HW7

Anya Conti

March 27, 2017

```
##
## Attaching package: 'effects'

## The following object is masked from 'package:car':
##
##     Prestige
```

6.10.1.

```
##
## Call:
## lm(formula = quality ~ gender + numYears + pepper + discipline +
##     easiness + raterInterest, data = rateprof)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.63978 -0.42534  0.03105  0.41535  1.26088
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -0.18066    0.24240  -0.745  0.45658
## gendermale          0.04678    0.06492   0.721  0.47162
## numYears            0.01760    0.01005   1.751  0.08085 .
## pepperyes           0.56166    0.09934   5.654 3.22e-08 ***
## disciplinePre-prof  0.09656    0.09139   1.057  0.29144
## disciplineSocSci    0.01865    0.08889   0.210  0.83393
## disciplineSTEM      0.29475    0.08148   3.618  0.00034 ***
## easiness            0.51288    0.04245  12.082  < 2e-16 ***
## raterInterest       0.54413    0.05937   9.165  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5892 on 357 degrees of freedom
## Multiple R-squared:  0.5158, Adjusted R-squared:  0.505
## F-statistic: 47.54 on 8 and 357 DF,  p-value: < 2.2e-16
```

[Test 1] $H_0 : \beta_2 = 0$ vs $H_A : \beta_2 \neq 0$ $\alpha = 0.05$

```
pt(q=coef(summary(lm.profqual))[3,3], df = 357, lower.tail = FALSE)*2
```

```
## [1] 0.08084827
```

Since the p-value of 0.0808 is greater than 0.05, the test is not statistically significant. As a result, we fail to reject the null hypothesis that $\beta_2 = 0$.

[Test 2] $H_0 : \beta_2 = 0$ vs $H_A : \beta_2 \leq 0$ $\alpha = 0.05$

```
pt(q=coef(summary(lm.profqual))[3,3], df = 357, lower.tail = TRUE)
```

```
## [1] 0.9595759
```

Since the p-value of 0.9596 is greater than 0.05, the test is not statistically significant. As a result, we fail to reject the null hypothesis that $\beta_2 = 0$.

[Test 3] $H_0 : \beta_2 = 0$ vs $H_A : \beta_2 \geq 0$ $\alpha = 0.05$

```
pt(q=coef(summary(lm.profqual))[3,3], df = 357, lower.tail = FALSE)
```

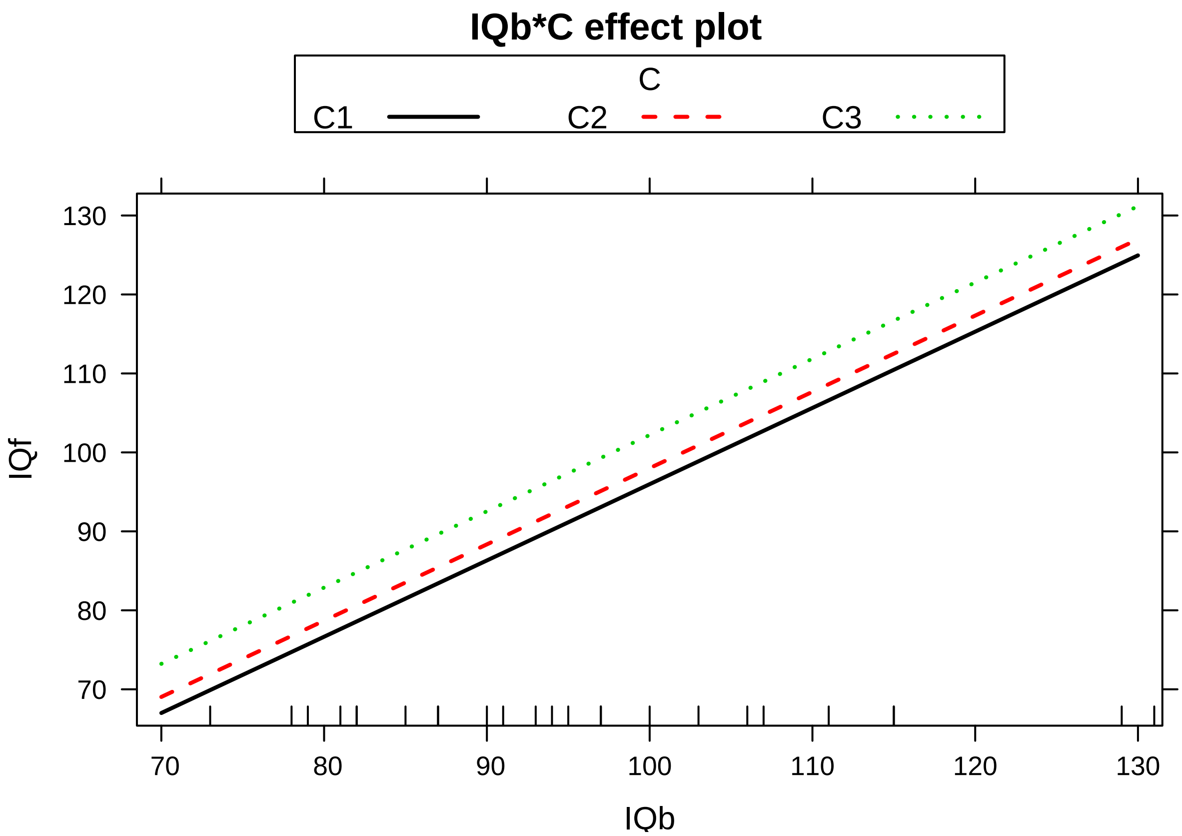
```
## [1] 0.04042413
```

Since the p-value of 0.0404 is less than 0.05, the test is statistically significant. As a result, we reject the null hypothesis that $\beta_2 = 0$ in favor of the alternative hypothesis that $\beta_2 \geq 0$.

6.12. Here is the linear model of $IQf = \beta_0 + \beta_1 IQb + \beta_2 CC2 + \beta_3 CC3$ where IQf is the IQ of the twin raised in a foster home, IQb is the IQ of the twin raised by the birth parents, CC2 is whether or not the birth parents are middle class (0 is no, 1 if yes), and CC3 is whether or not the birth parents are upper class (0 if no, 1 if yes).

```
##
## Call:
## lm(formula = IQf ~ IQb + C, data = twins)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.8235  -5.2366  -0.1111   4.4755  13.6978
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.6076    11.8551  -0.051    0.960
## IQb           0.9658     0.1069   9.031 5.05e-09 ***
## CC2           2.0353     4.5908   0.443    0.662
## CC3           6.2264     3.9171   1.590    0.126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.571 on 23 degrees of freedom
## Multiple R-squared:  0.8039, Adjusted R-squared:  0.7784
## F-statistic: 31.44 on 3 and 23 DF,  p-value: 2.604e-08
```

Here is the effects plot of the relationship between IQb and IQf, separated by class categories.



Here is an F-test for whether or not class has a significant effect on IQf given IQb.
$H_0 : \beta_2 = \beta_3 = 0$ vs H_A : at least one (β_2 or β_3) is $\neq 0$ Let $\alpha = 0.05$

```
## Linear hypothesis test
##
## Hypothesis:
## CC2 = 0
## CC3 = 0
##
## Model 1: restricted model
## Model 2: IQf ~ IQb + C
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     25 1493.5
## 2     23 1318.4  2    175.13 1.5276 0.2383
```

Since the p-value of this test of 0.2383 is greater than 0.05, we fail to reject the null hypothesis that $\beta_2 = \beta_3 = 0$. As such, the new model would be the following.

```
##
## Call:
## lm(formula = IQf ~ IQb, data = twins)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.3512  -5.7311   0.0574   4.3244  16.3531
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  9.20760    9.29990   0.990    0.332
## IQb          0.90144    0.09633   9.358  1.2e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.729 on 25 degrees of freedom
## Multiple R-squared:  0.7779, Adjusted R-squared:  0.769
## F-statistic: 87.56 on 1 and 25 DF,  p-value: 1.204e-09
```

6.14.

6.14.1. Here is the model $ln(acrePrice) = \beta_0 + \beta_1 year$ where year is a continuous variable. According to this model, for every increase in year by 1, there will be 10.05% increase in acre price.

```
##
## Call:
## lm(formula = log(acrePrice) ~ year, data = MN)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1008 -0.3773  0.1285  0.4365  2.2624
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.939e+02  3.984e+00  -48.67   <2e-16 ***
## year         1.005e-01  1.985e-03   50.60   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6808 on 18698 degrees of freedom
## Multiple R-squared:  0.1204, Adjusted R-squared:  0.1204
## F-statistic:  2560 on 1 and 18698 DF,  p-value: < 2.2e-16
```

6.14.2. Here is the model where year is not treated as a continuous variable, but rather a factor, where there is a individual coefficient estimate for the effect of each year on the log of acre price. According to this model, for every increase in year by 1, there will be 10.05% increase in acre price.

```
##
## Call:
## lm(formula = log(acrePrice) ~ 1 + fyear, data = MN)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9499 -0.3785  0.1301  0.4354  2.3456
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.27175    0.02848 255.345  < 2e-16 ***
## fyear2003   -0.00155    0.03207  -0.048    0.961
## fyear2004    0.14794    0.03155   4.689 2.76e-06 ***
## fyear2005    0.36026    0.03176  11.343  < 2e-16 ***
## fyear2006    0.39392    0.03195  12.329  < 2e-16 ***
## fyear2007    0.47682    0.03186  14.965  < 2e-16 ***
## fyear2008    0.68364    0.03162  21.620  < 2e-16 ***
## fyear2009    0.71407    0.03355  21.284  < 2e-16 ***
## fyear2010    0.75733    0.03260  23.231  < 2e-16 ***
```

```
## fyear2011    0.72071    0.03526  20.437  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6775 on 18690 degrees of freedom
## Multiple R-squared:  0.1293, Adjusted R-squared:  0.1289
## F-statistic: 308.5 on 9 and 18690 DF,  p-value: < 2.2e-16
```

6.14.3. Show that model A is a special case of model B, and so a hypothesis test of NH: model A versus AH: model B is reasonable Model A = Model B if in Model B, $2\beta_{b1} = \beta_{b2}, 3\beta_{b1} = \beta_{b3}, ..., 9\beta_{b9} = \beta_{b9}$ This is because In Model B, β_{b1} corresponds to the difference between 2003 and a base year given as 2002. Thus in model A, this would correspond to a 1 year increase in year. In Model B, β_{b2} corresponds to the difference between 2004 and a base year given as 2002. Thus in model A, this would correspond to a 2 year increase in year. Thus $2\beta_{b1} = \beta_{b2}$. Something similar can be shown for β_{b3} through β_{b9} as well. If these relationships hold, then it is a linear effect by the year. Thus we can test this restriction on Model B to see if Model A is appropriate.

6.14.4. Here the test discussed above is performed. Perform an F-test for the restrictions.
H_0 : model A OR
$2\beta_{b1} = \beta_{b2}, 3\beta_{b1} = \beta_{b3}, ..., 9\beta_{b9} = \beta_{b9}$ OR
$R\beta_b = r$ where r is the first matrix shown below and R is the second matrix shown below.
H_A : model B OR
at least one of those equations above is not true so there is not a linear relationship OR
$R\beta_b \neq r$ where r is the first matrix shown below and R is the second matrix shown below.
$\alpha = 0.05$

```
##      [,1]
## [1,]    0
## [2,]    0
## [3,]    0
## [4,]    0
## [5,]    0
## [6,]    0
## [7,]    0
## [8,]    0
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## [1,]    0    2   -1    0    0    0    0    0    0     0
## [2,]    0    3    0   -1    0    0    0    0    0     0
## [3,]    0    4    0    0   -1    0    0    0    0     0
## [4,]    0    5    0    0    0   -1    0    0    0     0
## [5,]    0    6    0    0    0    0   -1    0    0     0
## [6,]    0    7    0    0    0    0    0   -1    0     0
## [7,]    0    8    0    0    0    0    0    0   -1     0
## [8,]    0    9    0    0    0    0    0    0    0    -1
```

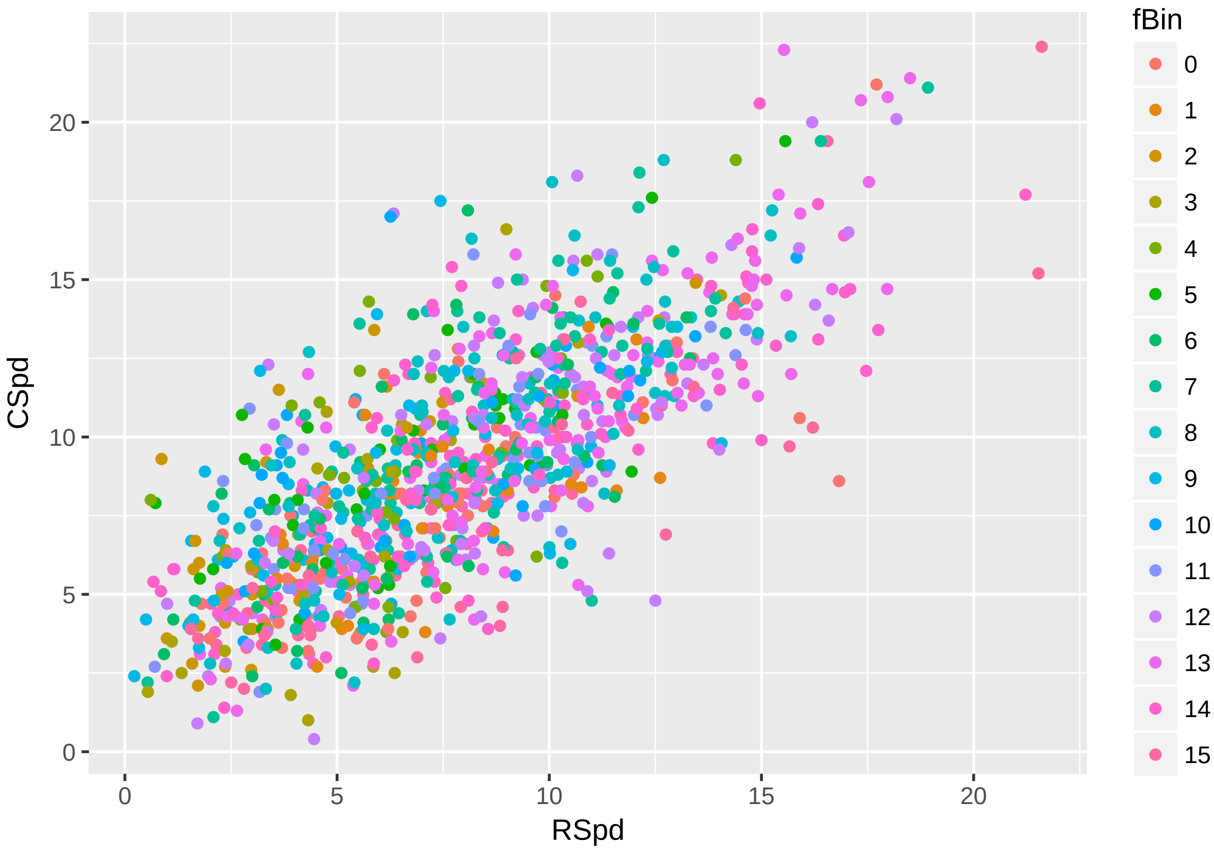
```
## Linear hypothesis test
##
## Hypothesis:
## 2 fyear2003 - fyear2004 = 0
## 3 fyear2003 - fyear2005 = 0
## 4 fyear2003 - fyear2006 = 0
## 5 fyear2003 - fyear2007 = 0
## 6 fyear2003 - fyear2008 = 0
## 7 fyear2003 - fyear2009 = 0
## 8 fyear2003 - fyear2010 = 0
```

```
## 9 fyear2003 - fyear2011 = 0
##
## Model 1: restricted model
## Model 2: log(acrePrice) ~ 1 + fyear
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1  18698 8666.9
## 2  18690 8579.2  8    87.686 23.878 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Because the p-value for the test is less than 2.2e-16 which means it is far below 0.05, we reject the null hypothesis that model A is a reasonable model, in favor of the null hypothesis that model B is the true model and that there is not a linear relationship easily shown between years.

6.18.1. Here is a graphical representation of the data, where CSpd is on the y-axis, RSpd is on the x-axis, and the bin is represented by color.

```
##
## Call:
## lm(formula = CSpd ~ RSpd + fBin, data = wm)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.7698 -1.4613 -0.1402  1.4415  9.2319
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.30396    0.33677   6.841  1.3e-11 ***
## RSpd         0.76223    0.02011  37.902  < 2e-16 ***
## fBin1        0.09692    0.49671   0.195 0.845331
## fBin2        1.26317    0.50253   2.514 0.012092 *
## fBin3        0.33772    0.51083   0.661 0.508668
## fBin4        1.78884    0.50525   3.541 0.000416 ***
## fBin5        1.68542    0.48894   3.447 0.000588 ***
## fBin6        1.07114    0.45102   2.375 0.017723 *
## fBin7        1.38259    0.39951   3.461 0.000559 ***
## fBin8        1.29759    0.39602   3.277 0.001084 **
## fBin9        1.23018    0.42199   2.915 0.003626 **
## fBin10       1.16865    0.46090   2.536 0.011364 *
## fBin11       0.67576    0.42502   1.590 0.112128
## fBin12       0.73804    0.38681   1.908 0.056649 .
## fBin13       0.76149    0.37146   2.050 0.040603 *
## fBin14       0.50096    0.37836   1.324 0.185767
## fBin15      -0.59391    0.40609  -1.462 0.143891
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.411 on 1099 degrees of freedom
## Multiple R-squared:  0.5955, Adjusted R-squared:  0.5896
## F-statistic: 101.1 on 16 and 1099 DF,  p-value: < 2.2e-16
```



fBin
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
CSpd
RSpd
0
5
10
15
20