3. Use the estimates from the toxicity problem. Generate an ROC curve and find the area under the curve. You have summarized data and I would like for you to generate the ROC curve "by hand." Hint: there are $g = 6$ values of $x = 1, \ldots, 6$. Let $\hat{p}_x$ be the predicted probability for x using the logistic regression model.

(a) Complete the following table, showing work. Note TRP is true positive rate and FPR is false positive rate.

(b) Plot TPR against FPR and find the area assuming a trapezoid between successive values.

Cut value	TPR	FPR
$0 \le c < \hat{p}_1$		
$\hat{p}_1 \le c < \hat{p}_2$		
$\hat{p}_2 \le c < \hat{p}_3$		
$\hat{p}_3 \le c < \hat{p}_4$		
$\hat{p}_4 \le c < \hat{p}_5$		
$\hat{p}_5 \le c < \hat{p}_6$		
$\hat{p}_6 \le c \le 1$		

a) From toxicity problem: $\beta_0 = -2.64367$ and $\beta_1 = 0.67399$

Firstly:

$\because \hat{p}_x = \dfrac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$

$\therefore \hat{p}_1 = \dfrac{\exp(-2.64367 + 0.67399 \cdot 1)}{1 + \exp(-2.64367 + 0.67399 \cdot 1)} = 0.122$

$\hat{p}_2 = 0.215$ $\quad \hat{p}_5 = 0.674$

$\hat{p}_3 = 0.349$ $\quad \hat{p}_6 = 0.802$

$\hat{p}_4 = 0.513$

$\Delta TPR = \dfrac{TP}{FN+TP}$ $\quad \Delta FPR = \dfrac{FP}{TN+FP}$; $\{28, 53, 93, 126, 172, 197\}$

$\begin{cases} \text{Actual good} = 28+53+\ldots+19) = 669 = FN+TP \\ \text{Actual bad} = 250 \times 6 - 669 = 831 = TN+FP \end{cases}$

	Predicted bad	Predicted good
Actual Bad	TN=(# True Negatives)	FP=(# False Positives)
Actual Good	FN=(# False Negatives)	TP=(# True Positives)

TP for $\hat{p}_6$ = 197 ← highest threshold.

$\hat{p}_5$ = 197 + 172 = 369

$\hat{p}_4$ = 197 + 172 + 126 = 495

$\hat{p}_3$ = 197 + 172 + 126 + 93 = 588

$\hat{p}_2$ = 197 + 172 + 126 + 93 + 53 = 641

$\hat{p}_1$ = 197 + 172 + 126 + 93 + 53 + 28 = 669

→ $\frac{TP}{669}$

∴

Cut value	TPR	
$0 \le c < \hat{p}_1$	$\frac{669}{669}$	= 1
$\hat{p}_1 \le c < \hat{p}_2$	$\frac{641}{669}$	= 0.958
$\hat{p}_2 \le c < \hat{p}_3$	$\frac{588}{669}$	= 0.879
$\hat{p}_3 \le c < \hat{p}_4$	$\frac{495}{669}$	= 0.740
$\hat{p}_4 \le c < \hat{p}_5$	$\frac{369}{669}$	= 0.552
$\hat{p}_5 \le c < \hat{p}_6$	$\frac{197}{669}$	= 0.294
$\hat{p}_6 \le c \le 1$	$\frac{0}{669}$	= 0

FP for $\hat{p}_6$ = 53 ← these are 250 − {28, 53, 93, 126, 172, 197}

$\hat{p}_5$ = 53 + 78 = 131

$\hat{p}_4$ = 53 + 78 + 124 = 255

$\hat{p}_3$ = 53 + 78 + 124 + 157 = 412

$\hat{p}_2$ = 53 + 78 + 124 + 157 + 197 = 609

$\hat{p}_1$ = 53 + 78 + 124 + 157 + 197 + 222 = 831

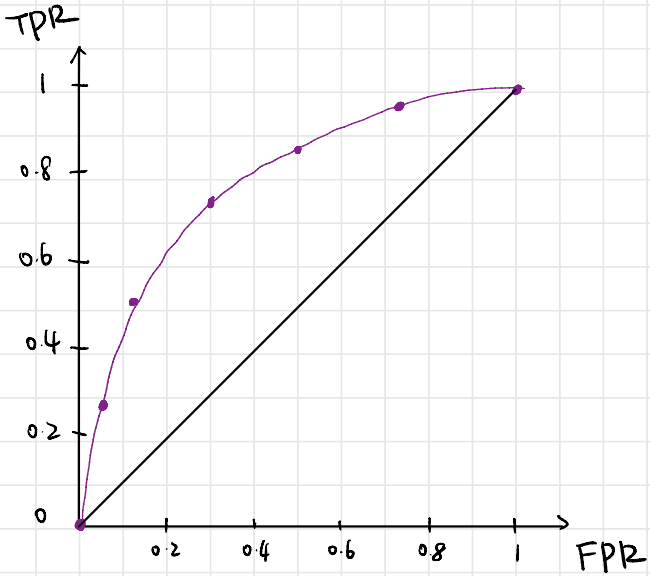
∴

Cut value	TPR	FPR
$0 \le c < \hat{p}_1$	$\frac{669}{669}$	$\frac{831}{831}$
$\hat{p}_1 \le c < \hat{p}_2$	$\frac{641}{669}$	$\frac{609}{831}$
$\hat{p}_2 \le c < \hat{p}_3$	$\frac{588}{669}$	$\frac{412}{831}$
$\hat{p}_3 \le c < \hat{p}_4$	$\frac{495}{669}$	$\frac{255}{831}$
$\hat{p}_4 \le c < \hat{p}_5$	$\frac{369}{669}$	$\frac{131}{831}$
$\hat{p}_5 \le c < \hat{p}_6$	$\frac{197}{669}$	$\frac{53}{831}$
$\hat{p}_6 \le c \le 1$	$\frac{0}{669}$	$\frac{0}{831}$

or:

Cut value	TPR	FPR
$0 \le c < \hat{p}_1$	1	1
$\hat{p}_1 \le c < \hat{p}_2$	0.958	0.733
$\hat{p}_2 \le c < \hat{p}_3$	0.879	0.496
$\hat{p}_3 \le c < \hat{p}_4$	0.740	0.307
$\hat{p}_4 \le c < \hat{p}_5$	0.552	0.157
$\hat{p}_5 \le c < \hat{p}_6$	0.294	0.064
$\hat{p}_6 \le c \le 1$	0	0

b)




```r
### Problem 3b

```{r p3b}
# Define the FPR (False Positive Rate) and TPR (True Positive Rate) values
fpr <- c(0.000, 0.064, 0.157, 0.307, 0.496, 0.733, 1.000)
tpr <- c(0.000, 0.294, 0.552, 0.740, 0.879, 0.958, 1.000)

# Calculate the differences in FPR (x-axis)
delta_fpr <- diff(fpr)

# Calculate the average TPR values between consecutive points (y-axis)
average_tpr <- (tpr[-length(tpr)] + tpr[-1]) / 2

# Apply the trapezoidal rule: Sum of (delta_fpr * average_tpr)
auc <- sum(delta_fpr * average_tpr)

# Print the AUC value
cat("The Area Under the Curve (AUC) is:", auc, "\n")
```
```

```
The Area Under the Curve (AUC) is: 0.77772
```

AUC = 0.778