

University of Central Florida
College of Business

QMB 6911
Capstone Project in Business Analytics

Solutions: Problem Set #7

1 Regression Models with Nonlinear Interactions

For the regression analysis below, I create some new variables. The first is the volume of each reel, calculated as the volume of a cylinder: the value π times the square of the radius of the reel, times the width of the reel. The density is then calculated as the weight divided by the volume.

```
R> flyreels[, 'Volume'] <-  
  pi * (flyreels[, 'Diameter']/2)^2 * flyreels[, 'Width']  
R> flyreels[, 'Density'] <-  
  flyreels[, 'Weight'] / flyreels[, 'Volume']
```

The idea is that a consumer will want to use a reel of a particular weight, if the weight corresponds favorably to the size of the reel.

In Table 1, a series of models is estimated including the new density variable. Model 1 contains all the variables but many of the coefficients are insignificant. Notice that density, diameter and weight are highly correlated because they are functionally related. At least one should be excluded. Removing a series of variables leaves a few choices between similar models. It seems as though weight and density are close substitutes, so a model should include only one of these two. The final model, called Model 5, has all significant coefficients but Model 4 includes the width variable, which is marginally significant. We have already explored the models excluding density.

We can keep width in the model since the significance is marginal and the coefficient may become significant under a refined specification.

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	4.03*** (0.86)	3.26*** (0.40)	4.35*** (0.28)	2.09*** (0.26)	2.31*** (0.24)
Width	-0.31 (0.30)		-0.40 (0.21)	0.30 (0.16)	
Diameter	0.06 (0.15)	0.17 (0.11)		0.40*** (0.05)	0.46*** (0.04)
Density	0.02 (0.55)	0.48 (0.29)	-0.18 (0.25)	1.20*** (0.22)	1.07*** (0.21)
Weight	0.11* (0.04)	0.07** (0.02)	0.12*** (0.01)		
SealedYes	0.42*** (0.05)	0.42*** (0.05)	0.42*** (0.05)	0.42*** (0.05)	0.43*** (0.05)
MachinedYes	0.73*** (0.08)	0.74*** (0.07)	0.72*** (0.08)	0.76*** (0.08)	0.73*** (0.08)
made_in_USATRUE	0.51*** (0.05)	0.52*** (0.05)	0.51*** (0.05)	0.52*** (0.05)	0.51*** (0.05)
R ²	0.75	0.74	0.75	0.74	0.74
Adj. R ²	0.74	0.74	0.74	0.73	0.73
Num. obs.	248	248	248	248	248

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Tab. 1: Regression Models with Density Variable

	Model 1	Model 2
(Intercept)	3.35*** (0.30)	2.03*** (0.47)
Width	0.28 (0.22)	0.39 (0.23)
Diameter	0.44*** (0.07)	0.36*** (0.08)
Density	1.13*** (0.25)	1.32** (0.41)
SealedYes	0.32*** (0.06)	0.65*** (0.10)
MachinedYes		0.65*** (0.09)
R ²	0.54	0.75
Adj. R ²	0.52	0.73
Num. obs.	135	113

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Tab. 2: Regression Models by Country of Manufacture

1.1 Comparison by Country of Manufacture

Table 2 shows the results of a series of regression models on different samples by country of manufacture. Model 1 shows the results for the sample of fly reels made in the USA and Model 2 shows the remaining fly reels made in China or Korea.

The width of the reel is insignificant in both samples and the coefficients are qualitatively similar across the samples, as well as matching in significance. This suggests that one model might be sufficient.

To test this statistically, I conduct an F -test. This compares a single model with only an indicator for the country of manufacture (the restricted model) with a separate model for each country. In this case, the full, unrestricted model has $K = 2 \times 6 = 12$ parameters, one for each variable in two models. The test that all of the coefficients are the same has $M = 6 - 1 = 5$ restrictions. The one restriction fewer accounts for the made-in-USA indicator in the full model, which allows for two separate intercepts. The F -statistic has a value of

$$\frac{(RSS_M - RSS)/M}{RSS/(N - K)} = \frac{(26.24962 - 25.2235)/4}{25.2235/237} = 2.41035.$$

This value is greater than 1, so we can compare it to the critical value of the F -statistic at the specified degrees of freedom for a conventional level of significance. These critical values are 3.400, 2.410, and 1.969 at the 1%, 5%, and 10% levels of significance, respectively.

This places the F -statistic only slightly above the critical value for the 5 percent level of significance. We conclude that fly reel prices may have some difference by country of manufacture but the difference is marginal. This suggests little justification for separate models by country of manufacture. We can investigate small differences between the models.

In the next section, I will consider a hybrid model with some interaction terms with separate coefficients by country of manufacture.

	Model 1	Model 2	Model 3
(Intercept)	2.01*** (0.26)	2.13*** (0.30)	1.89*** (0.31)
Width	0.34* (0.16)	0.30 (0.16)	0.35* (0.16)
Diameter	0.40*** (0.05)	0.40*** (0.05)	0.40*** (0.05)
Density	1.21*** (0.22)	1.14*** (0.32)	1.40*** (0.33)
SealedYes	0.63*** (0.09)	0.42*** (0.05)	0.64*** (0.09)
MachinedYes	0.65*** (0.08)	0.75*** (0.08)	0.64*** (0.08)
made_in_USATrue	0.75*** (0.09)	0.47* (0.19)	0.91*** (0.24)
SealedYes:made_in_USATrue	−0.30** (0.10)		−0.32** (0.11)
Density:made_in_USATrue		0.09 (0.37)	−0.29 (0.39)
R ²	0.75	0.74	0.75
Adj. R ²	0.74	0.73	0.74
Num. obs.	248	248	248

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Tab. 3: Regression Models with Interaction Terms by Country of Manufacture

1.2 Models with Interaction Terms

Table 3 shows the results of a series of regression models with different specifications of interaction terms by country of manufacture. The interaction with sealed and country of manufacture is significant in both specifications Model 1 and Model 3. In contrast, the interaction with density was not significant. Furthermore, the width of fly reels is a significant predictor in this more refined model. Since all variables are significant in this model and it has the highest \bar{R}^2 , Model 1 is the recommended model.

In terms of the production decision, there exists a substantial premium for fly reels made in the USA but this premium is half as large for sealed reels, however, this reduced premium is still smaller than the additional

value for a sealed reel.

Note that, after introducing the density variable to replace weight, then developing a more refined specification with the interaction term, we arrived at a model that was precise enough to detect the effect of the width of the reels.