**Surviving third-degree burns**

These data refer to 435 adults who were treated for third-degree burns by the University of Southern California General Hospital Burn Center. The patients were grouped according to the area of third-degree burns on the body (measured in square cm). In the table below are recorded, for each midpoint of the groupings 'log(area +1)', the number of patients in the corresponding group who survived, and the number who died from the burns. http://statmaster.sdu.dk/courses/st111

```
> burndata <- data.frame(rbind(c(1.35,13,0,1),
+ c(1.6,19,0,1),c(1.75,67,2,0.971014493),
+ c(1.85,45,5,0.9),c(1.95,71,8,0.898734177),
+ c(2.05,50,20,0.714285714),c(2.15,35,31,0.53030303),
+ c(2.25,7,49,0.125),c(2.35,1,12,0.076923077)))
> names(burndata) <- c("logarea","surv","died","propsurv")
> attach(burndata)
> logitsurv <- log(propsurv / (1-propsurv))
> burnexpl <- rep(logarea, surv+died)
> burnresp <- c()
> for(i in 1:9){
+    burnresp<-c(burnresp,rep(0,died[i]),rep(1,surv[i]))}
```
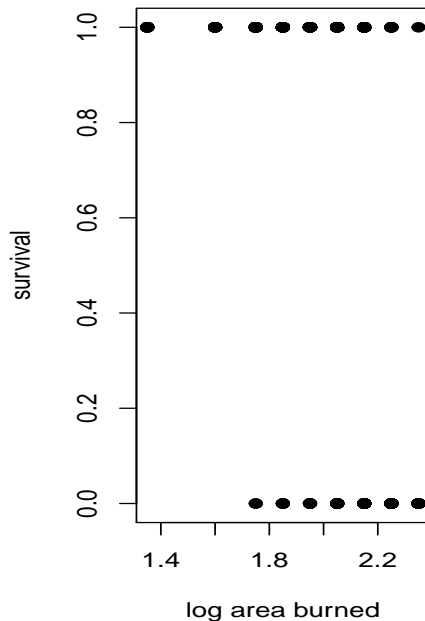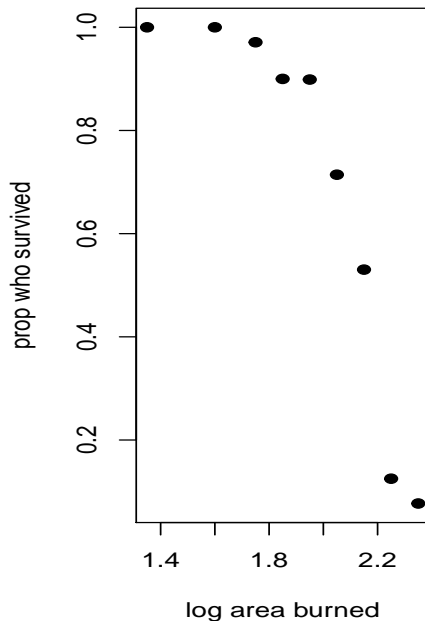
Wald & Likelihood Ratio tests

```
> summary(glm(burnresp~burnexpl, family="binomial"))

Call:
glm(formula = burnresp ~ burnexpl, family = "binomial")

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8518  -0.6998   0.1859   0.5239   2.2089

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   22.708      2.266  10.021   <2e-16 ***
burnexpl     -10.662      1.083  -9.849   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 525.39  on 434  degrees of freedom
Residual deviance: 335.23  on 433  degrees of freedom
AIC: 339.23

Number of Fisher Scoring iterations: 6
```

Wald: The Wald test is simply a way of running a z-test for the slope parameter, β_1, z-statistic: $z = \frac{b_1 - 0}{SE(b_1)}$.

or find a CI for β_1: $b_1 \pm z^* SE(b_1)$.

Note that the Wald test is for β_1 which represents the $\ln(OR)$ for a *one unit change in x*. A $\beta_1 = 0$ would indicate that your odds don't change (i.e., OR=1) regardless of how your X value changes. If you are interested in a CI for the odds ratio, you need to transform your CI for β_1 into a CI for e^{β_1}.

χ^2: The Likehood ratio test also tests whether the response is explained by the explanatory variable. We can output the deviance (2 * log-likelihood) for both the full (maximum likelihood!) and reduced (null) models.

$$
\begin{aligned}
G &= 2 \cdot \ln(L(\hat{p})) - 2 \cdot \ln(L(p_0)) \\
&= \text{null (restricted) deviance - residual (full model) deviance} \\
G &\sim \chi_1^2 \quad \text{when the null hypothesis is true}
\end{aligned}
$$

Measures of Association

With logistic regression, we don't have residuals, so we don't have a value like R^2. We can, however, measure whether or not the estimated model is consistent with the data. That is, is the model able to discriminate between successes and failures.

Consider looking at all the pairs of successes and failures. In the burn data we have 308 survivors and 127 deaths = 39,116 pairs of people. Given a particular pair, if the observation corresponding to a survivor has a *higher* probability of success than the observation corresponding to a death, we call the pair **concordant**. If the observation corresponding to a survivor has a *lower* probability of success than the observation corresponding to a death, we call the pair **discordant**. Tied pairs occur when the observed survivor has the same estimated probability as the observed death.

D_{xy}: Somers' D is the number of concordant pairs minus the number of discordant pairs divided by the total number of pairs.

gamma: Goodman-Kruskal gamma is the number of concordant pairs minus the number of discordant pairs divided by the total number of pairs excluding ties.

tau-a: Kendall's tau-a is the number of concordant pairs minus the number of discordant pairs divided by the total number of pairs of people (including pairs who both survived or both died).

For example: consider a pair of individuals with burn areas of 1.75 and 2.35.

$$p(x = 1.75) = \frac{e^{22.7083 - 10.6624 \cdot 1.75}}{1 + e^{22.7083 - 10.6624 \cdot 1.75}} = 0.983$$

$$p(x = 2.35) = \frac{e^{22.7083 - 10.6624 \cdot 2.35}}{1 + e^{22.7083 - 10.6624 \cdot 2.35}} = 0.087$$

The pairs would be concordant if the first individual survived and the second didn't. The pairs would be discordant if the first individual died and the second survived.

```
> install.packages("rms", repos="http://cran.r-project.org")

package rms successfully unpacked and MD5 sums checked

The downloaded packages are in
        C:\Users\jsh04747\AppData\Local\Temp\RtmpyEkMKa\downloaded_packages

> #you should be able to type simply: install.packages("rms") on your computer
> library(rms)
> burn.glm <- lrm(burnresp~burnexpl)
> print(burn.glm)

Logistic Regression Model

lrm(formula = burnresp ~ burnexpl)

                     Model Likelihood      Discrimination    Rank Discrim.
                        Ratio Test            Indexes           Indexes
Obs            435    LR chi2     190.15   R2      0.505     C      0.877
 0             127    d.f.             1   g       2.576     Dxy    0.753
 1             308    Pr(> chi2) <0.0001   gr     13.146     gamma  0.824
max |deriv| 8e-11                          gp      0.313     tau-a  0.312
                                           Brier   0.121


          Coef     S.E.    Wald Z Pr(>|Z|)
Intercept  22.7083 2.2661 10.02  <0.0001
burnexpl  -10.6624 1.0826 -9.85  <0.0001
```

The summary contains the following elements:

number of observations used in the fit, maximum absolute value of first derivative of log likelihood, model likelihood ratio chi2, d.f., P-value, c index (area under ROC curve), Somers' Dxy, Goodman-Kruskal gamma, Kendall's tau-a rank correlations between predicted probabilities and observed response, the Nagelkerke R^2 index, the Brier score computed with respect to Y > its lowest level, the g-index, gr (the g-index on the odds ratio scale), and gp (the g-index on the probability scale using the same cutoff used for the Brier score).