STAT3032_Homework3

Guizhen Yu

October 14, 2017

Answer for 5.3

Answer for 5.3.1

```
UN11$groupF <- factor(UN11$group, c("other", "oecd", "africa"))
m531 <- lm(lifeExpF ~ groupF, UN11)
summary(m531)
```

```
##
## Call:
## lm(formula = lifeExpF ~ groupF, data = UN11)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.8367  -3.3045   0.3635   2.7183  18.2277
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   75.3267     0.5856 128.627  < 2e-16 ***
## groupFoecd     7.1197     1.2709   5.602  7.1e-08 ***
## groupFafrica -15.5545     1.0426 -14.918  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.28 on 196 degrees of freedom
## Multiple R-squared:  0.6191, Adjusted R-squared:  0.6152
## F-statistic: 159.3 on 2 and 196 DF,  p-value: < 2.2e-16
```

```
lsmeans(m531, pairwise ~ groupF)
```

```
## $lsmeans
##  groupF   lsmean        SE  df lower.CL upper.CL
##  other  75.32674 0.5856229 196 74.17181 76.48167
##  oecd   82.44645 1.1279404 196 80.22199 84.67091
##  africa 59.77226 0.8626389 196 58.07102 61.47351
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast        estimate       SE  df t.ratio p.value
##  other - oecd   -7.119708 1.270907 196  -5.602  <.0001
##  other - africa 15.554479 1.042641 196  14.918  <.0001
##  oecd - africa  22.674187 1.419998 196  15.968  <.0001
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

The table in R gets the same conclusion as table 5.2.

Answer for 5.3.2

```
UN11$groupF <- factor(UN11$group, c("other", "oecd", "africa"))
m532 <- lm(lifeExpF ~ groupF + log(ppgdp), UN11)
lsmeans(m532, pairwise ~ groupF)
```

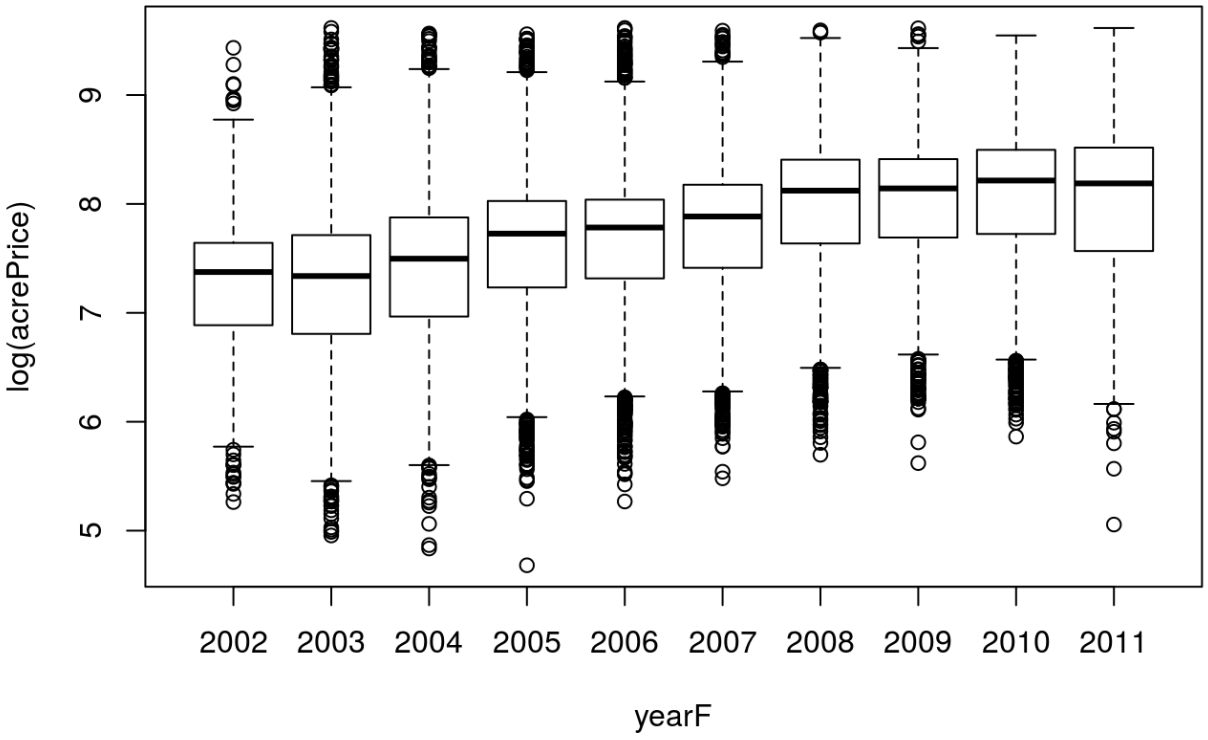
```
## $lsmeans
##  groupF   lsmean        SE  df lower.CL upper.CL
##  other  78.09529 0.5502648 195 77.01006 79.18053
##  oecd   79.62997 0.9594388 195 77.73777 81.52218
##  africa 67.45961 1.0377456 195 65.41296 69.50626
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast        estimate        SE  df t.ratio p.value
##  other - oecd   -1.534683 1.1736824 195  -1.308  0.3927
##  other - africa 10.635683 0.9791766 195  10.862  <.0001
##  oecd - africa  12.170365 1.5574486 195   7.814  <.0001
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

Compare to table 5.2, the mean of “other - oecd” changes sign when we use model in 5.3.2. Compare to table 5.2, the mean of “oecd - africa” decrease 10. Compare to table 5.2, the mean of “other - africa” decrease 2.

Answer for 5.4

Answer for 5.4.1

```
MinnLand$yearF <- factor(MinnLand$year)
plot(log(acrePrice) ~ yearF, MinnLand)
```



Based on the graph, the pattern described in question did not apparently repeated in Minnesota farm sales. Housing sales prices in Minnesota have a slight decrease in 2003, then keep increasing from 2003 to 2010, finally have a slight decrease in 2011.

Answer for 5.4.2

```
m542 <- lm(log(acrePrice) ~ yearF, MinnLand)
summary(m542)
```

```
##
## Call:
## lm(formula = log(acrePrice) ~ yearF, data = MinnLand)
##
## 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 ***
## yearF2003   -0.00155    0.03207  -0.048    0.961
## yearF2004    0.14794    0.03155   4.689 2.76e-06 ***
## yearF2005    0.36026    0.03176  11.343  < 2e-16 ***
## yearF2006    0.39392    0.03195  12.329  < 2e-16 ***
## yearF2007    0.47682    0.03186  14.965  < 2e-16 ***
## yearF2008    0.68364    0.03162  21.620  < 2e-16 ***
## yearF2009    0.71407    0.03355  21.284  < 2e-16 ***
## yearF2010    0.75733    0.03260  23.231  < 2e-16 ***
## yearF2011    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
```

Parameter Interpretation The mean of log(acrePrice) in 2002 is 7.27. The difference between 2003 and 2002's mean of log(acrePrice) is -0.0015 The difference between 2004 and 2002's mean of log(acrePrice) is 0.148 The difference between 2005 and 2002's mean of log(acrePrice) is 0.360 The difference between 2006 and 2002's mean of log(acrePrice) is 0.394 The difference between 2007 and 2002's mean of log(acrePrice) is 0.476 The difference between 2008 and 2002's mean of log(acrePrice) is 0.683 The difference between 2009 and 2002's mean of log(acrePrice) is 0.714 The difference between 2010 and 2002's mean of log(acrePrice) is 0.757 The difference between 2011 and 2002's mean of log(acrePrice) is 0.721

the coefficient estimate 7.27 is about 255.3 standard deviations from 0. Therefore, the intercept for 2002 is not equal to zero. the coefficient estimate -0.0015 is about -0.048 standard deviations from 0. Therefore, the intercept for 2003 is not differs from the intercept for 2002. the coefficient estimate 0.360 is about 4.68 standard deviations from 0. Therefore, the intercept for 2004 is differs from the intercept for 2002. the coefficient estimate 0.394 is about 11.34 standard deviations from 0. Therefore, the intercept for 2005 is differs from the intercept for 2002. the coefficient estimate 0.476 is about 12.32 standard deviations from 0. Therefore, the intercept for 2006 is differs from the intercept for 2002. the coefficient estimate 0.683 is about 14.97 standard deviations from 0. Therefore, the intercept for 2007 is differs from the intercept for 2002. the coefficient estimate 0.714 is about 21.62 standard deviations

from 0. Therefore, the intercept for 2008 is differs from the intercept for 2002. the coefficient estimate 0.757 is about 21.28 standard deviations from 0. Therefore, the intercept for 2009 is differs from the intercept for 2002. the coefficient estimate 0.721 is about 23.23 standard deviations from 0. Therefore, the intercept for 2010 is differs from the intercept for 2002. the coefficient estimate 0.148 is about 20.44 standard deviations from 0. Therefore, the intercept for 2011 is differs from the intercept for 2002.

```
m543 <- lm(log(acrePrice) ~ 0 + yearF, MinnLand)
summary(m543)
```

```
##
## Call:
## lm(formula = log(acrePrice) ~ 0 + yearF, data = MinnLand)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9499 -0.3785  0.1301  0.4354  2.3456
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## yearF2002  7.27175    0.02848   255.3   <2e-16 ***
## yearF2003  7.27020    0.01474   493.4   <2e-16 ***
## yearF2004  7.41969    0.01358   546.5   <2e-16 ***
## yearF2005  7.63201    0.01406   542.7   <2e-16 ***
## yearF2006  7.66567    0.01449   529.1   <2e-16 ***
## yearF2007  7.74857    0.01429   542.2   <2e-16 ***
## yearF2008  7.95539    0.01374   578.9   <2e-16 ***
## yearF2009  7.98582    0.01774   450.2   <2e-16 ***
## yearF2010  8.02908    0.01587   506.0   <2e-16 ***
## yearF2011  7.99246    0.02080   384.3   <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:  0.9923, Adjusted R-squared:  0.9923
## F-statistic: 2.417e+05 on 10 and 18690 DF,  p-value: < 2.2e-16
```

The standard errors of the regression coefficients are not the same as these standard errors because the standard error of the regression coefficients are estimates of the standard deviation in populations.

Answer for 5.8

Answer for 5.8.1

```
m581 <- lm(Y ~ 1 + X1 + I(X1^2) + X2 + I(X2^2) + I(X1*X2), data = cakes)
summary(m581)
```

```
##
## Call:
## lm(formula = Y ~ 1 + X1 + I(X1^2) + X2 + I(X2^2) + I(X1 * X2),
##     data = cakes)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4912 -0.3080  0.0200  0.2658  0.5454
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.204e+03  2.416e+02  -9.125 1.67e-05 ***
## X1           2.592e+01  4.659e+00   5.563 0.000533 ***
## I(X1^2)     -1.569e-01  3.945e-02  -3.977 0.004079 **
## X2           9.918e+00  1.167e+00   8.502 2.81e-05 ***
## I(X2^2)     -1.195e-02  1.578e-03  -7.574 6.46e-05 ***
## I(X1 * X2)  -4.163e-02  1.072e-02  -3.883 0.004654 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4288 on 8 degrees of freedom
## Multiple R-squared:  0.9487, Adjusted R-squared:  0.9167
## F-statistic:  29.6 on 5 and 8 DF,  p-value: 5.864e-05
```

The significance levels for X1^2, X2^2, X1*X2 are 0.004, 6.46e-05, 0.004 seperately. all these values are smaller than 0.005

Answer for 5.8.2

```
cakes$blockF <- factor(cakes$block)
m582 <- lm(Y ~ 1 + blockF + X1 + I(X1^2) + X2 + I(X2^2) + I(X1*X2), data = ca
kes)
summary(m582)
```

```
##
## Call:
## lm(formula = Y ~ 1 + blockF + X1 + I(X1^2) + X2 + I(X2^2) + I(X1 *
##     X2), data = cakes)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4525 -0.3046  0.0200  0.2924  0.4883
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.205e+03  2.542e+02  -8.672 5.43e-05 ***
## blockF1      1.143e-01  2.412e-01   0.474 0.650014
## X1           2.592e+01  4.903e+00   5.287 0.001140 **
## I(X1^2)     -1.569e-01  4.151e-02  -3.779 0.006898 **
## X2           9.918e+00  1.228e+00   8.080 8.56e-05 ***
## I(X2^2)     -1.195e-02  1.660e-03  -7.197 0.000178 ***
## I(X1 * X2)  -4.163e-02  1.128e-02  -3.690 0.007754 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4512 on 7 degrees of freedom
## Multiple R-squared:  0.9503, Adjusted R-squared:  0.9077
## F-statistic: 22.31 on 6 and 7 DF,  p-value: 0.0003129
```

Answer for 5.10

Answer for 5.10.1

The first model is a main effect model, and the second model is a model with intercept term. Their main difference is the first model do not have intercept term, while the second model have intercept term. This means in the first model region is not affect by year.

Answer for 5.10.2

```
MinnLand$yearF <- as.factor(MinnLand$year)
MinnLand$regionF <- as.factor(MinnLand$region)
m5102 <- lm(log(acrePrice) ~ yearF*regionF, data = MinnLand)
summary(m5102)
```

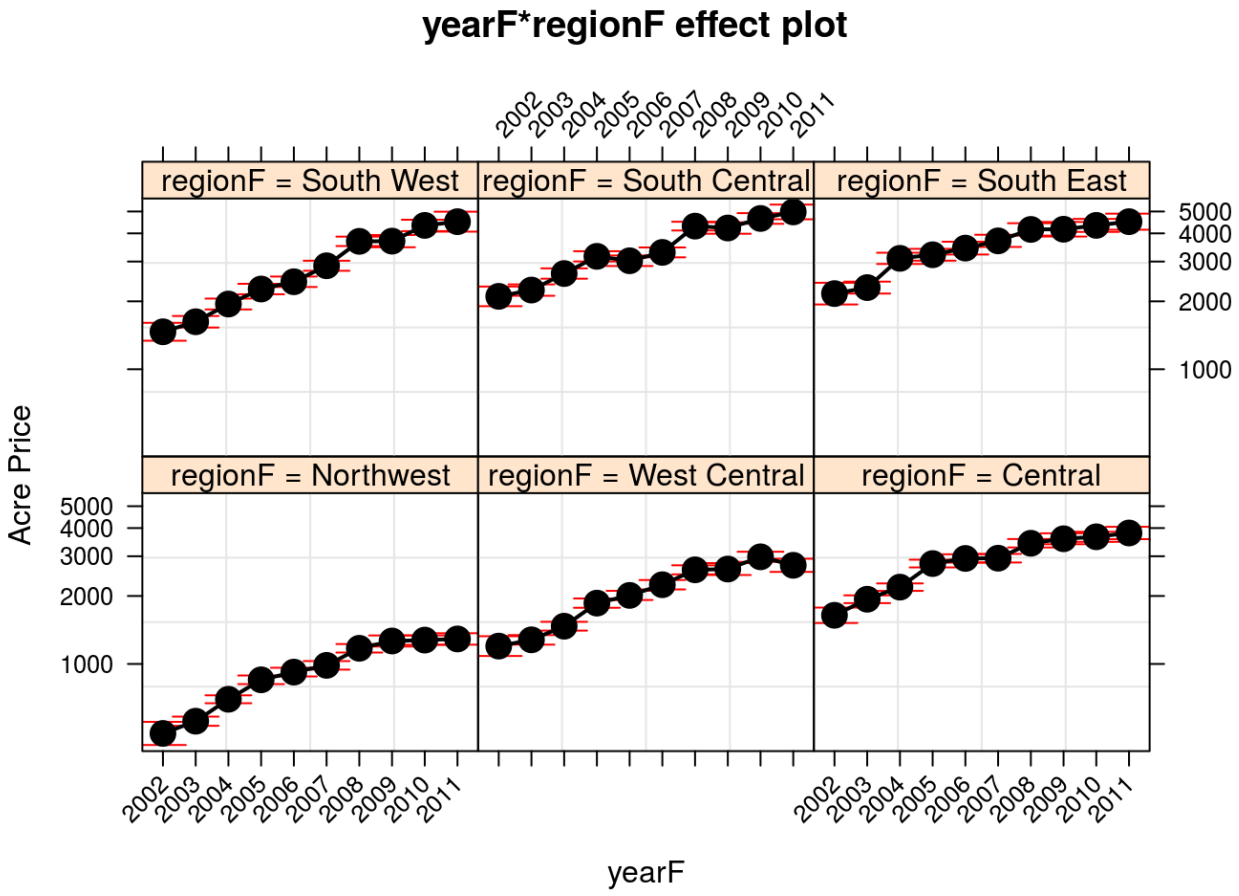
```
##
## Call:
## lm(formula = log(acrePrice) ~ yearF * regionF, data = MinnLand)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.73006 -0.27521  0.01157  0.25561  2.64607
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     6.19911    0.06022 102.939  < 2e-16 ***
## yearF2003                       0.12445    0.06476   1.922  0.05467 .
## yearF2004                       0.34837    0.06367   5.471 4.52e-08 ***
## yearF2005                       0.54665    0.06425   8.508  < 2e-16 ***
## yearF2006                       0.62531    0.06432   9.722  < 2e-16 ***
## yearF2007                       0.69422    0.06419  10.815  < 2e-16 ***
## yearF2008                       0.86828    0.06417  13.532  < 2e-16 ***
## yearF2009                       0.94283    0.06679  14.115  < 2e-16 ***
## yearF2010                       0.95188    0.06505  14.634  < 2e-16 ***
## yearF2011                       0.96351    0.06723  14.333  < 2e-16 ***
## regionFWest Central             0.89062    0.07915  11.253  < 2e-16 ***
## regionFCentral                  1.20484    0.07230  16.664  < 2e-16 ***
## regionFSouth West               1.09079    0.07613  14.328  < 2e-16 ***
## regionFSouth Central            1.45223    0.07896  18.392  < 2e-16 ***
## regionFSouth East               1.48043    0.08250  17.945  < 2e-16 ***
## yearF2003:regionFWest Central  -0.06041    0.08631  -0.700  0.48400
## yearF2004:regionFWest Central  -0.14535    0.08493  -1.711  0.08703 .
## yearF2005:regionFWest Central  -0.10822    0.08573  -1.262  0.20685
## yearF2006:regionFWest Central  -0.10811    0.08568  -1.262  0.20706
## yearF2007:regionFWest Central  -0.06810    0.08572  -0.794  0.42693
## yearF2008:regionFWest Central  -0.09024    0.08551  -1.055  0.29126
## yearF2009:regionFWest Central  -0.15673    0.08981  -1.745  0.08099 .
## yearF2010:regionFWest Central  -0.04117    0.08698  -0.473  0.63601
## yearF2011:regionFWest Central  -0.13921    0.09123  -1.526  0.12706
## yearF2003:regionFCentral        0.03938    0.07877   0.500  0.61715
## yearF2004:regionFCentral       -0.06105    0.07766  -0.786  0.43179
## yearF2005:regionFCentral       -0.01894    0.07830  -0.242  0.80887
## yearF2006:regionFCentral       -0.04535    0.07878  -0.576  0.56486
## yearF2007:regionFCentral       -0.11180    0.07888  -1.417  0.15636
## yearF2008:regionFCentral       -0.13345    0.07872  -1.695  0.09004 .
## yearF2009:regionFCentral       -0.16203    0.08284  -1.956  0.05049 .
## yearF2010:regionFCentral       -0.15092    0.08074  -1.869  0.06160 .
## yearF2011:regionFCentral       -0.12382    0.08462  -1.463  0.14342
## yearF2003:regionFSouth West    -0.02205    0.08522  -0.259  0.79580
## yearF2004:regionFSouth West    -0.06516    0.08377  -0.778  0.43671
## yearF2005:regionFSouth West    -0.11040    0.08394  -1.315  0.18842
```

```
## yearF2006:regionFSouth West    -0.11492    0.08399  -1.368  0.17127
## yearF2007:regionFSouth West    -0.02079    0.08352  -0.249  0.80339
## yearF2008:regionFSouth West     0.05438    0.08296   0.656  0.51215
## yearF2009:regionFSouth West    -0.01820    0.08741  -0.208  0.83508
## yearF2010:regionFSouth West     0.13339    0.08534   1.563  0.11805
## yearF2011:regionFSouth West     0.15965    0.09689   1.648  0.09942 .
## yearF2003:regionFSouth Central -0.06014    0.08766  -0.686  0.49271
## yearF2004:regionFSouth Central -0.11564    0.08592  -1.346  0.17838
## yearF2005:regionFSouth Central -0.13846    0.08626  -1.605  0.10848
## yearF2006:regionFSouth Central -0.26222    0.08656  -3.029  0.00246 **
## yearF2007:regionFSouth Central -0.24577    0.08595  -2.859  0.00425 **
## yearF2008:regionFSouth Central -0.15184    0.08513  -1.784  0.07451 .
## yearF2009:regionFSouth Central -0.24561    0.08923  -2.753  0.00592 **
## yearF2010:regionFSouth Central -0.15730    0.08721  -1.804  0.07129 .
## yearF2011:regionFSouth Central -0.10230    0.09262  -1.105  0.26938
## yearF2003:regionFSouth East    -0.06330    0.09128  -0.693  0.48806
## yearF2004:regionFSouth East     0.01118    0.08993   0.124  0.90108
## yearF2005:regionFSouth East    -0.15057    0.09083  -1.658  0.09741 .
## yearF2006:regionFSouth East    -0.16132    0.09183  -1.757  0.07897 .
## yearF2007:regionFSouth East    -0.15648    0.09113  -1.717  0.08598 .
## yearF2008:regionFSouth East    -0.21373    0.09131  -2.341  0.01926 *
## yearF2009:regionFSouth East    -0.28636    0.09530  -3.005  0.00266 **
## yearF2010:regionFSouth East    -0.25598    0.09252  -2.767  0.00567 **
## yearF2011:regionFSouth East    -0.22985    0.09718  -2.365  0.01803 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4818 on 18640 degrees of freedom
## Multiple R-squared:  0.5609, Adjusted R-squared:  0.5596
## F-statistic: 403.6 on 59 and 18640 DF,  p-value: < 2.2e-16
```

```
allEffects(m5102, transformation = list(link = log, inverse = exp))
```

```
##  model: log(acrePrice) ~ yearF * regionF
##
##  yearF*regionF effect
##       regionF
## yearF  Northwest West Central  Central South West South Central South East
##   2002  492.3127     1199.594 1642.465   1465.431      2103.470   2163.642
##   2003  557.5555     1278.931 1934.837   1623.433      2243.190   2300.079
##   2004  697.4860     1469.621 2189.146   1945.191      2654.671   3099.805
##   2005  850.4446     1859.686 2784.039   2266.846      3163.784   3215.132
##   2006  920.0490     2012.111 2933.398   2441.336      3024.319   3441.075
##   2007  985.6775     2243.626 2940.585   2873.612      3293.775   3704.436
##   2008 1173.0899     2611.757 3424.759   3687.007      4306.080   4163.453
##   2009 1263.8830     2632.896 3585.855   3694.265      4224.122   4171.430
##   2010 1275.3692     2982.294 3658.874   4338.009      4656.058   4339.191
##   2011 1290.2947     2735.426 3803.377   4505.536      4976.876   4506.179
```

```
plot(allEffects(m5102,transformation = list(link = log, inverse = exp)),
     x.var = "yearF", grid = TRUE, rotx = 45, ylab = "Acre Price")
```



```
plot(allEffects(m5102, transformation = list(link = log, inverse = exp)),
     x.var = 'yearF', z.var = 'regionF', multiline = TRUE,
     grid = TRUE, rotx = 45, ylab = 'Acre Price')
```