Exam 3 Review

Suppose that \( X_i = \mathbf{x} = (x_1, ..., x_k)^T \) is observed and that \( Y_i | X_i = \mathbf{x}_i \sim \text{Binomial}(n_i, \pi(x_i)) \) for \( i = 1, ..., N \) where

\[
\hat{\pi}(\mathbf{x}) = \frac{\exp(\hat{\alpha} + \hat{\beta}^T \mathbf{x})}{1 + \exp(\hat{\alpha} + \hat{\beta}^T \mathbf{x})}.
\]

This is called the **full model** for **logistic regression** and the \( (k + 1) \) parameters \( \alpha, \beta_1, ..., \beta_k \) are estimated.

For the **saturated model**, the \( Y_i | X_i = \mathbf{x}_i \sim \text{Binomial}(n_i, \pi_i) \) for \( i = 1, ..., N \) where

\[
\hat{\pi}_i = \frac{Y_i}{n_i}.
\]

This model estimates the \( N \) parameters \( \pi_i \).

Let \( l_{SAT}(\pi_1, ..., \pi_n) \) be the likelihood function for the saturated model and let \( l_{FULL}(\alpha, \beta) \) be the likelihood function for the full model. Let \( L_{SAT} = \log l_{SAT}(\hat{\pi}_1, ..., \hat{\pi}_N) \) be the log likelihood function for the saturated model evaluated at the MLE \((\hat{\pi}_1, ..., \hat{\pi}_N)\) and let \( L_{FULL} = \log l_{FULL}(\hat{\alpha}, \hat{\beta}) \) be the log likelihood function for the full model evaluated at the MLE \((\hat{\alpha}, \hat{\beta})\).

Then the **deviance** \( D = G^2 = -2(L_{FULL} - L_{SAT}) \).

The degrees of freedom for the deviance = \( df_{FULL} = N - k - 1 \) where \( N \) is the number of parameters for the saturated model and \( k + 1 \) is the number of parameters for the full model.

The saturated model is usually not very good for binary data (all \( n_i = 1 \)) or if the \( n_i \) are small. The saturated model can be good if all of the \( n_i \) are large or if \( \pi_i \) is very close to 0 or 1 whenever \( n_i \) is small.

If \( X \sim \chi^2_d \) then \( E(X) = d \) and \( V(X) = 2d \). An observed value of \( x > d + 3\sqrt{d} \) is unusually large and an observed value of \( x < d - 3\sqrt{d} \) is unusually small.

When the saturated model is good, a rule of thumb is that the logistic regression model is ok if \( G^2 \leq N - k - 1 \) (or if \( G^2 \leq N - k - 1 + 3\sqrt{N - k - 1} \)).

An estimated sufficient summary or **ESS plot** is a plot of \( w_i = \hat{\alpha} + \hat{\beta}^T \mathbf{x}_i \) versus \( Y_i \) with the logistic curve of fitted proportions

\[
\hat{\pi}(w_i) = \frac{e^{w_i}}{1 + e^{w_i}}
\]

added to the plot along with a step function of observed proportions.

29) Suppose that \( w_i \) takes many values (eg the LR model has a continuous predictor) and that \( k + 1 << N \). Know that the LR model is good if the step function tracks the logistic curve of fitted proportions in the ESS plot. Also know that you should check that the LR model is good before doing inference with the LR model. See HW6 4.
Response = Y
Terms = (X_1, ..., X_k)

## Sequential Analysis of Deviance

<table>
<thead>
<tr>
<th>Predictor</th>
<th>df</th>
<th>Deviance</th>
<th>df</th>
<th>Change</th>
<th>Deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ones</td>
<td>(N - 1 = df_o)</td>
<td>(G_o^2)</td>
<td>1</td>
<td>(G^2)</td>
<td>1</td>
</tr>
<tr>
<td>(X_1)</td>
<td>(N - 2)</td>
<td></td>
<td></td>
<td>(G_{FULL}^2)</td>
<td></td>
</tr>
<tr>
<td>(X_2)</td>
<td>(N - 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(X_k)</td>
<td>(N - k - 1 = df_{FULL})</td>
<td>(G_{FULL}^2)</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Data set = cbrain, Name of Fit = B1
Response = sex
Terms = (cephalic size log[size])

## Sequential Analysis of Deviance

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Total Deviance</th>
<th>df</th>
<th>Change Deviance</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ones</td>
<td>366.820</td>
<td>266</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cephalic</td>
<td>363.605</td>
<td>265</td>
<td>0.214643</td>
<td>1</td>
</tr>
<tr>
<td>size</td>
<td>315.793</td>
<td>264</td>
<td>47.8121</td>
<td>1</td>
</tr>
<tr>
<td>log[size]</td>
<td>305.045</td>
<td>263</td>
<td>10.7484</td>
<td>1</td>
</tr>
</tbody>
</table>

Know how to use the above output for the following test. Assume that the ESS plot has been made and that the observed proportions track the logistic curve. If the logistic curve looks like a line with small positive slope, then the predictors may not be useful. The following test asks whether \(\hat{\pi}(x_i)\) from the logistic regression should be used to estimate \(P(Y_i = 1 | x_i)\) or if none of the predictors should be used and

\[
P(Y_i = 1) \equiv \pi \approx \frac{\sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} n_i}
\]

for all \(i = 1, ..., N\).

30) The 4 step (log–likelihood) **deviance test** is

i) \(H_o: \beta_1 = \cdots = \beta_k = 0\) \(H_A:\text{ not } H_o\)

ii) test statistic \(G^2(o|F) = G_o^2 - G_{FULL}^2\)

iii) The p–value = \(P(W > G^2(o|F))\) where \(W \sim \chi_k^2\) has a chi–square distribution with \(k\) degrees of freedom. Note that \(k = k + 1 - 1 = df_o - df_{FULL} = N - 1 - (N - k - 1)\).

iv) Reject \(H_o\) if the p–value < \(\delta\) and conclude that there is a LR relationship between \(Y\) and the predictors \(X_1, ..., X_k\). If p–value \(\geq \delta\), then fail to reject \(H_o\) and conclude that there is not a LR relationship between \(Y\) and the predictors \(X_1, ..., X_k\).

See HW6 6a.
After obtaining an acceptable full model where
\[ \logit(\pi(x_i)) = \alpha + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} = \alpha + \beta^T x, \]
try to obtain a reduced model \( Y_i | X_{Ri} = x_{Ri} \sim \text{independent Binomial}(n_i, \pi(x_{Ri})) \) where
\[ \logit(\pi(x_{Ri})) = \alpha + \beta_{R1} x_{Ri1} + \cdots + \beta_{Rm} x_{Rim} = \alpha_R + \beta^T_R x_{Ri} \]
and \( \{x_{Ri1}, \ldots, x_{Rim}\} \subset \{x_1, \ldots, x_k\} \).

Let \( x_{R,m+1}, \ldots, x_{Rk} \) denote the \( k - m \) predictors that are in the full model but not in the reduced model. We want to test \( H_o: \beta_{R,m+1} = \cdots = \beta_{Rk} = 0 \). For notational ease, we will often assume that the predictors have been sorted and partitioned so that \( x_i = x_{Ri} \) for \( i = 1, \ldots, k \). Then the reduced model uses predictors \( x_1, \ldots, x_m \) and we test \( H_o: \beta_{m+1} = \cdots = \beta_k = 0 \). However, in practice this sorting is usually not done.

Assume that the ESS plot looks good. Then we want to test \( H_o: \) the reduced model can be used instead of the full model versus \( H_A: \) the full model is (significantly) better than the reduced model. Fit the full model and the reduced model to get the deviances \( G^2_{FULL} \) and \( G^2_{RED} \).

31) The 4 step change in deviance test is
i) \( H_o: \) the reduced model is good \( H_A: \) use the full model
ii) test statistic \( G^2(R|F) = G^2_{RED} - G^2_{FULL} \)
iii) The p–value = \( P(W > G^2(R|F)) \) where \( W \sim \chi^2_{k-m} \) has a chi–square distribution with \( k \) degrees of freedom. Note that \( k \) is the number of predictors in the full model while \( m \) is the number of predictors in the reduced model. Also notice that \( k - m = (k + 1) - (m + 1) = df_{RED} - df_{FULL} = N - m - 1 - (N - k - 1) \).
iv) Reject \( H_o \) if the p–value < \( \delta \) and conclude that the full model is (significantly) better than the reduced model.

If p–value \( \geq \delta \), then fail to reject \( H_o \) and conclude that the reduced model is good.

See HW6 6b.

32) If the reduced model leaves out a single variable \( X_i \), then the change in deviance test becomes \( H_o: \beta_i = 0 \) versus \( H_A: \beta_i \neq 0 \). This likelihood ratio is a competitor of the Wald test (see 28)). The likelihood ratio test is usually better than the Wald test if the sample size \( N \) is not large, but the Wald test is currently easier for software to produce. For large \( N \) the test statistics from the two test tend to be very similar (asymptotically equivalent tests).

33) If the reduced model is good, then the EE plot of \( \hat{\alpha}_R + \hat{\beta}^T_R x_{Ri} \) versus \( \hat{\alpha} + \hat{\beta}^T x_i \) should be highly correlated with the identity line with unit slope and zero intercept.

Know how to use the following output to test the reduced model versus the full model.
Response = Y  Terms = \((X_1, \ldots, X_k)\)  (Full Model)

<table>
<thead>
<tr>
<th>Label</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Est/SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>(\hat{\alpha})</td>
<td>(se(\hat{\alpha}))</td>
<td>(z_{0,0})</td>
<td>for Ho: (\alpha = 0)</td>
</tr>
<tr>
<td>(x_1)</td>
<td>(\hat{\beta}_1)</td>
<td>(se(\hat{\beta}_1))</td>
<td>(z_{0,1} = \hat{\beta}_1/se(\hat{\beta}_1))</td>
<td>for Ho: (\beta_1 = 0)</td>
</tr>
</tbody>
</table>
| \(
\vdots
\) | \(
\vdots
\) | \(
\vdots
\) | \(
\vdots
\) | \(
\vdots
\) |
| \(x_k\) | \(\hat{\beta}_k\) | \(se(\hat{\beta}_k)\) | \(z_{0,k} = \hat{\beta}_k/se(\hat{\beta}_k)\) | for Ho: \(\beta_k = 0\) |

Degrees of freedom: \(N - k - 1 = df_{FULL}\)

Deviance: \(D = G^2_{FULL}\)

Response = Y  Terms = \((X_1, \ldots, X_m)\)  (Reduced Model)

<table>
<thead>
<tr>
<th>Label</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Est/SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>(\hat{\alpha})</td>
<td>(se(\hat{\alpha}))</td>
<td>(z_{0,0})</td>
<td>for Ho: (\alpha = 0)</td>
</tr>
<tr>
<td>(x_1)</td>
<td>(\hat{\beta}_1)</td>
<td>(se(\hat{\beta}_1))</td>
<td>(z_{0,1} = \hat{\beta}_1/se(\hat{\beta}_1))</td>
<td>for Ho: (\beta_1 = 0)</td>
</tr>
</tbody>
</table>
| \(
\vdots
\) | \(
\vdots
\) | \(
\vdots
\) | \(
\vdots
\) | \(
\vdots
\) |
| \(x_m\) | \(\hat{\beta}_m\) | \(se(\hat{\beta}_m)\) | \(z_{0,m} = \hat{\beta}_m/se(\hat{\beta}_m)\) | for Ho: \(\beta_m = 0\) |

Degrees of freedom: \(N - m - 1 = df_{RED}\)

Deviance: \(D = G^2_{RED}\)

Data set = Banknotes, Name of Fit = B1 (Full Model)
Response  = Status
Terms  = (Diagonal Bottom Top)
Coefficient Estimates

<table>
<thead>
<tr>
<th>Label</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Est/SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2360.49</td>
<td>5064.42</td>
<td>0.466</td>
<td>0.6411</td>
</tr>
<tr>
<td>Diagonal</td>
<td>-19.8874</td>
<td>37.2830</td>
<td>-0.533</td>
<td>0.5937</td>
</tr>
<tr>
<td>Bottom</td>
<td>23.6950</td>
<td>45.5271</td>
<td>0.520</td>
<td>0.6027</td>
</tr>
<tr>
<td>Top</td>
<td>19.6464</td>
<td>60.6512</td>
<td>0.324</td>
<td>0.7460</td>
</tr>
</tbody>
</table>

Degrees of freedom: 196
Deviance: 0.009

Data set = Banknotes, Name of Fit = B2 (Reduced Model)
Response  = Status
Terms  = (Diagonal)
Coefficient Estimates

<table>
<thead>
<tr>
<th>Label</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Est/SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>989.545</td>
<td>219.032</td>
<td>4.518</td>
<td>0.0000</td>
</tr>
<tr>
<td>Diagonal</td>
<td>-7.04376</td>
<td>1.55940</td>
<td>-4.517</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Degrees of freedom: 198
Deviance: 21.109
34) Let \( \pi(x) = P(\text{success}|x) = 1 - P(\text{failure}|x) \) where a “success” is what is counted and a “failure” is what is not counted (so if the \( Y_i \) are binary, \( \pi(x) = P(Y_i = 1|x) \)). Then the estimated odds of success is

\[
\hat{\Omega}(x) = \frac{\hat{\pi}(x)}{1 - \hat{\pi}(x)}.
\]

35) In logistic regression, increasing a predictor \( x_i \) by 1 unit (while holding all other predictors fixed) multiplies the estimated odds of success by a factor of \( \exp(\hat{\beta}_i) \).

36) Suppose that the binary response variable \( Y \) is conditionally independent of \( x \) given a single linear combination \( \beta^T x \) of the predictors, written

\[ Y \perp x | \beta^T x. \]

If the LR model holds and if the first SIR predictor \( \hat{\beta}_{SIR1}^T x \) and \( \hat{\alpha} + \hat{\beta}^T x \) are highly correlated, then (to a good approximation)

\[ Y \perp x | \hat{\alpha} + \hat{\beta}^T x \quad \text{and} \quad Y \perp x | \hat{\beta}_{SIR1}^T x. \]

To make a binary response plot for logistic regression, fit SIR and the LR model and assume that the above conditions hold. Place the first SIR predictor on the horizontal axis and the 2nd SIR predictor \( \hat{\beta}_{SIR2}^T x \) on the vertical axis. If \( Y = 0 \) use symbol “0” and if \( Y = 1 \) use symbol “X.” If the LR model is good then consider the symbol density of X’s and ‘0’s in a narrow vertical slice. This symbol density should be approximately constant (up to binomial variation) from the bottom to the top of the slice. (Hence the X’s and 0’s should be mixed in the slice). The symbol density may change greatly as the slice is moved from the left to the right of the plot, eg from 0% to 100%. If there are slices where the symbol density is not constant from top to bottom, then the LR model may not be good (eg a more complicated model may be needed).

37) Given a predictor \( x \), sometimes \( x \) is not used by itself in the full LR model. Suppose that \( Y \) is binary. Then to decide what functions of \( x \) should be in the model, look at the conditional distribution of \( x|Y = i \) for \( i = 0, 1 \). These rules are used if \( x \) is an indicator variable or if \( x \) is a continuous variable.

| distribution of \( x|y = i \) | functions of \( x \) to include in the full LR model |
|-------------------------------|-----------------------------------------------|
| \( x|y = i \) is an indicator | \( x \)                                      |
| \( x|y = i \sim N(\mu_i, \sigma_i^2) \) | \( x \)                                      |
| \( x|y = i \sim N(\mu_i, \sigma_i^2) \) | \( x \) and \( x^2 \)                      |
| \( x|y = i \) has a skewed distribution | \( x \) and \( \log(x) \)                  |
| \( x|y = i \) has support on \((0,1)\) | \( \log(x) \) and \( \log(1 - x) \)        |

38) If \( w \) is a nominal variable with \( J \) levels, use \( J - 1 \) (indicator or) dummy variables \( x_{1w}, \ldots, x_{J-1w} \) in the full model.

39) An interaction is a product of two or more predictor variables. Interactions are difficult to interpret. Often interactions are included in the full model and the reduced model without any interactions is tested. The investigator is hoping that the interactions are not needed.
40) A **scatterplot** of $x$ vs. $Y$ is used to visualize the conditional distribution of $Y|x$. A **scatterplot matrix** is an array of scatterplots. It is used to examine the marginal relationships of the predictors and response. Place $Y$ on the top or bottom of the scatterplot matrix and also mark the plotted points by a 0 if $Y = 0$ and by $X$ if $Y = 1$. Variables with outliers, missing values or strong nonlinearities may be so bad that they should not be included in the full model.

41) Suppose that all values of the variable $x$ are positive. The **log rule** says add $\log(x)$ to the full model if $\max(x_i)/\min(x_i) > 10$.

42) To make a full model, use points 37), 38), 40) and 41) and sometimes 39). The number of predictors in the full model should be much smaller than the number of data cases $N$. Make an ESS plot to check that the full model is good.

43) **Variable selection** is closely related to the change in deviance test for a reduced model. You are seeking a subset $I$ of the variables to keep in the model. The $AIC(I)$ statistic is used as an aid in backward elimination and forward selection. The full model and the model with the smallest AIC are always of interest. Create a full model. The full model has a deviance at least as small as that of any submodel.

44) **Backward elimination** starts with the full model with $k$ variables and the predictor that optimizes some criterion is deleted. Then there are $k - 1$ variables left and the predictor that optimizes some criterion is deleted. This process continues for models with $k - 2, k - 3, ..., 3$ and 2 predictors. **Forward selection** starts with the model with 0 variables and the predictor that optimizes some criterion is added. Then there is 1 variable in the model and the predictor that optimizes some criterion is added. This process continues for models with 2, 3, ..., $k - 2$ and $k - 1$ predictors. Both forward selection and backward elimination result in a sequence of $k$ models \{$x^*_1$, \{$x^*_1, x^*_2$\}, ..., \{$x^*_1, x^*_2, ..., x^*_{k-1}$\}, \{$x^*_1, x^*_2, ..., x^*_k$\} = full model.

45) For logistic regression, suppose that the $Y_i$ are binary for $i = 1, ..., N$. Let $N_1 = \sum Y_i$ = the number of 1’s and $N_0 = N - N_1$ = the number of 0’s. Rule of thumb: the final submodel should have $m$ predictor variables where $m$ is small with $m \leq \min(N_1, N_0)/10$.

46) Know how to find good models from output. A good submodel $I$ will use a small number of predictors, have a good ESS plot, and have a good EE plot. A good LR submodel $I$ should have a deviance $G^2(I)$ close to that of the full model in that the change in deviance test 31) would not be rejected. Also the submodel should have a value of $AIC(I)$ close to that of the examined model that has the minimum $AIC$ value. The LR output for model $I$ should not have many variables with small Wald test p-values.

47) Heuristically, backward elimination tries to delete the variable that will increase the deviance the least. An increase in deviance greater than 4 (if the predictor has 1 degree of freedom) may be troubling in that a good predictor may have been deleted. In practice, the backward elimination program may delete the variable such that the submodel $I$ with $j$ predictors has 1) the smallest $AIC(I)$, 2) the smallest deviance $G^2(I)$ or 3) the biggest p-value (preferably from a change in deviance test but possibly from a Wald test) in the test $H_0 \beta_i = 0$ versus $H_A \beta_i \neq 0$ where the current model with $j + 1$
48) Heuristically, forward selection tries to add the variable that will decrease the deviance the most. An increase in deviance less than 4 (if the predictor has 1 degree of freedom) may be troubling in that a bad predictor may have been added. In practice, the forward selection program may add the variable such that the submodel I with j predictors has 1) the smallest AIC(I), 2) the smallest deviance $G^2(I)$ or 3) the smallest p–value (preferably from a change in deviance test but possibly from a Wald test) in the test $H_0 \beta_i = 0$ versus $H_A \beta_i \neq 0$ where the current model with j terms plus the predictor $x_i$ is treated as the full model (for all variables $x_i$ not yet in the model).

49) For logistic regression, let $N_1 = \text{number of ones}$ and $N_0 = N - N_1 = \text{number of zeros}$. A rough rule of thumb is that the full model should use no more than $\min(N_0, N_1)/5$ predictors and the final submodel should use no more than $\min(N_0, N_1)/10$ predictors.

50) For loglinear regression, a rough rule of thumb is that the full model should use no more than $N/5$ predictors and the final submodel should use no more than $N/10$ predictors.

51) Variable selection is pretty much the same for logistic regression and loglinear regression. Suppose that the full model is good and is stored in M1. Let M2, M3, M4, and M5 be candidate submodels found after forward selection, backward elimination, etc. Make a scatterplot matrix of M2, M3, M4, M5 and M1. Good candidates should have estimated linear predictors that are highly correlated with the full model estimated linear predictor (the correlation should be at least 0.9 and preferably greater than 0.95). For binary logistic regression, mark the symbols using the response variable $Y$. See HW7 1, HW8 1, HW9 1 and HW 10 1.

52) The final submodel I should have few predictors, few variables with large Wald p–values (0.01 to 0.05 is borderline), a good ESS plot and an EE plot that clusters tightly about the identity line. Do not use more predictors than the min AIC model $I_{min}$ and want $AIC(I) \leq AIC(I_{min}) + 7$. For the change in deviance test, want pvalue $\geq 0.01$ for variable selection (instead of $\delta = 0.05$). If a factor has J-1 dummy variables, either keep all I-1 dummy variables or delete all J-1 dummy variables, do not delete some of the dummy variables.

53) Know that when there is perfect classification in the binary logistic regression model, the LR MLE estimator does not exist and the output is suspect. However, often the full model deviance is close to 0 and the deviance test correctly rejects Ho.

54) Suppose that $X_i = x = (x_1, \ldots, x_k)^T$ is observed and that $Y_i | X_i = x_i \sim \text{independent Poisson}(\mu(x_i))$ for $i = 1, \ldots, N$ where

$$\hat{\mu}(x) = \exp(\hat{\alpha} + \hat{\beta}^T x).$$

This is called the full model for loglinear regression and the $(k + 1)$ parameters $\alpha, \beta_1, \ldots, \beta_k$ are estimated. Know how to predict $\hat{\mu}(x)$. Also $\hat{Y} = \hat{\mu}(x)$. See HW9 2, Q8.
For the **saturated model**, the $Y_i|X_i = x_i \sim \text{independent Poisson}(\mu_i)$ for $i = 1, ..., N$

where

$$\hat{\mu}_i = Y_i.$$ 

This model estimates the $N$ parameters $\mu_i$. The saturated model is usually bad. An exception is when all $N$ $Y_i$ are large.

The comments on the deviance in the middle of p. 1 still hold.

55) An estimated sufficient summary or **EY plot** is a plot of $w_i = \hat{\alpha} + \hat{\beta}^T x_i$ versus $Y_i$ with the exponential curve of estimated means

$$\hat{\mu}(w_i) = e^{w_i}$$

added to the plot along with a lowess curve.

56) Suppose that $w_i$ takes many values (eg the LLR model has a continuous predictor) and that $k + 1 << N$. Know that the LLR model is good if the lowess tracks the exponential curve of estimated means in the ESS plot. Also know that you should check that the LLR model is good before doing inference with the LLR model. See HW9 2.

57) Know how to perform the 4 step **deviance test**. This test is almost exactly the same as that in 30), but replace LR by LLR in the conclusion. The output looks almost like that shown on p. 2. See HW9 2, Q8.

The deviance test for LLR asks whether $\hat{\mu}(x_i)$ from LLR should be used to estimate $\mu(x_i)$ or should none of the predictors be used so $\hat{\mu} = \bar{Y} = \sum_{i=1}^N Y_i/N$.

58) Know how to perform the 4 step **Wald test**. This test is the same as 28) except replace LR by LLR.

59) Know that a (Wald) 95% CI for $\beta_i$ is $\hat{\beta}_i \pm 1.96SE(\hat{\beta}_i)$.

60) Know how to perform the 4 step **change in deviance test**. The output is almost the same as that on p. 4 and the test is exactly the same as that given in 31). For Ho, the parameters set to 0 are those that are in the full model but not the reduced model.

61) Know what a lurking variable is.

62) Know the difference between an observational study and an experiment. A clinical trial is a randomized controlled experiment performed on humans.

Exam 3 is on Wednesday, April 19 and covers Agresti material including points 23) through 28) on the Exam 2 review. 7 pages of notes. You should know how to use a random number table to draw a simple random sample in order to divide units into 2 groups.

In Agresti, we have covered ch. 1, 2.1, 2.2, 2.3, 2.4, 5.1, 5.2, 5.3, 5.4, and 5.5 but have skipped subsections 2.4.5, 2.4.6, 2.4.7, 5.3.3, 5.3.4, and 5.5.6.