

# Assignment 03 Regressions

BQOM 2578 | Data Mining

Theresa Wohlever

Sunday, October 5, 2025

## Table of contents

<b>Executive Summary</b>	<b>1</b>
<b>Data Preparation</b>	<b>2</b>
Importing Data, Cleaning, & Wrangling . . . . .	2
Preliminary Analysis . . . . .	7
<b>Regression</b>	<b>9</b>
Linear Regression by R Value . . . . .	9
Stepwise Linear Regression . . . . .	11
Logistic Regression . . . . .	14
<b>Logistic Regression Model 2 (Multi-Variable) : Beta Coefficients Discussion</b>	<b>18</b>
Statistical Significance Assessment . . . . .	18
Key Insights . . . . .	19
Practical Implications . . . . .	20
<b>References</b>	<b>21</b>

## Executive Summary

How well can we predict county Suicide Rates from the hospital information on a per county basis within Pennsylvania? Combine both county Suicide rate data with all PA hospital data to address this question. Dependent Variable is the County Suicide rate. The Data preparation includes removing a large number of features expected to be unrelated to suicide rates, changing the representation of categorical variables to integers, and joining hospital data and county level suicide data.

There are Linear regression models to predict county suicide rates. These include single feature models as well as stepwise models. There are also Logistic regression models.

Single Feature Linear Regression performed for every Independent Variable (Feature) selected as relevant for this analysis. Of those, the highest  $R^2$  (best) single-predictor model uses the variable discharges1864 with  $R^2 = 0.06481609$

Of the three Stepwise Linear Regression models attempted, all three performed better than the Single Feature Models. Of these three, the backwards stepwise model predicted County Suicide Rates best with an Adjusted R-squared value of 0.1634. In all linear regression cases, these  $R^2$  values are quite low; even though the explanatory power for the backwards stepwise linear regression is the best of all linear regression models, it can only explain 16% of the variance in county suicide rates.

Logistic regression provides insight into the impact of included features. Selected features were those shown to demonstrate significant impact on the linear regression. Using these features in a Logistic Regression we can better perceive their impact on County Suicide Rates.

1. children\_hospital: Very large effect ( $\beta = 2.849$ )
2. psych\_over17: Small effect, but not significant ( $\beta = 0.184$ )
3. discharges1864: Very small but significant effect ( $\beta = 0.0001$ )
4. psych\_over17\_beds\_lic: Negligible effect ( $\beta = 0.001$ )

These findings are discussed in detail in [Logistic Regression Model 2 \(Multi-Variable\) : Beta Coefficients Discussion](#).

## Data Preparation

### Importing Data, Cleaning, & Wrangling

#### Data Review

Numeric variables are continuous.

Integer variables are categorical.

```
##  
## Variable TYPES  
##  
  
# sapply(df, typeof)  
sapply(df, class)
```

facility_id	type_of_organization
"integer"	"integer"
children_hospital	hospital_ltc
"numeric"	"numeric"
on_site_ltc	privateroomexist
"numeric"	"numeric"
semiprivateroomexist	discharges1864
"numeric"	"integer"
alcohol_drug_detox	alcoholdetox_patient_days
"integer"	"integer"
alcohol_drug_treat	alcoholtreat_beds_lic
"integer"	"integer"
alcoholtreat_patient_days	comprehensive_rehab
"integer"	"integer"
comprehensive_rehab_beds_lic	Comprehensive_rehab_patient_days
"integer"	"integer"
psych_0to17	psych_0to17_beds_lic
"integer"	"integer"
psych_0to17_patient_days	psych_over17
"integer"	"integer"
psych_over17_beds_lic	psych_over17_patient_days
"integer"	"integer"
detox	clinpsyc
"numeric"	"numeric"
clinic_psychiatric	psychiatrists
"integer"	"integer"
target	
"numeric"	

```
##
## Visualize target values
##

# Histogram of target values
target_density <- density(df$target)

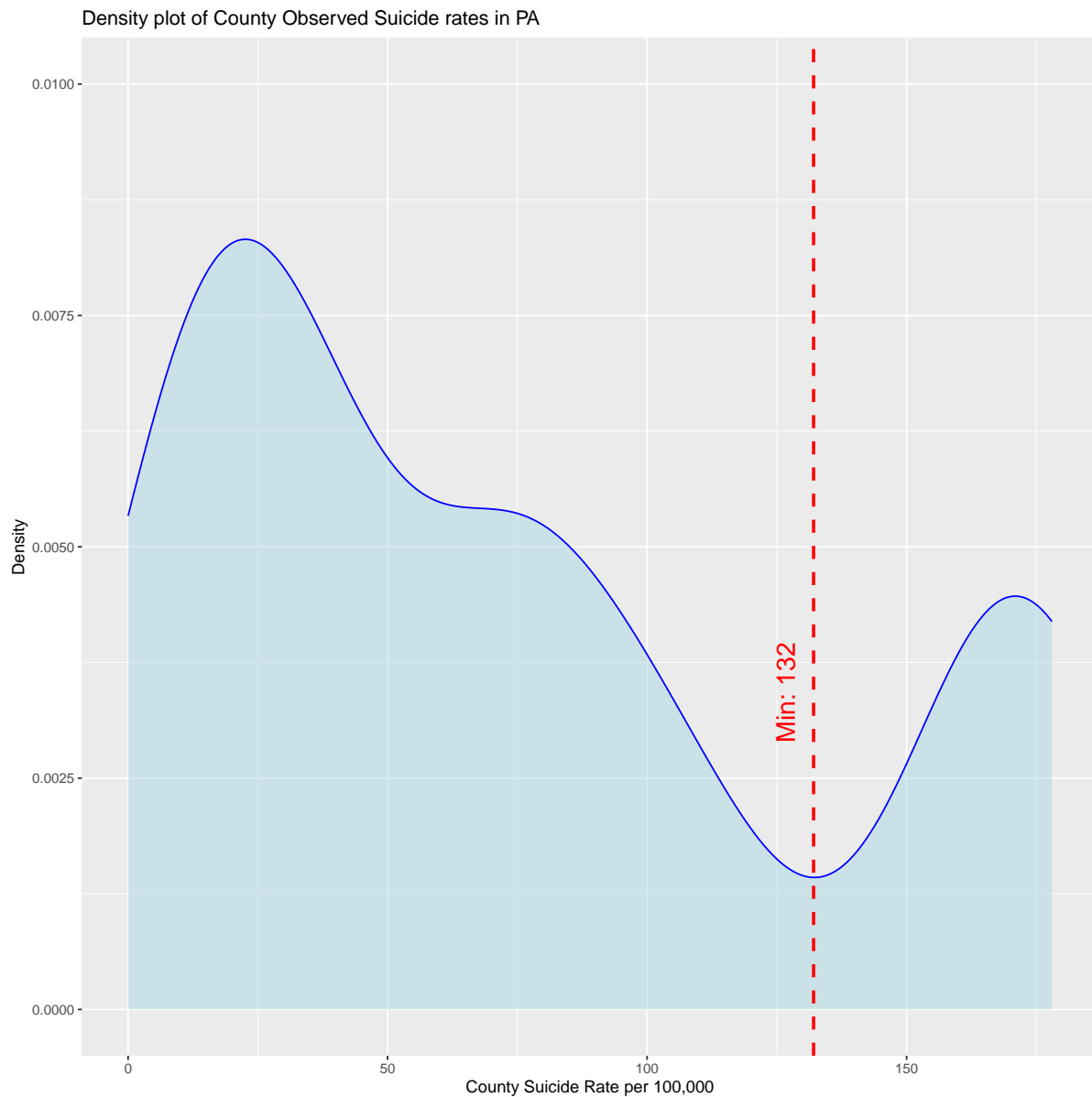
# Convert the density estimate to a function
dens_func <- approxfun(target_density$x, target_density$y)

# Use optimize() to find the minimum in a specified interval (choose based on your data)
result <- optimize(dens_func, interval = c(min(df$target), max(df$target)))
local_min_x <- result$minimum      # The x value where local minimum occurs
local_min_y <- result$objective    # The minimum density value
```

```

# Create density plot with ggplot2 and add vertical line at minimum
df_density <- data.frame(x = df$target)
ggplot(df_density, aes(x = df$target)) +
  geom_density(fill = "lightblue", color = "blue", alpha = 0.5) +
  geom_vline(xintercept = local_min_x, color = "red", linetype = "dashed", size = 1) +
  annotate("text", x = local_min_x, y = local_min_y + 0.002,
    label = sprintf("Min: %.0f", local_min_x), color = "red", angle = 90, vjust = -1, size = 6) +
  labs(title = "Density plot of County Observed Suicide rates in PA", x = "County Suicide Rate per 100,000", y = "Density") +
  coord_cartesian(ylim = c(0, 0.01))

```



```
## Make TARGET binary for logistic regression  
target_bin_cutoff <- local_min_x
```

Set the cut-off value for 1 or 0 (binary) for Logistic regression is the local minimum of county Suicide Rates.

```
##  
## Visualize ALL data across target values  
## Normalized Scatter
```

```
##
minMax <- function(x) { (x - min(x)) / (max(x) - min(x))}
df_norm <- as.data.frame(lapply(df, minMax))
df_long <- pivot_longer(df_norm, cols = -target, names_to = "variable", values_to = "value")
ggplot(df_long, aes(x = target, y = value, color = variable)) +
  geom_point(size = 2) +
  theme_minimal() +
  theme(legend.position = "bottom") +
  guides(color = guide_legend(nrow = 7, title = "Feature")) +
  labs(x = "", y = "", color = "Feature")
```



No clear clustering patterns across all variables vs. County Suicide Rates (Regression Target).

## Preliminary Analysis

Evaluate Correlation Matrix

```

## Prep for correlation
df_cor <- df

cor_mat <- cor(df)
cor_threshold <- 0
cor_threshold_count <- 2

cols_above_threshold <- which( colSums(abs(cor_mat) > cor_threshold, na.rm = TRUE) >= cor_threshold_count)
df <- subset(df, select = colnames(cor_mat)[cols_above_threshold] )
cor_mat <- cor(df)

cat(paste(colnames(cor_mat)[cols_above_threshold], collapse = "\n"))

```

```

facility_id
type_of_organization
children_hospital
hospital_ltc
alcohol_drug_treat
comprehensive_rehab
psych_0to17
psych_over17
detox
psychiatrists
NA
NA
NA
NA
NA
NA
NA
NA
NA

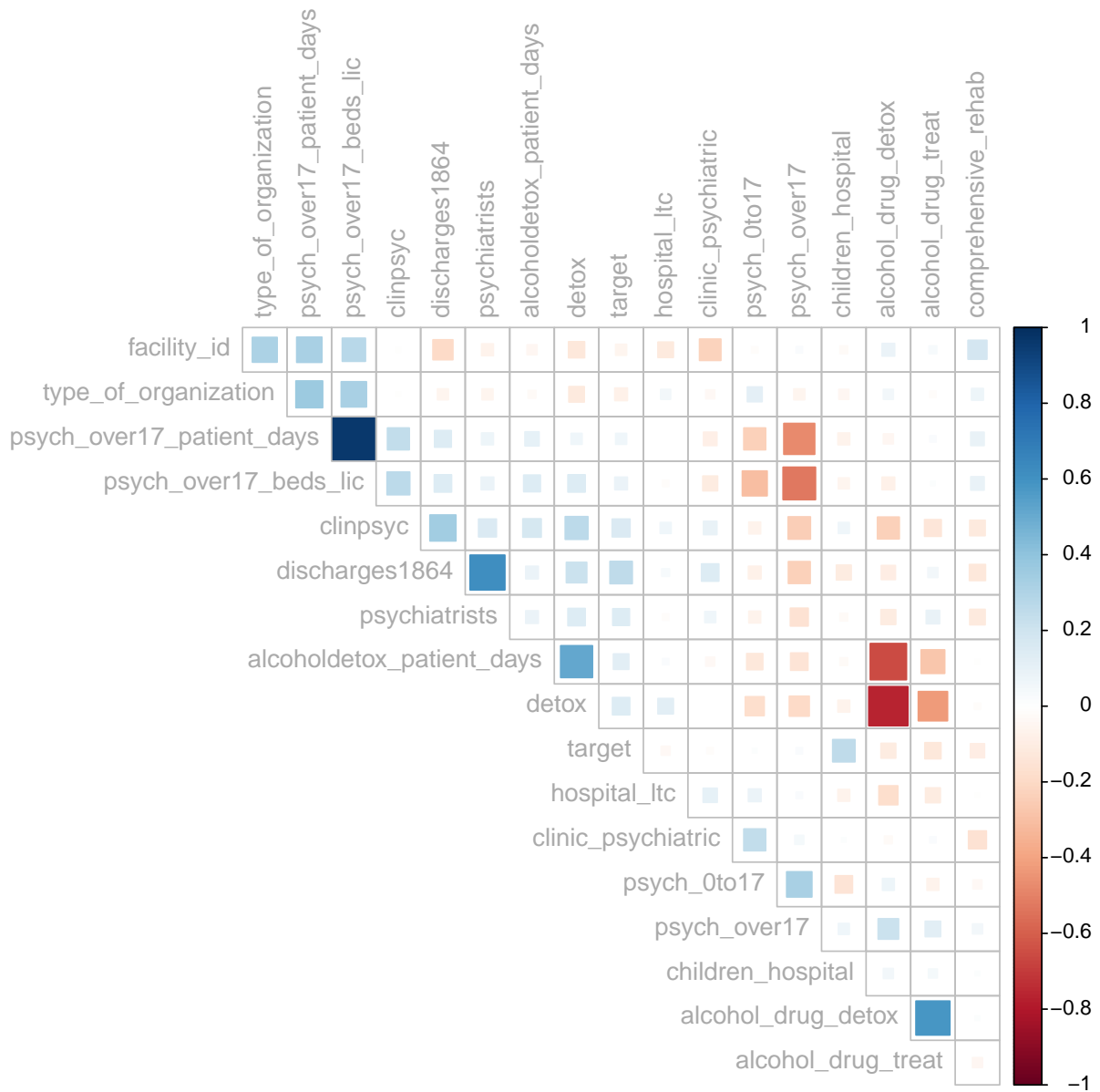
```

```

cor_mat_plot <- round(cor_mat, 2)
cor_mat_plot[is.na(cor_mat_plot)] <- 0 # Replace all NA values with zero
corrplot(cor_mat_plot,
  method="square",
  type="upper",
  order="AOE",
  tl.col="darkgrey",
  cl.align.text = "r",
  diag=FALSE,
  number.cex=0.6)

```





## Regression

### Linear Regression by R Value

Bake off by R squared

```

response <- "target"
predictors <- setdiff(names(df), response)

best_r2 <- -Inf
best_model <- NULL
best_predictor <- NULL

for (predictor in predictors) {
  formula <- as.formula(paste(response, "^", predictor))
  model <- lm(formula, data = df)
  r2 <- summary(model)$r.squared

  if (r2 > best_r2) {
    best_r2 <- r2
    best_model <- model
    best_predictor <- predictor
  }
}

cat("Best single-predictor model uses:", best_predictor, "with R^2 =", best_r2, "\n")

```

Best single-predictor model uses: discharges1864 with R<sup>2</sup> = 0.06481609

```

# To view details of best_model:
summary(best_model)

```

Call:

```
lm(formula = formula, data = df)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-109.26	-48.72	-15.47	32.77	114.25

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	6.374e+01	4.723e+00	13.494	< 2e-16 ***
discharges1864	3.279e-03	8.474e-04	3.869	0.000145 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 57.83 on 216 degrees of freedom  
Multiple R-squared: 0.06482, Adjusted R-squared: 0.06049  
F-statistic: 14.97 on 1 and 216 DF, p-value: 0.0001446

## Stepwise Linear Regression

```
# For more info: https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/step

model <- lm(target ~ ., data = df)
summary(model)
```

Call:

```
lm(formula = target ~ ., data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-98.80	-40.27	-14.39	26.85	132.00

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	9.837e+01	5.475e+01	1.797	0.073863 .
facility_id	5.033e-09	2.712e-07	0.019	0.985212
type_of_organization	-4.438e+00	3.035e+00	-1.462	0.145204
children_hospital	9.625e+01	2.053e+01	4.688	5.1e-06 ***
hospital_ltc	-8.800e+00	1.261e+01	-0.698	0.485998
discharges1864	4.268e-03	1.134e-03	3.763	0.000221 ***
alcohol_drug_detox	6.592e+00	1.572e+01	0.419	0.675419
alcoholdetox_patient_days	3.616e-03	3.944e-03	0.917	0.360344
alcohol_drug_treat	-2.762e+01	1.356e+01	-2.037	0.042938 *
comprehensive_rehab	-7.007e+00	4.397e+00	-1.594	0.112576
psych_0to17	1.157e+01	6.609e+00	1.750	0.081633 .
psych_over17	1.091e+01	4.858e+00	2.245	0.025842 *
psych_over17_beds_lic	3.940e-01	2.758e-01	1.429	0.154690
psych_over17_patient_days	-5.272e-04	8.691e-04	-0.607	0.544840
detox	8.600e+00	2.020e+01	0.426	0.670717
clinpsyc	-8.861e-01	8.850e+00	-0.100	0.920344
clinic_psychiatric	-5.928e+00	4.407e+00	-1.345	0.180073
psychiatrists	-6.321e-02	1.256e-01	-0.503	0.615476

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 54.57 on 200 degrees of freedom

Multiple R-squared: 0.229, Adjusted R-squared: 0.1634

F-statistic: 3.494 on 17 and 200 DF, p-value: 1.028e-05

```
# Perform stepwise regression
step_model_back <- step(model, direction = "backward", trace=0)
summary(step_model_back)
```

Call:

```
lm(formula = target ~ type_of_organization + children_hospital +
    discharges1864 + alcohol_drug_treat + comprehensive_rehab +
    psych_0to17 + psych_over17 + psych_over17_beds_lic + clinic_psychiatric,
    data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-106.02	-40.33	-10.98	28.88	134.42

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.295e+02	3.853e+01	3.361	0.000924 ***
type_of_organization	-5.079e+00	2.881e+00	-1.763	0.079397 .
children_hospital	9.551e+01	1.995e+01	4.788	3.20e-06 ***
discharges1864	4.001e-03	8.419e-04	4.753	3.74e-06 ***
alcohol_drug_treat	-2.928e+01	1.057e+01	-2.772	0.006083 **
comprehensive_rehab	-6.924e+00	4.281e+00	-1.617	0.107307
psych_0to17	9.837e+00	6.380e+00	1.542	0.124614
psych_over17	1.085e+01	4.695e+00	2.312	0.021772 *
psych_over17_beds_lic	2.496e-01	9.096e-02	2.744	0.006595 **
clinic_psychiatric	-6.337e+00	4.253e+00	-1.490	0.137781

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 53.87 on 208 degrees of freedom

Multiple R-squared: 0.2186, Adjusted R-squared: 0.1848

F-statistic: 6.466 on 9 and 208 DF, p-value: 4.224e-08

```
step_model_forward <- step(model, direction = "forward", trace=0)
summary(step_model_forward)
```

Call:

```
lm(formula = target ~ facility_id + type_of_organization + children_hospital +
    hospital_ltc + discharges1864 + alcohol_drug_detox + alcoholdetox_patient_days +
    alcohol_drug_treat + comprehensive_rehab + psych_0to17 +
    psych_over17 + psych_over17_beds_lic + psych_over17_patient_days +
    detox + clinpsyc + clinic_psychiatric + psychiatrists, data = df)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-98.80 -40.27 -14.39  26.85 132.00
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	9.837e+01	5.475e+01	1.797	0.073863 .
facility_id	5.033e-09	2.712e-07	0.019	0.985212
type_of_organization	-4.438e+00	3.035e+00	-1.462	0.145204
children_hospital	9.625e+01	2.053e+01	4.688	5.1e-06 ***
hospital_ltc	-8.800e+00	1.261e+01	-0.698	0.485998
discharges1864	4.268e-03	1.134e-03	3.763	0.000221 ***
alcohol_drug_detox	6.592e+00	1.572e+01	0.419	0.675419
alcoholdetox_patient_days	3.616e-03	3.944e-03	0.917	0.360344
alcohol_drug_treat	-2.762e+01	1.356e+01	-2.037	0.042938 *
comprehensive_rehab	-7.007e+00	4.397e+00	-1.594	0.112576
psych_0to17	1.157e+01	6.609e+00	1.750	0.081633 .
psych_over17	1.091e+01	4.858e+00	2.245	0.025842 *
psych_over17_beds_lic	3.940e-01	2.758e-01	1.429	0.154690
psych_over17_patient_days	-5.272e-04	8.691e-04	-0.607	0.544840
detox	8.600e+00	2.020e+01	0.426	0.670717
clinpsyc	-8.861e-01	8.850e+00	-0.100	0.920344
clinic_psychiatric	-5.928e+00	4.407e+00	-1.345	0.180073
psychiatrists	-6.321e-02	1.256e-01	-0.503	0.615476

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 54.57 on 200 degrees of freedom

Multiple R-squared: 0.229, Adjusted R-squared: 0.1634

F-statistic: 3.494 on 17 and 200 DF, p-value: 1.028e-05

## Logistic Regression

```
# Update target to be binomial
df$target <- ifelse(df$target < target_bin_cutoff, 0, 1)

# Logistic Regression with ONLY discharges1864
logreg_ch <- glm(target ~ discharges1864, data=df, family="binomial")
summary(logreg_ch)
```

Call:

```
glm(formula = target ~ discharges1864, family = "binomial", data = df)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.600e+00	2.084e-01	-7.677	1.63e-14 ***
discharges1864	9.391e-05	3.215e-05	2.921	0.00349 **

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 229.83 on 217 degrees of freedom  
Residual deviance: 221.17 on 216 degrees of freedom  
AIC: 225.17

Number of Fisher Scoring iterations: 4

```
# Logistic Regression with discharges1864 +
# children_hospital + psych_over17 + psych_over17_beds_lic
logreg_chdisch <- glm(target ~ discharges1864 + children_hospital + psych_over17
+ psych_over17_beds_lic, data=df, family="binomial")
summary(logreg_chdisch)
```

Call:

```
glm(formula = target ~ discharges1864 + children_hospital + psych_over17 +
    psych_over17_beds_lic, family = "binomial", data = df)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
--	----------	------------	---------	----------

```

(Intercept)      -2.271e+00  6.144e-01  -3.697 0.000218 ***
discharges1864    1.180e-04  3.465e-05   3.404 0.000663 ***
children_hospital  2.849e+00  8.489e-01   3.356 0.000790 ***
psych_over17      1.836e-01  2.194e-01   0.837 0.402656
psych_over17_beds_lic 1.335e-03  3.948e-03   0.338 0.735195

```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 229.83 on 217 degrees of freedom

Residual deviance: 206.84 on 213 degrees of freedom

AIC: 216.84

Number of Fisher Scoring iterations: 4

```

logreg <- logreg_chdisch
coeftable <- data.frame(col1=coef(logreg),col2=exp(coef(logreg)))
colnames(coeftable)<-c('Coefficient (log-odds)','e^coefficient (odds)')
coeftable

```

	Coefficient (log-odds)	e^coefficient (odds)
(Intercept)	-2.2712892545	0.1031791
discharges1864	0.0001179614	1.0001180
children_hospital	2.8490898212	17.2720540
psych_over17	0.1835961539	1.2015305
psych_over17_beds_lic	0.0013351614	1.0013361

```

#
# Confusion Matrix
#
df$PredLogOdds <- df$PredProbs <- predict(logreg, newdata=df)
df$PredProbs <- predict(logreg, newdata=df, type="response")
# type="response" gives the probability
summary(df$PredProbs)

```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.1110	0.1530	0.1606	0.2202	0.2018	0.8103

```

# transform prediction into either 1 (Elevated Suicide Rate) or 0 (Suicide Rate) using a cutoff point
for (cutoff in c(0.25, 0.15, 0.1, 0.05))

```

```

{
  df$PredHighSuicide <- ifelse(df$PredProbs >= cutoff,1,0)
  summary(df$PredHighSuicide)
  cat(paste("For a cutoff point of",
    cutoff, "the proportion of Counties predicted to have high suicide rates is",
    round(mean(df$PredHighSuicide),2)), "\n\n")
}

```

For a cutoff point of 0.25 the proportion of Counties predicted to have high suicide rates is 0.18

For a cutoff point of 0.15 the proportion of Counties predicted to have high suicide rates is 0.83

For a cutoff point of 0.1 the proportion of Counties predicted to have high suicide rates is 1

For a cutoff point of 0.05 the proportion of Counties predicted to have high suicide rates is 1

Of the two Logistic Regression models, **Model 2 (Multi-Variable)** demonstrates superior statistical and explanatory power.

### Logistic Regression Model Comparison

Metric	Model 1 (Single Variable)	Model 2 (Multi-Variable)
<b>Variables</b>	discharges1864 only	discharges1864 + children_hospital + psych_over17 + psych_over17_beds_lic
<b>AIC</b>	225.17	216.84
<b>Deviance</b>	3.77%	10.00%
<b>Ex-plained</b>		
<b>Significant</b>	1 out of 1	2 out of 4
<b>Predic-tors</b>		
<b>Residual Deviance</b>	221.17	206.84

### Model 2 (Multi-Variable) Demonstrates Superior Statistical Power

#### Lower AIC Score



Model 2 achieves an AIC of 216.84 compared to Model 1's 225.17, representing an **8.33-point improvement**. The AIC penalizes model complexity while rewarding goodness of fit, so this reduction indicates that the additional variables provide meaningful explanatory value that outweighs their complexity cost.

### **Greater Deviance Explained**

Model 2 explains **10.00% of the deviance** compared to Model 1's 3.77%. This represents a **2.65-fold increase** in explanatory power, indicating that the multi-variable model captures significantly more variance in county suicide rates.

### **Stronger Statistical Significance**

Model 2 includes two highly significant predictors ( $p < 0.001$ ): - **discharges1864**:  $z = 3.404$ ,  $p = 0.000663$  - **children\_hospital**:  $z = 3.356$ ,  $p = 0.000790$

The children\_hospital variable emerges as particularly powerful with an **odds ratio of 17.27**, meaning counties with children's hospitals have over 17 times higher odds of being classified as high suicide rate counties.

### **Model 2's Key Strengths**

#### **Clinical Relevance**

The model incorporates healthcare infrastructure variables that have logical connections to suicide rates: - Hospital discharge volume (discharges1864) reflects overall healthcare capacity - Children's hospital presence may indicate regional healthcare centralization - Psychiatric services availability directly relates to mental health resources

### **Limitations and Considerations**

#### **Overall Predictive Capacity**

Both models show *relatively low explanatory power* overall. Even the superior Model 2 only explains 10% of the deviance in suicide rates, indicating that: - Hospital infrastructure variables alone are insufficient predictors - Suicide rates are influenced by complex socioeconomic, demographic, and environmental factors not captured in this dataset - Additional variables (unemployment rates, population density, mental health resources) might improve model performance

## Binary Classification Performance

The confusion matrix analysis reveals classification challenges across different probability cut-offs: - At 0.25 cutoff: 18% of counties predicted as high suicide rate - At 0.15 cutoff: 83% of counties predicted as high suicide rate - This suggests the model has limited practical utility for classification

## Logistic Regression Models Comparison: Conclusion

**Model 2 (Multi-Variable) demonstrates the strongest statistical and explanatory power** due to its: - Significantly lower AIC score (8.33-point improvement) - Higher deviance explained (10.00% vs 3.77%) - Multiple significant predictors with meaningful effect sizes - Better model fit while maintaining reasonable complexity

However, both models reveal the **fundamental challenge** of predicting county suicide rates from hospital infrastructure data alone. The low overall explanatory power suggests that effective suicide rate prediction requires incorporating broader socioeconomic, demographic, and mental health resource variables beyond hospital characteristics. The `children_hospital` variable emerges as the most powerful single predictor, with its massive odds ratio of 17.27 suggesting that healthcare centralization patterns may serve as important indicators of regional suicide risk patterns.

## Logistic Regression Model 2 (Multi-Variable) : Beta Coefficients Discussion

The beta coefficients ( $\beta$ ) in logistic regression Model 2 represent the **change in log-odds of the outcome** (high suicide rate classification) for each one-unit increase in the predictor variable, holding all other variables constant.

## Statistical Significance Assessment

### Significant Predictors ( $p < 0.001$ )

#### `children_hospital` (Beta = 2.849)

- **Z-score:** 3.356, **p-value:** 0.000790
- **95% Confidence Interval:** [1.185, 4.513]
- **Interpretation:** This is the **strongest predictor** in the model with a very large effect size. Counties with children's hospitals have **17.27 times higher odds** of being classified as high suicide rate counties

- **Practical Significance:** The large beta coefficient suggests children's hospitals may serve as markers for healthcare centralization in counties with higher population density or urban characteristics

#### discharges1864 (Beta = 0.0001180)

- **Z-score:** 3.404, **p-value:** 0.000663
- **95% Confidence Interval:** [0.000050, 0.000186]
- **Interpretation:** Though the coefficient appears small, it's highly statistically significant. Each additional hospital discharge increases the odds by a factor of 1.0001
- **Practical Significance:** The cumulative effect becomes meaningful given the large range of discharge volumes across counties

#### Non-Significant Predictors (p > 0.05)

##### psych\_over17 (Beta = 0.1836)

- **Z-score:** 0.837, **p-value:** 0.403
- **95% Confidence Interval:** [-0.246, 0.614] (includes zero)
- **Interpretation:** Despite having a reasonable effect size (OR = 1.20), this variable lacks statistical significance, suggesting **no reliable relationship** with suicide rates

##### psych\_over17\_beds\_lic (Beta = 0.001335)

- **Z-score:** 0.338, **p-value:** 0.735
- **95% Confidence Interval:** [-0.006, 0.009] (includes zero)
- **Interpretation:** This variable shows **no meaningful effect** (OR  $\approx$  1.00) and is clearly non-significant

## Key Insights

### Model Reliability

- **Two out of four predictors** are statistically significant, indicating the model has some predictive validity but also contains potentially redundant variables
- The **large standard errors** for non-significant variables (0.219 and 0.004 respectively) relative to their coefficients suggest these relationships are unstable

## Effect Size Hierarchy

1. `children_hospital`: Very large effect ( $\beta = 2.849$ )
2. `psych_over17`: Small effect, but not significant ( $\beta = 0.184$ )
3. `discharges1864`: Very small but significant effect ( $\beta = 0.0001$ )
4. `psych_over17_beds_lic`: Negligible effect ( $\beta = 0.001$ )

## Confidence Intervals

The **95% confidence intervals** reveal critical information about coefficient reliability: - **Significant variables**: Confidence intervals exclude zero, confirming reliability - **Non-significant variables**: Confidence intervals include zero, indicating the true effect could be null

## Practical Implications

### Model Parsimony Considerations

The presence of two non-significant variables suggests the model could be **simplified** by removing `psych_over17` and `psych_over17_beds_lic` without meaningful loss of predictive power. This would likely: - Reduce model complexity - Improve interpretability - Potentially improve AIC score further

## Clinical and Policy Relevance

The **highly significant** `children_hospital` **coefficient** ( $\beta = 2.849$ ) indicates that healthcare infrastructure patterns, particularly pediatric care centralization, may serve as important **proxy indicators** for broader county characteristics that influence suicide rates. This finding warrants further investigation into the underlying mechanisms connecting pediatric healthcare access and community mental health outcomes.

The significant but small `discharges1864` **coefficient** suggests that overall hospital utilization patterns provide incremental predictive value, potentially reflecting broader healthcare access or community health status differences across counties.

## References

Hospital Data: <https://www.pa.gov/agencies/health/health-statistics/health-facilities/hospital-reports>

Suicide by County Data: <https://www.phaim.health.pa.gov/EDD/WebForms/DeathCntySt.aspx>