BINOMIAL CONFIDENCE INTERVALS AND DIAGNOSTICS FOR BINOMIAL REGRESSION

by

Jessica Bayer

Bachelor of Science in Mathematics, Southern Illinois University, 2005

A Research Paper
Submitted in Partial Fulfillment of the Requirements for the
Master of Science Degree

Department of Mathematics
in the Graduate School
Southern Illinois University Carbondale
July, 2007

ACKNOWLEDGMENTS

I would like to thank Dr. David Olive for his invaluable assistance and insights leading to the writing of this paper. He has been very supportive and patient throughout the production of this paper. I have found that he is not only an outstanding professor, but an excellent advisor as well. I would also like to thank the authors of whose books and research papers were used in the writing of this paper. On that note, another thanks to Dr. Olive for allowing me to use his books: "Applied Robust Statistics" and "A Course in Statistical Inference."

I would also like to express my gratitude to the members of my graduate committee: Dr. Sakthivel Jeyaratnam and Dr. Thomas Porter. I would also like to thank Amanuel Teweldemedhin and Comlan DeSouza for their support and encouragement. A special thanks goes to my parents and my sister for their love and support.

PREFACE

This paper will look at confidence intervals for the binomial distribution and the binomial regression model. There are three chapters that follow. In Chapter 1, we will consider three confidence intervals for the binomial parameter. In Chapter 2, we will examine graphical diagnostics for the binomial regression model. Chapter 3 examines a method of generating binomial regression data and checking whether OLS tests have correct p-values for large samples.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

A binomial experiment consists of a fixed number n of independent and identical trials, where each trial results in one of two outcomes. One outcome will be labeled a "success", while the other will be called a "failure". The probability of a "success" in a single trial is equal to some value ρ, while the probability of a "failure" is equal to $(1-\rho)$. We are interested in the random variable Y, the number of successes observed during n trials.

Some examples of a binomial experiment would be:

- Tossing a die 10 times and counting the number of times a three is observed.
- Selecting 500 refridgerators at random and observing the number that are not defective.
- Shooting a gun at a target and counting the number of "hits".

Suppose we conduct n trials and count the number of successes, denoted S, and the number of failures, denoted F. Let the probability of a success be ρ and the probability of failure be $(1-\rho)$. For some n trials the sequence of successes and failures could be

SSFSFSF...SSF

Let y be the number of successes, then $(n-y)$ is the number of failures. Since the trials are independent then any point has probability

$$\rho^y(1-\rho)^{n-y},$$

and since the number of n-tuples that contain y S's and $n - y$ F's is

$$\frac{n!}{y!(n-y)!},$$

then the random variable Y is said to have a *binomial distribution* based on n trials with success probability ρ if and only if

$$P(Y = y) = \frac{n!}{y!(n-y)!}\rho^y(1-\rho)^{n-y},\ y = 0, 1, 2, ..., n.$$

This research paper will provide information about finding an appropriate confidence interval for ρ, and about diagnostics for binomial regression.

Chapter 1 deals with selecting the best confidence interval (CI) for ρ. Three confidence intervals will be considered, namely; classical CI, Agresti-Coull CI, and exact CI.

Chapter 2 deals with diagnostics for binomial regression.

Chapter 3 examines a method of generating binomial regression data and checking whether OLS tests have correct p-values for large samples.

CHAPTER 1

CONFIDENCE INTERVALS

1.1 INTRODUCTION TO CONFIDENCE INTERVALS

Definition 1.1.1. Let the data $Y_1, Y_2, \ldots, Y_n$ have pdf or pmf f($\mathbf{y} \mid \theta$) with parameter space Θ and support $\mathcal{Y}$. Let $L_n(\mathbf{Y})$ and $U_n(\mathbf{Y})$ be statistics such that $L_n(\mathbf{y}) \le U_n(\mathbf{y})$, for all $\mathbf{y} \in \mathcal{Y}$. Then $(L_n(\mathbf{y}),\, U_n(\mathbf{y}))$ is a 100 $(1-\alpha)$% confidence interval (CI) for θ if

$$P_\theta(L_n(\mathbf{Y}) < \theta < U_n(\mathbf{Y}))=1-\alpha$$

for all $\theta \in \Theta$. The interval $(L_n(\mathbf{y}),\, U_n(\mathbf{y}))$ is a large sample 100 $(1-\alpha)$% CI for θ if

$$P_\theta(L_n(\mathbf{Y}) < \theta < U_n(\mathbf{Y}))\rightarrow 1-\alpha$$

for all $\theta \in \Theta$ as n$\rightarrow \infty$. (Olive 2007a)

We will consider three types of confidence intervals for the binomial distribution: classical, Agresti-Coull, and exact. First we will define the three CIs.

Let $Y_1, ..., Y_n$ be iid binomial$(1, \rho)$. Let $\hat{\rho} = \sum_{i=1}^n Y_i/n =$ number of "successes"$/n$.

Definition 1.1.2. The classical large sample 100 $(1-\alpha)$% CI for ρ is

$$\hat{\rho} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{\rho}(1-\hat{\rho})}{n}}$$

where $P(Z \le z_{1-\alpha/2}) = 1-\alpha/2$ if $Z \sim N(0,1)$. (Olive 2007a)

The classical interval should only be used if it agrees with the Agresti Coull interval.

The Agresti Coull CI takes $\tilde{n} = n + z^2_{1-\alpha/2}$ and

$$\tilde{\rho} = \frac{n\hat{\rho} + 0.5 z^2_{1-\alpha/2}}{n + z^2_{1-\alpha/2}}.$$

(The method adds $0.5 z^2_{1-\alpha/2}$ "0's and $0.5 z^2_{1-\alpha/2}$ "1's" to the sample, so that $\tilde{n}$ increases by $z^2_{1-\alpha/2}$.)

Definition 1.1.3. The large sample 100 $(1-\alpha)\%$ Agresti Coull CI for ρ is

$$\tilde{\rho} \pm z_{1-\alpha/2}\sqrt{\frac{\tilde{\rho}(1-\tilde{\rho})}{\tilde{n}}}.$$

(Olive 2007a)

Now let $Y_1, ..., Y_n$ be independent bin(m_i, ρ) random variables, let $W = \sum_{i=1}^n Y_i \sim \text{bin}(\sum_{\text{i=1}}^{\text{n}} \text{m}_\text{i}, \rho)$ and let $n_w = \sum_{i=1}^n m_i$. Often $m_i \equiv 1$ and then $n_w = n$. Let $P(F_{d_1,d_2} \leq F_{d_1,d_2}(\alpha)) = \alpha$ where F_{d_1,d_2} has an F distribution with d_1 and d_2 degrees of freedom. Assume $W = w$ is observed.

Definition 1.1.4. The Clopper Pearson "exact" 100 $(1-\alpha)\%$ CI for ρ is

$$\left(0, \frac{1}{1 \ + \ n_w \ F_{2n_w,2}(\alpha)}\right) \quad \text{for} \quad \text{w} = 0,$$

$$\left(\frac{n_w}{n_w \ + \ F_{2,2n_w}(1-\alpha)}, 1\right) \quad \text{for} \quad \text{w} = \text{n}_\text{w},$$

and (ρ_L, ρ_U) for $0 < w < n_w$ with

$$\rho_L = \frac{w}{w + (n_w - w + 1)F_{2(n_w - w+1),2w}(1-\alpha/2)}$$

and

$$\rho_U = \frac{w+1}{w+1+(n_w - w)F_{2(n_w-w),2(w+1)}(\alpha/2)}.$$

(Olive 2007a)

The "exact" CI is conservative: the actual coverage $(1-\delta_n) \geq 1-\alpha =$ the nominal coverage. This interval performs well if ρ is very close to 0 or 1.

Simulation of the confidence intervals is included in the following tables. The simulation gives coverage and scaled length for the three confidence intervals, where scaled length= $\sqrt{n}(U_n - L_n) \approx 2(1.96)\sqrt{\rho(1-\rho)}$ for large n. For each value of ρ, the probability of success, there are simulations for $n =$ 10, 50, 100, and 5000 each with $\alpha = 0.05$ and 5000 runs. We will use ccov, accov, and ecov to represent the coverage of the classical, Agresti-Coull, and exact confidence intervals, respectively. Clen, alen, elen will be used for the scaled lengths of the classical, Agresti-Coull, and exact confidence intervals, respectively. The confidence interval performs well when the coverage is between 0.92 and 0.98 and the scaled lengths are short.

We can make the following observations from the tables:

1. The exact coverage was good for all $n(min(\rho, 1-\rho))$.

2. The classical coverage was good for all $n(min(\rho, 1-\rho)) > 50$. In the simulation the classical CI performs well when $n = 100$ and 5000 and $0.1 \leq \rho \leq 0.9$. In general, the classical CI performs well when n is large and ρ is not close to 0 or 1.

3. The Agresti-Coull coverage was good for all $n(min(\rho, 1-\rho))$ combinations, but for $n(min(\rho, 1-\rho))$ small, the length of the exact interval was shorter.

n	ρ	*ccov*	*clen*	*accov*	*aclen*	*ecov*	*elen*
10	.0001	.0006	.0005	1	1.0149	.9994	.8189
50	.0001	.0052	.0022	1	.6037	.9948	.4129
100	.0001	.0106	.0031	1	.4458	.9894	.2978
5000	.0001	.3998	.0193	.986	.0768	.986	.0588
10	.001	.0108	.0098	1	1.0183	.9892	.8249
50	.001	.0472	.0198	.9994	.6125	.9994	.4273
100	.001	.101	.0304	.9964	.4603	.9964	.3206
5000	.001	.8794	.1189	.9642	.1439	.982	.1394
10	.01	.096	.0895	.9952	1.0475	.9952	.8758
50	.01	.3956	.1906	.986	.7022	.986	.5639
100	.01	.632	.2469	.9816	.5842	.9816	.4965
5000	.01	.9516	.3891	.9476	.3962	.9554	.4044
10	.1	.6508	.7602	.9244	1.2815	.9836	1.2883
50	.1	.8804	1.1224	.972	1.2432	.972	1.2767
100	.1	.935	1.1626	.9736	1.2169	.9584	1.2599
5000	.1	.9584	1.1756	.9552	1.1768	.959	1.1898
10	.2	.8864	1.2664	.9654	1.4451	.9938	1.5803
50	.2	.9388	1.5387	.9492	1.547	.97	1.6481
100	.2	.9308	1.5528	.9414	1.5568	.9696	1.6372
5000	.2	.9546	1.5678	.9542	1.5679	.9562	1.5816

Table 1.1. Results for simulation of CIs when nruns = 5000.

n	ρ	*ccov*	*clen*	*accov*	*aclen*	*ecov*	*elen*
10	.3	.834	1.5881	.9496	1.5445	.9632	1.7596
50	.3	.9374	1.7787	.9548	1.7396	.9674	1.8746
100	.3	.948	1.7863	.949	1.7659	.9598	1.8629
5000	.3	.9508	1.7936	.9502	1.7959	.9524	1.8101
10	.4	.9054	1.7856	.983	1.6052	.983	1.8696
50	.4	.9432	1.8994	.9432	1.8386	.9724	1.9892
100	.4	.948	1.9099	.948	1.8779	.9576	1.9847
5000	.4	.955	1.9202	.9526	1.9196	.9564	1.9339
10	.5	.8886	1.8329	.9782	1.6201	.9782	1.8967
50	.5	.9356	1.9401	.9356	1.8723	.9636	2.0279
100	.5	.937	1.9498	.937	1.9141	.9608	2.0237
5000	.5	.9546	1.9598	.9546	1.9590	.9546	1.9734
10	.6	.9032	1.7783	.9802	1.6029	.9802	1.8655
50	.6	.941	1.8985	.941	1.8379	.9686	1.9883
100	.6	.9492	1.9113	.9492	1.8791	.9576	1.9861
5000	.6	.9526	1.9203	.9508	1.9196	.9534	1.9339
10	.7	.8324	1.5875	.9546	1.5446	.9618	1.7597
50	.7	.931	1.7753	.9546	1.7369	.9664	1.8714
100	.7	.9466	1.7859	.9422	1.7657	.9594	1.864
5000	.7	.9528	1.7961	.9526	1.7956	.955	1.8098

Table 1.2. Continuation of Table 1.1.

n	ρ	*ccov*	*clen*	*accov*	*aclen*	*ecov*	*elen*
10	.8	.8874	1.2584	.9654	1.4428	.9926	1.5761
50	.8	.9362	1.5408	.9516	1.5488	.9662	1.6502
100	.8	.9342	1.5561	.9394	1.5597	.967	1.6404
5000	.8	.9522	1.5677	.952	1.5678	.9536	1.5816
10	.9	.654	.7613	.9276	1.2821	.9882	1.2892
50	.9	.8778	1.1191	.9702	1.2412	.9702	1.2739
100	.9	.931	1.1622	.9716	1.2167	.955	1.2597
5000	.9	.9504	1.1759	.9474	1.1771	.95	1.19
10	.99	.0894	.0828	.9962	1.0450	.9962	.8716
50	.99	.3934	.1897	.9866	.7018	.9866	.5633
100	.99	.6354	.2496	.9814	.5859	.9814	.4987
5000	.99	.9464	.3881	.9452	.3952	.9528	.4034
10	.999	.013	.0118	1	1.0191	.987	.8262
50	.999	.0532	.0227	.998	.6140	.998	.4296
100	.999	.0924	.0279	.9964	.4589	.9964	.3184
5000	.999	.8626	.1175	.963	.1429	.9794	.1383
10	.9999	.0006	.0005	1	1.0149	.9994	.8189
50	.9999	.0048	.0019	1	.6036	.9952	.4128
100	.9999	.0102	.0030	.9998	.4457	.9898	.2977
5000	.9999	.3916	.0191	.984	.0767	.984	.0586

Table 1.3. Continuation of Table 1.1.

1.2 CONFIDENCE INTERVALS FOR FINITE POPULATIONS

Let $\hat{\rho}$ = number of "successes"/n. Consider taking a simple random sample of size n from a finite population of known size N.

Definition 1.2.1. The classical finite population large sample 100 $(1-\alpha)$% CI for ρ is

$$\hat{\rho} \pm z_{1-\alpha/2}\sqrt{\tfrac{\hat{\rho}(1-\hat{\rho})}{n-1}\left(\tfrac{N-n}{N}\right)} = \hat{\rho} \pm z_{1-\alpha/2}SE(\hat{\rho})$$

where $P(Z \leq z_{1-\alpha/2}) = 1-\alpha/2$ if $Z \sim N(0,1)$. (Olive 2007a)

The Agresti-Coull CI takes $\tilde{n} = n + z^2_{1-\alpha/2}$ and

$$\tilde{\rho} = \frac{n\hat{\rho} + 0.5z^2_{1-\alpha/2}}{n + z^2_{1-\alpha/2}}.$$

Definition 1.2.2. The large sample 100 $(1-\alpha)$% Agresti Coull type finite population CI for ρ is

$$\tilde{\rho} \pm z_{1-\alpha/2}\sqrt{\tfrac{\tilde{\rho}(1-\tilde{\rho})}{\tilde{n}}\left(\tfrac{N-n}{N}\right)} = \tilde{\rho} \pm z_{1-\alpha/2}SE(\tilde{\rho}).$$

(Olive 2007a)

(This method adds $0.5z^2_{1-\alpha/2}$ "0's" and $0.5z^2_{1-\alpha/2}$ "1's" to the sample, so $\tilde{n}$ increases by $z^2_{1-\alpha/2}$.)

Notice that a 95% CI uses $z_{1-\alpha/2} = 1.96 \approx 2$.

For data from a finite population, large sample theory gives useful approximations as N and $n \to \infty$ and $n/N \to 0$. Theory suggests that the Agresti Coull CI should have better coverage than the classical CI if ρ is near 0 or 1, if the sample size n is moderate, and if n is small compared to the population size N. If n is large, but small compared to N, the coverage of the classical and Agresti Coull CIs should be similar. As n increases to N, $\hat{\rho}$ goes to ρ, SE($\hat{\rho}$) goes to 0, and the classical CI may perform well. SE($\tilde{\rho}$) also goes to 0, but $\tilde{\rho}$ is a biased estimator of ρ and the Agresti Coull CI will not perform well if n/N is too large.

Simulation of the CIs is included in the following tables. The simulation gives coverage and scaled length for the classical and Agresti-Coull CIs. For each value of ρ, the probability of success, there are simulations for $n = 50$, 100, 200, 300, 400, and 450 each with $N = 500$, $\alpha = 0.05$, and 5000 runs.

We can make the following observations from the tables:

1. The classical coverage was good for all values of ρ when n was near N.
2. The Agresti-Coull coverage was good for $n \leq 0.6N$.

n	ρ	*ccov*	*clen*	*accov*	*aclen*
50	.01	.4072	.2324	.9912	.7350
100	.01	.6666	.2764	.9528	.5603
200	.01	.9208	.2814	.9174	.4076
300	.01	.9112	.2412	.9216	.3085
400	.01	.9374	.1734	.6744	.2091
450	.01	.9236	.1231	.4072	.1456
50	.1	.9036	1.0949	.9496	1.1818
100	.1	.95	1.0451	.962	1.0879
200	.1	.9374	.9081	.9412	.9273
300	.1	.9402	.7435	.9418	.7541
400	.1	.9348	.5261	.9500	.5317
450	.1	.9500	.3723	.898	.3758
50	.2	.9446	1.4797	.9608	1.4713
100	.2	.9454	1.3987	.9492	1.3948
200	.2	.9422	1.2144	.9626	1.2127
300	.2	.9578	.9922	.9384	.9912
400	.2	.9484	.7017	.9484	.7012
450	.2	.9636	.4962	.9546	.4959

Table 1.4. Results for simulation of finite CIs when nruns = 5000.

n	ρ	*ccov*	*clen*	*accov*	*aclen*
50	.3	.942	1.6990	.9408	1.6459
100	.3	.9478	1.6070	.948	1.5807
200	.3	.9486	1.3926	.9498	1.3809
300	.3	.9524	1.1374	.9514	1.1310
400	.3	.9496	.8042	.9498	.8008
450	.3	.9488	.5687	.9464	.5666
50	.4	.9258	1.8224	.9514	1.7461
100	.4	.9496	1.7182	.9496	1.6808
200	.4	.9532	1.4889	.9532	1.4724
300	.4	.9478	1.2158	.9478	1.2067
400	.4	.9398	.8596	.9398	.8548
450	.4	.9528	.6079	.9528	.6048
50	.5	.9512	1.8614	.9512	1.7780
100	.5	.9464	1.7548	.9464	1.7139
200	.5	.9480	1.5197	.9480	1.5017
300	.5	.9426	1.2408	.9426	1.2309
400	.5	.9452	.8774	.9452	.8721
450	.5	.9496	.6204	.9496	.6171

Table 1.5. Continuation of Table 1.4.

n	ρ	*ccov*	*clen*	*accov*	*aclen*
50	.6	.9342	1.8226	.9568	1.7463
100	.6	.9488	1.7195	.9488	1.6819
200	.6	.9528	1.4889	.9528	1.4724
300	.6	.9486	1.2156	.9486	1.2065
400	.6	.9456	.8595	.9456	.8547
450	.6	.9528	.6078	.9528	.6048
50	.7	.9502	1.7015	.9488	1.6479
100	.7	.9466	1.6062	.9498	1.5799
200	.7	.9546	1.3924	.9528	1.3808
300	.7	.9546	1.1369	.9574	1.1306
400	.7	.9452	.8039	.9476	.8006
450	.7	.9472	.5686	.9496	.5665
50	.8	.9534	1.4834	.962	1.4741
100	.8	.9448	1.3998	.9448	1.3958
200	.8	.9410	1.2154	.9598	1.2136
300	.8	.9606	.9921	.9416	.9911
400	.8	.9488	.7019	.9450	.9014
450	.8	.9588	.4963	.9440	.4959

Table 1.6. Continuation of Table 1.4.

n	ρ	*ccov*	*clen*	*accov*	*aclen*
50	.9	.9036	1.0989	.9464	1.1844
100	.9	.9486	1.0448	.9642	1.0876
200	.9	.9410	.9088	.9380	.9279
300	.9	.9412	.7433	.9432	.7538
400	.9	.9370	.5259	.9486	.5316
450	.9	.9494	.3722	.8896	.3758
50	.99	.4094	.2319	.9924	.7344
100	.99	.6738	.2806	.9476	.5621
200	.99	.9228	.2826	.9146	.4083
300	.99	.9164	.2416	.9200	.3089
400	.99	.9414	.1734	.6740	.2091
450	.99	.9176	.1229	.4120	.1455

Table 1.7. Continuation of Table 1.4.

CHAPTER 2

PLOTS FOR BINOMIAL REGRESSION

2.1 INTRODUCTION TO BINOMIAL REGRESSION

Regression models are used to study the conditional distribution $Y|\mathbf{x}$ given the $p \times 1$ vector of nontrivial predictors $\mathbf{x}$. In this chapter we will consider regression models for the binomial distribution. This section follows Olive (2007b) closely.

Definition 2.1.1. Let the sufficient predictor $SP = \alpha + \boldsymbol{\beta}^T\mathbf{x}$. The *binomial regression model* states that $Y_1, ..., Y_n$ are independent random variables with

$$Y_i \sim \mathrm{binomial}(\mathrm{m_i}, \rho(\alpha + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{\mathrm{i}})),$$

or

$$Y_i|SP_i \sim \mathrm{binomial}(\mathrm{m_i}, \rho(\mathrm{SP_i})).$$

The *binary regression model* is the special case where $m_i \equiv 1$ for $i = 1, ..., n$. (Olive 2007b)

The conditional mean function is $E(Y_i|SP_i) = m_i\rho(SP_i)$ and variance function is $V(Y_i|SP_i) = m_i\rho(SP_i)(1 - \rho(SP_i))$.

Definition 2.1.2. The *logistic regression (LR) model* is the special case of binomial regression where

$$P(\mathrm{success}|\mathbf{x}_{\mathrm{i}}) = \rho(\mathbf{x}_{\mathrm{i}}) = \tfrac{\exp(\alpha+\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{\mathrm{i}})}{1+\exp(\alpha+\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_{\mathrm{i}})}.$$

Equivalently,

$$\rho(SP) = \frac{\exp(SP)}{1 + \exp(SP)}.$$

(Olive 2007b)

The binary logistic regression model is important since for many data sets the response variable takes on two values: 0 or 1. The occurrence of an event is labelled as a 1 or a "success," while the nonoccurrence of an event is labelled as a 0 or a "failure." For binary data, if $P(Y = 1) = \rho$ then $Y \sim$ binomial(1,ρ). Hence if the Y_i are independent with $P(Y = 1|SP) = \rho(SP) = 1 - P(Y = 0|SP)$, then a binary regression model holds.

For the nonbinary case it is more difficult to check if the regression model holds because there are other distributions that are appropriate for data that takes on values $0, 1, ..., m$ if $m \geq 2$. Often the LR mean function is a good approximation to the data, the LR MLE is a consistent estimator of $\boldsymbol{\beta}$, but the LR model is not appropriate. The problem is that for many data sets where $E(Y_i|\mathbf{x}_i) = m_i\rho(SP_i)$, it turns out that $V(Y_i|\mathbf{x}_i) > m_i\rho(SP_i)(1 - \rho(SP_i))$. This phenomenon is called *overdispersion*.

The beta–binomial regression (BBR) model can be used as an alternative to the LR model. Let $\delta = \rho/\theta$ and $\nu = (1 - \rho)/\theta$, so $\rho = \delta/(\delta + \nu)$ and $\theta = 1/(\delta + \nu)$. Let

$$B(\delta, \nu) = \frac{\Gamma(\delta)\Gamma(\nu)}{\Gamma(\delta + \nu)}.$$

If Y has a beta–binomial distribution, $Y \sim \text{BB}(\text{m}, \rho, \theta)$, then the probability mass function of Y is

$$P(Y = y) = \binom{m}{y} \frac{B(\delta + y, \nu + m - y)}{B(\delta, \nu)}$$

for $y = 0, 1, 2, ..., m$ where $0 < \rho < 1$ and $\theta > 0$. Then $\delta > 0$ and $\nu > 0$. Then

$E(Y) = m\delta/(\delta+\nu) = m\rho$ and $\mathrm{V}(Y) = m\rho(1-\rho)[1+(m-1)\theta/(1+\theta)]$. If $Y|\pi \sim$ binomial(m, π) and $\pi \sim$ beta(δ, ν), then $Y \sim \mathrm{BB}(\mathrm{m}, \rho, \theta)$.

Definition 2.1.3. The BBR model states that $Y_1, ..., Y_n$ are independent random variables where $Y_i|SP_i \sim \mathrm{BB}(\mathrm{m_i}, \rho(\mathrm{SP_i}), \theta)$.

For the BBR model the conditional mean function is $E(Y_i|SP_i) = m_i\rho(SP_i)$ and the conditional variance function is $V(Y_i|SP_i) = m_i\rho(SP_i)(1-\rho(SP_i))[1+(m_i-1)\theta/(1+\theta)]$.

The BBR model has the same mean function as the binomial regression model, but allows for overdispersion. As $\theta \to 0$, it can be shown that $V(\pi) \to 0$ and the BBR model converges to the binomial regression model.

2.2 THE ESS PLOT AND THE OD PLOT

A useful plot to visualize the conditional distribution $Y|\mathbf{x}$ of the LR binary regression model is the estimated sufficient summary plot or ESS plot of the estimated sufficient predictor ESP $= \hat{\alpha} + \hat{\boldsymbol{\beta}}^{\boldsymbol{T}}\mathbf{x}$ versus Y with the estimated mean function

$$\hat{\rho}(ESP) = \frac{\exp(ESP)}{1+\exp(ESP)}$$

added as a visual aid. Since binomial regression is the study of $Y|\mathbf{x}$, the ESS plot is very important for analyzing LR models.

The ESS plot can be used to assess the adequacy of the binary LR model. Suppose that both the number of 0s and the number of 1s is large compared to the number of predictors p, that the ESP takes on many values, and that the binary LR model is a good approximation to the data. Then $Y|ESP \approx$ Binomial(1, $\hat{\rho}(ESP)$). If $-5 < ESP < 5$ then the estimated mean function has the characteristic "ESS" shape of the logistic curve.

This plot is useful as a goodness of fit diagnostic. Divide the ESP into J "slices" each containing approximately n/J cases. Compute the sample mean = sample proportion of the Y's in each slice and add the resulting step function to the ESS plot. This step function is a simple nonparametric estimator of the mean function $\rho(SP)$. If the step function follows the estimated LR mean function (the logistic curve) closely, then the LR model fits the data well. The lowess curve is a nonparametric estimator of the mean function called a "scatterplot smoother." The lowess curve may be more useful than the step function if the ESP does not take on many values.

For both the LR and BBR models with

$$\rho(SP) = \frac{\exp(SP)}{1 + \exp(SP)},$$

the conditional distribution of $Y|\mathbf{x}$ can be visualized with an ESS plot of the ESP versus Y_i/m_i with the logistic curve $\hat{\rho}(ESP)$ added as a visual aid.

Using graphical diagnostics to check the goodness of fit of the LR model would be useful since the binomial regression model is simpler than the BBR model. To check for overdispersion, the OD plot of $\hat{V}(Y|SP)$ versus $\hat{V} = [Y - \hat{E}(Y|SP)]^2$ should be used.

Using both the ESS plot and the OD plot we can assess the adequacy of the LR model. The ESS plot is used to visualize the conditional distribution $Y|\mathbf{x}$. The plotted points should follow the estimated parametric mean function $\hat{\rho}(ESP)$. If the lowess curve follows the logistic curve closely, then the LR mean function may be a useful approximation for $E(Y|\mathbf{x})$. The OD plot is used to check the variance function.

Recall that if a count Y is not too small, then a normal approximation is good for the binomial distribution. Notice that if $Y_i = E(Y|SP) + 2\sqrt{V(Y|SP)}$, then $[Y_i - E(Y|SP)]^2 = 4V(Y|SP)$. Then if both the estimated mean and estimated variance functions are good approximations, the plotted points in the OD plot will scatter about a wedge formed by the $\hat{V} = 0$ line and the line through the origin with slope 4: $\hat{V} = 4\hat{V}(Y|SP)$. Only about 5% of the plotted points should be above this line. The evidence of overdispersion increases as the scale of the vertical axis increases from 4 to 10 times the scale of the horizontal axis. If the scale of the vertical axis is more than 10 times that of the horizontal then there is evidence of overdispersion.

If the binomial LR OD plot is used but the data follows a beta–binomial regression model, then $\hat{V}_{mod} = \hat{V}(Y_i|ESP) \approx m_i\rho(ESP)(1 - \rho(ESP))$ while $\hat{V} = [Y_i - m_i\rho(ESP)]^2 \approx (Y_i - E(Y_i))^2$. Hence $E(\hat{V}) \approx V(Y_i) \approx m_i\rho(ESP)(1 - \rho(ESP))[1 + (m_i - 1)\theta/(1 + \theta)]$, so the plotted points with $m_i = m$ should scatter about a line with slope $\approx$

$$1 + (m-1)\frac{\theta}{1+\theta} = \frac{1+m\theta}{1+\theta}.$$

Numerical summaries are also available. The deviance G^2 is a statistic used to assess the goodness of fit of the logistic regression model much as R^2 is used for multiple linear regression. If the ESS and OD plots look good and the deviance G^2 satisfies $G^2/(n-p-1) \approx 1$, then the LR model is likely useful. If $G^2 > (n-p-1) + 3\sqrt{n-p+1}$, then a more complicated count model may be needed.

The following three pages are examples of the ESS plot and OD plot for specific data sets. Explanation of each data set is provided with the plots.

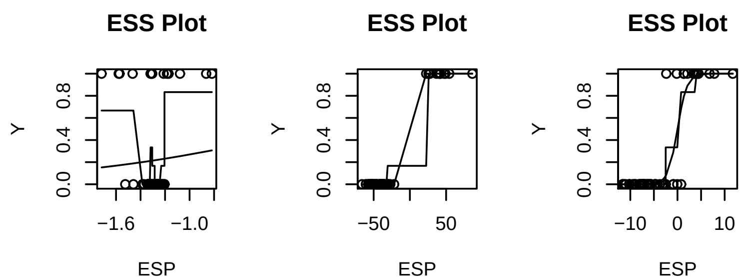

Figure 2.1. Plots for Museum Data

Example 1. Schaaffhausen (1878) gives data on skulls at a museum. The 1st 47 skulls are humans while the remaining 13 are apes. The response variable *ape* is 1 for an ape skull. The left plot in Figure 2.1 uses the predictor *face length.* The model fits very poorly since the probability of a 1 decreases then increases. The middle plot uses the predictor *head height* and perfectly classifies the data since the ape skulls can be separated from the human skulls with a vertical line as ESP = 0. The right plot uses predictors *lower jaw length, face length,* and *upper jaw length.* None of the predictors is good individually, but together provide a good LR model since the observed proportions (the step function) track the model proportions (logistic curve) closely.

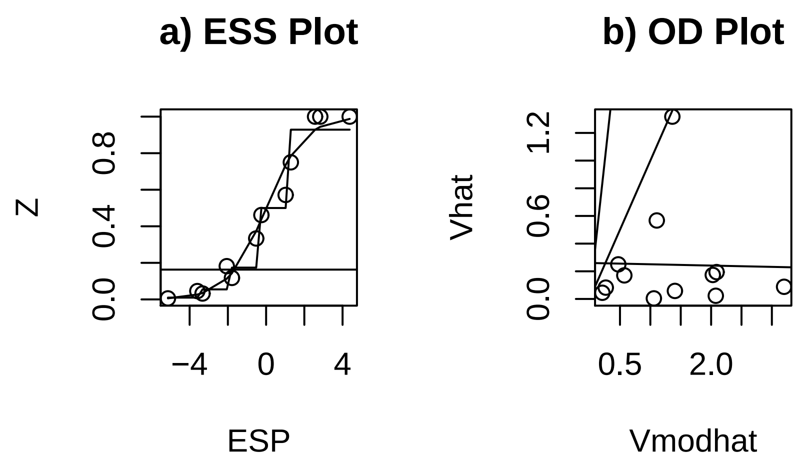

Figure 2.2. Plots for Death Penalty Data

Example 2. Abraham and Ledolter (2006) describe death penalty sentencing in Georgia. The predictors are *aggravation level* from 1 to 6 (treated as a continuous variable) and *race of victim* coded as 1 for white and 0 for black. There were 362 jury decisions and 12 level–race combinations. The response variable was the number of death sentences in each combination. The ESS plot in Figure 2.2a shows that the Y_i/m_i are close to the estimated LR mean function (the logistic curve). The step function based on 5 slices also tracks the logistic curve well. The OD plot is shown in Figure 2.2b with the identity, slope 4 and OLS lines added as visual aids. The vertical scale is less than the horizontal scale and there is no evidence of overdispersion.

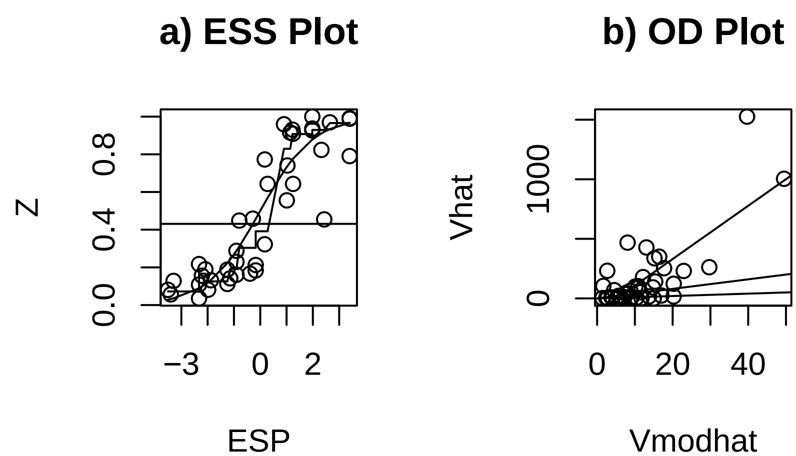

Figure 2.3. Plots for Rotifer Data

Example 3. Collett (1999) describes a data set where the response variable is the number of rotifers that remain in suspension in a tube. A rotifer is a microscopic invertebrate. The two predictors were the *density* of a stock solution of Ficolli and the *species* of rotifer coded as 1 for polyarthra major and 0 for keratella cochlearis. Figure 2.3a shows the ESS plot. Both the observed proportions and the step function track the logistic curve well, suggesting that the LR mean function is a good approximation to the data. The OD plot suggests that there is overdispersion since the vertical scale is about 30 times the horizontal scale. The OLS line has slope much larger than 4 and two outliers seem to be present.

2.3 SIMULATION OF BINOMIAL AND BETA-BINOMIAL REGRESSION DATA

Computer simulation was used to generate binomial and beta-binomial regression data to check for overdispersion. For type 1 a binomial distribution was used and for type 2 a beta-binomial distribution was used. For n = 50, 100, 200, 300, 400, and 500 the number of times $\hat{V}/\hat{V}(Y|SP) \geq 10$ was counted. This is labeled mr in the following tables. For the same values of n the number of times the deviance $G^2 > n - q - 1 + 3\sqrt{(}n - q - 1)$ was counted. This is labeled as dr in the following tables. The simulation used nruns = 1000, so mr and dr are listed as percentages out of 1000 runs.

We can make the following conclusions from the tables:

1. For the binomial distribution $mr < 0.06$ for all values of n. Also, $G^2 = 0$ for all values of n. The values of mr and G^2 that were obtained from simulation suggest the LR model holds.

2. For the beta-binomial distribution $mr > 0.06$ for all values of n. When $G^2 > 0.8$ then values of mr and dr suggest that the LR model does not hold.

n	$type$	θ	mr	dr
50	1	1	.001	0
100	1	1	.01	0
200	1	1	.025	0
300	1	1	.036	0
400	1	1	.047	0
500	1	1	.059	0

Table 2.1. Results for overdispersion using the binomial distribution.

n	*type*	θ	*mr*	*dr*
50	2	.1	.063	.163
50	2	.2	.247	.444
50	2	.3	.398	.632
50	2	.4	.531	.722
50	2	.5	.599	.773
50	2	.6	.697	.804
50	2	.7	.727	.84
50	2	.8	.762	.858
50	2	.9	.806	.872
50	2	1	.808	.882
100	2	.1	.205	.27
100	2	.2	.499	.682
100	2	.3	.77	.898
100	2	.4	.875	.945
100	2	.5	.939	.977
100	2	.6	.963	.98
100	2	.7	.98	.99
100	2	.8	.979	.99
100	2	.9	.984	.989
100	2	1	.993	.994

Table 2.2. Results for overdispersion using the beta-binomial distribution.

n	$type$	θ	mr	dr
200	2	.1	.449	.433
200	2	.2	.869	.925
200	2	.3	.964	.994
200	2	.4	.99	1
200	2	.5	.999	1
200	2	.6	.997	1
200	2	.7	1	1
200	2	.8	1	1
200	2	.9	1	1
200	2	1	1	1
300	2	.1	.61	.581
300	2	.2	.956	.98
300	2	.3	.998	.999
300	2	.4	1	1
300	2	.5	1	1
300	2	.6	1	1
300	2	.7	1	1
300	2	.8	1	1
300	2	.9	1	1
300	2	1	1	1

Table 2.3. Continuation of Table 2.2

n	$type$	θ	mr	dr
400	2	.1	.741	.651
400	2	.2	.984	.996
400	2	.3	1	1
400	2	.4	1	1
400	2	.5	1	1
400	2	.6	1	1
400	2	.7	1	1
400	2	.8	1	1
400	2	.9	1	1
400	2	1	1	1
500	2	.1	.84	.754
500	2	.2	.996	.999
500	2	.3	.999	1
500	2	.4	1	1
500	2	.5	1	1
500	2	.6	1	1
500	2	.7	1	1
500	2	.8	1	1
500	2	.9	1	1
500	2	1	1	1

Table 2.4. Continuation of Table 2.2

CHAPTER 3

OLS TESTS FOR BINOMIAL REGRESSION DATA

3.1 THE OLS ESTIMATOR

In this chapter we will simulate binary regression data to find whether the OLS tests have correct p-values for large samples. But first we will give some important results concerning the OLS estimator. This section follows Chang and Olive (2006) closely.

Let

$$\mathrm{Cov}(\boldsymbol{x}) = \mathrm{E}[(\boldsymbol{x} - \mathrm{E}(\boldsymbol{x}))(\boldsymbol{x} - \mathrm{E}(\boldsymbol{x}))^{\mathrm{T}}] = \boldsymbol{\Sigma}_{\boldsymbol{x}}$$

and $\mathrm{Cov}(\boldsymbol{x}, Y) = E[(\boldsymbol{x} - E(\boldsymbol{x}))(Y - E(Y))] = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. Let the OLS estimator be $(\hat{\alpha}_{OLS}, \hat{\boldsymbol{\beta}}_{OLS})$. Then the population coefficients from an OLS regression of Y on $\boldsymbol{x}$ are

$$\alpha_{OLS} = E(Y) - \boldsymbol{\beta}_{OLS}^T E(\boldsymbol{x}) \text{ and } \boldsymbol{\beta}_{\mathrm{OLS}} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x}\mathrm{Y}}. \tag{3.1}$$

Let the data be $(Y_i, \boldsymbol{x}_i)$ for $i = 1, ..., n$. Let the $p \times 1$ vector $\boldsymbol{\eta} = (\alpha, \boldsymbol{\beta}^T)^T$, let $\boldsymbol{X}$ be the $n \times p$ OLS design matrix with ith row $(1, \boldsymbol{x}_i^T)$, and let $\boldsymbol{Y} = (Y_1, ..., Y_n)^T$. Then the OLS estimator $\hat{\boldsymbol{\eta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$. The sample covariance of $\boldsymbol{x}$ is

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T \text{ where the sample mean } \overline{\boldsymbol{x}} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \boldsymbol{x}_{\mathrm{i}}.$$

Similarly, define the sample covariance of $\boldsymbol{x}$ and Y to be

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y}) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i Y_i - \overline{\boldsymbol{x}}\ \overline{Y}.$$

Following Seber and Lee (2003, p. 106),

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{pmatrix} \frac{1}{n} + \overline{\boldsymbol{x}}^T\boldsymbol{D}^{-1}\overline{\boldsymbol{x}} & -\overline{\boldsymbol{x}}^T\boldsymbol{D}^{-1} \\ -\boldsymbol{D}^{-1}\overline{\boldsymbol{x}} & \boldsymbol{D}^{-1} \end{pmatrix}$$

where the $(p-1) \times (p-1)$ matrix

$$\boldsymbol{D}^{-1} = [(n-1)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}]^{-1} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}/(n-1). \tag{3.2}$$

The first result shows that $\hat{\boldsymbol{\eta}}$ is a consistent estimator of $\boldsymbol{\eta}$.

i) Suppose that $(Y_i, \boldsymbol{x}_i^T)^T$ are iid random vectors such that $\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}$ and $\boldsymbol{\Sigma}_{\boldsymbol{x}Y}$ exist. Then

$$\hat{\alpha}_{OLS} = \overline{Y} - \hat{\boldsymbol{\beta}}_{OLS}^T\overline{\boldsymbol{x}} \xrightarrow{D} \alpha_{OLS}$$

and

$$\hat{\boldsymbol{\beta}}_{OLS} = \frac{n}{n-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} \xrightarrow{D} \boldsymbol{\beta}_{OLS} \text{ as } \text{n} \to \infty.$$

The following results will be for 1D regression and some notation is needed. Many 1D regression models have an error e with

$$\sigma^2 = \text{Var(e)} = \text{E(e}^2). \tag{3.3}$$

Let $\hat{e}$ be the error residual for e. Let the population OLS residual

$$v = Y - \alpha_{OLS} - \boldsymbol{\beta}_{OLS}^T\boldsymbol{x} \tag{3.4}$$

with

$$\tau^2 = E[(Y - \alpha_{OLS} - \boldsymbol{\beta}_{OLS}^T\boldsymbol{x})^2] = E(v^2), \tag{3.5}$$

and let the OLS residual be

$$r = Y - \hat{\alpha}_{OLS} - \hat{\boldsymbol{\beta}}_{OLS}^T\boldsymbol{x}. \tag{3.6}$$

Typically the OLS residual r is not estimating the error e and $\tau^2 \neq \sigma^2$, but the following results show that the OLS residual is of great interest for 1D regression models.

Assume that a 1D model holds, $Y \perp\!\!\!\perp \boldsymbol{x}|\alpha + \boldsymbol{\beta}^T\boldsymbol{x}$, which is equivalent to $Y \perp\!\!\!\perp \boldsymbol{x}|\boldsymbol{\beta}^T\boldsymbol{x}$. Then under regularity conditions, results ii) – iv) below hold.

ii) Li and Duan (1989): $\boldsymbol{\beta}_{OLS} = c\boldsymbol{\beta}$ for some constant c.

iii) Li and Duan (1989) and Chen and Li (1998):

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{OLS} - c\boldsymbol{\beta}) \xrightarrow{D} N_{p-1}(\mathbf{0}, \boldsymbol{C}_{OLS}) \tag{3.7}$$

where

$$\boldsymbol{C}_{OLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} E[(Y - \alpha_{OLS} - \boldsymbol{\beta}_{OLS}^T\boldsymbol{x})^2(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{x} - E(\boldsymbol{x}))^T]\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}. \tag{3.8}$$

iv) Chen and Li (1998): Let $\boldsymbol{A}$ be a known full rank constant $k \times (p-1)$ matrix. If the null hypothesis Ho: $\boldsymbol{A\beta} = \mathbf{0}$ is true, then

$$\sqrt{n}(\boldsymbol{A}\hat{\boldsymbol{\beta}}_{OLS} - c\boldsymbol{A\beta}) = \sqrt{n}\boldsymbol{A}\hat{\boldsymbol{\beta}}_{OLS} \xrightarrow{D} N_k(\mathbf{0}, \boldsymbol{AC}_{OLS}\boldsymbol{A}^T)$$

and

$$\boldsymbol{AC}_{OLS}\boldsymbol{A}^T = \tau^2\boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{A}^T. \tag{3.9}$$

Notice that $\boldsymbol{C}_{OLS} = \tau^2\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}$ if $v = Y - \alpha_{OLS} - \boldsymbol{\beta}_{OLS}^T\boldsymbol{x} \perp\!\!\!\perp \boldsymbol{x}$ or if the MLR model holds. If the MLR model holds, $\tau^2 = \sigma^2$.

To create test statistics, the estimator

$$\hat{\tau}^2 = \mathrm{MSE} = \frac{1}{\mathrm{n-p}}\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{r}_\mathrm{i}^2 = \frac{1}{\mathrm{n-p}}\sum_{\mathrm{i=1}}^{\mathrm{n}}(\mathrm{Y_i} - \hat{\alpha}_{\mathrm{OLS}} - \hat{\boldsymbol{\beta}}_{\mathrm{OLS}}^{\mathrm{T}}\boldsymbol{x}_\mathrm{i})^2$$

will be useful. The estimator $\hat{\boldsymbol{C}}_{OLS} =$

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1} \left[\frac{1}{n} \sum_{i=1}^{n} [(Y_i - \hat{\alpha}_{OLS} - \hat{\boldsymbol{\beta}}_{OLS}^T \boldsymbol{x}_i)^2 (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T] \right] \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1} \tag{3.10}$$

can also be useful. Notice that for general 1D regression models, the OLS MSE estimates τ^2 rather than the error variance σ^2.

v) Result iv) suggests that a test statistic for $Ho : \boldsymbol{A\beta} = \boldsymbol{0}$ is

$$W_{OLS} = n \hat{\boldsymbol{\beta}}_{OLS}^T \boldsymbol{A}^T [\boldsymbol{A} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1} \boldsymbol{A}^T]^{-1} \boldsymbol{A} \hat{\boldsymbol{\beta}}_{OLS} / \hat{\tau}^2 \xrightarrow{D} \chi_k^2, \tag{3.11}$$

the chi–square distribution with k degrees of freedom.

Before presenting the main theoretical result, some results from OLS MLR theory are needed. Let the $p \times 1$ vector $\boldsymbol{\eta} = (\alpha, \boldsymbol{\beta}^T)^T$, the known $k \times p$ constant matrix $\tilde{\boldsymbol{A}} = [\boldsymbol{a} \ \boldsymbol{A}]$ where $\boldsymbol{a}$ is a $k \times 1$ vector, and let $\boldsymbol{c}$ be a known $k \times 1$ constant vector. Following Seber and Lee (2003), the usual F statistic for testing $Ho : \tilde{\boldsymbol{A}}\boldsymbol{\eta} = \boldsymbol{c}$ is

$$F_0 = \frac{(SSE(H) - SSE)/k}{SSE/(n-p)} = \tag{3.12}$$

$$(\tilde{\boldsymbol{A}}\hat{\boldsymbol{\eta}} - \boldsymbol{c})^T [\tilde{\boldsymbol{A}} (\boldsymbol{X}^T \boldsymbol{X})^{-1} \tilde{\boldsymbol{A}}^T]^{-1} (\tilde{\boldsymbol{A}}\hat{\boldsymbol{\eta}} - \boldsymbol{c}) / (k \hat{\tau}^2)$$

where $MSE = \hat{\tau}^2 = SSE/(n-p)$, $SSE = \sum_{i=1}^n r_i^2$ and

$$SSE(H) = \sum_{i=1}^{n} r_i^2(H)$$

is the minimum sum of squared residuals subject to the constraint $\tilde{\boldsymbol{A}}\boldsymbol{\eta} = \boldsymbol{c}$. Recall that if Ho is true, the MLR model holds and the errors e_i are iid $N(0, \sigma^2)$, then $F_o \sim F_{k,n-p}$, the F distribution with k and $n-p$ degrees of freedom. Also recall that if $Z_n \sim F_{k,n-p}$, then

$$Z_n \xrightarrow{D} \chi_k^2 / k \tag{3.13}$$

as $n \to \infty$.

Theorem 3.1.1 and (3.13) suggest that OLS output, originally meant for testing with the MLR model, can also be used for testing with many 1D regression data sets. Without loss of generality, let the 1D model $Y \perp\!\!\!\perp \boldsymbol{x}|\alpha + \boldsymbol{\beta}^T \boldsymbol{x}$ be written as

$$Y \perp\!\!\!\perp \boldsymbol{x}|\alpha + \boldsymbol{\beta}_R^T \boldsymbol{x}_R + \boldsymbol{\beta}_O^T \boldsymbol{x}_O$$

where the reduced model is $Y \perp\!\!\!\perp \boldsymbol{x}|\alpha + \boldsymbol{\beta}_R^T \boldsymbol{x}_R$ and $\boldsymbol{x}_O$ denotes the terms outside of the reduced model. Notice that OLS ANOVA F test corresponds to Ho: $\boldsymbol{\beta} = \boldsymbol{0}$ and uses $\boldsymbol{A} = \boldsymbol{I}_{p-1}$. The tests for Ho: $\beta_i = 0$ use $\boldsymbol{A} = (0, ..., 0, 1, 0, ..., 0)$ where the 1 is in the ith position and are equivalent to the OLS t tests. The test Ho: $\boldsymbol{\beta}_O = \boldsymbol{0}$ uses $\boldsymbol{A} = [\boldsymbol{0}\ \boldsymbol{I}_j]$ if $\boldsymbol{\beta}_O$ is a $j \times 1$ vector, and the test statistic (3.12) can be computed by running OLS on the full model to obtain SSE and on the reduced model to obtain $SSE(R) \equiv SSE(H)$.

In the theorem below, it is crucial that Ho: $\boldsymbol{A\beta} = \boldsymbol{0}$. Tests for Ho: $\boldsymbol{A\beta} = \boldsymbol{1}$, say, may not be valid even if the sample size n is large. Also, confidence intervals corresponding to the t tests are for $c\boldsymbol{\beta}_i$, and are usually not very useful when c is unknown.

Theorem 3.1.1. *Assume that a 1D regression model holds and that Equation (3.11) holds when $Ho: \boldsymbol{A\beta} = \boldsymbol{0}$ is true. Then the test statistic (3.12) satisfies*

$$F_0 = \frac{n-1}{kn} W_{OLS} \xrightarrow{D} \chi_k^2 / k$$

as $n \to \infty$.

Proof. See Olive (2007c). □

3.2 SIMULATION OF THE OLS F STATISTIC

In this section, simulation is used to generate the OLS F statistic for binary logistic regression.

For atype = 1 the partial F test is used. That is, we test $Ho : \beta_i = 0, i = q/2, ..., q$ where q is the number of predictors.

For atype = 2 the t test is used. That is, we test $Ho : \beta_q = 0$.

For atype = 3 we test $Ho : \boldsymbol{\beta} = \mathbf{0}$.

For each atype and each value of n, nruns = 1000. For each table folscov is the proportion of 1000 runs where $F_{OLS} > F_{(0.95, dfNum, dfDenom)}$.

In Table 3.1, where atype = 1, $F_{OLS} > F_{(0.95, q/2, n-q-1)}$.

In Table 3.2, where atype = 2, $F_{OLS} > F_{(0.95, 1, n-q-1)}$.

In Table 3.3, where atype = 3, $F_{OLS} > F_{(0.95, q, n-q-1)}$

n	$atype$	$folscov$
10	1	NA
50	1	0.04
100	1	0.054
200	1	0.054
300	1	0.054
400	1	0.044
500	1	0.051
600	1	0.062
700	1	0.056
800	1	0.051
900	1	0.047
1000	1	0.047
2000	1	0.051
3000	1	0.049
4000	1	0.047
5000	1	0.051

Table 3.1. Results for simulation of *folscov* when $atype = 1$.

n	$type$	$folscov$
10	2	0.044
50	2	0.045
100	2	0.046
200	2	0.047
300	2	0.042
400	2	0.052
500	2	0.049
600	2	0.045
700	2	0.064
800	2	0.04
900	2	0.062
1000	2	0.064
2000	2	0.04
3000	2	0.052
4000	2	0.057
5000	2	0.054

Table 3.2. Results for simulation of *folscov* when $atype = 2$.

n	*type*	*folscov*
10	3	NA
50	3	0.035
100	3	0.038
200	3	0.056
300	3	0.042
400	3	0.05
500	3	0.041
600	3	0.034
700	3	0.053
800	3	0.042
900	3	0.055
1000	3	0.065
2000	3	0.054
3000	3	0.057
4000	3	0.043
5000	3	0.06

Table 3.3. Results for simulation of *folscov* when $atype = 3$.

Conclusions from Tables 3.1, 3.2, and 3.3.

1. For atype = 1, folscov is around 0.05.

2. For atype = 2, folscov is around 0.05.

3. For atype = 3, folscov is around 0.05.

We can conclude that the OLS p-values are approximately correct for some binary regression models.

REFERENCES

[1] Abraham, B., and Ledolter, J. (2006), *Introduction to Regression Modeling*, Thomson Brooks/Cole, Belmont, CA.

[2] Chang, J. and Olive, D.J. (2006), "Resistant Dimension Reduction," Preprint.

[3] Chen, C.H. and Li, K.C. (1998), "Can SIR be as Popular as Multiple Linear Regression?," Statistica Sinica, 8, 289–316.

[4] Collett, D. (1999), *Modelling Binary Data,* Chapman & Hall/CRC, Boca Raton, Florida.

[5] Li, K.C. and Duan, N. (1989), "Slicing Regression: A link-free regression method," *The Annals of Statistics*, 19, 505–530.

[6] Olive, D.J. (2007a), *A Course in Statistical Inference*, for preprint, see http://www.math.siu.edu/olive/infbook.htm.

[7] Olive, D.J. (2007b), "Plots for Binomial Regression," for preprint, see http://www.math.siu.edu/olive/ppbreg.pdf.

[8] Olive, D.J. (2007c), *Applied Robust Statistics*, for preprint, see http://www.math.siu.edu/olive/ol-bookp.htm.

[9] Schaaffhausen, H. (1878), "Die Anthropologische Sammlung Des Anatomischen Der Universitat Bonn," *Archiv fur Anthropologie,* 10, 1-65, Appendix.

[10] Seber, G.A.F. and Lee, A.J. (2003), *Linear Regression Analysis*, Wiley, NY.

APPENDICES

VITA

Graduate School
Southern Illinois University

JESSICA BAYER Date of Birth: July 25, 1983

77 TONEY-CORT LANE APT 12, MURPHYSBORO, ILLINOIS 62966

11415 BAYER LANE, EQUALITY, ILLINOIS, 62934

Southern Illinois University at Carbondale
Bachelor of Science, Mathematics, August 2005

Special Honors and Awards:

June Rice Malan Math Award, April 2003

John Olmstead Outstanding Teaching Assistant Award, April 2007

Research Paper Title:
BINOMIAL CONFIDENCE INTERVALS AND DIAGNOSTICS FOR BINOMIAL REGRESSION

Major Professor: Dr. David Olive