Problem 1. Agresti 5.19.

\( R = 1 : \logit(\pi) = 6.7 + .1A + 1.4S \). \( R = 0 : \logit(\pi) = 7.0 + .1A + 1.2S \)

The YS conditional odds ratio is \( \exp(1.4) = 4.1 \) for blacks and \( \exp(1.2) = 3.3 \) for whites. Note that .2, the coeff. of the cross-product term, is the difference between the log odds ratios 1.4 and 1.2. The coeff. of S of 1.2 is the log odds ratio between Y and S when R = 0 (whites), in which case the RS interaction does not enter the equation. The P-value of \( P < .01 \) for smoking represents the result of the test that the log odds ratio between Y and S for whites is 0.

Problem 2. Agresti 5.20

Part a. The estimated log odds ratio between race and driving after consuming a substantial amount of alcohol was −.72 in Grade 12 (i.e., for each gender, the estimated odds for blacks of driving after consuming a substantial amount of alcohol were \( e^{-0.72} = .49 \) times the estimated odds for whites. The corresponding estimated log odds ratio was −.72 + .74 = .02 for Grade 9, −.72 + .38 = .34 for Grade 10, and −.72 + .01 = −.71 for Grade 11. i.e. there is essentially no association in Grade 9, but the association changes to an odds ratio of about .5 in Grades 11 and 12.

Problem 3. Agresti 5.24

Are people with more social ties less likely to get colds? Use logistic models to analyze the 2x2x2x2 contengency table on pp. 1943 of article by S. Cohen et al., J.Am.Med.Assoc. 277 (24).

See next several pages of SAS output:
HW5 Problem 3

Residuals vs predicted eta_i with LOESS Overlay

The LOESS Procedure
Selected Smoothing Parameter: 1
Dependent Variable: res

Stepwise Model Selection

<table>
<thead>
<tr>
<th>Response Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordered Value</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

Stepwise Selection Procedure

<table>
<thead>
<tr>
<th>Class Level Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
</tr>
<tr>
<td>titer</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>virus</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>social</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-1.5530</td>
<td>0.2880</td>
<td>29.0752</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>titer</td>
<td></td>
<td>1.9280</td>
<td>0.2923</td>
<td>43.5231</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>virus</td>
<td></td>
<td>-0.6051</td>
<td>0.2805</td>
<td>4.6533</td>
<td>0.0310</td>
</tr>
<tr>
<td>social</td>
<td></td>
<td>0.6538</td>
<td>0.2782</td>
<td>5.5228</td>
<td>0.0188</td>
</tr>
</tbody>
</table>

Odds Ratio Estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>titer f&lt;=2f vs f&gt;=4f</td>
<td>6.876</td>
<td>3.878</td>
</tr>
<tr>
<td>virus fHanks vs fRV39f</td>
<td>0.546</td>
<td>0.315</td>
</tr>
<tr>
<td>social f1-5f vs f&gt;=6f</td>
<td>1.923</td>
<td>1.115</td>
</tr>
</tbody>
</table>
HW5 Problem 3
Residuals vs predicted eta_i with LOESS Overlay

The LOESS Procedure
Selected Smoothing Parameter: 1
Dependent Variable: res

Hosmer and Lemeshow
Goodness-of-Fit Test

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1753</td>
<td>6</td>
<td>0.9029</td>
</tr>
</tbody>
</table>
HW5 Problem 3
Residuals vs predicted eta_i with LOESS Overlay

The LOESS Procedure
Selected Smoothing Parameter: 1
Dependent Variable: res
HW5 Problem 3

Std. Pearson residual plots
Problem 4. Agresti 5.25
The derivative equals \[ \frac{\beta e^{\alpha + \beta x}}{[1+e^{\alpha + \beta x}]^2} = \beta \pi(x)(1-\pi(x)) \]

Problem 5. Agresti 5.26
The odds ratio \( e^\beta \) is approximately equal to the relative risk when the probability is near 0 and the complement is near 1, since
\[
e^\beta = \frac{\pi(x + 1)/(1-(\pi(x + 1)/[\pi(x)/(1-\pi(x + 1))])} \approx \pi(x + 1)/\pi(x)
\]

Problem 6. Finney and Pregibon vasoconstriction

a. The AICs for the binary regressions using logit, probit and complimentary log-log links are 35.227, 35.287 and 32.622, respectively. Therefore, the complimentary log-log model has the smallest AIC.

b. In the fitted complimentary log-log model, we have \( \hat{\beta}_0 = -2.9736, \hat{\beta}_1 = 3.9702, \) and \( \hat{\beta}_2 = 4.3361. \) Therefore:
\[
P(V = 1) = 1 - \exp[-\exp\{-2.9736 + 3.9702\log(volume) + 4.3361\log(rate)\}]
\]
\[
P(V = 1) = 1 - \exp[-0.051 * volume^{3.9702} * rate^{4.3361}]
\]
Thus, increasing volume or rate increases the probability of vasoconstriction.

c. The Hosmer-Lemeshow test statistic is 10.0705 (with d.f. = 8). This corresponds to a p-value of 0.2601. Therefore, there is no evidence of lack-of-fit.

d. An approach to detect ill-fit observations is to consider observations where the absolute value of the Pearson Residual is greater than 3. Using this approach, there are 2 observations that are considered ill-fit.

e. There are two influential observations according to the Dfbetas and Cis, and they are the same observations that have large Pearson residuals.

f. When we remove observations 4 and 18, the AIC value for the model drops from 32.622 to 13.32. None of the standardized Pearsons residuals are greater than 2.5, suggesting that no observations are especially ill-fitting. However, the coefficients change significantly. In particular, before we had \( \hat{\beta}_0 = -2.9736, \hat{\beta}_1 = 3.9702, \) and \( \hat{\beta}_2 = 4.3361. \) Now we have \( \hat{\beta}_0 = -16.58384, \hat{\beta}_1 = 21.02387, \) and \( \hat{\beta}_2 = 25.30425. \)

See next several pages of SAS output:
HW5, problem 6
Vaso Data Analysis [Cloglog]

Probability modeled is cons=1.

<table>
<thead>
<tr>
<th>Model Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion</td>
</tr>
<tr>
<td>AIC</td>
</tr>
<tr>
<td>SC</td>
</tr>
<tr>
<td>-2 Log L</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analysis of Maximum Likelihood Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>Ivol</td>
</tr>
<tr>
<td>lrate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hosmer and Lemeshow Goodness-of-Fit Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
</tr>
<tr>
<td>10.0705</td>
</tr>
</tbody>
</table>
**HW5, problem 6**  
*Vaso Data Analysis [Cloglog]*

**The LOGISTIC Procedure**

<table>
<thead>
<tr>
<th>Case Number</th>
<th>Covariates</th>
<th>Pearson Residual</th>
<th>Deviance Residual</th>
<th>DfBeta</th>
<th>DfBeta</th>
<th>95% Confidence Interval</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.4055</td>
<td>-0.2877</td>
<td>3.6225</td>
<td>-0.8722</td>
<td>-0.9173</td>
<td>1.0287</td>
<td>0.9587</td>
</tr>
<tr>
<td>18</td>
<td>0.3471</td>
<td>-0.1625</td>
<td>3.0796</td>
<td>-0.7663</td>
<td>-0.7754</td>
<td>0.7999</td>
<td>0.7419</td>
</tr>
</tbody>
</table>

**Influence Diagnostics**

- Pearson Residual vs. Case Number
- Deviance Residual vs. Case Number
- Leverage vs. Case Number
- 95% Confidence Interval vs. Case Number

---

*Note: The plots depict the relationship between various diagnostic measures and case numbers. The plots indicate how influential each case number is based on the Pearson Residual, Deviance Residual, Leverage, and 95% Confidence Interval.*
HW5, problem 6
Vaso Data Analysis [Cloglog] Without Observations 4 and 18

Number of Observations Read 37
Number of Observations Used 37

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>cons</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>19</td>
</tr>
</tbody>
</table>

Probability modeled is cons=1.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Intercept Only</th>
<th>Intercept and Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>53.266</td>
<td>13.325</td>
</tr>
<tr>
<td>SC</td>
<td>54.877</td>
<td>18.158</td>
</tr>
<tr>
<td>-2 Log L</td>
<td>51.266</td>
<td>7.325</td>
</tr>
</tbody>
</table>

Testing Global Null Hypothesis: BETA=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>43.9410</td>
<td>2</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Score</td>
<td>18.7712</td>
<td>2</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Wald</td>
<td>3.1492</td>
<td>2</td>
<td>0.2071</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-16.5838</td>
<td>10.3652</td>
<td>2.5598</td>
<td>0.1096</td>
</tr>
<tr>
<td>lvol</td>
<td>1</td>
<td>21.0237</td>
<td>13.0935</td>
<td>2.5782</td>
<td>0.1083</td>
</tr>
<tr>
<td>lrate</td>
<td>1</td>
<td>25.3041</td>
<td>16.7511</td>
<td>2.2819</td>
<td>0.1309</td>
</tr>
</tbody>
</table>

Hosmer and Lemeshow Goodness-of-Fit Test

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6550</td>
<td>5</td>
<td>0.9853</td>
</tr>
</tbody>
</table>
Problem 7. Agresti 6.4

For table 10.1, treating marijuana; Set the value as 1 if the predictor use of alcohol is YES and 0 otherwise; 1 if the predictor use of cigarette is YES and 0 otherwise; female as 1 and male as 0; white as 1 and 0 other race. Using a backwards elimination, the final model is composed of the predictors alcohol, cigarette and gender. All interaction terms were non-significant. Therefore, the fitted model is:

\[
\text{logit}(\hat{\pi}) = -5.1883 + 3.0201 \ast \text{alcohol} + 2.8591 \ast \text{cigarette} - 0.3279 \ast \text{gender}
\]

The Pearson GOF statistic yields a p-value of 0.8781, which indicates there is no evidence of gross lack of fit in this model. The odds of marijuana use among alcohol users is \( \exp(3.0201) = 20.494 \) times the odds among non-alcohol users when keeping the remaining parameters constant; the odds of marijuana use among smokers is \( \exp(2.8591) = 17.446 \) times the odds among non-smokers when keeping the remaining parameters constant. And the odds of marijuana use among males is \( 1/\exp(0.3279) = 1.388 \) times the odds among females when keeping the remaining parameters constant.

See next several pages of SAS output:
Problem 7, Agresti 6.4
Stepwise Regression on Marijuana Data

Final Model Selected

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-5.1883</td>
<td>0.4769</td>
<td>118.3642</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>a</td>
<td>1</td>
<td>3.0201</td>
<td>0.4653</td>
<td>42.1249</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td>2.8591</td>
<td>0.1642</td>
<td>303.0914</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>g</td>
<td>1</td>
<td>-0.3279</td>
<td>0.1026</td>
<td>10.2200</td>
<td>0.0014</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>a 1 vs 2</td>
<td>20.494</td>
<td>8.233 51.016</td>
</tr>
<tr>
<td>c 1 vs 2</td>
<td>17.446</td>
<td>12.645 24.071</td>
</tr>
<tr>
<td>g 1 vs 2</td>
<td>0.720</td>
<td>0.589 0.881</td>
</tr>
</tbody>
</table>

Hosmer and Lemeshow Goodness-of-Fit Test

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.8966</td>
<td>3</td>
<td>0.5941</td>
</tr>
</tbody>
</table>
Problem 7, Agresti 6.4
Residuals vs predicted eta_i

Fit Plot for res

Value of the Linear Predictor

Smooth = 0.969
Problem 7, Agresti 6.4
Std. Pearson residual plots

Pearson Residual

Standardized Pearson Residual

Value of the Linear Predictor
Problem 8. Dixon and Massey

First, we fit the model with all main effects and 2-way interactions. Choosing the predictors that have p-value greater than 0.05, we get a model with the predictors age and weight. The fitted model is given by:

\[
\logit(\hat{\pi}) = \hat{\beta}_0 + \hat{\beta}_1 \ast Age + \hat{\beta}_2 \ast weight
\]

\[
\logit(\hat{\pi}) = -7.5128 + 0.0636 \ast Age + 0.0160 \ast weight
\]

A backward elimination can also be done to find the best model. The Hosmer-Lemeshow test has a p-value equal to 0.7941, which indicates that there is no gross lack of fit in this model. The odds of having an incident increases by a multiplicative factor of \(\exp(0.0636) = 1.066\) (95% C.I is (1.025,1.108)) for every unit increase in age while holding weight constant; and the odds of having an incident increases by a multiplicative factor of 1.016 (95% C.I is (1.000,1.032)) for every unit increase in weight while holding age constant.

See next several pages of SAS output:
### Problem 8, heart.sas
#### Stepwise Regression on Heart Data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-7.5128</td>
<td>1.7093</td>
<td>19.3176</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>AG</td>
<td>1</td>
<td>0.0636</td>
<td>0.0197</td>
<td>10.4389</td>
<td>0.0012</td>
</tr>
<tr>
<td>W</td>
<td>1</td>
<td>0.0160</td>
<td>0.00795</td>
<td>4.0464</td>
<td>0.0443</td>
</tr>
</tbody>
</table>

#### Odds Ratio Estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG</td>
<td>1.066</td>
<td>1.025 - 1.108</td>
</tr>
<tr>
<td>W</td>
<td>1.016</td>
<td>1.000 - 1.032</td>
</tr>
</tbody>
</table>

#### Hosmer and Lemeshow Goodness-of-Fit Test

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6510</td>
<td>8</td>
<td>0.7941</td>
</tr>
</tbody>
</table>

### Problem 8, heart.sas
#### Backward Regression on Heart Data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-7.5128</td>
<td>1.7093</td>
<td>19.3176</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>AG</td>
<td>1</td>
<td>0.0636</td>
<td>0.0197</td>
<td>10.4389</td>
<td>0.0012</td>
</tr>
<tr>
<td>W</td>
<td>1</td>
<td>0.0160</td>
<td>0.00795</td>
<td>4.0464</td>
<td>0.0443</td>
</tr>
</tbody>
</table>

#### Odds Ratio Estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG</td>
<td>1.066</td>
<td>1.025 - 1.108</td>
</tr>
<tr>
<td>W</td>
<td>1.016</td>
<td>1.000 - 1.032</td>
</tr>
</tbody>
</table>

#### Hosmer and Lemeshow Goodness-of-Fit Test

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6510</td>
<td>8</td>
<td>0.7941</td>
</tr>
</tbody>
</table>
Cook's distance and Std. Pearson residuals

Problem 8, heart.sas
Problem 8, heart.sas

Cook's distance and Std. Pearson resids

<table>
<thead>
<tr>
<th>Obs</th>
<th>ID</th>
<th>AG</th>
<th>S</th>
<th>D</th>
<th>Ch</th>
<th>H</th>
<th>W</th>
<th>eta</th>
<th>phat</th>
<th>lcl</th>
<th>ucl</th>
<th>p</th>
<th>h2</th>
<th>q</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>21</td>
<td>37</td>
<td>110</td>
<td>70</td>
<td>312</td>
<td>71</td>
<td>170</td>
<td>2.43996</td>
<td>0.08017</td>
<td>0.04468</td>
<td>0.13973</td>
<td>3.38713</td>
<td>0.007438</td>
<td>3.39979</td>
<td>0.08662</td>
</tr>
<tr>
<td>29</td>
<td>29</td>
<td>40</td>
<td>110</td>
<td>74</td>
<td>336</td>
<td>68</td>
<td>166</td>
<td>2.31323</td>
<td>0.09003</td>
<td>0.05377</td>
<td>0.14694</td>
<td>3.17915</td>
<td>0.0065553</td>
<td>3.18962</td>
<td>0.06713</td>
</tr>
<tr>
<td>41</td>
<td>41</td>
<td>40</td>
<td>130</td>
<td>90</td>
<td>520</td>
<td>68</td>
<td>169</td>
<td>2.26522</td>
<td>0.09404</td>
<td>0.05666</td>
<td>0.15210</td>
<td>3.10375</td>
<td>0.0066363</td>
<td>3.11410</td>
<td>0.06478</td>
</tr>
<tr>
<td>86</td>
<td>86</td>
<td>34</td>
<td>110</td>
<td>80</td>
<td>214</td>
<td>67</td>
<td>139</td>
<td>3.12677</td>
<td>0.04201</td>
<td>0.01770</td>
<td>0.09646</td>
<td>4.77495</td>
<td>0.0082924</td>
<td>4.79487</td>
<td>0.19223</td>
</tr>
<tr>
<td>126</td>
<td>126</td>
<td>28</td>
<td>120</td>
<td>86</td>
<td>386</td>
<td>70</td>
<td>189</td>
<td>2.70816</td>
<td>0.06249</td>
<td>0.02557</td>
<td>0.14481</td>
<td>3.87319</td>
<td>0.0132553</td>
<td>3.89912</td>
<td>0.20423</td>
</tr>
<tr>
<td>159</td>
<td>159</td>
<td>40</td>
<td>110</td>
<td>70</td>
<td>244</td>
<td>70</td>
<td>161</td>
<td>2.39324</td>
<td>0.08369</td>
<td>0.04887</td>
<td>0.13966</td>
<td>3.30891</td>
<td>0.0066023</td>
<td>3.31989</td>
<td>0.07325</td>
</tr>
<tr>
<td>184</td>
<td>184</td>
<td>43</td>
<td>138</td>
<td>94</td>
<td>320</td>
<td>65</td>
<td>159</td>
<td>2.23450</td>
<td>0.09669</td>
<td>0.05937</td>
<td>0.15365</td>
<td>3.05643</td>
<td>0.0063453</td>
<td>3.06618</td>
<td>0.06003</td>
</tr>
</tbody>
</table>

ROC Curve for Model

Area Under the Curve = 0.7349
Predicted Probabilities for CNT=1 with 95% Confidence Limits
At W=165.2

R-Square 0.0759  Max-rescaled R-Square 0.1410
**ROC Model: Age**

### Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-4.7925</td>
<td>0.9595</td>
<td>24.9481</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>AG</td>
<td>1</td>
<td>0.0633</td>
<td>0.0192</td>
<td>10.8270</td>
<td>0.0010</td>
</tr>
</tbody>
</table>

### Odds Ratio Estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG</td>
<td>1.065</td>
<td>1.026 1.106</td>
</tr>
</tbody>
</table>
**ROC Model: Weight**

### Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-4.7244</td>
<td>1.3692</td>
<td>11.9064</td>
<td>0.0006</td>
</tr>
<tr>
<td>W</td>
<td>1</td>
<td>0.0167</td>
<td>0.00783</td>
<td>4.5536</td>
<td>0.0328</td>
</tr>
</tbody>
</table>

### Odds Ratio Estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>1.017</td>
<td>1.001 - 1.033</td>
</tr>
</tbody>
</table>
ROC Curve for Weight
Area Under the Curve = 0.6367

ROC Curves for Comparisons

- ROC Curve (Area)
  - Age (0.6985)
  - Weight (0.6367)
### ROC Association Statistics

<table>
<thead>
<tr>
<th>ROC Model</th>
<th>Mann-Whitney</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Area</td>
<td>Standard Error</td>
<td>95% Wald Confidence Limits</td>
<td>Somers' D</td>
<td>Gamma</td>
<td>Tau-a</td>
</tr>
<tr>
<td>Age</td>
<td>0.6985</td>
<td>0.0504</td>
<td>0.5998</td>
<td>0.7972</td>
<td>0.3970</td>
<td>0.4065</td>
</tr>
<tr>
<td>Weight</td>
<td>0.6367</td>
<td>0.0551</td>
<td>0.5287</td>
<td>0.7447</td>
<td>0.2734</td>
<td>0.2758</td>
</tr>
</tbody>
</table>

### ROC Contrast Coefficients

<table>
<thead>
<tr>
<th>ROC Model</th>
<th>Row1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1</td>
</tr>
<tr>
<td>Weight</td>
<td>-1</td>
</tr>
</tbody>
</table>

### ROC Contrast Test Results

<table>
<thead>
<tr>
<th>Contrast</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference = Weight</td>
<td>1</td>
<td>0.6822</td>
<td>0.4088</td>
</tr>
</tbody>
</table>

### ROC Contrast Estimation and Testing Results by Row

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% Wald Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age - Weight</td>
<td>0.0618</td>
<td>0.0748</td>
<td>-0.0848</td>
<td>0.2084</td>
<td>0.6822</td>
</tr>
</tbody>
</table>
/* HW5, Problem 3 */
/*ods rtf file="SAS Output_HW5 Prob 3_files.rtf";*/

title1 'HW5 Problem 3';

data colds;
input colds total titer$ virus$ social$;
datalines;
25 33 f<=2f fRV39f f1-5f
20 38 f<=2f fRV39f f>=6f
18 30 f<=2f fHanksf f1-5f
21 43 f<=2f fHanksf f>=6f
11 34 f>=4f fRV39f f1-5f
8 42 f>=4f fRV39f f>=6f
3 26 f>=4f fHanksf f1-5f
3 30 f>=4f fHanksf f>=6f
;
run;
proc print;
run;

proc logistic data=colds outest=betas covout;
title2 'Stepwise Regression on Cold Data';
class titer virus social/param=ref;
model colds/total = titer virus social titer*virus virus*social titer*social
   / selection=stepwise
   slentry=0.05
   slstay=0.05
   details
   lackfit;
output out=pred p=phat lower=lcl upper=ucl stdreschi = q reschi=p h=h xbeta=eta predprob=(individual c
run;

/* Creating the standardized residuals from the "pred" dataset */
data pred2; set pred; res = p/sqrt(1-h); run;

/* Usual plot of stand. Residuals vs predicted eta_i */
proc sgplot data=pred2;
title2 "Residuals vs predicted eta_i";
scatter y=res x=eta;
run;

/* Now, doing the LOESS overlay on the r_i vs eta_i fit. You can learn about smoothing parameter selectio
Lot to learn!   */
proc loess data=pred2;
title2 "Residuals vs predicted eta_i with LOESS Overlay";
model res = eta;
run;
/* Also, you can try this one below */

proc sgscatter data=pred;
title2 "Std. Pearson residual plots";
plot p*eta q*eta / loess;
run;
/*ods rtf close;*/

/* These are some extras you can ignore */

proc print data=betas;
title2 'Parameter Estimates and Covariance Matrix';
run;
proc print data=pred;
title2 'Predicted Probabilities and 95% Confidence Limits';
run;

/* HW5, problem 6 */
/*ods rtf bodytitle file='SAS Output HW5 Problem 6.rtf';*/
title 'HW5, problem 6';
data vaso;
  input cons volume rate;
datalines;
/* DATALINES not shown here. */
.
.
.
;
run;

data vaso1; set vaso;
lvol = log(volume);
lrate = log(rate);
run;

/*proc print data=vaso1;*/
/*run;*/

proc logistic data=vaso1 descending;
title2 'Vaso Data Analysis [Logit]';
model cons = lvol lrate/link = logit aggregate lackfit;
output out=pred p=phat lower=lcl upper=ucl reschi=p h=h xbeta=eta;
run;

proc logistic data=vaso1 descending;
title2 'Vaso Data Analysis [Probit]';
model cons = lvol lrate/link = probit aggregate lackfit;
output out=pred p=phat lower=lcl upper=ucl reschi=p h=h xbeta=eta;
run;
proc logistic data=vaso1 descending;
title2 'Vaso Data Analysis [Cloglog]';
model cons = lvol lrate/link = cloglog aggregate lackfit influence iplots;
output out=pred p=phat lower=lcl upper=ucl reschi=p h=h xbeta=eta;
run;

/* Note that the 2 aberrant observations are 4 and 18. Both have high Pearson, Dbeta, and Cis. */
data vaso2; set vaso1;
  if _n_ = 4 then delete; if _n_ = 18 then delete;
proc logistic descending data=vaso2;
  model cons = lvol lrate / link = cloglog lackfit;
run;
ods rtf close;
/* Aside */
/* proc logistic data=vaso1 descending;
  model cons = lvol lrate / link = cloglog aggregate lackfit influence;
  output out=pred reschi = u stdreschi = r xbeta=eta p=p;
  run; */
data pred1; set pred; keep r u;
run;
proc print data=pred1;
run;
title;
```sas
/* Problem 7, Agresti 6.4 */
/*a = alcohol; c = cigarette use; m = marijuana use; r = race [1 = white, 2 = black]; g = gender [1 = F */

title "Problem 7, Agresti 6.4";
/*ods rtf bodytitle file='SAS Output HW5 Problem 7.rtf';*/
data mari;
input a c m r g count total;
datalines;
1 1 1 1 1 405 673
1 1 1 2 1 23 46
1 2 1 1 1 13 231
1 2 1 2 1 2 21
2 1 1 1 1 1 18
2 1 1 2 1 0 1
2 2 1 1 1 1 118
2 2 1 2 1 0 12
1 1 1 1 2 453 681
1 1 1 2 2 30 49
1 2 1 1 2 28 229
1 2 1 2 2 1 19
2 1 1 1 2 1 18
2 1 1 2 2 1 9
2 2 1 1 2 1 134
2 2 1 2 2 0 17;
run;

proc logistic data= mari outest=betas covout;
title2 'Stepwise Regression on Marijuana Data';
class a c r g/param=ref;
model count/total = a|c|r|g@2 / selection=stepwise slentry=0.05 slstay=0.05 details lackfit;
output out=pred p=p hat lower=lcl upper=ucl stdreschi = q reschi=p h=h xbeta=eta predprob=(individual c
run;

/* Creating the standardized residuals from the "pred" dataset */
data pred2; set pred; res = p/sqrt(1-h);

/* Usual plot of stand. Residuals vs predicted eta_i */
proc sgplot data=pred2;
   title2 "Residuals vs predicted eta_i";
   scatter y=res x=eta;
run;

proc loess data=pred2;
title2 'residuals vs. eta';
model res = eta;
run;

/* Also, you can try this one below */
proc sgscatter data=pred;
title2 "Std. Pearson residual plots";
plot p*eta q*eta / loess;
run;
/*ods rtf close;*/
```
title "Problem 8, heart.sas";
datalines;
/* DATALINES not shown here (very long) */.
.
.
;
run;

proc logistic data=heart descending;
title2 'Stepwise Regression on Heart Data';
model CNT = AG|S|D|Ch|H|W@2
   / selection=stepwise
      slentry=0.05
      slstay=0.05
details
      lackfit
   scale = none;
output out=predict p=phat lower=lcl upper=ucl stdreschi = q reschi=p h=h2 xbeta=eta c = c predprob=(in
run;

proc logistic data=heart descending;
title2 'Backward Regression on Heart Data';
model CNT = AG|S|D|Ch|H|W@2
   / selection=B fast ctable
      slentry=0.05
      slstay=0.05
details
      lackfit;
output out=predict p=phat lower=lcl upper=ucl stdreschi = q reschi=p h=h2 c = c xbeta=eta predprob=(in
run;

/* Plots, and also comment from the Table on Outliers. Interpretation in details, wrt odds ratios */
proc sgscatter data=predict;
title2 "Cook's distance and Std. Pearson resids";
plot q*eta q*id c*id/loess;
proc print data=predict(where=(c>0.2 or q>3 or q<-3));
*var y width color c r;
run;
/* Now, predictive accuracy */
proc logistic data = heart descending plots(only)=(roc(id=obs) effect);
model CNT = AG W/scale = none details lackfit rsquare;
run;

proc logistic data = heart descending plots;
model CNT = AG W/scale = none details lackfit rsquare outroc=ROCcurve ctable;
run;

/* Below find individual ROC curves for each covariate */

/* The GAM below is some extras */
*proc gam plots(clm);
proc gam data=heart;  
model CNT(EVENT='1') = spline(AG) spline(Ch) spline(W) / dist=binomial link=logit;
run;
*ods graphics off;
*ods html close;