STA 6384, Report 6.2

Carson Slater *Baylor University*

This report reproduces the analysis from Section 6.3, which examines logistic regression with multiple binary covariates. The example is based on a convenience sample of patients visiting a clinic for evaluation of coronary artery (CA) disease. The researchers were interested in whether electrocardiogram (ECG) measurements, specifically ST segment depression levels, were associated with disease status, while controlling for gender as a potential confounding variable.

The ST segment is a portion of the electrocardiogram trace, and a "depressed" ST segment may indicate coronary ischemia—a condition where there is insufficient blood flow in the coronary arteries. This analysis investigates how both gender and ECG ST segment depression levels jointly predict the presence of coronary artery disease.

Data Description and Setup

The dataset contains 78 patients with 5 variables. The binary response variable is the presence (1) or absence (0) of coronary artery disease. Two binary predictor variables are included:

- x_1 **(Gender)**: Coded as 0 for female patients and 1 for male patients
- x_2 **(ECG ST segment depression)**: Coded as 0 for depression < 0.1 mm and 1 for depression $\geq$ 0.1 mm

Exploratory Data Analysis

Table 6.2: Coronary Artery Disease Data

Table 1: Coronary artery disease data

Sex	ECG	No Disease	Disease	Total
Female	< 0.1 ST segment depression	11	4	15
Female	$\geq$ 0.1 ST segment depression	10	8	18
Male	< 0.1 ST segment depression	9	9	18
Male	$\geq$ 0.1 ST segment depression	6	21	27

Table 6.3: Gender by Disease

Table 2: Gender by disease status

Gender	No Disease	Disease	Total
Female	21 (63.64%)	12 (36.36%)	33
Male	15 (33.33%)	30 (66.67%)	45
Total	36	42	78

Chi-square Test and Odds Ratios

Analyzing the basic 2×2 table using Pearson's Chi-square test reveals a clear association between gender and disease status. The test statistic is χ^2 = 5.8681 with a p-value of 0.0154. Male patients with symptoms are far more likely to have a diagnosis of CA disease than females, indicating that gender is a confounding variable that should be controlled for in the analysis.

Examining the ECG-disease association within each gender:

- **For females**: The odds ratio is estimated to be 2.2, indicating that women with higher ST segment depression levels had 2.2 times the odds of CA disease than those with lower levels.
- **For males**: The odds ratio is estimated to be 3.5, indicating that men with higher ST segment depression levels had 3.5 times the odds of CA disease than those with lower levels.

Logistic Regression Model

Model Specification

The logistic regression model we fit is:

$$\text{logit}(\pi_i) = \alpha + \beta_1 x_1 + \beta_2 x_2$$

where: - π_i is the probability of coronary artery disease for patient i - x_1 is the indicator for sex (female = 0, male = 1)
- x_2 is the indicator for ECG (< 0.1 ST depression = 0, $\geq$ 0.1 ST depression = 1)

Model Results

Table 3: Logistic regression parameter estimates

	Parameter	Estimate	SE	z value	Pr(>|z|)	Interpretation
(Intercept)	α (Intercept)	-1.1747	0.485	-2.420	0.0155	Log odds of CA for females with ECG < 0.1
x1	β_1 (Male vs Female)	1.2770	0.498	2.564	0.0103	Increment to log odds for males
x2	β_2 (ECG $\geq$ 0.1 vs < 0.1)	1.0545	0.498	2.118	0.0342	Increment to log odds for ECG $\geq$ 0.1

The fitted model equation is:

$$\text{logit}(\pi_i) = -1.1747 + 1.277x_1 + 1.0545x_2$$

Model Interpretation

Table 4: Model-predicted logits, odds, and probabilities by group

Group	Logit	Odds	Probability
Female, ECG < 0.1	-1.1747	0.3089	0.2360
Female, ECG $\geq$ 0.1	-0.1202	0.8868	0.4700
Male, ECG < 0.1	0.1023	1.1077	0.5255
Male, ECG $\geq$ 0.1	1.1568	3.1797	0.7607

Odds Ratios

The analysis reveals two key findings:

1. **Gender Effect**: The odds ratio for males compared to females is exp($\hat{\beta}_1$) = 3.586. This means that men have 3.6 times higher odds of CA disease compared to women, controlling for ECG status. The 95% Wald confidence interval is [1.379, 9.838].
2. **ECG Effect**: The odds ratio for ECG $\geq$ 0.1 versus ECG < 0.1 is exp($\hat{\beta}_2$) = 2.871. This indicates that subjects with higher ST segment depression have 2.9 times the odds of CA disease compared to those with lower ST segment depression, controlling for gender. The 95% Wald confidence interval is [1.1, 7.848].

Model Fit Assessment

Table 5: Model fit assessment

Fit Statistic	Value
AIC (intercept only)	109.6690
AIC (with covariates)	101.9000
Deviance (intercept only)	107.6690
Deviance (with covariates)	95.9000
Likelihood Ratio Test	11.7690
LR df	2.0000
LR p-value	0.0028

Both the AIC and deviance statistics indicate that the model with covariates provides a better fit than the intercept-only model. The likelihood ratio test comparing the two models yields a test statistic of 11.769 with 2 degrees of freedom and a p-value of 0.0028, indicating that the addition of the covariates significantly improves model fit.

Conclusion

This analysis successfully reproduces the results from Section 6.3, demonstrating the application of logistic regression with multiple binary covariates. The key findings are:

1. **Both predictors are statistically significant**: Gender and ECG ST segment depression levels both significantly predict coronary artery disease status ($p < 0.05$ for both).
2. **Substantial effect sizes**: Male gender increases the odds of disease by a factor of 3.6, while elevated ST segment depression increases the odds by a factor of 2.9.
3. **Model validity**: The likelihood ratio test confirms that including both covariates significantly improves model fit compared to the null model.
4. **Clinical relevance**: The findings suggest that both gender and ECG abnormalities are important independent risk factors for coronary artery disease, which aligns with established clinical knowledge.

The logistic regression framework effectively handles the binary nature of both the response and predictor variables, providing interpretable results in terms of odds ratios while controlling for potential confounding between gender and ECG status.