4.2

```
attach(Transact)
df = Transact
df['a'] = (t1+t2)/2
df['d'] = (t1-t2)
m1 = lm(df$time~t1+t2,df)
m1
m2 = lm(df$time~df$a+df$d, df)
m2
m3 = lm(df$time~df$t2+df$d, df)
m3
m4 = lm(df$time~df$t1+df$t2+df$a+df$d, df)
m4
```

4.2.1

They are aliased because there are more regressor terms in the model m4 than there were linearly independent quantities for the original model. The slope is uninterpretable in this model for a and d because they are defined solely in terms of t1 and t2.

4.2.2

Constant:

- The intercept stays the same because it represents the total minutes of labor when there are no orders, which is again essentially uninterpretable.
- The coefficient of determination and residuals.
- Degrees of freedom

Differs:

- Slope
- T-value, p-value
- Std. error

4.2.3

Since d is the difference between the number of t1 and t2 orders, t2's relationship to time is significantly affected when the new regressor d is present in m3. If the difference between t1 and t2(d) is held constant while t2 is changed, then the effect that t2 has on time will be significantly different since t1 will be changed as the difference is being held constant.

4.6

This coefficient means that for every unit increase pctUrban, fertility (Number of children per woman) decreases by about 1%.

*((e^-.01)-1)*100 = -.995*

5.4

5.4.1
No, this pattern is not repeated in Minnesota farm sales, the mean farm sales increases every year except for 2010 – 2011, the increase from year to year does seem to slow after 2006 however.

5.4.2
- The intercept is the estimated price of an acre in 2002.
- The slope represents the change in estimate from 2002 to the year in question.
- The t-statistic and p-value show that the change from 2002-2003 is not very significant.

attach(MinnLand)
boxplot(log(acrePrice)~year, MinnLand)

5.4.3
R Output:

```
fyear2002 fyear2003 fyear2004 fyear2005 fyear2006 fyear2007 fyear2008 fyear2009 fyear2010 fyear2011
 7.271749  7.270199  7.419694  7.632009  7.665669  7.748572  7.955386  7.985819  8.029081  7.992459
> ?with
> with(MinnLand, tapply(log(acrePrice), MinnLand$year, mean))
    2002     2003     2004     2005     2006     2007     2008     2009     2010     2011
7.271749 7.270199 7.419694 7.632009 7.665669 7.748572 7.955386 7.985819 8.029081 7.992459
```

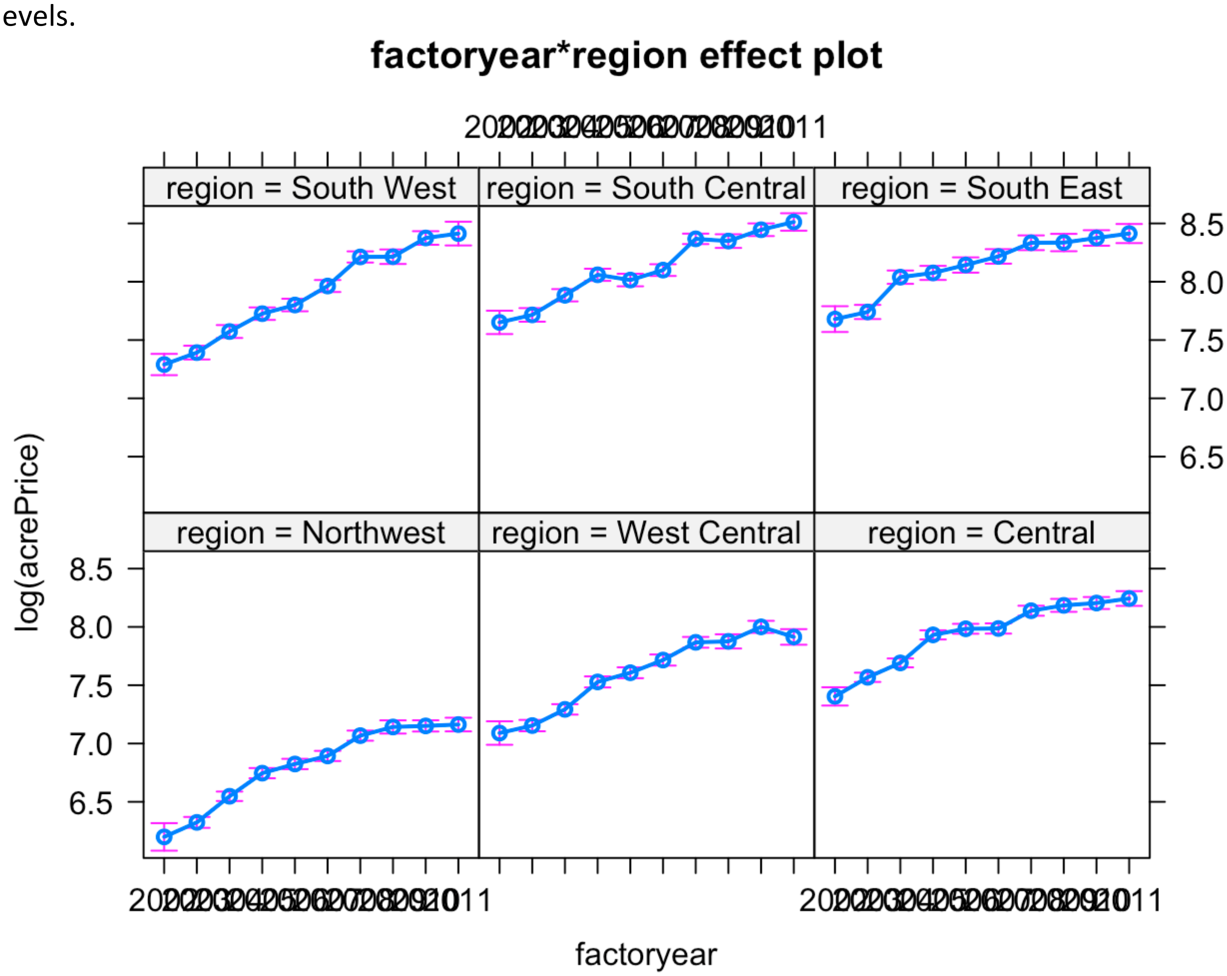
The parameter estimates for year as a factor are identical to the means for each group.

m1 = lm(log(acrePrice) ~ fyear-1, data=MinnLand)
coefficients(m1)
with(MinnLand, tapply(log(acrePrice), MinnLand$year, mean))

The standard errors are different because the standard deviation remains the same for each year in our regression model but the standard deviation changes from year to year for the sample mean.

5.10
5.10.1
In model b, it is a second order mean function so the model is accounting for year matched up with every category of region but in model a, it treats them completely separately.
a = lm(log(acrePrice) ~ factoryear + region)
b = lm(log(acrePrice) ~ factoryear + region + factoryear:region)
summary(a)
summary(b)

5.10.2
It appears that there is a main effect present and the response mean is different across factor levels.



plot(allEffects(b))

3. Problem 5.4 (of textbook) [40pts]

4. Problem 5.10 (of textbook) [20pts]