Question 6 Solutions

This is only an outline solution. Necessary requirements of this analysis are listed below.

1. EDA
2. Model development documentation
3. Final model diagnostics
4. Cross-validation exercise
5. Answers to each of the below
    - Is the change in the average revenue different from 95 cents when the planned revenue increases by $1?
    - Explain what interaction terms in your model mean in context supported by data visualizations
    - Give two reasons why the OLS model coefficients may be biased and/or not consistent, be specific.
    - Propose (but do not actually implement) a plan for an IV approach to improve your forecasting model.

Here, we focus on using the data from 2004 and 2005 to train and build the model, and then the data from 2006 and 2007 is used for the backtesting or cross-validation exercise. A completely valid approach is to build the model using data from 2004-2007, but a comparison of fits from 2004-2005 (at a minimum) and 2006-2007 would also be a good exercise. These are omitted here for brevity.

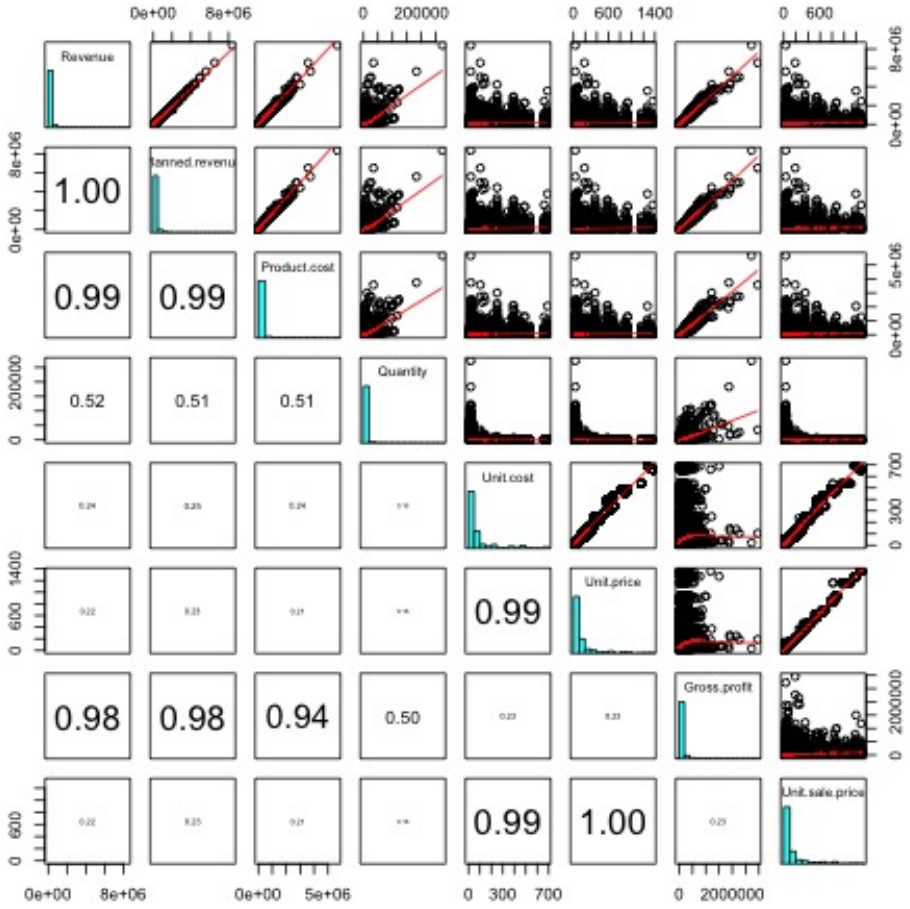
First, I want to set the United States as the first level of the retailer country variable, this way all indicators/coefficients will be in comparison to the United States (as opposed to Australia).

```
> retailSales$Retailer.country = factor(retailSales$Retailer.country,
+         levels = rev(levels(retailSales$Retailer.country)));

> rs2 = na.omit(retailSales);
> rs2Train = rs2[ (rs2$Year==2004)|(rs2$Year==2005), ];
> rs2Test = rs2[ (rs2$Year==2006)|(rs2$Year==2007), ];
```

EDA

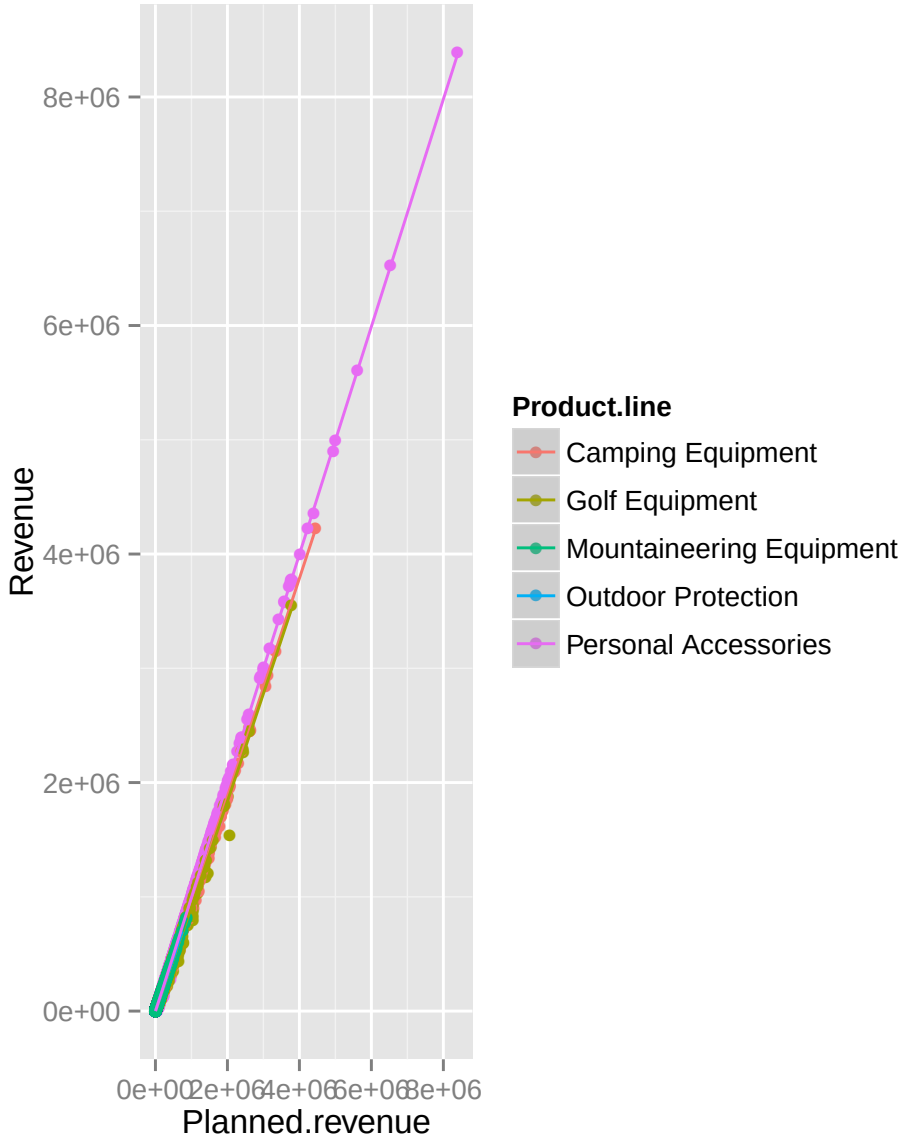
First, we focus on bivariate relationships among the possible numeric predictor variables. Remember to define any needed functions.

```
> panel.hist <- function(x, ...)
+ {
+     usr <- par("usr"); on.exit(par(usr))
+     par(usr = c(usr[1:2], 0, 1.5) )
+     h <- hist(x, plot = FALSE)
+     breaks <- h$breaks; nB <- length(breaks)
+     y <- h$counts; y <- y/max(y)
+     rect(breaks[-nB], 0, breaks[-1], y, col = "cyan", ...)
+ }
> panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...)
+ {
+     usr <- par("usr"); on.exit(par(usr))
+     par(usr = c(0, 1, 0, 1))
+     r <- abs(cor(x, y))
+     txt <- format(c(r, 0.123456789), digits = digits)[1]
+     txt <- paste0(prefix, txt)
+     if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
+     text(0.5, 0.5, txt, cex = cex.cor * r)
+ }

> pairs(Revenue~Planned.revenue + Product.cost + Quantity + Unit.cost + Unit.price +
+         Gross.profit + Unit.sale.price, data=rs2Train,
+         upper.panel=panel.smooth, lower.panel=panel.cor, diag.panel=panel.hist)
```



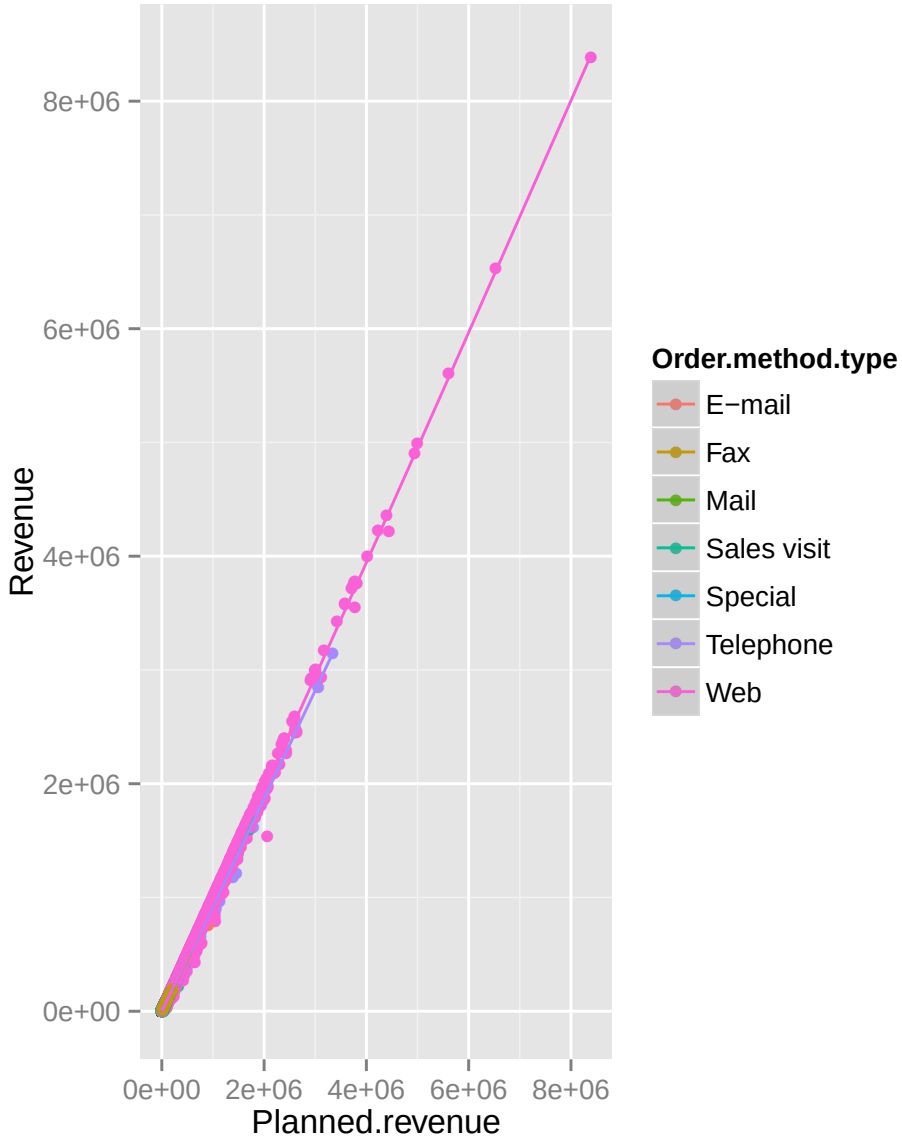
There is a clear and strong linear association between the (actual) revenue and the planned revenue. There are also strong linear associations with quantity and product cost/sales/price information. However, it doesn't make sense to use these variables in a model unless they represent forecasts. The next plots explore three-way association between the planned revenue and actual revenues by the levels of product line, ordering method, and country variables.

```
> qplot(Planned.revenue, Revenue, colour=Product.line,
+         data=rs2Train, geom=c("point", "smooth"));
```



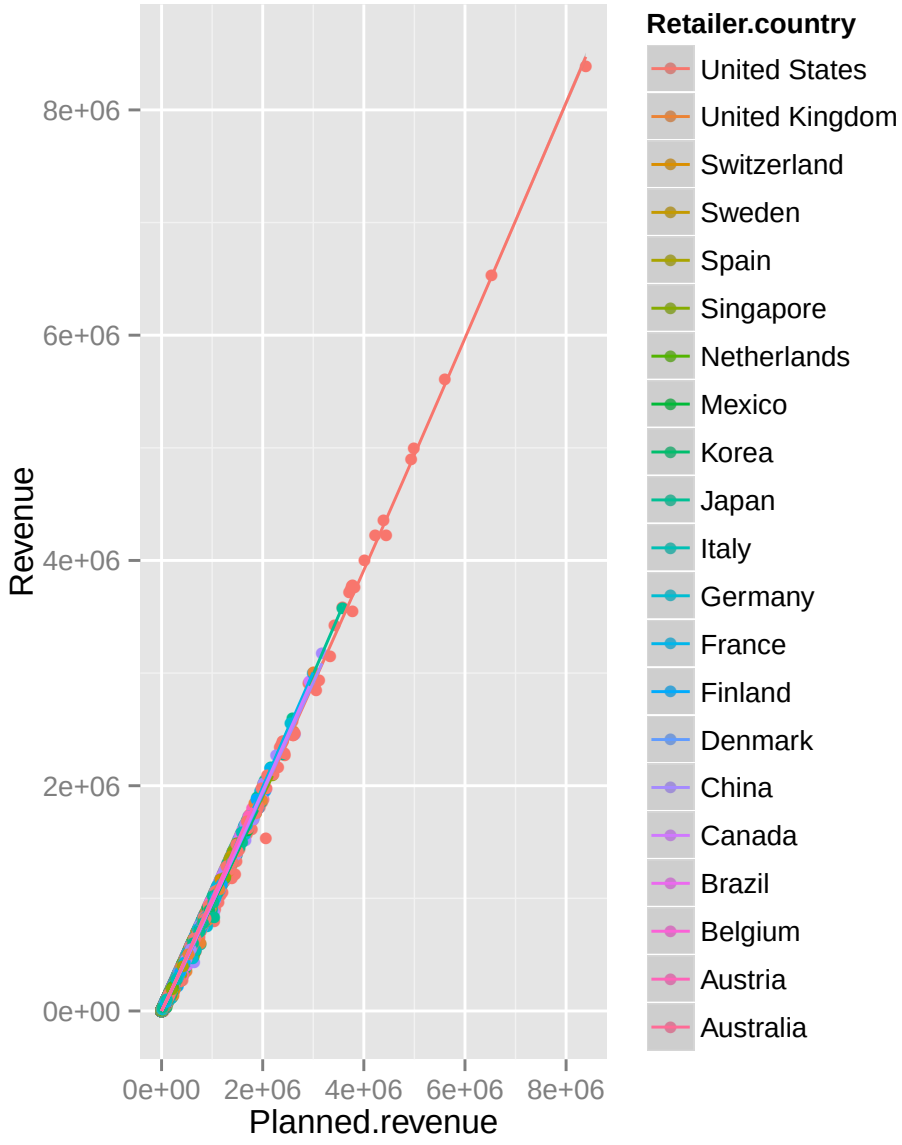
This might suggest a potential interaction between product line and planned revenue; the slope between the actual revenue and planned revenue depends on product line.

```
> qplot(Planned.revenue, Revenue, colour=Order.method.type,
+         data=rs2Train, geom=c("point", "smooth"))
```



This might suggest a potential interaction between order method and planned revenue; visually we can see how many more web orders there are.

```
> qplot(Planned.revenue, Revenue, colour=Retailer.country,
+         data=rs2Train, geom=c("point", "smooth"))
```



This might suggest a potential interaction between retailer country and planned revenue; visually we can see how much comes from the US.

Modeling

The first model we explore is demonstrative, showing the result of blindly including all of the numeric predictor variables.

```
> m1 = lm(Revenue~Planned.revenue + Product.cost + Quantity + Unit.cost + Unit.price +
+         Gross.profit + Unit.sale.price, data=rs2Train);
> coeftest(m1);
> vif(m1);
```

```
t test of coefficients:

                   Estimate  Std. Error     t value  Pr(>|t|)
(Intercept)      8.5527e-10  1.4481e-11  5.9060e+01 < 2.2e-16 ***
Planned.revenue  3.7915e-15  7.8540e-16  4.8275e+00 1.397e-06 ***
Product.cost      1.0000e+00  8.5058e-16  1.1757e+15 < 2.2e-16 ***
Quantity          6.5454e-15  1.8194e-15  3.5976e+00 0.0003223 ***
Unit.cost        -8.9877e-13  6.9282e-13 -1.2973e+00 0.1945603
Unit.price       -6.4288e-14  1.2843e-12 -5.0100e-02 0.9600787
Gross.profit      1.0000e+00  8.3877e-16  1.1922e+15 < 2.2e-16 ***
Unit.sale.price   5.5862e-13  1.4237e-12  3.9240e-01 0.6947780
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Planned.revenue    Product.cost        Quantity       Unit.cost      Unit.price
     512.896681      220.546144        1.588458       73.247958      861.970719
   Gross.profit Unit.sale.price
      83.561166      924.793078
```

Not only does it really not make sense to use any other variable other than the planned revenue, but doing so introduces an enormous amount of collinearity which we would expect given the EDA.

The product and product type variables are not considered because their coefficients cannot all be estimated and/or introduces singularity issues. Our initial model is below.

```
> m2 = lm(Revenue~Planned.revenue + Product.line + Order.method.type +
+         Retailer.country, data=rs2Train);
> coeftest(m2);
```

```
t test of coefficients:

                                     Estimate  Std. Error   t value
(Intercept)                       -1.1537e+04  5.9982e+02  -19.2348
Planned.revenue                    9.7228e-01  4.0427e-04 2405.0095
Product.lineGolf Equipment        -5.6686e+03  3.9432e+02  -14.3758
Product.lineMountaineering Equipment  4.6453e+03  4.3977e+02   10.5630
Product.lineOutdoor Protection     6.4417e+03  3.7987e+02   16.9579
Product.linePersonal Accessories   9.2832e+03  2.8165e+02   32.9594
Order.method.typeFax               1.6632e+03  5.6408e+02    2.9486
Order.method.typeMail              3.0335e+03  6.0601e+02    5.0056
Order.method.typeSales visit       4.6529e+02  5.0458e+02    0.9221
Order.method.typeSpecial           1.7086e+03  6.6284e+02    2.5776
Order.method.typeTelephone        -1.3629e+03  4.9850e+02   -2.7340
Order.method.typeWeb              -1.1372e+03  4.5240e+02   -2.5137
Retailer.countryUnited Kingdom     3.4056e+03  6.7645e+02    5.0345
Retailer.countrySwitzerland        6.7918e+03  8.4146e+02    8.0714
Retailer.countrySweden             6.5340e+03  7.8572e+02    8.3160
Retailer.countrySpain              5.3530e+03  6.8696e+02    7.7923
Retailer.countrySingapore          3.9150e+03  6.5610e+02    5.9670
```

```
Retailer.countryNetherlands           5.3034e+03  6.4128e+02    8.2701
Retailer.countryMexico                4.0893e+03  6.7167e+02    6.0883
Retailer.countryKorea                 2.9490e+03  6.8380e+02    4.3127
Retailer.countryJapan                 3.4996e+03  5.9985e+02    5.8342
Retailer.countryItaly                 5.3090e+03  6.6425e+02    7.9925
Retailer.countryGermany               3.9078e+03  6.2445e+02    6.2579
Retailer.countryFrance                4.8301e+03  6.0111e+02    8.0353
Retailer.countryFinland               3.4792e+03  7.2044e+02    4.8293
Retailer.countryDenmark               7.2412e+03  7.5354e+02    9.6096
Retailer.countryChina                 1.3502e+03  7.1472e+02    1.8892
Retailer.countryCanada                4.6338e+03  6.1440e+02    7.5420
Retailer.countryBrazil                5.3546e+03  7.7363e+02    6.9214
Retailer.countryBelgium               6.4055e+03  7.3561e+02    8.7078
Retailer.countryAustria               7.6598e+03  6.9874e+02   10.9624
Retailer.countryAustralia             7.0110e+03  8.1314e+02    8.6221
                                     Pr(>|t|)
(Intercept)                         < 2.2e-16 ***
Planned.revenue                     < 2.2e-16 ***
Product.lineGolf Equipment          < 2.2e-16 ***
Product.lineMountaineering Equipment < 2.2e-16 ***
Product.lineOutdoor Protection      < 2.2e-16 ***
Product.linePersonal Accessories    < 2.2e-16 ***
Order.method.typeFax                 0.003198 **
Order.method.typeMail               5.636e-07 ***
Order.method.typeSales visit         0.356482
Order.method.typeSpecial             0.009958 **
Order.method.typeTelephone           0.006265 **
Order.method.typeWeb                 0.011957 *
Retailer.countryUnited Kingdom      4.851e-07 ***
Retailer.countrySwitzerland         7.512e-16 ***
Retailer.countrySweden              < 2.2e-16 ***
Retailer.countrySpain               7.038e-15 ***
Retailer.countrySingapore           2.474e-09 ***
Retailer.countryNetherlands         < 2.2e-16 ***
Retailer.countryMexico              1.171e-09 ***
Retailer.countryKorea               1.624e-05 ***
Retailer.countryJapan               5.525e-09 ***
Retailer.countryItaly               1.425e-15 ***
Retailer.countryGermany             4.014e-10 ***
Retailer.countryFrance              1.008e-15 ***
Retailer.countryFinland             1.385e-06 ***
Retailer.countryDenmark             < 2.2e-16 ***
Retailer.countryChina                0.058887 .
Retailer.countryCanada              4.910e-14 ***
Retailer.countryBrazil              4.666e-12 ***
Retailer.countryBelgium             < 2.2e-16 ***
Retailer.countryAustria             < 2.2e-16 ***
Retailer.countryAustralia           < 2.2e-16 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

The baseline group is the combination of camping equipment product line, email ordering method, and United States retailer country.

Consider the important of the coefficient for the planned revenue: the change in the average revenue is \$0.97228 when the planned revenue increases by \$1; so this is giving information about the actual revenue per \$1 of planned revenue adjusting for product line, ordering method and retailer country. Notice the country effects are all relative to the United Stations. This model is probably a good model to address if the change in the average revenue is 95 cents when the planned revenue increases by 1 dollar adjusting for the product line, order method and country.

```
> linearHypothesis(m2, "Planned.revenue=0.95");

Linear hypothesis test
```

```
Hypothesis:
Planned.revenue = 0.95

Model 1: restricted model
Model 2: Revenue ~ Planned.revenue + Product.line + Order.method.type +
    Retailer.country

  Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
1  14056 3.1618e+12
2  14055 2.5998e+12  1 5.6203e+11 3038.4 < 2.2e-16 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

Here we are testing $H_0 : \beta_1 = 0.95; H_A : \beta_1 \neq 0.95$. The p-value is small so the null hypothesis is rejected. There is evidence the change in the average revenue is different from \$0.95 when the planned revenue increases by \$1 adjusting for product line, ordering method and retailer country. The sample coefficient is \$0.97 so this is actually a good thing!

Next, we consider some smaller models that allow us to understand how each of product line, ordering method and retailer affects the relationship between actual and planned revenues by considering the interaction terms.

```
> m2a = lm(Revenue~Planned.revenue + Product.line +
+          Planned.revenue:Product.line, data=rs2Train);
> coeftest(m2a);

t test of coefficients:

                                                 Estimate  Std. Error
(Intercept)                                   -1.4131e+03  1.4947e+02
Planned.revenue                                9.3605e-01  4.5317e-04
Product.lineGolf Equipment                     7.3858e+02  3.2599e+02
Product.lineMountaineering Equipment           1.5963e+03  3.7074e+02
Product.lineOutdoor Protection                 1.8675e+03  2.9767e+02
Product.linePersonal Accessories              -1.5131e+03  2.1104e+02
Planned.revenue:Product.lineGolf Equipment    -1.9755e-02  9.1199e-04
Planned.revenue:Product.lineMountaineering Equipment  6.7444e-03  2.3263e-03
Planned.revenue:Product.lineOutdoor Protection        2.2315e-03  3.8159e-03
Planned.revenue:Product.linePersonal Accessories      6.1615e-02  5.5896e-04
                                                  t value  Pr(>|t|)
(Intercept)                                       -9.4541 < 2.2e-16 ***
Planned.revenue                                 2065.5540 < 2.2e-16 ***
Product.lineGolf Equipment                         2.2657  0.023487 *
Product.lineMountaineering Equipment               4.3056 1.677e-05 ***
Product.lineOutdoor Protection                     6.2737 3.629e-10 ***
Product.linePersonal Accessories                  -7.1700 7.872e-13 ***
Planned.revenue:Product.lineGolf Equipment       -21.6614 < 2.2e-16 ***
Planned.revenue:Product.lineMountaineering Equipment    2.8992  0.003746 **
Planned.revenue:Product.lineOutdoor Protection          0.5848  0.558688
Planned.revenue:Product.linePersonal Accessories      110.2312 < 2.2e-16 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

In comparison to the camping equipment product line, the mountaineering equipment and personal accessories products have a significantly higher actual per planned revenue and gold equipment has a significantly lower actual per planned revenue. Product line impacts actual per planned revenue.

```
> m2b = lm(Revenue~Planned.revenue + Order.method.type +
+          Planned.revenue:Order.method.type, data=rs2Train);
> coeftest(m2b);

t test of coefficients:

                                            Estimate  Std. Error  t value
(Intercept)                                -4.2215e+02  4.4217e+02  -0.9547
Planned.revenue                             9.3795e-01  2.1813e-03 429.9852
Order.method.typeFax                        1.2595e+03  6.4554e+02   1.9510
Order.method.typeMail                       3.4332e+02  6.9058e+02   0.4972
Order.method.typeSales visit                2.8558e+02  5.5284e+02   0.5166
Order.method.typeSpecial                    6.0894e+02  7.6270e+02   0.7984
Order.method.typeTelephone                  3.1940e+02  5.3858e+02   0.5930
Order.method.typeWeb                       -5.2433e+03  4.9562e+02 -10.5792
Planned.revenue:Order.method.typeFax       -1.5921e-02  6.8521e-03  -2.3235
Planned.revenue:Order.method.typeMail      -1.2083e-02  6.5290e-03  -1.8506
Planned.revenue:Order.method.typeSales visit -6.9727e-03  2.9264e-03  -2.3827
Planned.revenue:Order.method.typeSpecial   -1.1206e-02  1.0063e-02  -1.1136
Planned.revenue:Order.method.typeTelephone -7.5497e-03  2.4894e-03  -3.0327
Planned.revenue:Order.method.typeWeb        4.0084e-02  2.2218e-03  18.0412
                                            Pr(>|t|)
(Intercept)                                 0.339733
Planned.revenue                            < 2.2e-16 ***
Order.method.typeFax                        0.051073 .
Order.method.typeMail                       0.619089
Order.method.typeSales visit                0.605468
Order.method.typeSpecial                    0.424653
Order.method.typeTelephone                  0.553165
Order.method.typeWeb                       < 2.2e-16 ***
Planned.revenue:Order.method.typeFax        0.020168 *
Planned.revenue:Order.method.typeMail       0.064247 .
Planned.revenue:Order.method.typeSales visit  0.017199 *
Planned.revenue:Order.method.typeSpecial    0.265463
Planned.revenue:Order.method.typeTelephone  0.002428 **
Planned.revenue:Order.method.typeWeb       < 2.2e-16 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

In comparison to the email ordering method, the fax, sales visit and telephone ordering methods have significantly less actual per planned revenue and web ordering has significantly more actual per planned revenue. Ordering method impacts actual per planned revenue.

```
> m2c = lm(Revenue~Planned.revenue + Retailer.country +
+          Planned.revenue:Retailer.country, data=rs2Train);
> coeftest(m2c);

t test of coefficients:

                                             Estimate  Std. Error
(Intercept)                                -8.6553e+03  4.6717e+02
Planned.revenue                             9.7540e-01  5.9399e-04
Retailer.countryUnited Kingdom              4.5164e+03  8.0368e+02
Retailer.countrySwitzerland                 6.9403e+03  9.8134e+02
Retailer.countrySweden                      7.0887e+03  8.9727e+02
Retailer.countrySpain                       6.7777e+03  8.1857e+02
Retailer.countrySingapore                   6.9006e+03  7.7850e+02
Retailer.countryNetherlands                 6.3117e+03  7.6177e+02
Retailer.countryMexico                      8.2897e+03  8.0526e+02
Retailer.countryKorea                       8.0210e+03  8.3686e+02
Retailer.countryJapan                       4.6477e+03  6.9740e+02
Retailer.countryItaly                       5.9067e+03  7.7967e+02
Retailer.countryGermany                     5.4334e+03  7.3142e+02
Retailer.countryFrance                      4.4594e+03  7.0733e+02
Retailer.countryFinland                     6.4552e+03  8.5960e+02
Retailer.countryDenmark                     6.6897e+03  8.9329e+02
Retailer.countryChina                       3.5571e+03  8.8333e+02
Retailer.countryCanada                      6.8403e+03  7.1851e+02
Retailer.countryBrazil                      6.1160e+03  9.4258e+02
Retailer.countryBelgium                     7.6212e+03  8.6114e+02
Retailer.countryAustria                     5.2799e+03  8.0642e+02
Retailer.countryAustralia                   7.1258e+03  1.0131e+03
Planned.revenue:Retailer.countryUnited Kingdom -8.1967e-03  1.7307e-03
Planned.revenue:Retailer.countrySwitzerland  5.8880e-03  5.5463e-03
Planned.revenue:Retailer.countrySweden      -3.8622e-03  4.1436e-03
Planned.revenue:Retailer.countrySpain       -1.2380e-02  3.1685e-03
Planned.revenue:Retailer.countrySingapore   -2.1879e-02  2.5663e-03
Planned.revenue:Retailer.countryNetherlands -7.8858e-03  2.9704e-03
Planned.revenue:Retailer.countryMexico      -3.4042e-02  3.3624e-03
Planned.revenue:Retailer.countryKorea       -3.5478e-02  2.6230e-03
Planned.revenue:Retailer.countryJapan       -4.1741e-03  1.5028e-03
Planned.revenue:Retailer.countryItaly       -3.5532e-03  2.9238e-03
Planned.revenue:Retailer.countryGermany     -5.8048e-03  1.9710e-03
Planned.revenue:Retailer.countryFrance       2.9145e-03  1.8499e-03
Planned.revenue:Retailer.countryFinland     -1.8211e-02  2.4123e-03
Planned.revenue:Retailer.countryDenmark      1.1097e-02  6.3124e-03
Planned.revenue:Retailer.countryChina       -1.2525e-02  1.6108e-03
Planned.revenue:Retailer.countryCanada      -9.6010e-03  1.9191e-03
Planned.revenue:Retailer.countryBrazil      -6.1424e-03  3.1627e-03
Planned.revenue:Retailer.countryBelgium     -8.9160e-03  3.9707e-03
Planned.revenue:Retailer.countryAustria      1.6757e-02  2.8467e-03
Planned.revenue:Retailer.countryAustralia   -6.3085e-03  7.5527e-03
                                              t value  Pr(>|t|)
(Intercept)                                 -18.5270 < 2.2e-16 ***
Planned.revenue                            1642.1148 < 2.2e-16 ***
Retailer.countryUnited Kingdom                5.6197 1.950e-08 ***
Retailer.countrySwitzerland                   7.0723 1.596e-12 ***
Retailer.countrySweden                        7.9003 2.988e-15 ***
Retailer.countrySpain                         8.2799 < 2.2e-16 ***
Retailer.countrySingapore                     8.8639 < 2.2e-16 ***
Retailer.countryNetherlands                   8.2856 < 2.2e-16 ***
Retailer.countryMexico                       10.2944 < 2.2e-16 ***
Retailer.countryKorea                         9.5846 < 2.2e-16 ***
Retailer.countryJapan                         6.6643 2.758e-11 ***
Retailer.countryItaly                         7.5758 3.791e-14 ***
Retailer.countryGermany                       7.4286 1.161e-13 ***
Retailer.countryFrance                        6.3046 2.976e-10 ***
Retailer.countryFinland                       7.5096 6.289e-14 ***
Retailer.countryDenmark                       7.4888 7.364e-14 ***
Retailer.countryChina                         4.0270 5.679e-05 ***
```

```
Retailer.countryCanada                              9.5200 < 2.2e-16 ***
Retailer.countryBrazil                              6.4886 8.957e-11 ***
Retailer.countryBelgium                             8.8501 < 2.2e-16 ***
Retailer.countryAustria                             6.5474 6.059e-11 ***
Retailer.countryAustralia                           7.0335 2.107e-12 ***
Planned.revenue:Retailer.countryUnited Kingdom     -4.7361 2.199e-06 ***
Planned.revenue:Retailer.countrySwitzerland         1.0616  0.288435
Planned.revenue:Retailer.countrySweden             -0.9321  0.351309
Planned.revenue:Retailer.countrySpain              -3.9073 9.375e-05 ***
Planned.revenue:Retailer.countrySingapore          -8.5255 < 2.2e-16 ***
Planned.revenue:Retailer.countryNetherlands        -2.6548  0.007943 **
Planned.revenue:Retailer.countryMexico            -10.1244 < 2.2e-16 ***
Planned.revenue:Retailer.countryKorea             -13.5256 < 2.2e-16 ***
Planned.revenue:Retailer.countryJapan              -2.7775  0.005485 **
Planned.revenue:Retailer.countryItaly              -1.2153  0.224289
Planned.revenue:Retailer.countryGermany            -2.9452  0.003233 **
Planned.revenue:Retailer.countryFrance              1.5755  0.115163
Planned.revenue:Retailer.countryFinland            -7.5493 4.646e-14 ***
Planned.revenue:Retailer.countryDenmark             1.7580  0.078777 .
Planned.revenue:Retailer.countryChina              -7.7761 7.998e-15 ***
Planned.revenue:Retailer.countryCanada             -5.0029 5.715e-07 ***
Planned.revenue:Retailer.countryBrazil             -1.9422  0.052137 .
Planned.revenue:Retailer.countryBelgium            -2.2455  0.024754 *
Planned.revenue:Retailer.countryAustria             5.8865 4.035e-09 ***
Planned.revenue:Retailer.countryAustralia          -0.8353  0.403583
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

In comparison to the United States, all countries other than Switzerland, Sweden, Denmark, Brazil and Australia have significantly less actual per planned revenue; and Austria has significantly more actual per planned revenue. Retailer country impacts actual per planned revenue.

Refer back to model #2. Note that all variables in the model are highly significant , although not every level has a coefficient that is significant. One way to improve the model would be to specify the indicators to include (e.g. all product types with the exception sales visit). This would be a preferred approach for the subsequent models but is omitted here to be brief. Consider all possible interactions. This model takes a bit of time to run so be careful and patient. The output is omitted here.

```
> m3 = lm(Revenue~Planned.revenue*Product.line*Order.method.type*Retailer.country,
+         data=rs2Train);
```

Most of the three- and four-way interactions are not needed in the model, so we refit the model specifying interactions that are significant.

```
> m4 = lm(Revenue~Planned.revenue + Product.line + Order.method.type +
+         Retailer.country + Planned.revenue:Product.line + Planned.revenue:Order.meth
+         Planned.revenue:Retailer.country + Order.method.type:Retailer.country +
+         Planned.revenue:Product.line:Order.method.type +
+         Planned.revenue:Order.method.type:Retailer.country, data=rs2);
```

Again, most of the three-way interactions, while significant, are only so because of certain specific levels and similarly for the two-say interactions between the categorical predictions. Our last step is remove all interactions with the exception of interactions between the categorical variables and the planned revenue. This modeling choice is supported by our EDA.

```
> m5 = lm(Revenue~Planned.revenue + Product.line + Order.method.type + Retailer.countr
> coeftest(m5);
```

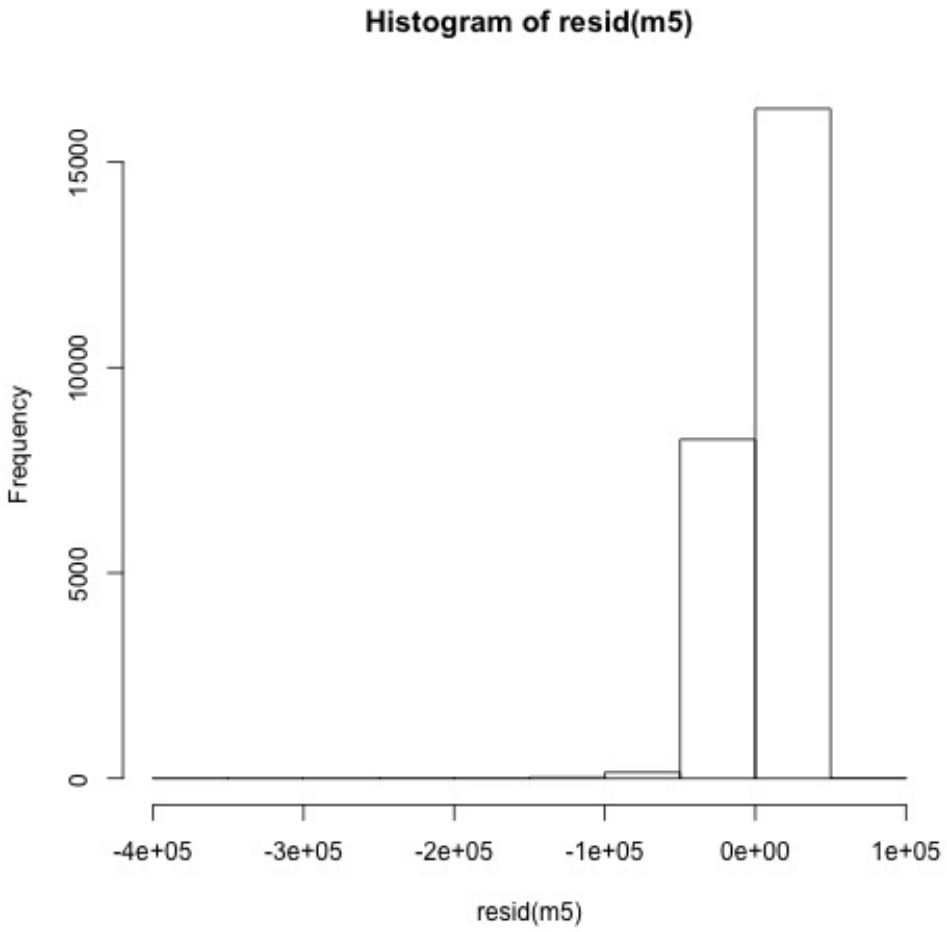
```
t test of coefficients:

                                                 Estimate  Std. Error
(Intercept)                                   -3.6099e+03  4.1064e+02
Planned.revenue                                9.3239e-01  1.6978e-03
Product.lineGolf Equipment                    -1.5987e+02  2.7984e+02
Product.lineMountaineering Equipment           2.2210e+03  2.7686e+02
Product.lineOutdoor Protection                 2.4199e+03  2.6500e+02
Product.linePersonal Accessories              -8.1281e+02  1.9188e+02
Order.method.typeFax                           8.2535e+02  4.3484e+02
Order.method.typeMail                          3.7750e+02  4.7485e+02
Order.method.typeSales visit                  -1.5320e+02  3.5649e+02
Order.method.typeSpecial                       1.9164e+02  5.6554e+02
Order.method.typeTelephone                    -5.1976e+01  3.5167e+02
Order.method.typeWeb                          -3.9949e+03  3.2384e+02
Retailer.countryUnited Kingdom                 2.0476e+03  4.5491e+02
Retailer.countrySwitzerland                    3.8237e+03  5.3260e+02
Retailer.countrySweden                         4.1295e+03  4.7707e+02
Retailer.countrySpain                          2.9672e+03  4.6621e+02
Retailer.countrySingapore                      2.4551e+03  4.6328e+02
Retailer.countryNetherlands                    2.6861e+03  4.5798e+02
Retailer.countryMexico                         2.9114e+03  4.8258e+02
Retailer.countryKorea                          2.5019e+03  4.9292e+02
Retailer.countryJapan                          2.3123e+03  4.0415e+02
Retailer.countryItaly                          3.0068e+03  4.5760e+02
Retailer.countryGermany                        2.6856e+03  4.1392e+02
Retailer.countryFrance                         2.5905e+03  4.1264e+02
Retailer.countryFinland                        2.3261e+03  5.3830e+02
Retailer.countryDenmark                        4.6814e+03  5.3595e+02
Retailer.countryChina                          1.0464e+03  5.2367e+02
```

```
Retailer.countryCanada                                3.1862e+03  4.1027e+02
Retailer.countryBrazil                                3.5736e+03  5.6713e+02
Retailer.countryBelgium                               3.8186e+03  4.9551e+02
Retailer.countryAustria                               3.7748e+03  4.6907e+02
Retailer.countryAustralia                             3.3534e+03  5.0547e+02
Planned.revenue:Product.lineGolf Equipment           -1.6265e-02  5.6444e-04
Planned.revenue:Product.lineMountaineering Equipment  8.1224e-03  1.0147e-03
Planned.revenue:Product.lineOutdoor Protection        1.0289e-02  4.1458e-03
Planned.revenue:Product.linePersonal Accessories      5.1557e-02  4.0388e-04
Planned.revenue:Order.method.typeFax                  7.0569e-04  4.7650e-03
Planned.revenue:Order.method.typeMail                 3.5082e-03  4.9753e-03
Planned.revenue:Order.method.typeSales visit          5.4893e-04  2.1439e-03
Planned.revenue:Order.method.typeSpecial              2.8400e-03  7.5752e-03
Planned.revenue:Order.method.typeTelephone           -1.4672e-03  1.8892e-03
Planned.revenue:Order.method.typeWeb                  1.5091e-02  1.6658e-03
Planned.revenue:Retailer.countryUnited Kingdom        1.4093e-03  8.0451e-04
Planned.revenue:Retailer.countrySwitzerland           5.5360e-03  1.9021e-03
Planned.revenue:Retailer.countrySweden                5.2248e-03  1.9731e-03
Planned.revenue:Retailer.countrySpain                 2.6714e-03  1.3728e-03
Planned.revenue:Retailer.countrySingapore             1.8908e-03  1.1588e-03
Planned.revenue:Retailer.countryNetherlands           4.3823e-03  1.2522e-03
Planned.revenue:Retailer.countryMexico                2.1835e-03  1.3594e-03
Planned.revenue:Retailer.countryKorea                -2.4910e-04  1.1793e-03
Planned.revenue:Retailer.countryJapan                 1.6086e-03  6.5733e-04
Planned.revenue:Retailer.countryItaly                 2.8596e-03  1.2687e-03
Planned.revenue:Retailer.countryGermany               1.7473e-03  9.4460e-04
Planned.revenue:Retailer.countryFrance                2.1124e-03  8.5053e-04
Planned.revenue:Retailer.countryFinland               1.4559e-03  1.1585e-03
Planned.revenue:Retailer.countryDenmark               4.8521e-03  3.0745e-03
Planned.revenue:Retailer.countryChina                -5.9377e-04  7.7463e-04
Planned.revenue:Retailer.countryCanada                3.7238e-03  8.3663e-04
Planned.revenue:Retailer.countryBrazil                2.7732e-03  1.5551e-03
Planned.revenue:Retailer.countryBelgium               4.6931e-03  1.8234e-03
Planned.revenue:Retailer.countryAustria               4.8668e-03  1.4643e-03
Planned.revenue:Retailer.countryAustralia             3.6819e-03  1.8268e-03
                                                      t value  Pr(>|t|)
(Intercept)                                           -8.7910 < 2.2e-16 ***
Planned.revenue                                      549.1840 < 2.2e-16 ***
Product.lineGolf Equipment                            -0.5713 0.5678158
Product.lineMountaineering Equipment                   8.0221 1.085e-15 ***
Product.lineOutdoor Protection                         9.1317 < 2.2e-16 ***
Product.linePersonal Accessories                      -4.2360 2.284e-05 ***
Order.method.typeFax                                   1.8981 0.0577010 .
Order.method.typeMail                                  0.7950 0.4266279
Order.method.typeSales visit                          -0.4298 0.6673773
Order.method.typeSpecial                               0.3389 0.7347152
Order.method.typeTelephone                            -0.1478 0.8825035
Order.method.typeWeb                                 -12.3358 < 2.2e-16 ***
Retailer.countryUnited Kingdom                         4.5011 6.790e-06 ***
Retailer.countrySwitzerland                            7.1793 7.205e-13 ***
Retailer.countrySweden                                 8.6559 < 2.2e-16 ***
Retailer.countrySpain                                  6.3646 1.993e-10 ***
Retailer.countrySingapore                              5.2994 1.172e-07 ***
Retailer.countryNetherlands                            5.8651 4.547e-09 ***
Retailer.countryMexico                                 6.0330 1.633e-09 ***
Retailer.countryKorea                                  5.0758 3.887e-07 ***
Retailer.countryJapan                                  5.7215 1.068e-08 ***
Retailer.countryItaly                                  6.5708 5.105e-11 ***
Retailer.countryGermany                                6.4881 8.859e-11 ***
Retailer.countryFrance                                 6.2780 3.488e-10 ***
Retailer.countryFinland                                4.3212 1.558e-05 ***
Retailer.countryDenmark                                8.7348 < 2.2e-16 ***
Retailer.countryChina                                  1.9982 0.0457064 *
Retailer.countryCanada                                 7.7661 8.409e-15 ***
Retailer.countryBrazil                                 6.3012 3.004e-10 ***
Retailer.countryBelgium                                7.7065 1.342e-14 ***
Retailer.countryAustria                                8.0475 8.830e-16 ***
Retailer.countryAustralia                              6.6342 3.330e-11 ***
```

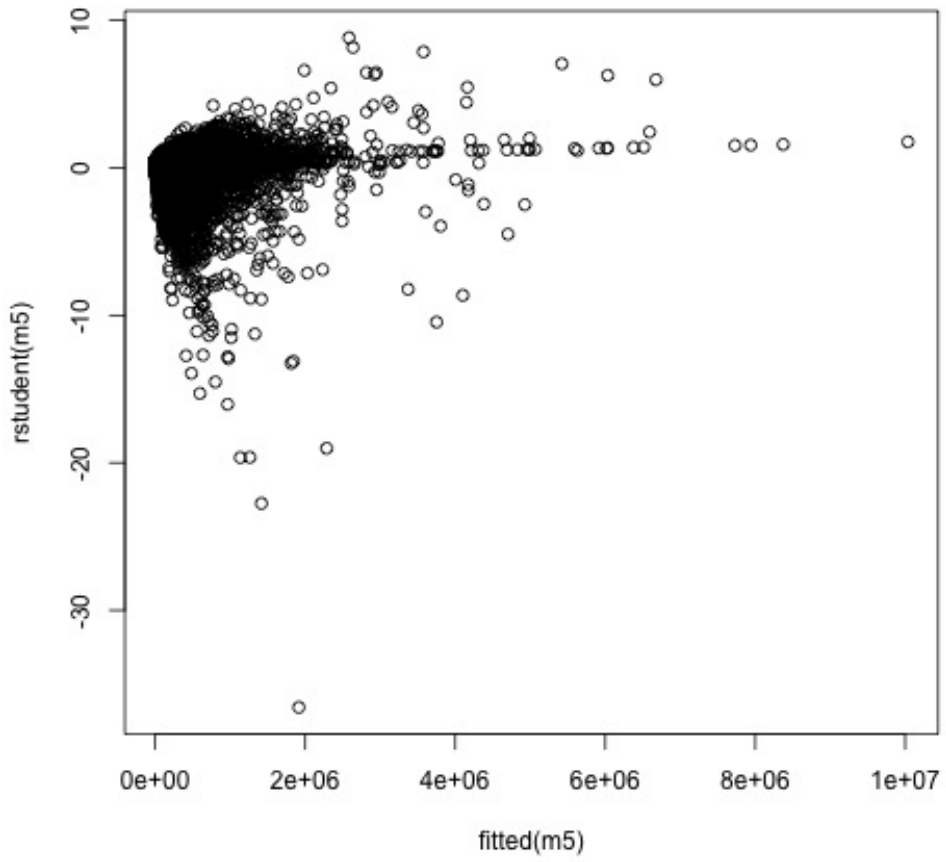
```
Planned.revenue:Product.lineGolf Equipment           -28.8168 < 2.2e-16 ***
Planned.revenue:Product.lineMountaineering Equipment   8.0050 1.247e-15 ***
Planned.revenue:Product.lineOutdoor Protection         2.4819 0.0130748 *
Planned.revenue:Product.linePersonal Accessories     127.6525 < 2.2e-16 ***
Planned.revenue:Order.method.typeFax                   0.1481 0.8822654
Planned.revenue:Order.method.typeMail                  0.7051 0.4807373
Planned.revenue:Order.method.typeSales visit           0.2560 0.7979186
Planned.revenue:Order.method.typeSpecial               0.3749 0.7077323
Planned.revenue:Order.method.typeTelephone            -0.7766 0.4373857
Planned.revenue:Order.method.typeWeb                   9.0594 < 2.2e-16 ***
Planned.revenue:Retailer.countryUnited Kingdom         1.7518 0.0798171 .
Planned.revenue:Retailer.countrySwitzerland            2.9105 0.0036119 **
Planned.revenue:Retailer.countrySweden                 2.6481 0.0081006 **
Planned.revenue:Retailer.countrySpain                  1.9460 0.0516639 .
Planned.revenue:Retailer.countrySingapore              1.6317 0.1027536
Planned.revenue:Retailer.countryNetherlands            3.4996 0.0004667 ***
Planned.revenue:Retailer.countryMexico                 1.6062 0.1082378
Planned.revenue:Retailer.countryKorea                 -0.2112 0.8327056
Planned.revenue:Retailer.countryJapan                  2.4472 0.0144031 *
Planned.revenue:Retailer.countryItaly                  2.2539 0.0242082 *
Planned.revenue:Retailer.countryGermany                1.8498 0.0643540 .
Planned.revenue:Retailer.countryFrance                 2.4836 0.0130132 *
Planned.revenue:Retailer.countryFinland                1.2567 0.2088615
Planned.revenue:Retailer.countryDenmark                1.5782 0.1145384
Planned.revenue:Retailer.countryChina                 -0.7665 0.4433737
Planned.revenue:Retailer.countryCanada                 4.4510 8.585e-06 ***
Planned.revenue:Retailer.countryBrazil                 1.7832 0.0745597 .
Planned.revenue:Retailer.countryBelgium                2.5739 0.0100626 *
Planned.revenue:Retailer.countryAustria                3.3238 0.0008894 ***
Planned.revenue:Retailer.countryAustralia              2.0155 0.0438594 *
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

Diagnostics

```
> hist(resid(m5));
```

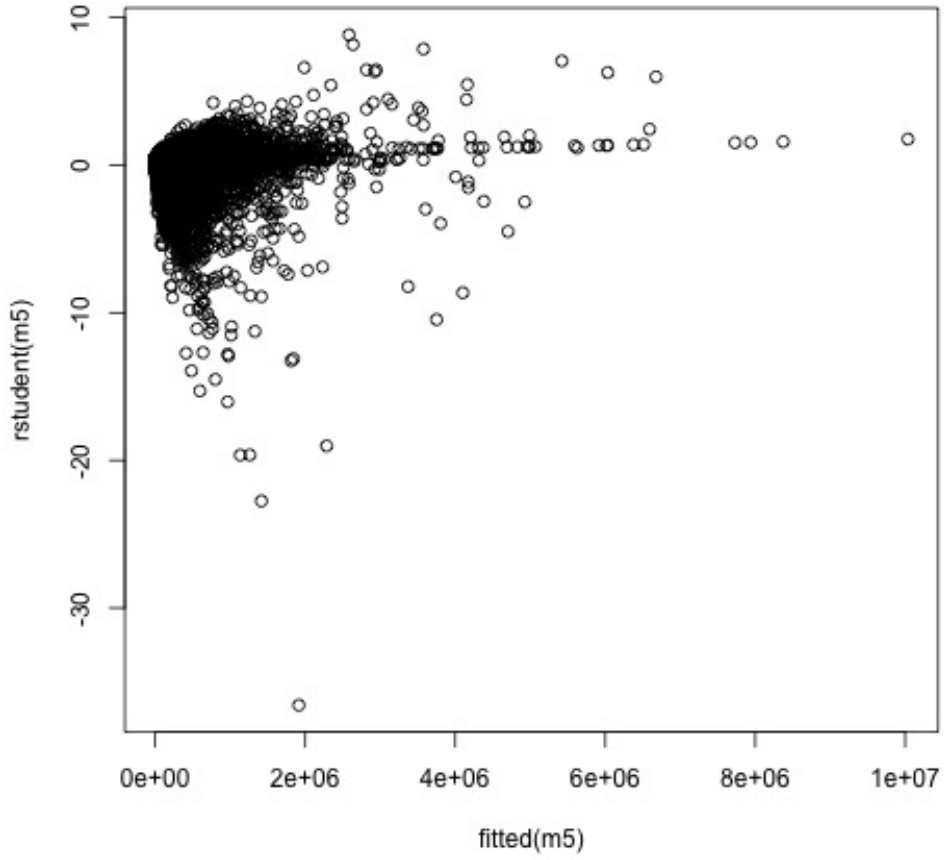


```
> plot(fitted(m5), rstudent(m5));
```




```
> ncvTest(m5);

Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 34919.94    Df = 1     p = 0
```

```
> influencePlot(m5);
```




```
         StudRes         Hat      CookD
38263 -36.582379 0.007784885 0.40081267
43857   1.081115 0.106303934 0.04735376
```

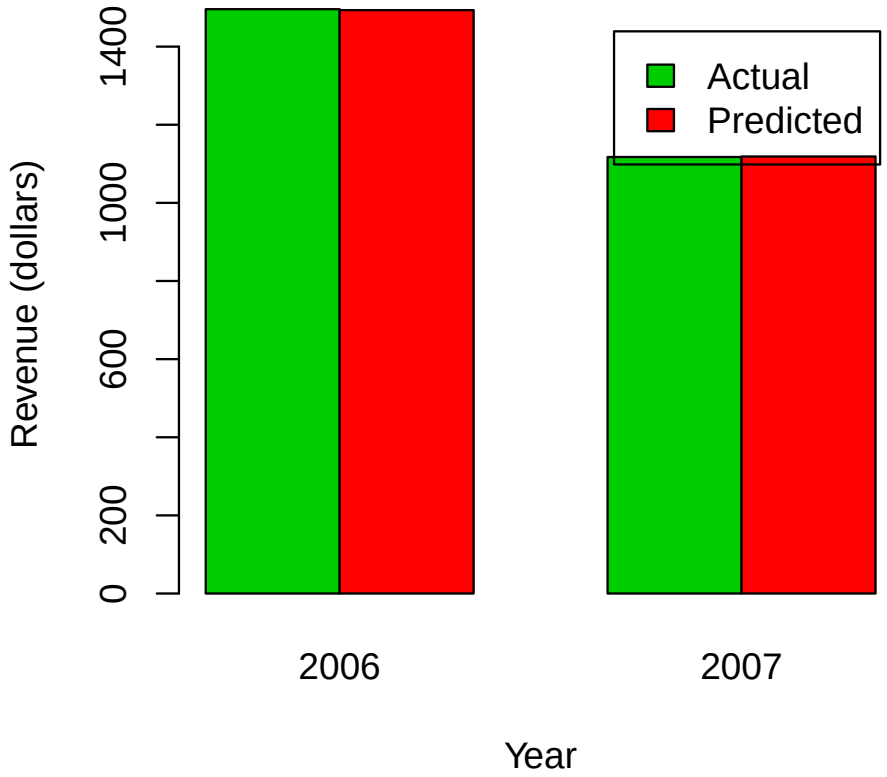
```
> outlierTest(m5);
```

```
       rstudent unadjusted p-value Bonferonni p
38263 -36.58238        2.2557e-285  5.5812e-281
80599 -22.75415        1.9201e-113  4.7508e-109
80893 -19.64331         2.9569e-85   7.3162e-81
43555 -19.63176         3.6991e-85   9.1527e-81
59431 -19.01793         4.5341e-80   1.1219e-75
17095 -16.03316         1.4651e-57   3.6250e-53
38269 -15.29494         1.4404e-52   3.5639e-48
69280 -14.51817         1.4604e-47   3.6134e-43
81481 -13.93109         6.0167e-44   1.4887e-39
59725 -13.25201         6.0157e-40   1.4885e-35
```

Backtesting

We have already build the model using data from 2004 and 2005; below we forecast and aggregate the predicted revenues by year. The model slightly underpredicts the revenues for 2006 and slightly overpredicts the revenues for 2007.

```
> pred = predict(m5, rs2Test);
> rs2Test = data.frame(rs2Test, pred);
> actual = tapply(rs2Test$Revenue, rs2Test$Year, sum) / 1000000;
> pred = tapply(rs2Test$pred, rs2Test$Year, sum) / 1000000;

> barplot(rbind(actual, pred), beside=TRUE,
+ names.arg=c("2006", "2007"), col=c(3,2), legend.text=c("Actual", "Predicted"),
+ xlab="Year",
+ ylab="Revenue (dollars)")
```



Other Questions

- Estimation issues: the zero mean condition is not satisfied, there are likely endogeneity issues (e.g. marketing), and there are time varying-effects since this is panel data.
- Suggested IV: Could use the planned quantity, cost and price information to model planned revenue.