Mixed Measures ANOVA - Expectation of Distance

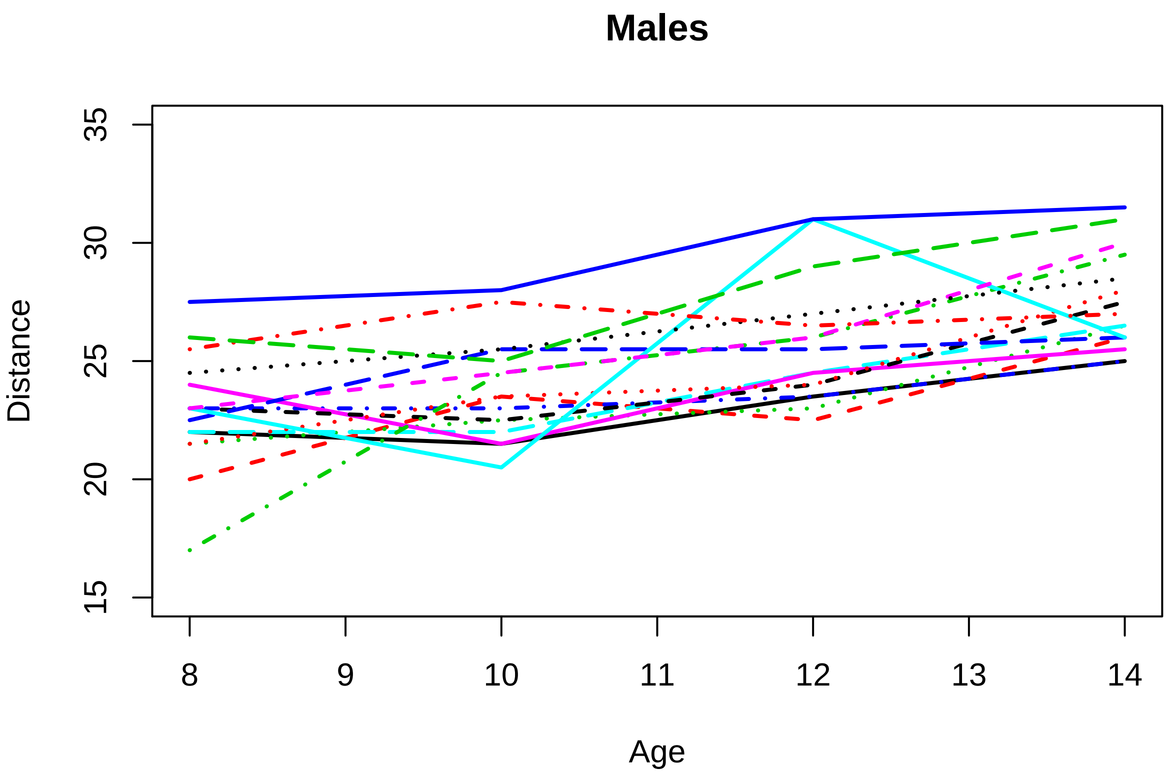
Greg Johnson

8/14/2017

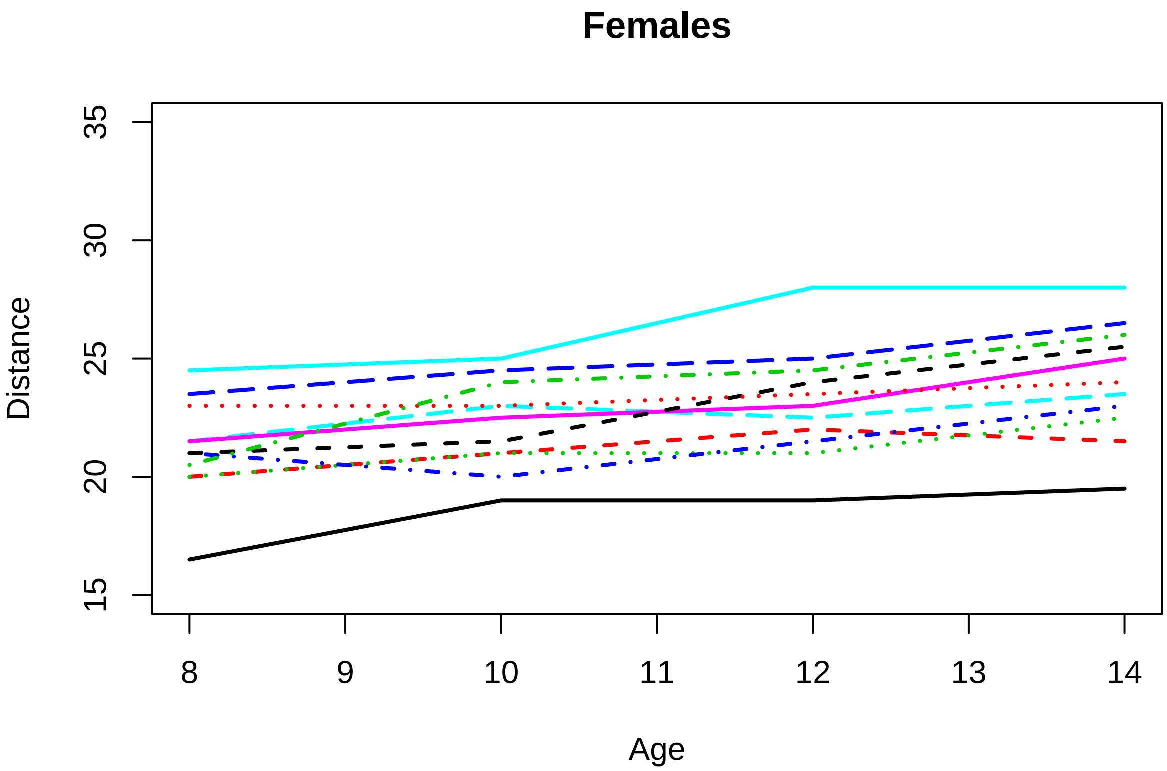
We have data on expectation of distance for 108 subjects, along with their gender and age.

We can explore the data first by making two separate plots, one for each sex, in which every line represents one subject's relationship between age and distance.

```
# separate plots for sexs, separate lines for subjects
wide <- dcast(ortho, formula = Subject + Sex ~ age, value.var = "distance")
y.lim <- c(15, 35)
matplot(c(8, 10, 12, 14), t(wide[1:16, 3:6]), xlab = "Age", ylab = "Distance",
    main = "Males", type = "l", lwd = rep(2, 16), ylim = y.lim)
```



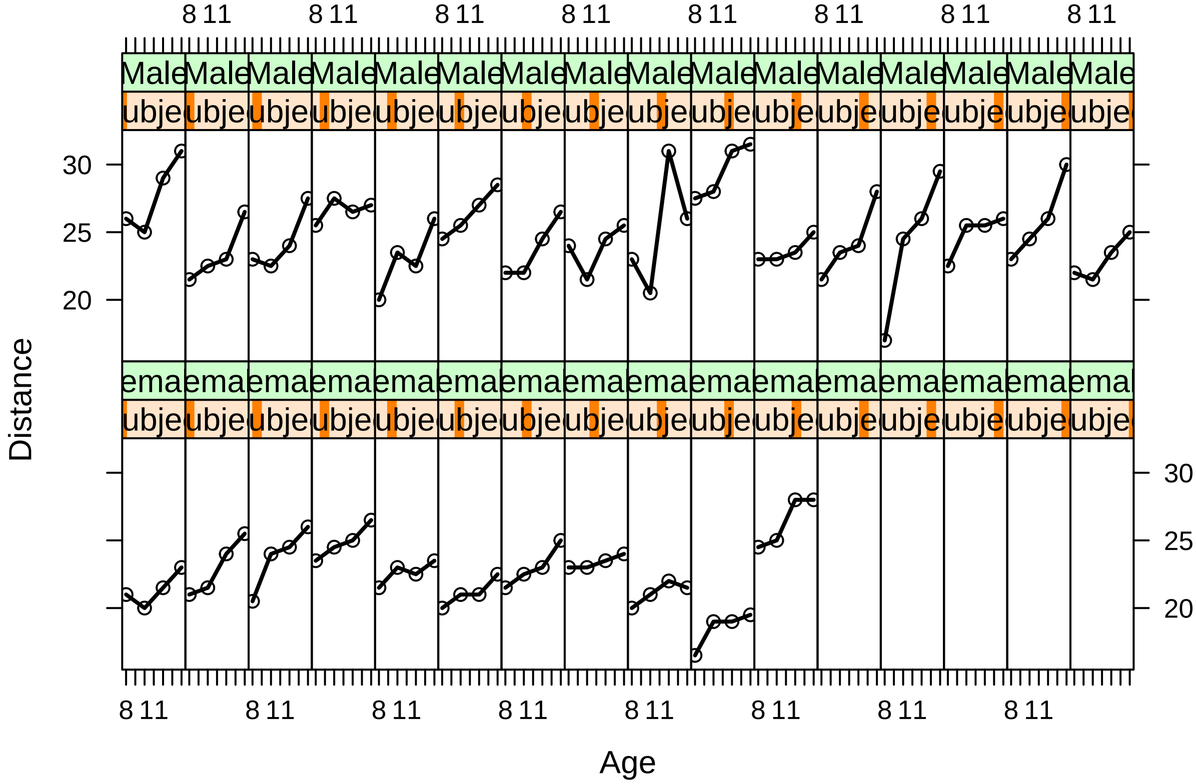

```
matplot(c(8, 10, 12, 14), t(wide[17:27, 3:6]), xlab = "Age", ylab = "Distance",
    main = "Females", type = "l", lwd = rep(2, 10), ylim = y.lim)
```



It looks like there's a general linear trend upwards in both sexes and that males, on average, see larger increases in distance for an increase in age than females do.

We can also create plots for subjects and group these plots by sex.

```
# separate plots for subjects and sexes
ortho$Subject <- c(rep(1:16, each = 4), rep(1:11, each = 4))
colr <- "black"
xyplot(distance ~ age | Subject * Sex, data = ortho, type = "b", lwd = 2, ylab = "Distance",
    xlab = "Age", col = colr, as.table = TRUE)
```



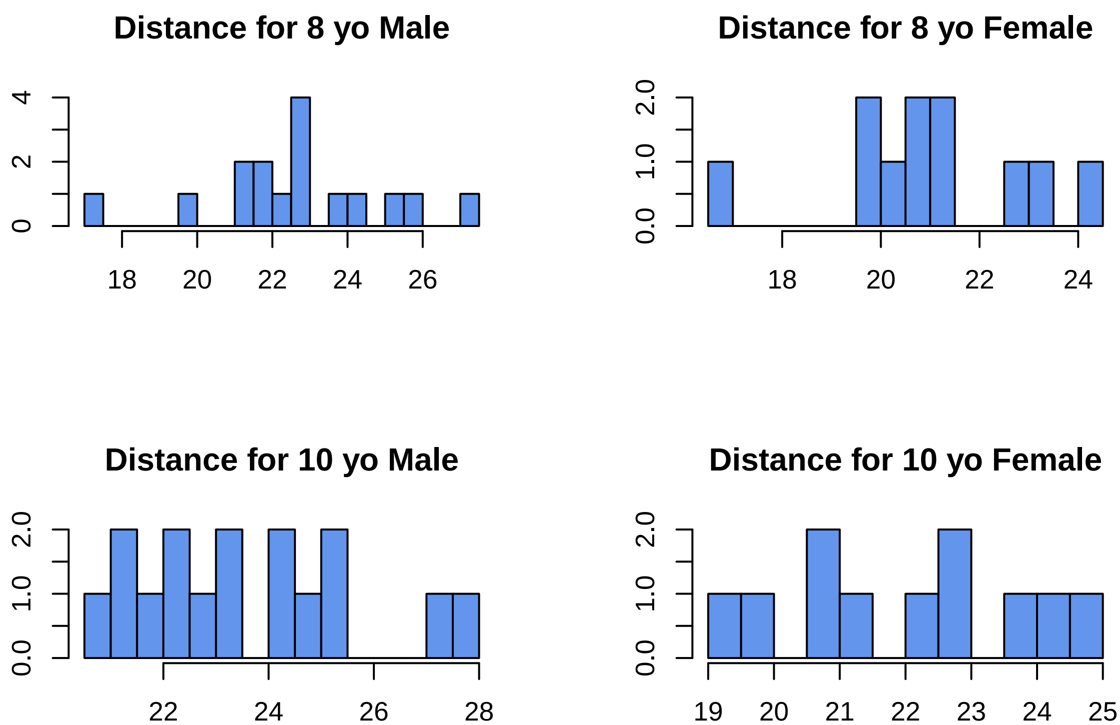
It’s hard to pick out any general trends since every subject is represented separately. However there does seem to be a general trend upwards.

We can formally model whether sex, age, or an interaction affects our expectation of distance by fitting a Mixed Measures ANOVA.

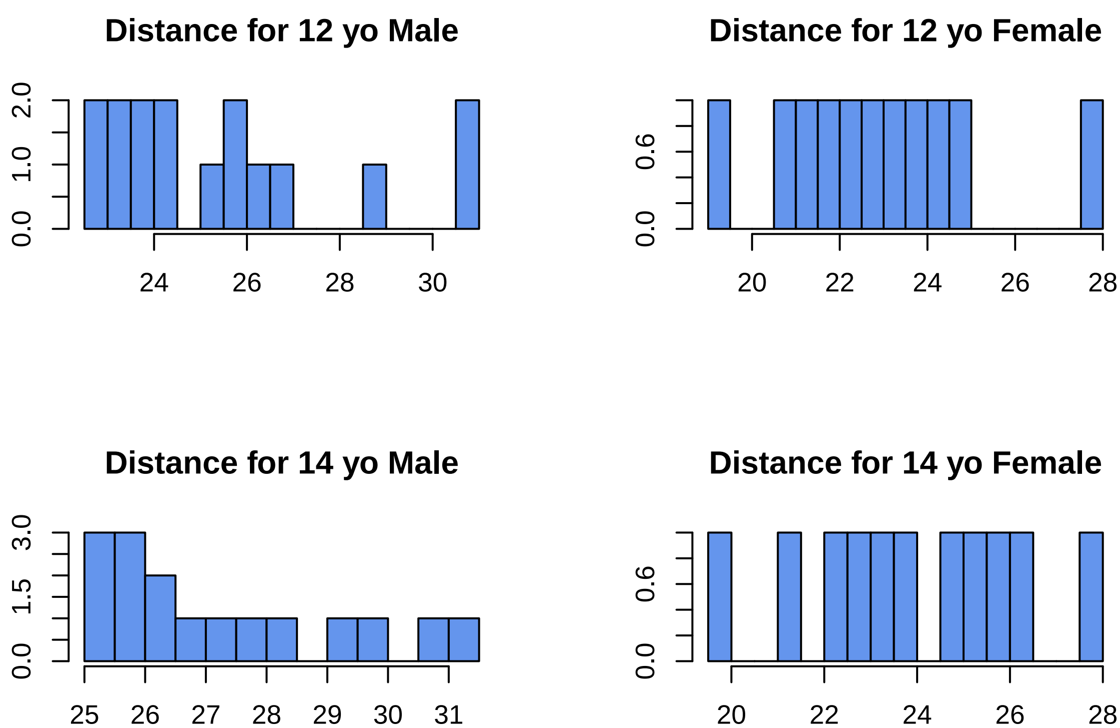
Sex is our between-subjects factor; age is our within-subjects factor. I used the Laerd Statistics webpage as a guide to evaluating the assumptions for this design.

```
ortho <- Orthodont
ortho$age <- factor(ortho$age, levels = c(8, 10, 12, 14))

par(mfrow = c(2, 2))
hist(ortho[ortho$age == 8 & ortho$Sex == "Male", 1], breaks = 20, col = "cornflowerblue",
    xlab = NULL, ylab = NULL, main = "Distance for 8 yo Male")
hist(ortho[ortho$age == 8 & ortho$Sex == "Female", 1], breaks = 20, col = "cornflowerblue",
    xlab = NULL, ylab = NULL, main = "Distance for 8 yo Female")
hist(ortho[ortho$age == 10 & ortho$Sex == "Male", 1], breaks = 20, col = "cornflowerblue",
    xlab = NULL, ylab = NULL, main = "Distance for 10 yo Male")
hist(ortho[ortho$age == 10 & ortho$Sex == "Female", 1], breaks = 20, col = "cornflowerblue",
    xlab = NULL, ylab = NULL, main = "Distance for 10 yo Female")
```




```r
hist(ortho[ortho$age == 12 & ortho$Sex == "Male", 1], breaks = 20, col = "cornflowerblue",
    xlab = NULL, ylab = NULL, main = "Distance for 12 yo Male")
hist(ortho[ortho$age == 12 & ortho$Sex == "Female", 1], breaks = 20, col = "cornflowerblue",
    xlab = NULL, ylab = NULL, main = "Distance for 12 yo Female")
hist(ortho[ortho$age == 14 & ortho$Sex == "Male", 1], breaks = 20, col = "cornflowerblue",
    xlab = NULL, ylab = NULL, main = "Distance for 14 yo Male")
hist(ortho[ortho$age == 14 & ortho$Sex == "Female", 1], breaks = 20, col = "cornflowerblue",
    xlab = NULL, ylab = NULL, main = "Distance for 14 yo Female")
```



First we assess normality within each group defined by the factors. However ANOVA designs are robust to deviations from normality so we're really just assessing by eye for extreme non-normality which we don't see here.

Next we want to assess homogeneity of covariance matrices (formed from the repeated-measures factor) between the between-subjects factor i.e. we want the covariance matrix of different ages to be the same for the males and females. We can use Box's M-test (G., E., P. Box, 1949). Since Box's is highly sensitive, we use an $\alpha = .001$ (Tabachnick & Fidell, 2013).

```
boxM(wide[,3:6],grouping=wide$Sex)
```

```
##
##  Box's M-test for Homogeneity of Covariance Matrices
##
## data:  wide[, 3:6]
## Chi-Sq (approx.) = 17.335, df = 10, p-value = 0.06727
```

Since $p > \alpha$, we cannot reject the null hypothesis that there is homogeneity of covariance matrices.

Our last assumption is compound symmetry, a very restrictive assumption about the structure of the covariance matrices of age. It must be evaluated separately for the males and the females. We'll evaluate using Mauchly's test of sphericity.

```
#pass sample covariance matrix of group and size of group
Mauchly<-function(A,n){
  #based on Mauchly (1940) and Abdi (2010)
  #double center covariance matrix
  A_mean<-mean(A)
  row_mean<-rowSums(A)/nrow(A)
  col_mean<-colSums(A)/ncol(A)
  R<-rbind(row_mean,row_mean,row_mean,row_mean)
  C<-cbind(col_mean,col_mean,col_mean,col_mean)
  DC<-A-R-C+A_mean

  lambda<-eigen(DC)$values[c(1,2,3)]
  W<-prod(lambda)/(1/3*sum(lambda))^3

  #chi-square approximate to W
  f <- (2*(4-1)^2+4+2)/(6*(4-1)*(n-1))
  chisq <- -(1-f)*(n-1)*log(W)
  df <- .5*4*(4-1)

  #return p-value
  pchisq(chisq,df,lower.tail=FALSE)
}

cov_male<-cov(wide[wide$Sex=="Male",3:6])
cov_female<-cov(wide[wide$Sex=="Female",3:6])
Mauchly(cov_male,sum(ortho$Sex=="Male"))
```

```
## [1] 0.0006556121
```

```
Mauchly(cov_female,sum(ortho$Sex=="Female"))
```

```
## [1] 0.01893096
```

Compound symmetry does not hold. We can use the Huyn-Feldt (1976) correction to the df of the F statistics produced from the RMANOVA model. First we need to compute an estimate of Box's measure of sphericity (1954 a&b), specifically we'll use Huynh-Feldt's approximation of ϵ.

```r
A<-cov(wide[,3:6])
A_mean<-mean(A)
row_mean<-rowSums(A)/nrow(A)
col_mean<-colSums(A)/ncol(A)
R<-rbind(row_mean,row_mean,row_mean,row_mean)
C<-cbind(col_mean,col_mean,col_mean,col_mean)
DC<-A-R-C+A_mean
n<-nrow(wide)

ehat <- (sum(diag(DC)))^2/((4-1)*sum(DC^2))

e <- (n*(4-1)*ehat-2)/((4-1)*(n-1-(4-1)*ehat))
e
```

```
## [1] 0.9843975
```

We can now use our estimate of ϵ to correct our F-statistics.

Now let's run the model.

```r
output = aov(distance~age*Sex + Error(Subject),data=ortho)
summary(output)
```

```
##
## Error: Subject
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Sex        1  140.5  140.46   9.292 0.00538 **
## Residuals 25  377.9   15.12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
##           Df Sum Sq Mean Sq F value   Pr(>F)
## age        3 237.19   79.06  40.032 1.49e-15 ***
## age:Sex    3  13.99    4.66   2.362   0.0781 .
## Residuals 75 148.13    1.98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Let's recompute p-values using the F-values of the model but with the corrected degrees of freedom (rounded to integers)

```r
#Sex
df1<-round(ehat*(2-1))
df2<-round(ehat*(2-1)*25)
pf(9.292,df1,df2,lower.tail=FALSE)
```

```
## [1] 0.005895127
```

```r
#Age
df1<-round(ehat*(4-1))
df2<-round(ehat*(4-1)*75)
pf(40.032,df1,df2,lower.tail=FALSE)
```

```
## [1] 3.028028e-20
```

```r
#Age:Sex
df1<-round(ehat*(4-1))
```

```r
df2<-round(ehat*(4-1)*75)
pf(2.362,df1,df2,lower.tail=FALSE)
```

```
## [1] 0.07256643
```

Based on the output, sex and age are statistically significant predictors of distance. However the data don't support the existence of an interaction term between sex and age. In other words, the effect of sex on distance does not (statistically) significantly change when age is varied. This is corroborated by our first figure where the positive linear trend between distance and age is practically the same between the males and females.