1. The following data is from a study on infant respiratory disease, namely the proportions
of children developing bronchitis or pneumonia in their first year of life by type of
feeding and sex.

<table>
<thead>
<tr>
<th></th>
<th>disease</th>
<th>nondisease</th>
<th>sex</th>
<th>food</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>1</td>
<td>77</td>
<td>381</td>
<td>Boy</td>
</tr>
<tr>
<td>#2</td>
<td>2</td>
<td>19</td>
<td>128</td>
<td>Boy</td>
</tr>
<tr>
<td>#3</td>
<td>3</td>
<td>47</td>
<td>447</td>
<td>Boy</td>
</tr>
<tr>
<td>#4</td>
<td>4</td>
<td>48</td>
<td>336</td>
<td>Girl</td>
</tr>
<tr>
<td>#5</td>
<td>5</td>
<td>16</td>
<td>111</td>
<td>Girl</td>
</tr>
<tr>
<td>#6</td>
<td>6</td>
<td>31</td>
<td>433</td>
<td>Girl</td>
</tr>
</tbody>
</table>

Below is the summary of the GLM fit with the logit link.

```r
# Call:
glm(formula = cbind(disease, nondisease) ~ sex + food, family = binomial,  
    # data = babyfood)

# Coefficients:
# Estimate Std. Error z value Pr(>|z|)
# (Intercept) -1.6127 0.1124 -14.347 < 2e-16 ***
# sexGirl    -0.3126 0.1410 -2.216  0.0267 *
# foodBreast -0.6693 0.1530 -4.374 1.22e-05 ***
# foodSuppl  -0.1725 0.2056 -0.839  0.4013
# ---
# Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# (Dispersion parameter for binomial family taken to be 1)

# Null deviance: 26.37529 on 5 degrees of freedom
# Residual deviance: 0.72192 on 2 degrees of freedom
# AIC: 40.24

# Number of Fisher Scoring iterations: 4
```

Please answer the following questions based on the results.

(a) Interpret the coefficient estimate of \(-0.1725\) for variable “foodSuppl”.
(b) Predict the chance of disease for a breastfed girl.
(c) Find the Pearson’s residual for a breastfed boy.

2. Explain the over-dispersion in the following Poisson-Gamma model:

\[
\begin{align*}
Y|Z & \sim \text{Poisson}(Z), \\
Z & \sim \text{Gamma}(\mu/\phi, 1/\phi),
\end{align*}
\]
3. Suppose for the $i$-th observation, $i = 1, \ldots, I$, $y_i$ is the sample proportions of successes in $n_i$ i.i.d. Bernoulli trials, and the explanatory variables are denoted by $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$. Consider the binomial regression with the probit link. Derive the explicit form of the updating step in the Fisher scoring algorithm.

4. Consider the two-way ANOVA model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

(1)

where $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$, and the identifiability constraints are $\sum_{i=1}^{I} \alpha_i = 0$, $\sum_{j=1}^{J} \beta_j = 0$, $\sum_{i=1}^{I} \gamma_{ij} = 0$, $\forall j$ and $\sum_{j=1}^{J} \gamma_{ij} = 0$, $\forall i$. Consider the test $H_0 : \alpha_i = 0, \forall i$. Derive the $F$-statistic for this test.