Chap 2, problem 13

```
> options(prompt = " ", continue = " ")
 c.table <- array(data = c(251, 48, 34, 5), dim = c(2,2),
                 dimnames = list(c("made", "missed"),c("made", "missed")))
 data <- data.frame(First = c('made','missed'),
                    c.table, trials = rowSums(c.table))
 fit.glm <- glm(made/trials ~ First, data = data,
                family= binomial, weights = trials)
 c <- round(fit.glm$coefficients,4)
 summary(fit.glm)

Call:
glm(formula = made/trials ~ First, family = binomial, data = data,
    weights = trials)

Deviance Residuals:
[1]  0  0

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.9991     0.1827  10.939   <2e-16 ***
Firstmissed   0.2627     0.5042   0.521    0.602
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance:  2.8575e-01  on 1  degrees of freedom
Residual deviance: -1.6431e-14  on 0  degrees of freedom
AIC: 12.624

Number of Fisher Scoring iterations: 3

 anova(fit.glm, test = 'Chisq')

Analysis of Deviance Table

Model: binomial, link: logit

Response: made/trials

Terms added sequentially (first to last)


      Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                      1    0.28575
First  1  0.28575         0    0.00000    0.593
```

(a) aggregate(yi, xi, FUN = sum) is to collect all the yi values with the same xi values, then take sum of the yi values.

(b)

$$\log \frac{\texttt{prob(success.second | missed.first)}}{1-\texttt{prob(success.second | made.first}}) = \texttt{1.9991} + \texttt{0.2627} \times \texttt{First.missed}$$

(c)

$$\texttt{Odds ratio} = 1.3004$$

Wald: $0.48 < \text{OR} < 3.49$
profile LR: $0.52 < \text{OR} < 3.94$

(d) The Wald test p-value is 0.602 and the LRT p-value is 0.593. The LRT p-value is the same as found in Chapter 1. We did not perform the Wald test in Chapter 1.

(e) The likelihood ratio test is the same as compared to Chap 1. The maximized likelihood ratio doesn't change if we reparameterized.

$$\frac{\max_\theta f(\theta \mid H_0)}{\max_\theta f(\theta \mid H_a)} = \frac{\max_\theta f(g(\theta) \mid H_0)}{\max_\theta f(g(\theta) \mid H_a)}$$

Chap 3, problem 3

(a) The marginals of a multinomial is a binomial. Suppose (x_1, x_2, x_3) is multinomial (p_1, p_2, p_3).

$$\begin{aligned}
\Pr[x_1 = k] &= \sum_{n_2,n_3} \binom{n}{k, n_2, n_3} p_1^k p_2^{n_2} p_3^{n_3} \\
&= \sum_{n_2,n_3} \frac{n!}{k! n_2! n_3!} p_1^k p_2^{n_2} p_3^{n_3} \\
&= \frac{n!}{k!(n_2+n_3)!} p_1^k (p_2+p_3)^{n-k} \\
&= \frac{n!}{k!(n-k)!} p_1^k (1-p_1)^{n-k}
\end{aligned}$$

(b) By definition

$$f(n_1, n_2, n_3 \mid n_1+) = f(n_1, n_2, n_3)/f(n_1+)$$

By definition

$$f(n_1, n_2, n_3) = \binom{n}{n_1, n_2, n_3, n_4} p_1^{n_1} p_2^{n2} p_3^{n_3} p_4^{n_4}$$

From part a) The marginal n_1+ is binomial(n, p_1+p_2) , so

$$f(n_1+) = \binom{n}{n_1+} (p_1+p_2)^{n_1+} (1-p_1-p_2)^{n-n_1+} = \binom{n}{n_1+} (p_1+p_2)^{n_1+} (p_3+p_4)^{n-n_1+}$$

Hence

$$\begin{aligned}
f(n_1, n_2, n_3 \mid n_1+) &= f(n_1, n_2, n_3)/f(n_1+) \\
&= \Big\{ \binom{n}{n_1, n_2, n_3, n_4} p_1^{n_1} p_2^{n2} p_3^{n_3} p_4^{n_4} \Big\} / \Big\{ \binom{n}{n_1+} (p_1+p_2)^{n_1+} (p_3+p_4)^{n-n_1+} \Big\} \\
&= \Big\{ \binom{n}{n_1, n_2, n_3, n_4} [p_1/(p_1+p_2)]^{n_1} [p_2/(p_1+p2)]^{n2} [p_3/(p_3+p_4]^{n_3} [p_4/(p_3+p_4)]^{n_4} \Big\} / \Big\{ \binom{n}{n_1+} \Big\} \\
&= \frac{n_1+! n_2+!}{n_1! n_2! n_3! n_4!} [\frac{p_1}{p_1+p_2}]^{n_1} [\frac{p_2}{p_1+p2}]^{n2} [\frac{p_3}{p_3+p_4}]^{n_3} [\frac{p_4}{p_3+p_4}]^{n_4} \\
&= \frac{n_1+!}{n_1! n_2!} [\frac{p_1}{p_1+p_2}]^{n_1} [\frac{p_2}{p_1+p2}]^{n2} \times \frac{n_2+!}{n_3! n_4!} [\frac{p_3}{p_3+p_4}]^{n_3} [\frac{p_4}{p_3+p_4}]^{n_4}
\end{aligned}$$

This probability is the product of conditional probabilities $n_1, n_2 \,|\, n_1 + n_2 = n_1+$, and $n_3, n_4 \,|\, n_3 + n_4 = n_2+$, which are binomials.

(c) From part b)

$$f(n_1, n_2, n_3, n_4) = f(n_1, n_2 | n_1 + n_2 = n_1+) f(n_3, n_4 | n_3 + n_4 = n_2+) \times f(n_1+) f(n_2+)$$

Chap 3, problem 8

```
hep <- read.csv('http://www.chrisbilder.com/categorical/Chapter2/healthcare_worker.csv')
data <- data.frame(hep, no_hep = hep[, "Size"] - hep[, "Hepatitis"])
n_ij <- data[,c("Hepatitis", "no_hep") ]
n_i_plus <- hep[,"Size"]
n_plus_j <- colSums(data[,c("Hepatitis", "no_hep") ])
n_ij_indep <- outer(n_i_plus, n_plus_j)
n <- sum(n_ij)
Peason_X_sq <- sum( (n_ij - n_ij_indep / n)^2/ (n_ij_indep/n) )
neg_two_Lambda <- 2 * sum( n_ij * log( n_ij / (n_ij_indep/n)))
p <- 1 - pchisq(c(Peason_X_sq, neg_two_Lambda),
                df = (nrow(n_ij)-1) * (ncol(n_ij)-1))
T <- c(Peason_X_sq, neg_two_Lambda)
t <- round(T,4)
p <- round(p, 4)
```

$$X^2 = \sum_{ij} n_{ij} \frac{(n_{ij} - n_{i+}n_{+j}/n)^2}{n_{i+}n_{+j}/n} = 4.5043, \;\; p = 0.342$$

$$-2\log(\Lambda) = 2\sum_{ij} n_{ij} \log(\frac{n_{ij}}{n_{i+}n_{+j}/n}) = 3.735, \quad p = 0.4431$$