Exam 2 is Wednesday, March 25. 7 sheets of notes

The material for categorical data follows Agresti closely.

A *categorical variable* is one for which the measurement scale consists of a set of categories.

Categorical variables having an ordered scales are called *ordinal variables* while categorical variables having unordered scales are called *nominal variables.*

A *response variable* y is the variable of interest. *Explanatory variables* $x_1, ..., x_p$ are used to predict (or explain) y.

Agresti is concerned with categorical response variables. The explanatory variable may be categorical or quantitative (*quantitative variables* take values on a numerical scale).

The following topic is not in Agresti but is in (p. 682 5th ed.) the Math 483 text.

20) χ^2 *goodness of fit test.* Suppose that there is a single categorical variable Y that has k categories $A_1, ..., A_k$. Let $p_i = P(A_i) =$ probability that Y falls in category A_i. Let the experiment be performed n times (ie randomly selected from some population, or the outcome of n independent identical experiments). Let y_i be the observed counts that n trials resulted in category i for $i = 1, ..., k$. Let $n = \sum_{i=1}^k y_i$ and let $E_i = n\pi_i$.

a) If the π_i are chosen before collecting data, then the four step test is

i) $H_o : p_i = \pi_i$ for $i = 1, ..., k$, $\quad H_A :$ *not* H_o

ii) test statistic

$$X^2 = \sum_{i=1}^k \frac{(y_i - n\pi_i)^2}{n\pi_i} = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

iii) The p–value $= P(W > X^2)$ where $W \sim \chi^2_{k-1}$ has a chi–square distribution with $k - 1$ degrees of freedom.

iv) Reject H_o if the p–value $< \delta$, otherwise fail to reject H_o and give a nontechnical conclusion.

b) If the π_i are computed after estimating r parameters from the data with the MLE, then steps i) and ii) are the same but iii) and iv) change slightly: if $X^2 > \chi^2_{\delta,k-1}$ then the p–value $< \delta$ so reject Ho. If $X^2 < \chi^2_{\delta,k-1-r}$ then the p–value $> \delta$ so fail to reject Ho. If $\chi^2_{\delta,k-1-r} < X^2 < \chi^2_{\delta,k-1}$, the test is inconclusive. Here $P(W > \chi^2_{\delta,d}) = \delta$ if W has a chi–square distribution with d degrees of freedom. Still give a nontechnical conclusion in step iv).

The p–value is either given in output or approximated using a table. If δ is not given, use $\delta = 0.05$. If H_o is rejected, then conclude that there is strong evidence that the model $p_i = \pi_i$, $i = 1, ..., k$ does not hold. If H_o is not rejected, then there is not enough evidence to conclude that the model $p_i = \pi_i$, $i = 1, ..., k$ does not hold.

21) The *chi–square test for independence or homogeneity*: Suppose that there are two categorical variables: the row variable with I categories and the column variable with J categories. Know how to perform the 4 step test:
i) Ho: there is no relationship between the two categorical variables
Ha: there is a relationship.
ii) test statistic =

$$X^2 = \sum \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

iii) p-value = $P(W > X^2)$ where $W \sim \chi^2_{(I-1)(J-1)}$,
(the degrees of freedom = $(I-1)(J-1)$).
iv) Reject Ho if the p–value $\leq \delta$, and conclude that there is a relationship between the two categorical variables. If the p-value $> \delta$, fail to reject Ho and conclude that there is no relationship between the two variables.

See HW4 1,2, HW5 1, Q4.

Sometimes X^2 is given by output but sometimes you need to compute the expected count and the chi-square contribution. Recall that the expected cell count E = (row total)(column total)/(table total). The chi-square cell contribution = $(O-E)^2/E$ where O and E are the observed and expected cell counts. The expected cell count and the cell chi-square contribution need to be computed for each of the IJ cells. Finally, X^2 is the sum of all IJ cell chi-square contributions.

Sometimes the p–value is given by output but sometimes it needs to be obtained from χ^2 table. The *df* $= (I-1)(J-1)$. Find the two values in the df row of χ^2 table that are closest to X^2. Then the p–value is between the values on the top row of the table. For example, if df = 5 and $X^2 = 13.00$ then 12.83 and 15.09 bracket X^2 and $0.010 < p - value < 0.025$. If X^2 is big and way off the χ^2 table , then p-value < 0.001. For example, if df = 5 and $X^2 = 57$, then p–value = 0. If X^2 is small and way off the χ^2 table , then p-value > 0.25. For example, if df = 5 and $X^2 = 4.33$, then p–value > 0.25.

22) The *likelihood ratio test for independence or homogeneity*: This test is exactly the same as 21) except the test statistic in step ii) replaces X^2 with

$$G^2 = 2 \sum n_{ij} \log(n_{ij}/\hat{\mu}_{ij}) = 2 \sum O_{ij} \log(O_{ij}/E_{ij}).$$

The **simple logistic regression** (SLR) model is $Y|X = x \sim$ independent Bernoulli($\pi(x)$) random variables where

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}.$$

The Y_i are **random variables** while the X_i are treated as known **constants**. The parameters α and β are **unknown constants** that need to be estimated.

(If the X_i are random variables, then the model is conditional on the $X_i's$. Hence the $X_i's$ are still treated as constants.)

The response variable Y is the variable that you want to predict while the independent (or predictor or explanatory) variable X is the variable used to predict the response.

For the exam and final know the meaning of the simple logistic regression output. Shown next are an actual ARC output and an output only using symbols.
Response = Y
Coefficient Estimates

Label	Estimate	Std. Error	Est/SE	p-value
Constant	$\hat{\alpha}$	$se(\hat{\alpha})$	z	for Ho: $\alpha = 0$
x	$\hat{\beta}$	$se(\hat{\beta})$	$z_o = \hat{\beta}/se(\hat{\beta})$	for Ho: $\beta = 0$

```
Number of cases:            N
Degrees of freedom:         N-2
-----------------------------------------------------------------
Binomial Regression
Kernel mean function = Logistic
Response      = Status
Terms         = (Bottom)
Trials        = Ones
Coefficient Estimates
Label      Estimate        Std. Error     Est/SE    p-value
Constant  -21.5922         2.93038        -7.368     0.0000
Bottom     2.33378         0.319705        7.300     0.0000

Number of cases:            200
Degrees of freedom:         198
```

Know: A **scatter plot** is a plot of W vs Z is a plot with W on the horizontal axis and Z on the vertical axis and is used to display the conditional distribution of Z given W.

For SLR the scatterplot of X vs Y is often used.

The following problems are important for both exam 2 and the final. Suppose computer output is given.

23) Given a value $X = x$ of the explanatory variable, and given computer output, find

$$\hat{\pi}(x) = \frac{\exp(\hat{\alpha} + \hat{\beta}x)}{1 + \exp(\hat{\alpha} + \hat{\beta}x)}.$$

See HW5 2b.

24) The large sample 100 $(1 - \delta)$ % CI for β is $\hat{\beta} \pm z_{\delta/2}\ se(\hat{\beta})$. See HW5 2c.

Note that a 90% CI uses $z_{\delta/2} = 1.645$, a 95% CI uses $z_{\delta/2} = 1.96$, and a 99% CI uses $z_{\delta/2} = 2.576$,

25) Be able to perform the 4 step *Wald test* of hypotheses:
i) State the hypotheses Ho: $\beta = 0$ Ha: $\beta \neq 0$.
ii) Find the test statistic $zo = \hat{\beta}/se(\hat{\beta})$ or obtain it from output.
iii) p–value $= 2P(Z < -|zo|) = 2P(Z > |zo|)$. Find the p–value from output or use the standard normal table.
iv) State whether you reject Ho or fail to reject Ho and give a nontechnical sentence restating your conclusion in terms of the story problem.

Recall that Ho is rejected if the p–value $< \delta$, and use $\delta = 0.05$ if δ is not given. If Ho is rejected, then conclude that X is a useful SLR predictor for Y. If you fail to reject Ho, then conclude that X is not a useful SLR predictor for Y. Note that "SLR" is crucial. It could be that X is a very useful predictor for Y, but not a good SLR predictor.

The Wald test is good if the SLR model holds and if the sample size is large. It is better to use the output to get the test statistic and p–value than to use formulas and the tables, but I may not give the relevant output. Expect to get at least two testing of hypotheses problems, one where Ho is rejected and one where Ho is not rejected. See Q5, HW5 2d.

Logistic regression can also be used for binomial data with predictors $\boldsymbol{X}_i = (X_{i1}, ..., X_{ik})^T$. Suppose that $\boldsymbol{X}_i = \boldsymbol{x} = (x_1, ..., x_k)^T$ is observed. Then $Y_i|\boldsymbol{X}_i = \boldsymbol{x}_i \sim$ independent Binomial$(n_i, \pi(\boldsymbol{x}_i))$ for $i = 1, ..., N$ where

$$\pi(\boldsymbol{x}) = \frac{\exp(\alpha + \boldsymbol{\beta}^T \boldsymbol{x})}{1 + \exp(\alpha + \boldsymbol{\beta}^T \boldsymbol{x})}.$$

Here $\boldsymbol{\beta}^T \boldsymbol{x} = \beta_1 x_1 + \cdots + \beta_k x_k$. Binary regression is a special case if $n_i = 1$ for $i = 1, ..., N$.

Notice that $E(Y_i/n_i|\boldsymbol{x}_i) = \pi(\boldsymbol{x}_i)$.

The following two questions are important for the exam and final.

26) Given the values $\boldsymbol{x}$ of the k explanatory variables, and given computer output, find

$$\hat{\pi}(\boldsymbol{x}) = \frac{\exp(\hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x})}{1 + \exp(\hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x})}.$$

See HW5 3f.

27) The large sample 100 $(1 - \delta)$ % CI for β_i is $\hat{\beta}_i \pm z_{\delta/2}\ se(\hat{\beta}_i)$. See HW5 3gh.

For the exam and final know the meaning of the (multiple) logistic regression output. Next are shown an actual ARC output and an output only using symbols.

Response = Y
Coefficient Estimates

Label	Estimate	Std. Error	Est/SE	p-value
Constant	$\hat{\alpha}$	$se(\hat{\alpha})$	$z_{o,0}$	for Ho: $\alpha = 0$
x_1	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$z_{o,1} = \hat{\beta}_1/se(\hat{\beta}_1)$	for Ho: $\beta_1 = 0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	$\hat{\beta}_k$	$se(\hat{\beta}_k)$	$z_{o,k} = \hat{\beta}_k/se(\hat{\beta}_k)$	for Ho: $\beta_k = 0$

```
Scale factor:                1.
Number of cases:             N
Degrees of freedom:          N - k - 1
Pearson X2:
Deviance:                    D = G^2
-------------------------------------

Binomial Regression
Kernel mean function = Logistic
Response      = Status
Terms         = (Bottom Left)
Trials        = Ones
Coefficient Estimates
Label      Estimate        Std. Error     Est/SE    p-value
Constant  -389.806         104.224        -3.740     0.0002
Bottom     2.26423         0.333233        6.795     0.0000
Left       2.83356         0.795601        3.562     0.0004

Scale factor:                  1.
Number of cases:             200
Degrees of freedom:          197
Pearson X2:              179.809
Deviance:                 99.169
```

28) Be able to perform the 4 step Wald test of hypotheses:
i) State the hypotheses Ho: $\beta_j = 0$ Ha: $\beta_j \neq 0$.
ii) Find the test statistic $z_{o,j} = \hat{\beta}_j/se(\hat{\beta}_j)$ or obtain it from output.
iii) p–value $= 2P(Z < -|z_{oj}|) = 2P(Z > |z_{oj}|)$. Find the p–value from output or use the standard normal table.
iv) State whether you reject Ho or fail to reject Ho and give a nontechnical sentence restating your conclusion in terms of the story problem.

Recall that Ho is rejected if the p–value $< \delta$, and use $\delta = 0.05$ if δ is not given. If Ho is rejected, then conclude that X_j is needed in the LR model for Y given that the other $p-1$ predictors are in the model. If you fail to reject Ho, then conclude that X_j is not needed in the LR model for Y given that the other $p-1$ predictors are in the model. Note that X_j could be a very useful SLR predictor, but may not be needed if other predictors are added to the model. See HW5 3ij.

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

$$\hat{\pi}(\boldsymbol{x}) = \frac{\exp(\hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x})}{1 + \exp(\hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x})}.$$

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 $Y_i|\boldsymbol{x}_i \sim$ independent Binomial(n_i, π_i) for $i = 1, ..., N$ where

$$\hat{\pi}_i = Y_i/n_i.$$

This model estimates the N parameters π_i.

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

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 π_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 plot or **response plot** is a plot of the estimated sufficient predictor $ESP_i = \hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i$ versus Y_i with the logistic curve of fitted proportions

$$\hat{\pi}(ESP_i) = \frac{e^{ESP_i}}{1 + e^{ESP_i}}$$

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

29) Suppose that ESP_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 response plot. Also know that you should check that the LR model is good before doing inference with the LR model. See HW6 3,4.

Response = Y Terms = $(X_1, ..., X_k)$ Sequential Analysis of Deviance

Predictor	df	Total Deviance	df	Change Deviance
Ones	$N-1 = df_o$	G_o^2		
X_1	$N-2$		1	
X_2	$N-3$		1	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	
X_k	$N-k-1 = df_{FULL}$	G^2_{FULL}	1	

```
-----------------------------------------
Data set = cbrain, Name of Fit = B1
Response      = sex
Terms         = (cephalic size log[size])
Sequential Analysis of Deviance
                     Total             Change
Predictor      df   Deviance      |    df   Deviance
Ones          266   363.820       |
cephalic      265   363.605       |     1   0.214643
size          264   315.793       |     1   47.8121
log[size]     263   305.045       |     1   10.7484
```

Know how to use the above output for the following test. Assume that the response 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 $\hat{\pi}(\boldsymbol{x}_i)$ from the logistic regression should be used to estimate $P(Y_i = 1|\boldsymbol{x}_i)$ or if none of the predictors should be used and

$$P(Y_i = 1) \equiv \pi \approx \sum_{i=1}^{N} Y_i / \sum_{i=1}^{N} n_i \text{ for all } i = 1, ..., N.$$

30) The 4 step **deviance test** is

i) $H_o : \boldsymbol{\beta} = 0 \quad H_A : \boldsymbol{\beta} \neq 0$

ii) test statistic $G^2(o|F) = G^2_o - G^2_{FULL}$

iii) The p–value $= P(W > G^2(o|F))$ where $W \sim \chi^2_k$ 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 $x_1, ..., x_k$.

Also use R output from the full model and the null model. See HW6 5b 6a.

```
outf <- glm(Y~x1 + x2 + ... + xk, family = binomial)
outn <- glm(Y~1,family = binomial); anova(outn,outf,test="Chi")
  Resid. Df Resid. Dev  Df  Deviance    P(>|Chi|)
1       ***    ****
2       ***    ****     k   G^2(O|F)     pvalue
```

The sufficient predictor SP $= \alpha + \boldsymbol{\beta}^T \boldsymbol{x}$. After obtaining an acceptable full model where

$$SP = SP(Full) = \alpha + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} = \alpha + \boldsymbol{\beta}^T \boldsymbol{x},$$

try to obtain a **reduced model** $Y_i | \boldsymbol{X}_{Ri} = \boldsymbol{x}_{Ri} \sim$ independent Binomial$(n_i, \pi(\boldsymbol{x}_{Ri}))$ where

$$SP(Red) = \alpha + \beta_{R1} x_{Ri1} + \cdots + \beta_{Rm} x_{Rim} = \alpha_R + \boldsymbol{\beta}_R^T \boldsymbol{x}_{Ri}$$

and $\{x_{Ri1}, ..., x_{Rim}\} \subset \{x_1, ..., x_k\}$.

Let $x_{R,m+1}, ..., x_{Rk}$ 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 $i = 1, ..., k$. Then the reduced model uses predictors $x_1, ..., x_m$ and we test $H_o : \beta_{m+1} = \cdots = \beta_k = 0$. However, in practice this sorting is usually not done.

Assume that the response plot looks good. Then we want to test Ho: 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 $\quad H_A$: 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 $k - m$ 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 HW 5c, 6b. Also use R output

```
outf <- glm(Y~x1 + x2 + ... + xk, family = binomial)
outr <- glm(Y~ x3 + x5 + x7,family = binomial); anova(outr,outf,test="Chi")
  Resid. Df Resid. Dev  Df  Deviance    P(>|Chi|)
1       ***    ****
2       ***    ****      k-m  G^2(R|F)     pvalue
```

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 often easier for software to produce. For large N the test statistics from the two test tend to be very similar (asymptotically equivalent tests). The "drop1(outf,test="Chi")" command works in *R*. In *Arc*, select "Examine Submodels" from the *B1* menu, then click on the circle for *Change in deviance for fitting each term last.*

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

Label	Estimate	Std. Error	Est/SE	p-value
Constant	$\hat{\alpha}$	$se(\hat{\alpha})$	$z_{o,0}$	for Ho: $\alpha = 0$
x_1	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$z_{o,1} = \hat{\beta}_1/se(\hat{\beta}_1)$	for Ho: $\beta_1 = 0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	$\hat{\beta}_k$	$se(\hat{\beta}_k)$	$z_{o,k} = \hat{\beta}_k/se(\hat{\beta}_k)$	for Ho: $\beta_k = 0$

Degrees of freedom: N - k - 1 = df_{FULL}
Deviance: $D = G^2_{FULL}$

Response = Y Terms = $(X_1, ..., X_m)$ (Reduced Model)

Label	Estimate	Std. Error	Est/SE	p-value
Constant	$\hat{\alpha}$	$se(\hat{\alpha})$	$z_{o,0}$	for Ho: $\alpha = 0$
x_1	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$z_{o,1} = \hat{\beta}_1/se(\hat{\beta}_1)$	for Ho: $\beta_1 = 0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_m	$\hat{\beta}_m$	$se(\hat{\beta}_m)$	$z_{o,m} = \hat{\beta}_k/se(\hat{\beta}_m)$	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
Label      Estimate        Std. Error     Est/SE    p-value
Constant   2360.49         5064.42         0.466     0.6411
Diagonal  -19.8874         37.2830        -0.533     0.5937
Bottom     23.6950         45.5271         0.520     0.6027
Top        19.6464         60.6512         0.324     0.7460

Degrees of freedom:          196
Deviance:                  0.009

Data set = Banknotes, Name of Fit = B2  (Reduced Model)
Response      = Status
Terms         = (Diagonal)
Coefficient Estimates
Label      Estimate        Std. Error     Est/SE    p-value
Constant   989.545         219.032         4.518     0.0000
Diagonal  -7.04376         1.55940        -4.517     0.0000

Degrees of freedom:          198
Deviance:                 21.109
```

33) If the reduced model is good, then the **EE plot** of $ESP(R) = \hat{\alpha}_R + \hat{\boldsymbol{\beta}}_R^T \boldsymbol{x}_{Ri}$ versus $ESP = \hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i$ should be highly correlated with the identity line with unit slope and zero intercept.

34) Let $\pi(\boldsymbol{x}) = P(\text{success}|\boldsymbol{x}) = 1 - \text{P}(\text{failure}|\boldsymbol{x})$ where a "success" is what is counted and a "failure" is what is not counted (so if the Y_i are binary, $\pi(\boldsymbol{x}) = P(Y_i = 1|\boldsymbol{x})$). Then the **estimated odds of success** is

$$\hat{\Omega}(\boldsymbol{x}) = \frac{\hat{\pi}(\boldsymbol{x})}{1 - \hat{\pi}(\boldsymbol{x})} = \exp(ESP).$$

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 $\exp(\hat{\beta}_i)$.