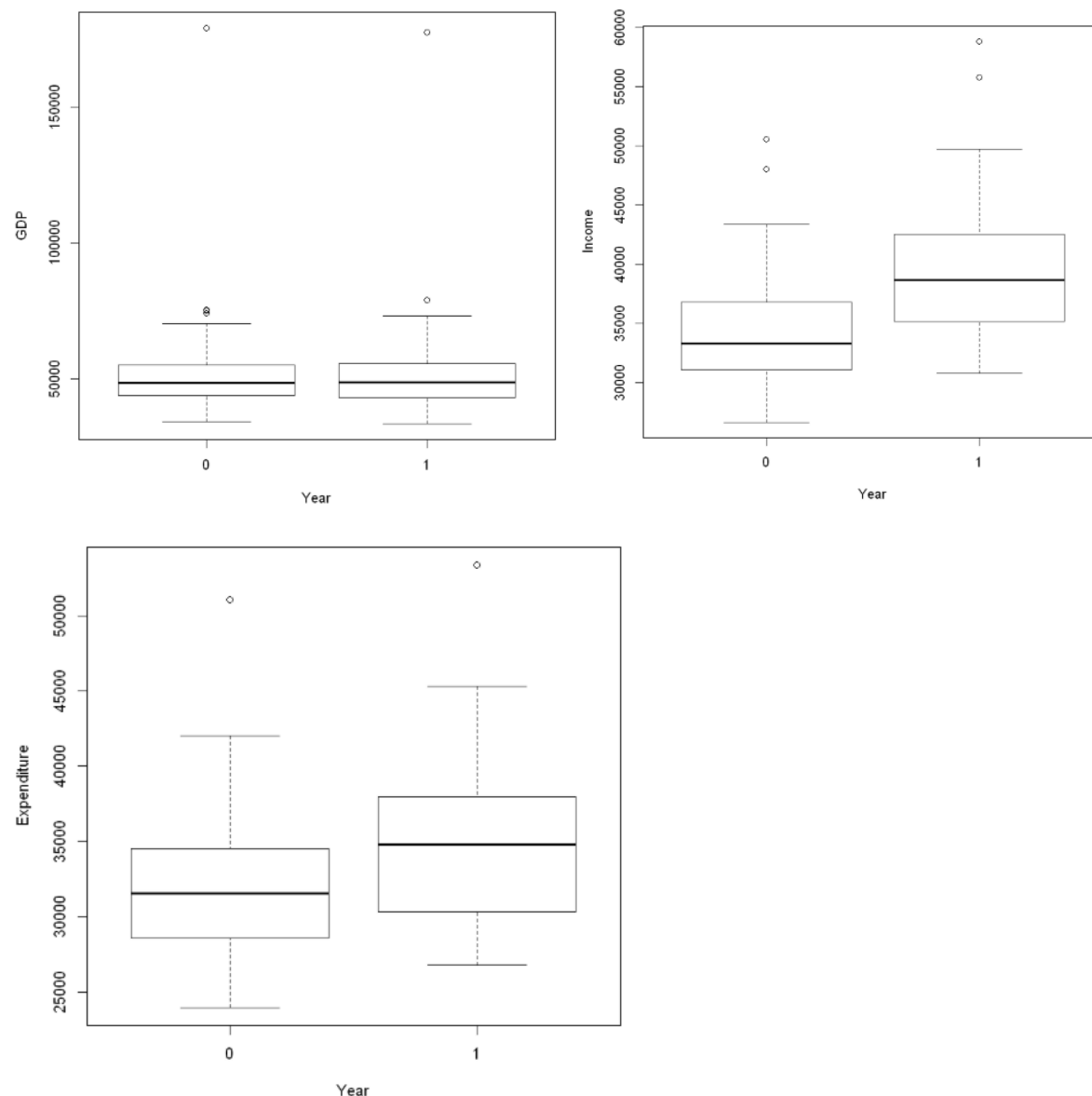


Hursh Desai

Question:

The question I wanted to look at was based off the per capita real gdp, per capita disposable personal income, and per capita personal consumption expenditures of each state in the US could be correctly classify them into the year 2007 and 2012. Coincidentally this question can also be looked upon as very giving somewhat of an answer to the question of if the first term of Obama's presidency helped the US economy get back on its feet. This question arose from my curiosity of how long it took for America to recover after the 2008 recession. Some cursory searches on the internet revealed many conflicted opinions on this subject some saying that the recession ended in 2009 but the effects can be felt on the economy till this day. I wanted to see the level of association was present between the common economic indicators such as gdp, income, and expenditure and the two years 2007 and 2012.

Analysis:



Judging based solely off the boxplots there does not seem to be that much separation when it comes to GDP, however for income and expenditure there is a slight increase by 2012 maybe indicating that the economy had gotten back on track by that time. However, D.C. does seem to be an outlier for all three with Connecticut being an additional outlier for Income. We will see if taking them out improves the regression later. This also shows us that the economy has improved even if just slightly since the pre-recession era. However, now the question becomes is it clearly separable from the numbers of 2007.

```
Call:
glm(formula = Year ~ gdp + inc + exp, family = binomial, maxit = 500)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.46487  -0.70390   0.00667   0.65876   3.11967
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -9.444e+00  2.387e+00  -3.957 7.59e-05 ***
gdp          -1.129e-04  3.932e-05  -2.872 0.00407 **
inc           6.843e-04  1.633e-04   4.191 2.77e-05 ***
exp          -2.926e-04  1.455e-04  -2.010 0.04440 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 141.402  on 101  degrees of freedom
Residual deviance: 91.711  on 98  degrees of freedom
AIC: 99.711
```

```
Number of Fisher Scoring iterations: 6
```

It seems like the regression is statistically significant overall.

A anova: 4 × 5					
	Df	Deviance	AIC	LRT	Pr(> Chi)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
<none>	NA	57.21855	65.21855	NA	NA
gdp	1	109.77153	115.77153	52.552975	4.187978e-13
inc	1	111.31067	117.31067	54.092112	1.913077e-13
exp	1	63.87376	69.87376	6.655204	9.886664e-03

A matrix: 1 × 2 of type dbl

```
gstat
49.69057 9.298162e-11
```

The likelihood ratio test also shows that the overall regression is significant. And the likelihood ratio tests for each slope are also highly significant.

gdp: 2.04223429623792 inc: 6.23034226969484 exp: 4.93184409062193

The VIFs are not too high at all meaning that we are okay on collinearity.

Hosmer and Lemeshow goodness of fit (GOF) test

```
data: Year, fitted(recession1)
X-squared = 19.192, df = 8, p-value = 0.01386
```

The Hosmer-Lemeshow GOF test shows that the data is well fit by this model.

Dxy: 0.9096

The Somers' D for this model was a .9006 which means that the proportion of concordant pairs was around 95% and discordant pairs was around 5% which is showing excellent separation.

```
A matrix: 5 × 3 of type lgl
  1  2  3
1 FALSE TRUE FALSE
1 FALSE FALSE TRUE
2 TRUE TRUE FALSE
2 FALSE TRUE TRUE
3 TRUE TRUE TRUE

$label
' (Intercept)' - '1' - '2' - '3'

$size
2 × 2 × 3 × 3 × 4

$Cp
69.7953944129596 - 89.3412303905765 - 7.94378091549939 - 64.2861350985678 - 4

A data.frame: 4 × 6
  Intercept  gdp  inc  exp logLikelihood  AIC
  <lgl> <lgl> <lgl> <lgl> <dbl> <dbl>
0 TRUE FALSE FALSE FALSE -69.31472 138.62944
1 TRUE FALSE TRUE FALSE -57.69595 117.39190
2 TRUE TRUE TRUE FALSE -31.93688 67.87376
3* TRUE TRUE TRUE TRUE -28.60928 63.21855
```

Both ways of finding the best subset show that all three predictors are useful in seeing if the two years are seperable.

\$rp 60: 12.0495222213773 7: -4.54830962832895

\$cook 9: 9.68962212969426e-09 60: 4.61252624910127

Going back to the outliers now even though the cook's distances and residuals are the highest for 60: Washington D.C. there doesn't seem to be any reason to take them out since our regression here is highly statistically significant. It also doesn't make sense to take them out since it is necessary to take into account all states to see how the economy fared after the recession. However, if we are to check what happens to the regression if we take Washington D.C. and Connecticut out we see a very little increase in the p-values of the likelihood ratio tests for the predictors and the overall regression and the Somers' D does not even change in the slightest. The goodness of fit test even shows that the model

isn't fitting the data as well as it did before.

1:	0.0205265428218214 2:	0.0236913481414833 3:	0.0227348807066492 4:	0.0168757527556049 5:	0.0833746341410491 6:
	0.0505284871795477 7:	0.696831256691125 8:	0.0100286426634517 9:	5.21726687008219e-06 10:	0.122597818394301 11:
	0.0181104216230554 12:	0.0536758588935016 13:	0.0159857711770321 14:	0.0784724340326166 15:	0.00929454382467777 16:
	0.0322611348535454 17:	0.0544165292409907 18:	0.0100663190596861 19:	0.024468965747086 20:	0.0117452789051964 21:
	0.391809308258041 22:	0.111559014406288 23:	0.013893592058833 24:	0.0417943184125346 25:	0.0173093624192581 26:
	0.0176041518244695 27:	0.00874259119513988 28:	0.0415391158042062 29:	0.0406042589142739 30:	0.19818801586489 31:
	0.346798761906034 32:	0.00720305676075936 33:	0.0715477677547693 34:	0.042568957643331 35:	0.0179025792424903 36:
	0.0114110706953232 37:	0.0714852565449474 38:	0.0101649022762387 39:	0.0605164676757418 40:	0.0840798839570457 41:
	0.0143126968472371 42:	0.141276745872237 43:	0.0293847971799666 44:	0.0256162565025574 45:	0.00476739004008157 46:
	0.0433889724479867 47:	0.179035660979473 48:	0.125558183959113 49:	0.00708798614441173 50:	0.0322419737147751 51:
	0.0523045970582372 52:	0.124281881998625 53:	0.751332899007844 54:	0.0662257287342732 55:	0.198115598828587 56:
	0.650784147076801 57:	0.405317677618279 58:	0.997752130814256 59:	0.0365756588958625 60:	0.000897654668497625 61:
	0.342943913518515 62:	0.0720555815841627 63:	0.30485603090065 64:	0.0985552102737365 65:	0.316938558443601 66:
	0.168251747661658 67:	0.380259726263471 68:	0.808120781255635 69:	0.0712268830686008 70:	0.283020737191591 71:
	0.0973392685644666 72:	0.940508014926504 73:	0.835589237314547 74:	0.100724277548146 75:	0.460678569091423 76:
	0.105899100705094 77:	0.141364500246727 78:	0.171577367282051 79:	0.692402987551783 80:	0.0734596979330191 81:
	0.84736041385948 82:	0.920538945472861 83:	0.0621811569334562 84:	0.618201427424371 85:	0.23584179844349 86:
	0.83034969771603 87:	0.119065139035105 88:	0.609119510390063 89:	0.0745012214663632 90:	0.499957794298974 91:
	0.616600732981688 92:	0.0995811216372502 93:	0.610229816106131 94:	0.39875357148092 95:	0.448768673797371 96:
	0.0427791644889953 97:	0.36751300774332 98:	0.775221081436133 99:	0.652075730882591 100:	0.0591808618496191 101:
	0.307519886088614 102:	0.937522350663779			

```

recession.predict
Year  0  1
      0 42  9
      1 14 37

```

Looking at the numbers that can be calculated using the Classification Table for the estimate of expected proportion of observations correctly classified if the actual and predicted memberships were independent, we see that if we compare it to the observed correctly classified proportion that we have a good number of correctly classified observations. Thus we can say that using the predictors gdp, expenditure and income, there is significant seperablilty between 2007 and 2012 and thus the economy has recovered since the recession in that given time.