Regression Analysis - STAT 482 - Probem Set 10

Alex Towell (atowell@siue.edu)

A company wishes to study the effects of three different types of promotions on sales of its crackers. For each store in the sample, the sales for the promotion period (y) and the sales for the preceeding period (x) are observed. The data is available from Table 22.1 on the course website. Response variable y is listed in the first column, continuous input x is listed in the second column, categorical input promotion type (1,2,3) is listed in the third column. The fourth column, the observation number, can be ignored.

Preliminary

We load the data in R with:

```
data = read.table('CH22TA01.txt')[,1:3]
colnames(data) = c("sales","pre.sales","type")
data$type = as.factor(data$type)
head(data)
```

sales	pre.sales	type
38	21	1
39	26	1
36	22	1
45	28	1
33	19	1
43	34	2

Problem 1

Define indicator variables I_1 and I_2 using promotion type 3 as the baseline level.

We let

$$I_1 = \begin{cases} 1, & \text{if } \texttt{type} = 1, \\ 0, & \text{otherwise} \end{cases}$$

and

$$I_2 = \begin{cases} 1, & \text{if } \texttt{type} = 2, \\ 0, & \text{otherwise}. \end{cases}$$

We do this transformation in R with:

```
contrasts(data$type) = contr.treatment(3,base=3)
contrasts(data$type)
```

```
##   1 2
## 1 1 0
## 2 0 1
## 3 0 0
```

Problem 2

Write an additive model for response y using continuous input variable x and indicator variables I_1, I_2.

The model of the mean sales given pre-sales and type is given by

$$E(\texttt{sales}) = \beta_0 + \beta_1\texttt{pre-sales} + \beta_2 I_1 + \beta_3 I_2.$$

Comments

The model of the data generating process given the predictors is given by

$$\texttt{sales}_i = \beta_0 + \beta_1\texttt{pre-sales} + \beta_2 I_1 + \beta_3 I_2 + \epsilon_i,$$

where ϵ_i is zero mean i.i.d. normal.

Problem 3

Write a regression function for each of the promotion types.

$$E(\texttt{sales}) = \begin{cases} (\beta_0 + \beta_2) + \beta_1\texttt{pre-sales} & \text{if type} = 1, \\ (\beta_0 + \beta_3) + \beta_1\texttt{pre-sales} & \text{if type} = 2, \\ \beta_0 + \beta_1\texttt{pre-sales} & \text{if type} = 3. \end{cases}$$

Problem 4

Define predictor effect parameters as partial differences in means.

Letting y denote sales, x denote the continuous predictor pre-sales, and t_j denote type $= j$.

Then,

$$\begin{aligned} \beta_1 &= \frac{\partial E(y)}{\partial x} \\ \beta_2 &= E(y|x, t_1) - E(y|x, t_3) \\ \beta_3 &= E(y|x, t_2) - E(y|x, t_3) \\ \beta_2 - \beta_3 &= E(y|x, t_1) - E(y|x, t_2). \end{aligned}$$

Problem 5

Provide an interpretation for each effect parameter, stated in the context of the problem.

β_1 is the difference in mean sales from a 1 unit increase in pre-sales, with the type held constant.

β_2 is the difference in mean sales between type 1 and type 3, with pre-sales held constant.

β_3 is the difference in mean sales between type 2 and type 3, with pre-sales held constant.

$\beta_2 - \beta_3$ is the difference in mean sales between type 1 and type 2, with pre-sales held constant.

Problem 6

Compute interval estimates for each of the effect parameters.

To derive $\text{CI}(\beta_2 - \beta_3)$, we replicate a bit of distribution theory. Let $a = (0\ 0\ 1\ -1)'$ and $b = (b_0\ b_1\ b_2\ b_3)$, then

$$a'b = b_2 - b_3,$$

for which

$$E(a'b) = a'E(b) = (0\ 0\ 1\ -1)(\beta_0\ \beta_1\ \beta_2\ \beta_3)' = \beta_2 - \beta_3$$

and

$$\text{var}(a'b) = a'\,\text{cov}(b)a = \sigma^2 a'(X'X)^{-1}a$$

where

$$X = \begin{Bmatrix} 1 & x_1 & I_{11} & I_{12} \\ 1 & x_2 & I_{21} & I_{22} \\ \vdots & \vdots & \vdots & \\ 1 & x_n & I_{n1} & I_{n2}. \end{Bmatrix}$$

Note that we do not know σ^2, so we estimate it with MS_E.

We compute the estimates of the coefficients in the additive model with:

```
add.mod = lm(sales ~ pre.sales+type,data=data)

b.hat = coef(add.mod)
dfe = nrow(model.matrix(add.mod)) - ncol(model.matrix(add.mod))
V = vcov(add.mod)
a = c(0,0,1,-1)

b.hat.12 = a %*% b.hat
se.12 = sqrt(a %*% V %*% a)

#t.stat.12 = b.hat.12 / se.12
#p.value.12 = 2*(1-pt(abs(t.stat.12),dfe))
#print(c(t.stat.12,p.value.12))

b.hat.12.lower = b.hat.12 - qt(.975,dfe) * se.12
b.hat.12.upper = b.hat.12 + qt(.975,dfe) * se.12

print(c(b.hat.12.lower,b.hat.12.upper))
```

```
## [1] 2.370456 7.780324
```

```
confint(add.mod)
```

```
##                    2.5 %     97.5 %
## (Intercept) -1.6473329 10.400514
## pre.sales    0.6727716  1.124347
## type1       10.3232717 15.630390
## type2        5.2850286 10.517853
```

We see that

$$\text{CI}(\beta_1) = [.673, 1.124],$$
$$\text{CI}(\beta_2) = [10.323, 15.630],$$
$$\text{CI}(\beta_3) = [5.285, 10.518],$$
$$\text{CI}(\beta_2 - \beta_3) = [2.370, 7.780].$$

Problem 7

Create a scatterplot of the data with the estimated regression lines.

```r
intercept.3 = b.hat[1]
intercept.1 = b.hat[1]+b.hat[3]
intercept.2 = b.hat[1]+b.hat[4]
slope = b.hat[2]

cat(slope, intercept.1, intercept.2, intercept.3)
```

```
## 0.8985594 17.35342 12.27803 4.376591
```

```r
attach(data)
plot(pre.sales[type == 1],
     sales[type == 1],
     xlab='pre-sales',
     ylab='sales',
     pch=1,
     col='blue',
     xlim=c(min(pre.sales),max(pre.sales)),
     ylim=c(min(sales),max(sales)))
points(pre.sales[type == 2], sales[type == 2], pch=2, col='red')
points(pre.sales[type == 3], sales[type == 3], pch=15, col='green')

abline(intercept.1,slope,col='blue')
abline(intercept.2,slope,col='red')
abline(intercept.3,slope,col='green')
```