Exam 3 is on Wednesday, April 15 and covers Agresti material **including points 23) through 35) on the Exam 2 review, but more emphasis on Poisson regression.** 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.

36) 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
$x\|y = i \sim N(\mu_i, \sigma^2)$	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 $\log(1 - x)$

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

38) 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. If x1 x5 x7 is an interaction, keep the main effects x1, x5 and x7 in the model.

39) 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.

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

41) To make a full model, use points 36), 37), 39) and 40) and sometimes 38). The number of predictors in the full model should be much smaller than the number of data cases N. Make a response plot to check that the full model is good.

42) **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.

43) **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 opti-

mizes 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.

44) 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 $m \leq \min(N_1, N_0)/10$.

45) Know how to find good models from output. A good submodel I will use a small number of predictors, have a good response 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.

46) 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 Ho $\beta_i = 0$ versus H_A $\beta_i \neq 0$ where the current model with $j+1$ variables is treated as the full model.

47) Heuristically, forward selection tries to add the variable that will decrease the deviance the most. A decrease 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 Ho $\beta_i = 0$ 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).

48) For logistic regression, let N_1 = number of ones and $N_0 = N - N_1$ = number of zeroes. 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.

49) For Poisson 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.

50) Variable selection is pretty much the same for logistic regression and Poisson 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 are described in point 52). For the EE plot for binary logistic regression, mark the symbols using the response variable Y. See HW7 1,2, HW8 1, HW9 1 and HW 10 1.

51) If $\Delta(I) = AIC(I) - AIC(I_{min})$, then models with $\Delta(I) \leq 2$ are good, models with $4 \leq \Delta(I) \leq 7$ are borderline. Look at the submodel I_I with the smallest number of predictors such that $\Delta(I_I) \leq 2$, and also examine submodels I with fewer predictors than I_I with $\Delta(I) \leq 7$. I_I **is the initial submodel to examine**. If a factor has J-1 dummy variables, either keep all J-1 dummy variables or delete all J-1 dummy variables, do not delete some of the dummy variables.

52) Know how to find good models from output. The following rules of thumb (roughly in order of decreasing importance) may be useful. It is often not possible to have all 10 rules of thumb to hold simultaneously. Let submodel I have $k_I + 1$ predictors, including a constant. Do not use more predictors than submodel I_I, which has no more predictors than the minimum AIC model. It is possible that $I_I = I_{min} = I_{full}$. Assume the response plot for the full model is good. Then the submodel I is good if
i) the response plot for the submodel looks like the response plot for the full model.
ii) Want corr(ESP,ESP(I)) ≥ 0.95.
iii) The plotted points in the EE plot cluster tightly about the identity line.
iv) Want the p-value ≥ 0.01 for the change in deviance test that uses I as the reduced model.
v) Want $r_I + 1 \leq \min(N_1, N_0)/10$ for binary regression where
$N_1 = \#$ 1's and $N_0 = \#$ 0's.
vi) Want the deviance $G^2(I) \geq G^2(full)$ but close. ($G^2(I) \geq G^2(full)$ since adding predictors to I does not increase the deviance).
vii) Want AIC(I) $\leq AIC(I_{min}) + 7$ where I_{min} is the minimum AIC model found by the variable selection procedure.
viii) Want hardly any predictors with p-values > 0.05.
ix) Want few predictors with p-values between 0.01 and 0.05.
x) Want $G^2(I) \leq N - k_I - 1 + 3\sqrt{N - k_I - 1}$.

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 $\boldsymbol{x}_i = (x_1, ..., x_k)^T$ is observed and that $Y_i|\boldsymbol{x}_i \sim$ independent Poisson($\mu(\boldsymbol{x}_i)$) for $i = 1, ..., N$ where

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

This is called the **full model** for **Poisson regression** (PR) and the $(k+1)$ parameters $\alpha, \beta_1, ..., \beta_k$ are estimated. Know how to predict $\hat{\mu}(\boldsymbol{x})$. Also $\hat{Y} = \hat{\mu}(\boldsymbol{x})$. See HW9 2, Q8.

For the **saturated model**, the $Y_i|\boldsymbol{x}_i \sim$ independent Poisson(μ_i) for $i = 1, ..., N$ where

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

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

For deviance, $df = N-k-1$ and if the saturated model is good, a rule of thumb is that the Poisson regression model is ok if $G^2 \leq N-k-1$ (or if $G^2 \leq N-k-1+3\sqrt{N-k-1}$).

55) An estimated sufficient summary or **response plot** is a plot of $ESP_i = \hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i$ versus Y_i with the exponential curve of estimated means

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

added to the plot along with a lowess curve.

56) Suppose that ESP_i takes many values (eg the PR model has a continuous predictor) and that $k+1 << N$. Know that the PR model may be good if the lowess tracks the exponential curve of estimated means in the response plot. Also know that you should check that the PR model is good before doing inference with the PR 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 PR in the conclusion. The output looks almost like that shown on p. 2 except *Arc* output will use menu P1 and *R* output will use *family = poisson*. See HW9 2, Q8.

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

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

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

60) Know how to perform the 4 step **change in deviance test**. The output is almost the same as that on p. 4 (except *Arc* menus B1 and B2 are replaced by P1 and P2 while *R* output on p. 3 uses *family = poisson*) 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.

In Agresti for Exam 3, we have covered odds in 2.3 p. 28, 3.1, 3.2, 3.3, 3.4, 4.1, 4.2, 4.3, 4.4, 5.1, 5.2, 5.3.1 but have skipped subsections most of 2.3, 2.4.5, 2.4.6, 4.1.5, 4.3.3, 4.3.4, 4.3.5, 4.5, 5.1.6, 5.1.7, 5.1.8, 5.2.4, and 5.2.6.