Complex Data lab 3

Anna Zaleska

Data input and calculating means

```
wtloss <- read.table("../data/weightloss.dat",
header=F)

## Give names to variables
names(wtloss) <- c("id", paste("y", 1:4, sep=""), "program")

## Univariate format
wtloss.uni <- data.frame(id=rep(wtloss$id, each=4),
wgtloss=as.numeric(t(as.matrix(wtloss[,2:5]))),
program=rep(wtloss$program, each=4),
month=seq(0,9,3),
time.cat=rep(1:4))

wtloss.uni$prog.fac <- factor(wtloss.uni$program, labels=c("1:encourage", "2:none"))
attach(wtloss.uni)
wgt.mean <- tapply(wgtloss, list(month, program), mean)
wgt.sd <- tapply(wgtloss, list(month, program), sd)
wgt.mean
```

```
##          1        2
## 0 241.6088 244.9500
## 3 219.2059 237.0038
## 6 193.7706 227.1192
## 9 169.0441 218.0346
```

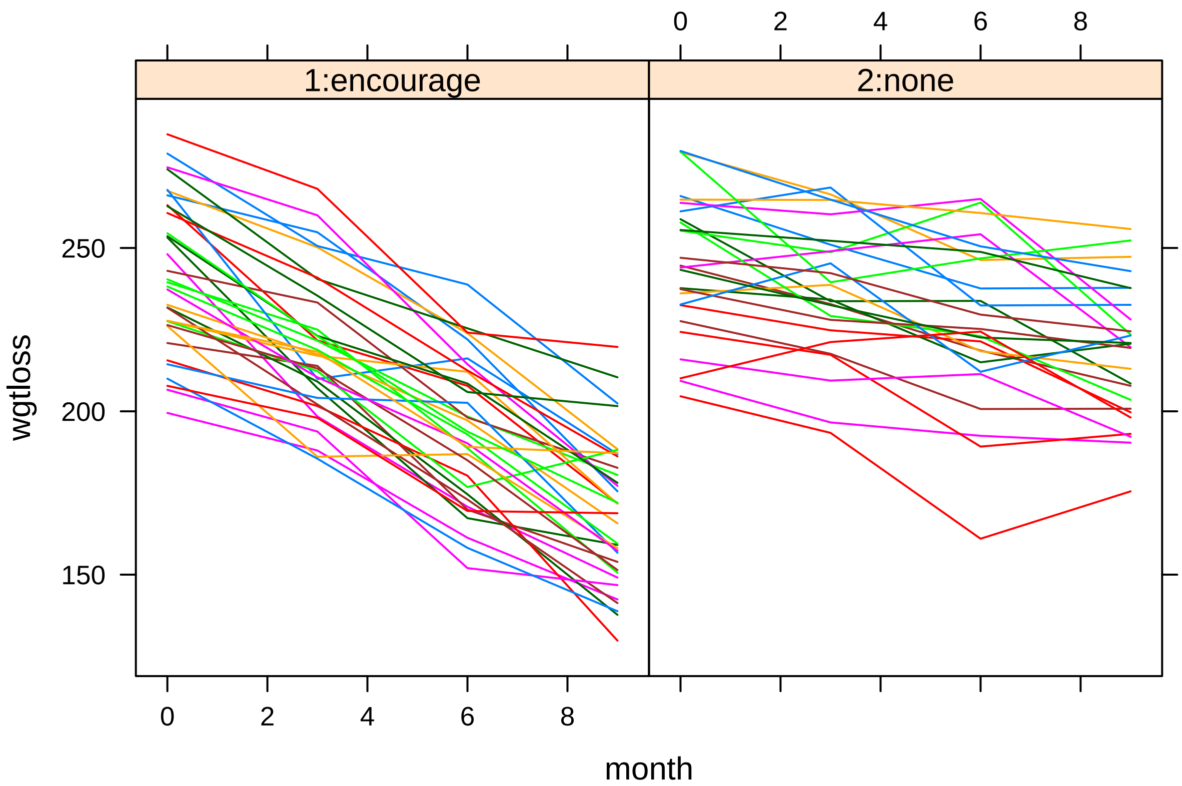
```
wgt.sd
```

```
##          1        2
## 0 23.01513 21.72067
## 3 21.19745 20.40583
## 6 21.92547 24.62901
## 9 21.99597 20.59980
```

```
detach(wtloss.uni)
```

Plotting response profiles

```
xyplot(wgtloss~month|prog.fac, type='l',groups=id,data=wtloss.uni)
```



The most general mean models with different covariance classes:

```
## CATEGORICAL time, UN, REML
wtloss.un.cat <- gls(wgtloss~factor(month)*prog.fac,
correlation=corSymm(form= ~1 | id),
weights=varIdent(form= ~1 | month),
data=wtloss.uni)
#summary(wtloss.un.cat)

## CATEGORICAL time, CS, REML
wtloss.cs.cat <- gls(wgtloss~factor(month)*prog.fac,
correlation=corCompSymm(form= ~1 | id),
weights=varIdent(form= ~1),
data=wtloss.uni)
#summary(wtloss.cs.cat)

## CATEGORICAL time, AR1, REML
wtloss.ar1.cat <- gls(wgtloss~factor(month)*prog.fac,
correlation=corAR1(form= ~1 | id),
weights=varIdent(form= ~1),
data=wtloss.uni)
#summary(wtloss.ar1.cat)
```

Unstructured vs compound symmetry

H_0 : compound symmetry is adequate for the data
H_1 : unstructred is required

```
anova(wtloss.un.cat, wtloss.cs.cat)
```

```
##               Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## wtloss.un.cat     1 18 1936.877 1998.918 -950.4383
## wtloss.cs.cat     2 10 1925.823 1960.291 -952.9116 1 vs 2 4.946568  0.7633
```

CS vs. UN

Structure	-2 REML Log-Likelihood	Number of Cov. Parameters
Compound Symmetry	1905.8232	10
Unstructured	1900.8766	18
Difference	4.9466	8

LRT yields $G^2 = 4.9466$ with 8 df (p =0.7633), so we do not reject the null hypothesis at $\alpha = 0.05$ and conclude that the assumption of compound symmetry covariance structure is adequate for the data.

Unstructured vs autoregressive

H_0 : autoregressive is adequate for the data
H_1 : unstructred is required

```
anova(wtloss.un.cat, wtloss.ar1.cat)
```

```
##                Model df      AIC      BIC    logLik   Test  L.Ratio
## wtloss.un.cat      1 18 1936.877 1998.918 -950.4383
## wtloss.ar1.cat     2 10 1951.247 1985.714 -965.6234 1 vs 2 30.37015
##                p-value
## wtloss.un.cat
## wtloss.ar1.cat   2e-04
```

AR-1 vs. UN

Structure	-2 REML Log-Likelihood	Number of Cov. Parameters
Autoregressive	1931.2468	10
Unstructured	1900.8766	18
Difference	30.3702	8

LRT yields $G^2 = 30.3702$ with 8 df (p =0.0002), so we reject the null hypothesis at $\alpha = 0.05$ and conclude that the assumption of autoregressive covariance structure is inappropriate when compared to unstructured

Autoregressive vs compound symmetry

Since CS and AR-1 have the same number of parameters = 10 , no LRT is necessary. We can directly compare their likelihoods, or -2*log(likelihood):

- $-2 * log(likelihood)$ for CS = 1905.8232
- $-2 * log(likelihood)$ for AR-1 = 1931.2468

Since $-2 * log(likelihood)$ for CS is smaller than for AR-1, CS has a higher likelihood and we conclude that CS is an adequate model for the covariance structure when compared to AR-1.

The most adequate covariance structure is compound symmetry model.

Are all these tests correct?
anova(wtloss.un.cat, wtloss.ar1.cat)
anova(wtloss.un.cat, wtloss.cs.cat)
anova(wtloss.ar1.cat, wtloss.cs.cat)

First two tests are correct. The last one is theoretically incorrect because autoregressive and compound symmetry models are not nested and we should not compare their log-likelihoods or perform LRT on them. In such a case we can compare AIC to find the best model but in our situation both models have the same number of parameters so comparing AICs is the same as comparing likelihoods and thus we can do it.

Structure	-2 REML Log-Likelihood	Number of Cov. Parameters	AIC
Compoud Symmetry	1905.8232	10	1925.823
Autoregressive	1931.2468	10	1951.247
Unstructured	1900.8766	18	1936.877

Thus, we will use a compound symmetry covariance structure for the remainder of the lab

AUC - test for equality of the area under the curve in two groups.

```
t <- wtloss.uni$month
L2 <- 0.5*c(t[1] + t[2] - 2*t[4], t[3] - t[1], t[4]-t[2], t[4]-t[3])
coefs <- summary(wtloss.cs.cat)$coefficients
AUC_enc <- sum(L2 * wgt.mean[,1])
AUC_none <- sum(L2 * wgt.mean[,2])
```

Estimated mean AUC in encouragement program is -319.5705882.
Estimated mean AUC in no encouragement program is -117.7038462.

Testing for the equality of AUCs:
We will use the contrast $(-L_2, L_2)$.

```
L <- c(-L2, L2)
anova(wtloss.cs.cat, L=L)
```

```
## Denom. DF: 232
##  F-test for linear combination(s)
##                   (Intercept)                factor(month)3
##                           7.5                          -3.0
##                factor(month)6                factor(month)9
##                          -3.0                          -1.5
##                prog.fac2:none factor(month)3:prog.fac2:none
##                          -7.5                           3.0
## factor(month)6:prog.fac2:none factor(month)9:prog.fac2:none
##                           3.0                           1.5
##   numDF  F-value p-value
## 1     1 759.3707  <.0001
```

Because of p-value which is < 0.0001 we reject the null hypothesis and conclude that response profile is the same for both treatments.

Parametric curves

Quadratic time trend

```
## QUADRATIC time, CS, REML,
wtloss.cs.quad <- gls(wgtloss~month*prog.fac + I(month^2)*prog.fac,
correlation=corCompSymm(form= ~1 | id),
weights=varIdent(form= ~1),
data=wtloss.uni,
method="ML")
#summary(wtloss.cs.quad)

## CATEGORICAL time, CS, ML
wtloss.cs.cat.ml <- gls(wgtloss~factor(month)*prog.fac,
correlation=corCompSymm(form= ~1 | id),
weights=varIdent(form= ~1),
data=wtloss.uni,
method="ML")
#summary(wtloss.cs.cat.ml)
```

H_0 : quadratic model
H_1 : saturated mode

```
anova(wtloss.cs.cat.ml, wtloss.cs.quad)
```

```
##                  Model df      AIC      BIC    logLik   Test   L.Ratio
## wtloss.cs.cat.ml     1 10 1955.331 1990.138 -967.6657
## wtloss.cs.quad       2  8 1951.700 1979.545 -967.8500 1 vs 2 0.3686901
##                  p-value
## wtloss.cs.cat.ml
## wtloss.cs.quad    0.8316
```

Testing the Quadratic trend

Structure	-2 Log Likelihood	Number of Cov. Parameters
Quadratic model	1935.7	8
Saturated model	1935.3314	10
Difference	0.3686	2

LRT yields G2 =0.369 (p = 0.8316), so we fail to reject the null hypothesis at $\alpha = 0.05$ and conclude that the model with Month as a quadratic effect seems to fit the data adequately. (Note: $chi^2_{2,0.95} = 5.99$)

Linear time trend

```
wtloss.cs.lin <- gls(wgtloss~month*prog.fac,
correlation=corCompSymm(form= ~1 | id),
weights=varIdent(form= ~1),
data=wtloss.uni,
method="ML")
#summary(wtloss.cs.lin)
```

H_0 : linear model
H_1 : quadratic model

```
anova(wtloss.cs.lin, wtloss.cs.quad)
```

```
##                Model df      AIC      BIC    logLik   Test   L.Ratio
```

```
## wtloss.cs.lin      1  6 1948.295 1969.179 -968.1477
## wtloss.cs.quad     2  8 1951.700 1979.545 -967.8500 1 vs 2 0.5954061
##                 p-value
## wtloss.cs.lin
## wtloss.cs.quad  0.7425
```

Linear vs. Quadratic

Structure	-2 Log Likelihood	Number of Cov. Parameters
Linear model	1936.2954	6
Quadratic model	1935.7	8
Difference	0.5954	2

LRT yields G2 =0.596 (p = 0.7425), so we fail to reject the null hypothesis at $\alpha = 0.05$ and conclude that the model with Month as a linear effect fits the data better than the model with Month as a quadratic effect. (Note: $chi^2_{2,0.95} = 5.99$)

Testing for intersections

```
## Linear model, NO interacations
wtloss.cs.lin.noint <- gls(wgtloss~month + prog.fac,
correlation=corCompSymm(form= ~1 | id),
weights=varIdent(form= ~1),
data=wtloss.uni,
method="ML")
#summary(wtloss.cs.lin.noint)
```

$H_0 : \beta_4 = 0$
$H_1 : \beta_4 \neq 0$

```
anova(wtloss.cs.lin, L= c(0,0,0,1))
```

```
## Denom. DF: 236
##  F-test for linear combination(s)
## [1] 1
##   numDF  F-value p-value
## 1     1 184.3676  <.0001
```

Since the p-value for month*program is < 0.0001, we reject the null hypothesis and conclude that there is an interaction between month and program.

Thus, our final model is the linear model with interaction.

```
wtloss.cs.lin.final <- gls(wgtloss~month*prog.fac,
correlation=corCompSymm(form= ~1 | id),
weights=varIdent(form= ~1),
data=wtloss.uni)
summary(wtloss.cs.lin.final)$coefficients
```

```
##          (Intercept)                month       prog.fac2:none
##           242.376765            -8.104314             2.994774
## month:prog.fac2:none
##             5.083288
```

What can you conclude about the two weight programs in terms of their effectiveness?
Let us look at estimated rate of change for both programs.

First program: $\hat{\beta}_2 =$ -8.1043137
Second program: $\hat{\beta}_2 + \hat{\beta}_4 =$ -3.0210256
We can conclude that program with receiving encouragement is much more effective than the one without.

Interpreting quadratic trends:

```
wtloss.cs.quad.noint <- gls(wgtloss~month + prog.fac + I(month^2),
correlation=corCompSymm(form= ~1 | id),
weights=varIdent(form= ~1),
data=wtloss.uni)

coefficients.m <- wtloss.cs.quad.noint$coefficients
coefficients <- c(coefficients.m[1], coefficients.m[2], coefficients.m[4], coefficients.m[3])
change_rates <- coefficients[2] + 2*coefficients[3]*t[1:4]
expected1 <- coefficients[1] + coefficients[2]*t[1:4] + coefficients[3]*t[1:4]*t[1:4]
```

$\hat{\beta}_1$ = 232.0118529 expected mean response at baseline of subjects in program 1.
$\hat{\beta}_4$= 25.8695701 change in expected response at baseline of subjects in program 2 vs. program 1.
Rate of change in program 1 is $\beta_2 + 2\beta_3 time_{ij}$.
Plugging in the above estimates,

Program 1

time	Rate of Change	Expected Response
0	-5.4490556	232.0118529
1	-5.7507222	215.2121863
2	-6.0523889	197.5075196
3	-6.3540556	178.8978529

Thus, the mean response for Program 1 decreases over time. For the second program effect is the same. We need to add $\hat{\beta}_4$ to the expected response at each time point but the trend remains.