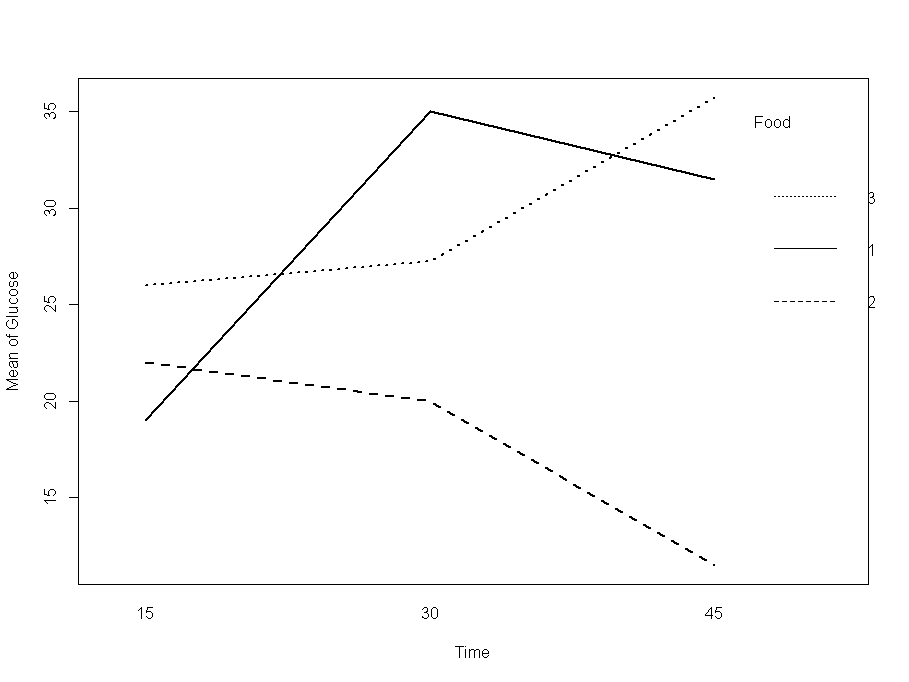
**Stat 510 Week 10 Homework**

**Submit answers in a Word document or a pdf to the designated ANGEL Drop Box in the Drop Boxes folder.**

1. Use the glucose.csv dataset from the Week 10 folder. The data are from a study in which the effects of three different foods on serum glucose levels in humans are compared. For each person, serum glucose level was measured at 15, 30, and 45 minutes after eating the food. There were four different persons per food, for a total n = 12 persons in the study.

Variables in the dataset are Food (1, 2, or 3), Subject (numbered 1 to 12), Time (15, 30, 45) and Gluc (short for Glucose, the response variable).

A. Create an interaction plot that gives mean response by food type and time. Give the plot and briefly describe the Food, Time, and interaction effects.



Above is the graph of mean glucose levels across time for the three types of food. The mean response varies for different type of food over time.

We see that the overall level first increased and then drops for food 1. While for food 2 it drops with time and for food 3 it increases with time.

B. Using the gls command in R, fit repeated measures models for each of the following assumptions about correlations among measurements at different times:

* Compound symmetry
* Completely unstructured with possibly unequal variances
* AR(1)
* AR(1) with unequal variances

Compare the AIC and BIC values for the four different models using the anova command. Give the results and explain which model looks to be the best.

Model df AIC BIC logLik Test L.Ratio p-value

fit.compsym 1 11 180.8759 195.1301 -79.43794

fit.nostruct 2 15 180.4815 199.9191 -75.24076 1 vs 2 8.394358 0.0782

fit.ar1 3 11 182.0474 196.3016 -80.02371 2 vs 3 9.565902 0.0484

fit.ar1het 4 13 183.5032 200.3491 -78.75158 3 vs 4 2.544248 0.2802

We see that the AIC for models 1 and 2 is approximately the same (slightly lower for 2) while the BIC is lowest for model 1. Based on the above figures, we predict that model 1 is the most suitable in this case.

C. For the best model, use the anova command to determine F-test results for Food, Time, and the Food\*Time interaction. Give the results and indicate which effects are statistically significant.

Compound symmetry for correlations:

numDF F-value p-value

(Intercept) 1 503.0710 <.0001

factor(Food) 2 11.1121 3e-04

factor(Time) 2 11.9493 2e-04

factor(Food):factor(Time) 4 30.5345 <.0001

We find that all Food, Time, and the Food\*Time interaction are significant. To go a step further we found that the 2nd model had approximately same AIC so let’s see its values:

Completely unstructured with possibly unequal variances

numDF F-value p-value

(Intercept) 1 543.2401 <.0001

factor(Food) 2 30.8966 <.0001

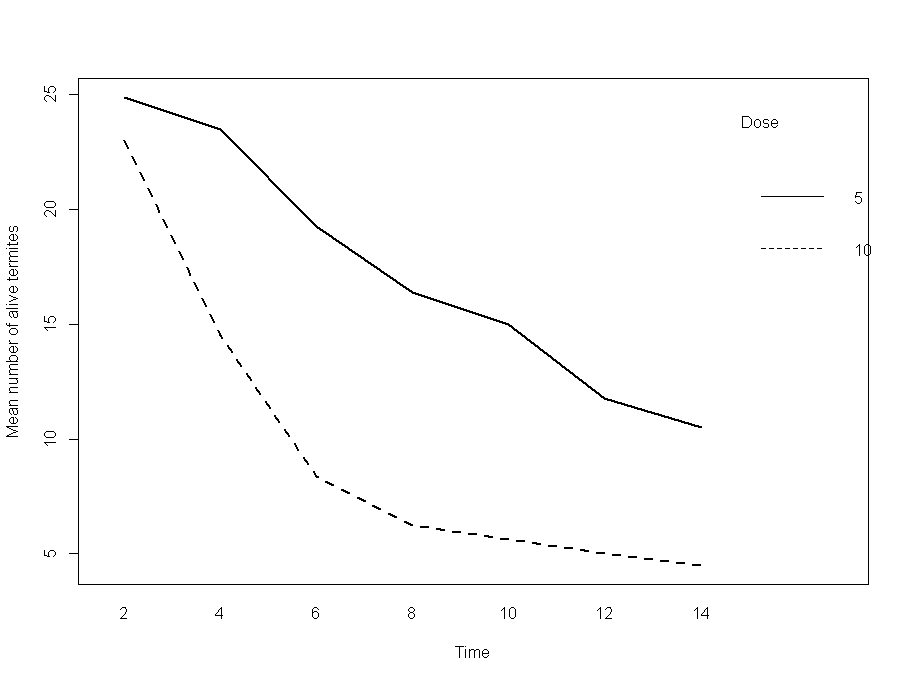
factor(Time) 2 6.7605 0.0042

factor(Food):factor(Time) 4 67.3480 <.0001

We see here too that all the factors are significant.

2. Use the dataset termites.csv from the Week 10 folder. The data are from an experiment to investigate the effects of certain tree resins on termites. A resin derived from the bark of tropical trees is dissolved in a solvent and placed on filter paper in different concentrations (5mg and 10mg). For each dosage level, eight dishes are set up with 25 termites in each dish. The Y variable is the number of termites still alive after a number of days. The y-variable is measured at 2, 4, 6, 8, 10, 12, and 14 days from the beginning of the experiment. Variable names in the dataset are Dish, Dose, Time, and Y.

A. Create an interaction plot that shows the mean response for each dose at each time. Time should be on the horizontal. Give the plot and write a brief description of the Dose, Time and Dose by Time interaction effects.



Above we have a graph of mean responses across time for the two doses. Note:

* We see that the overall level is higher for Dose=5.
* Generally the mean response (number of alive termites) decreases over time. This may point towards the behavior where interaction between Dose and Time is not significant.
* We also see the behavior that for Dose = 5 the response decreases almost linearly. On the other hand the response variable for Dose=10 decreases rapidly at first and gradually later. We will need further analysis to point out whether the interaction is significant or not.

B. Using gls in R, fit repeated measures models for each of the following assumptions about the correlations among observations at different times:

* Compound symmetry
* AR(1) with equal variances

Then, compare the AIC and BIC values using the anova command. Give the results and comment on which model provides the better fit.

The results of comparison:

Model df AIC BIC logLik

fit.compsym 1 16 549.8521 591.2115 -258.9260

fit.ar1 2 16 509.8885 551.2480 -238.9442

The AR(1) model for correlations among repeated measures gives the lowest AIC and BIC statistics, and therefore is the recommended model.

C. Use the anova command to get ANOVA F-test results for the model with the compound symmetry assumption. Give the results and briefly describe them.

The results for compound symmetry for correlations are:

anova(fit.compsym)

Denom. DF: 98

numDF F-value p-value

(Intercept) 1 140.3743 <.0001

factor(Dose) 1 11.5200 0.001

factor(Time) 6 107.7644 <.0001

factor(Dose):factor(Time) 6 7.2400 <.0001

We see that all the factors are significant. The various factors are:

* Dish is a “random” factor assumed to have mean 0 and an unknown constant variance. “Random” means that the dishes are considered to be a random sample from a larger population of dishes.
* The notation Dish(Dose) specifies that dishes were nested in doses meaning that different dishes were in the different treatment groups.
* The Dose factor measures whether the mean response differs for the 2 treatments when we average over all dishes and all times.
* The Time factor measures whether the mean response differs over time when we average over all dishes and all doses.
* The Time\*Dose interaction which is sensitive to whether the pattern across time depends upon the specific dose used.
* The errors are assumed to be independently normally distributed with mean 0 and constant variance.

From above we also see that the Time and Dose interaction is also significant.

D. Use the anova command to get ANOVA F-test results for the model with the AR(1) assumption about correlations between measurements at different times. Give the results and briefly describe how they differ from the results in part C.

The results for AR(1) for correlations are:

anova(fit.ar1)

Denom. DF: 98

numDF F-value p-value

(Intercept) 1 250.14551 <.0001

factor(Dose) 1 7.49346 0.0074

factor(Time) 6 42.36379 <.0001

factor(Dose):factor(Time) 6 8.34215 <.0001

We see that all the factors are significant. From above we also see that the Time and Dose interaction is also significant.

We also notice that there’s not much difference! For this example, we get essentially the same results for the important factors of interest regardless of the model used for correlations across time.

E. Now use a cubic orthogonal polynomial to describe the time trend along with an AR(1) assumption about correlations between measurements at different times. (Note: With seven time points it would be possible to fit a 6th degree polynomial, but powers past the 3rd power don’t seem very efficient, so we’ll use a cubic.

Use the summary command to summarize the coefficient estimates. Briefly describe what is shown about the overall time trend and how the doses differ with regard to the time trend.

The output is:

Generalized least squares fit by REML

Model: Y ~ factor(Dose) \* poly(Time, degree = 3)

Data: nestinginfo

AIC BIC logLik

498.0987 524.5426 -239.0493

Correlation Structure: AR(1)

Formula: ~1 | Dish

Parameter estimate(s):

Phi

0.8708797

Coefficients:

Value Std.Error t-value p-value

(Intercept) 17.31932 1.374190 12.603294 0.0000

factor(Dose)10 -7.71294 1.943398 -3.968792 0.0001

poly(Time, degree = 3)1 -53.56562 6.217058 -8.615911 0.0000

poly(Time, degree = 3)2 2.70761 3.571091 0.758203 0.4500

poly(Time, degree = 3)3 3.96439 2.659799 1.490483 0.1391

factor(Dose)10:poly(Time, degree = 3)1 -4.82441 8.792247 -0.548712 0.5844

factor(Dose)10:poly(Time, degree = 3)2 27.67688 5.050286 5.480259 0.0000

factor(Dose)10:poly(Time, degree = 3)3 -13.50800 3.761523 -3.591099 0.0005

We observe the following in the highlighted area:

* The first line gives the comparison of Dose = 10 to Dose = 5. Note that Dose = 10 differs from Dose = 5, as we saw in the graph too.
* The next three lines give significance results for the linear, quadratic and cubic components (in that order) of a cubic polynomial for the time trend. Note that the linear component is statistically significant, but the quadratic and cubic effects are not.
* The final 3 lines give significance results for interactions between Dose and the polynomial components. The following line has to do with the linear component:

factor(Dose)10:poly(Time, degree = 3)1 -4.82441 8.792247 -0.548712 0.5844

We see here that the interaction between Dose and linear trend of time is not significant. We can’t say here that the slope differs for Dose = 5 and Dose = 10.