same among all of the enclosures, the analysis of the empirical logits using a Randomized Complete Block (RCB) analysis is simple and relatively straightforward to do with modern statistical software. The key to this model is the ability to declare blocks as a random factor to ensure that the individual comparisons of the marginal logits to 0 captures the true extent of the uncertainty.

If the number of mosquitoes that makes a choice is highly variable and/or the empirical probabilities of selection are close to 0 or 1, then the more advanced methods will be preferred. Note that if some of the cages has 100% or 0% of mosquitoes choosing one of the traps, then many of the methods will fail because the empirical logit are \( \pm \infty \) respectively. In such cases, Bayesian methods can often work quite well.

22.7 Example: Reprise: Are mosquitoes choosy? A preference experiment with INCOMPLETE blocks.

This (fictitious) data is based on experiment conducted by Dan Peach in the Department of Biological Sciences, Simon Fraser University.

This third analysis continues that of the previous section but now deals with cases where there is missing data, i.e. not all oils are tested on all days. For example, you may have 5 different oils, but only three enclosures that can be used in a day. Consequently, not all oils can be tested in one day. Or, missing data (e.g. something went wrong with an enclosure) implies that the blocks are also not complete. We will (arbitrarily) delete some of the data from the previous section and redo the analyses.

As before, both sexes of mosquitoes consume sugar from flowers and from honeydew (no, not the fruit\(^\text{10}\)). Volatile chemicals emitted by flowers may be used by foraging mosquitoes to locate a source of sugar. In this experiment, three different oils (\(T_1, T_2, T_3\)) from a plant were extracted and applied to delta traps within one of three enclosures (see previous section for a picture) on each day with one trap having the extracted oils (the treatment) and the other delta trap serving as a control.

Then approximately 60 mosquitos were released in the enclosure. After 2 hours, the traps were removed and the mosquitos in each trap counted. Not all mosquitos choose one of the traps. These mosquitos can be ignored as they are non-informative about preferences. Each experiment took one day, i.e. the oils were tested on a single day in 3 enclosures and were randomly assigned to the three enclosures in a day similar to:

<table>
<thead>
<tr>
<th>Day</th>
<th>Chamber 1</th>
<th>Chamber 2</th>
<th>Chamber 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T1 vs. C</td>
<td>T2 vs. C</td>
<td>T3 vs. C</td>
</tr>
<tr>
<td>2</td>
<td>T2 vs. C</td>
<td>T3 vs. C</td>
<td>T1 vs. C</td>
</tr>
<tr>
<td>3</td>
<td>T2 vs. C</td>
<td>T1 vs. C</td>
<td>T3 vs. C</td>
</tr>
<tr>
<td>etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

However, as noted, we will (arbitrarily) remove some of the data and the revised design looks like:

Visit [https://en.wikipedia.org/wiki/Honeydew_(secretion)] if you are curious.
where a 1 indicates that the treatment was tested on that day and 0 indicates it was not. We see that not every oil is tested on every day and not every day has every oil tested. The design is INCOMPLETE.

A naive analysis might simply delete those blocks with missing data. This is inefficient and costly – there is information in the incomplete blocks and you should not simply discard it.

Incomplete block designs must be analyzed with care because:

- Simple marginal comparisons of treatment no longer have the same set of blocks and so some of the differences in the margins are confounded with treatment effects. The key is to do a block analysis to “adjust” the reading to make all treatments compared against a comparable set of blocks.
- There are two sources of information about comparisons among treatments. The intra-block analysis looks at differences among treatments WITHIN blocks. If the blocking design is complete, this is the entire information about treatment comparisons. However, with incomplete blocks, there is another (smaller) source of information, the inter-block analysis. This will automatically be included if you do the analysis with random blocks.

As before, there are now TWO different questions of interest

- Is there evidence for each oil \((T1, T2, T3)\) that mosquitos choose it with a probability different than 0.5, i.e. is there evidence of an attractiveness or repulsion for each oil
- Can we compare the efficacy of the three treatments

We will arbitrarily delete some data from the \texttt{mosquitos3.csv} in the Sample Program Library at: \url{http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms}. Consult the program code for details.

The data are imported into a \textit{JMP} datatable and the derived variables for the total number of mosquitos that chose, and the empirical proportion that chose the treatment \((p\_treatment)\) and the empirical logit of the proportion that chose the treatment \((\text{logit}\_treatment)\) are computed in the usual way:

As before, notice that the experimental unit is the trap, but the observational unit (what is actually measured) is the individual mosquito. The response variable for each individual mosquito is the either \textit{Treatment} or \textit{Control} trap. The fact that the data have been summarized to the replicate level doesn’t
change the observational unit. It is very tempting to think that the counts in each trap are the response variable, but that is incorrect – these are simply a summary of the data.

Consequently, this is yet another example of pseudo-replication (experiment units not equal to the observational unit) and any analysis must account for this mismatch. Now some care must be taken in the analysis:

- Compute a response variable at the trap (experimental unit) level. This is often the “average” response when the response variable is continuous. In this case, the “average” will be the empirical proportion that choose the Treatment trap in each replicate. **Now a block analysis that analyzes all the oils simultaneously MUST BE USED.** You should not analyze each oil separately because the set of blocks over which the oil was tested is NOT the same.

- Similarly, compute the logit of the proportion that chose the treatment, and **perform a block analysis on the empirical logits** to see if there is evidence that the mean logit is different from $0 = \logit(0.5)$. You should not analyze each oil separately because the set of blocks over which the oil was tested is NOT the same.

- Fit a generalized-linear model but adjust the results for overdispersion. This is known as a quasi-binomial or quasi-logit analysis. The goodness-of-fit statistics can be used to estimate the overdispersion factor which is then used to adjust the standard errors, test statistics, and $p$-values. **Unfortunately, this will NOT automatically incorporate the inter-block information because blocks must be declared as fixed.**

- Perform a generalized linear mixed model analysis where the ordinary logistic regression is augmented with random effects. The key problem here is the need to integrate over the random effects when computing the likelihood and this can be computationally challenging. **This automatically includes both the inter- and intra-block information.**

- Perform a Bayesian analysis. This analysis uses MCMC sampling to do the numerical integration over the likelihood. This analysis is beyond the scope of these notes.

As before, simple ch-square tests on the poole data are NOT appropriate because of the pseudo-replication issues.

Is there evidence of overdispersion? Here is a plot of the 95% confidence intervals for the proportion that chose the treatment trap compared to the overall proportion that chose the treatment trap.
Notice that several of the confidence intervals from the individual cages do not cover the overall proportion that chose each treatment for each oil. This indicates that there is some (random) day and enclosure effect that is influencing all of the mosquitos simultaneously (e.g. replicate humidity; replicate temperature, etc.).

22.7.1 Using the simple proportions as data

Because the design is an incomplete block design, a combined block analysis MUST be used to avoid having the oils using different sets of blocks when computing the marginal means. We fit an block model with RANDOM blocks to fully capture the variability in the data. If a block model with fixed blocks was fit, it would assume that the exact same set of days and enclosures would be used in the future. While we may have control over the enclosures, we have no control over environmental variables such as humidity that may change from day to day. The Analyze -> Fit Model platform is used to fit the RCB model with random blocks as shown in the previous section. This gives the marginal means and confidence limits, but does not let you test if the individual marginal proportions are equal to 0.5 except through a custom test (not shown, but see me for details). You could simply see if the confidence interval for the marginal proportion for each treatment contains the value 0.5.

<table>
<thead>
<tr>
<th>Level</th>
<th>Least Squares Mean</th>
<th>Std Error</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>0.60466676</td>
<td>0.03360717</td>
<td>0.53553867</td>
<td>0.67379496</td>
</tr>
<tr>
<td>T2</td>
<td>0.69559709</td>
<td>0.03360717</td>
<td>0.62748900</td>
<td>0.76572529</td>
</tr>
<tr>
<td>T3</td>
<td>0.56850615</td>
<td>0.04022913</td>
<td>0.46457135</td>
<td>0.64844094</td>
</tr>
</tbody>
</table>

Notice that not all of the marginal “means” all have the same standard error. This is because the
design was incomplete and each oil was tested in differing number of blocks. Once the block analysis is fit, it is then straightforward to compare the efficacy of the oils in the usual way, i.e. using a multiple comparison procedure: The test for the equality of treatment effect and the estimated difference between the margin values and the compact letter display are shown in the usual way:

The compact letter display (cld) is interpreted in the usual fashion.

A plot of the results can be made in the usual fashion (not shown).

### 22.7.2 Using the empirical logits as data

In an analogous fashion, we proceed by finding the empirical logit of the proportion of mosquitoes that choose the treatment trap in each replicate and then perform a block analysis analysis on the logits.

If the mosquitoes were choosing the traps simply at random, we would expect that $p = 0.5$ or the $logit = 0.0$ for each marginal logit. Hence our hypothesis is now

$$H_0 : \mu_{logit} = 0.0$$

$$H_A : \mu_{logit} \neq 0.0$$

for each marginal logit.
We again fit the block model to the logits to do the comparison of the marginal logits for all three oils simultaneously and to compare the efficacy of the three oils. The Analyze->Fit Model platform is used to fit the RCB model with random blocks in a similar fashion as before. This gives the marginal means and confidence limits, but does not let you test if the individual marginal logits are equal to 0.0 except through a custom test (not shown, but see me for details). You could simply see if the confidence interval for the marginal logit for each treatment contains the value 0.0.

Notice that not all of the marginal “logits” all have the same standard error, again because they use different sets of blocks. Once the block analysis is fit, it is then straightforward to compare the efficacy of the oils in the usual way, i.e. using a multiple comparison procedure: The test for the equality of treatment effect and the estimated difference between the margin values and the compact letter display are shown in the usual way:

The compact letter display (cld) is interpreted in the usual fashion. Notice that when we back transform the difference in logits, it is equivalent to the odds ratio (OR) for choosing the treated trap compared to the control trap and an \( OR = 1 \) indicates no evidence of a difference in efficacy between the oils.
A plot of the results can be make in the usual fashion (not shown).

### 22.7.3 Logistic regression using overdispersion

The enclosures serve as “cluster” in this experiment, so the ad hoc methods of correcting for overdispersion caused by “cluster” effects can also be done, i.e. estimating the overdispersion factor ($\hat{c}$) and multiplying the standard errors by $\sqrt{\hat{c}}$. This is known as a quasi-binomial or quasi-logit analysis.

Because data has been summarized to the replicate level, we can use, and it is highly recommended that the second way to specify the model be used where the $Y$ box as the number of “success” first, and then the total number of trials second. Then a generalized linear model is fit is used in the Analyze->Fit Model platform as shown in the previous section. Note that both the number of mosquitoes that chose the treatment trap and the number mosquitoes that chose either trap are specified in the $Y$ box. There are NO terms in the Effects box to give the intercept only model. Also don’t forget to check the box to adjust for over dispersion. **CAUTION:** If you stack the data and use the *Freq* button for the individual counts of the responses, your estimate of over dispersion will be too small as *JMP* then does not aggregate up to the replicate level to compute the over dispersion factor.

This gives:

![Generalized Linear Model Fit](image)

This analysis is not completely satisfactory because *Day* is treated as a fixed effect and so the estimated marginal logits will tend to have reported standard errors that are too small. See me for additional details.
Note that different packages may give slight different values for the \( p \)-values for the marginal logits depending on the way the overdispersion factor is computed (e.g. using the deviance or the Pearson chi-square) and the way the \( p \)-value is computed (e.g. using quasilikelihood or a Wald statistic). All are asymptotically equivalent but can differ in small sample sizes – contact me for more details.

Once the block analysis is fit, it is then straight forward to compare the efficacy of the oils in the usual way, i.e. using a multiple comparison procedure: The output includes the effect tests:

<table>
<thead>
<tr>
<th>Source</th>
<th>L-R</th>
<th>DF</th>
<th>ChiSquare</th>
<th>Prob&gt;ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day</td>
<td></td>
<td>14</td>
<td>46.743381</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>trt</td>
<td></td>
<td>2</td>
<td>11.43705</td>
<td>0.0033*</td>
</tr>
</tbody>
</table>

but does not automatically give the differences in marginal logits nor the multiple comparison (groan). We need to do a individual contrast for each pair of treatments and remember to adjust for multiple testing (e.g. using a Bonferonni correction). For example, here is the contrast between the first two treatments:

It is a pity that JMP doesn’t make this more convenient. The compact letter display (\textit{cld}) is interpreted in the usual fashion. Notice that when we back transform the difference in logits, it is equivalent to the odds ratio (OR) for choosing the treated trap compared to the control trap and an \( OR = 1 \) indicates no evidence of a difference in efficacy between the oils.

A plot of the results can be make in the usual fashion as well (but not shown).

The estimated overdispersion factor is indicating that standard errors from a simple logistic fit must be adjusted by the square root of the factor. This was automatically done above.

This analysis is not completely satisfactory because \textit{Day} is treated as a fixed effect and so the estimated marginal logits will tend to have reported standard errors that are too small. See me for additional details.
22.7.4 GLIMM modeling

Finally, a generalized linear mixed model logistic ANOVA can be done that explicitly models the random effects of days and enclosures (the combination of day and cage) directly. This analysis will automatically include the inter- and intra-block information. **Unfortunately this cannot be done in JMP (Version 11). Ignore the rest of this section.**

This corresponds to the model in the shorthand notation:

\[
\text{logit}(p) = \text{trt} + \text{Day}(R) + \text{Experiment}(R)
\]

or the generalized linear model:

\[
treatment_i \sim \text{Binomial}(n_i, p_i) \\
\theta_i = \text{logit}(p_i) \\
\theta_i = \text{trt} + \text{Day}(R) + \text{Experiment}(R)
\]

where the *Experiment* term represent the combination of day and cage. You should not use *Cage* as the random effect because cage 1 on day 1 is not necessarily the same “cage” on day 2, despite it having the same physical location etc. Perhaps something happened to the cage on day 2 that made it subtly different from the same cage on the previous day.

The usual comparison among treatments is made

The *p*-values may differ slightly between packages depending a likelihood ratio or Wald test is used (contact me for details).

The difference of logits corresponds to the \( \log(\text{Odds Ratio}) \) of the proportion choosing the treatment trap vs the control trap between two different oils. A back transformation can give estimates of the odds ratio (and confidence limits) directly. Notice that you do NOT take the anti-log of the *se* of the log(odds ratio) to get the *se* of the odds-ratio, but must do a delta-method transformation. Contact me for details.

The estimated variance components (on the logit scale) are found as:

### 22.7.5 Recommendation

As long as the number of mosquitos that select one of the traps is roughly the same among all of the enclosures, the analysis of the empirical logits using a incomplete block analysis is simple and relatively straightforward to do with modern statistical software. The key to this model is the ability to declare blocks as a random factor to ensure that the individual comparisons of the marginal logits to 0 captures the true extent of the uncertainty and to capture the intra- and inter-block information.

If the number of mosquitos that makes a choice is highly variable and/or the empirical probabilities of selection are close to 0 or 1, then the more advanced methods will be preferred. Note that if some of the cages has 100% or 0% of mosquitos choosing one of the traps, then many of the methods will fail because the empirical logit are \( \pm \infty \) respectively. In such cases, Bayesian methods can often work quite well.