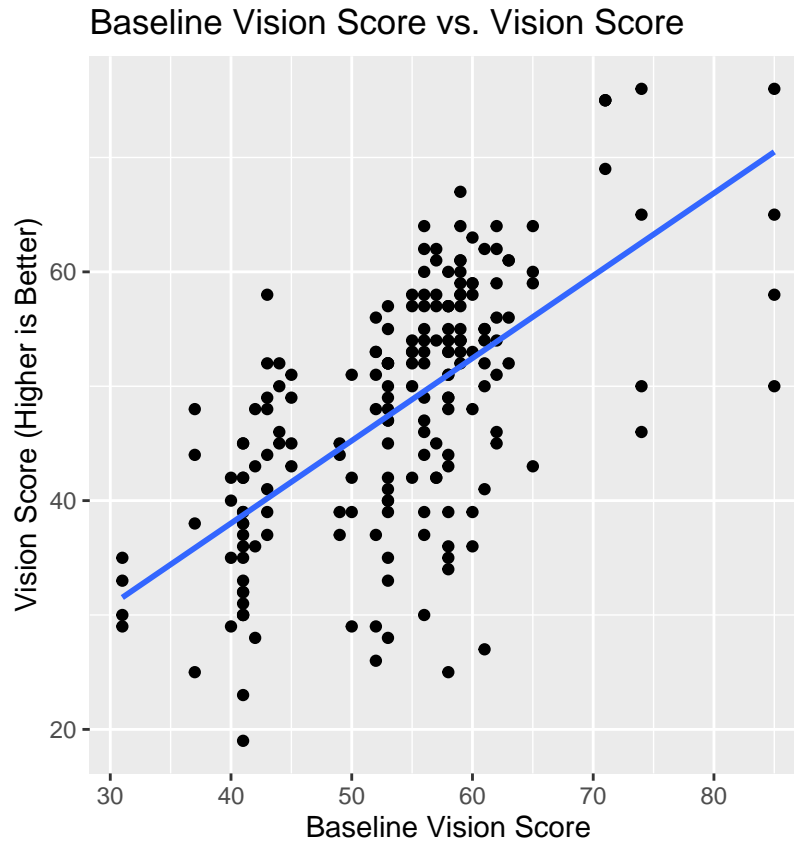# Macular Degeneration Code

## Jillian Warburton

### 2023-03-10

**Exploratory Data Analysis**

1. Change the `Trt` variable in the ARMD dataset to a factor.

2. Create of scatterplot of `Baseline` vs. `Vision`.

### Baseline Vision Score vs. Vision Score



3. Create a scatterplot `Time` by `Vision` where dots are colored according to `Trt`. Include fitted regression lines for both those who have a treatment and those who don't.

## Time vs. Vision
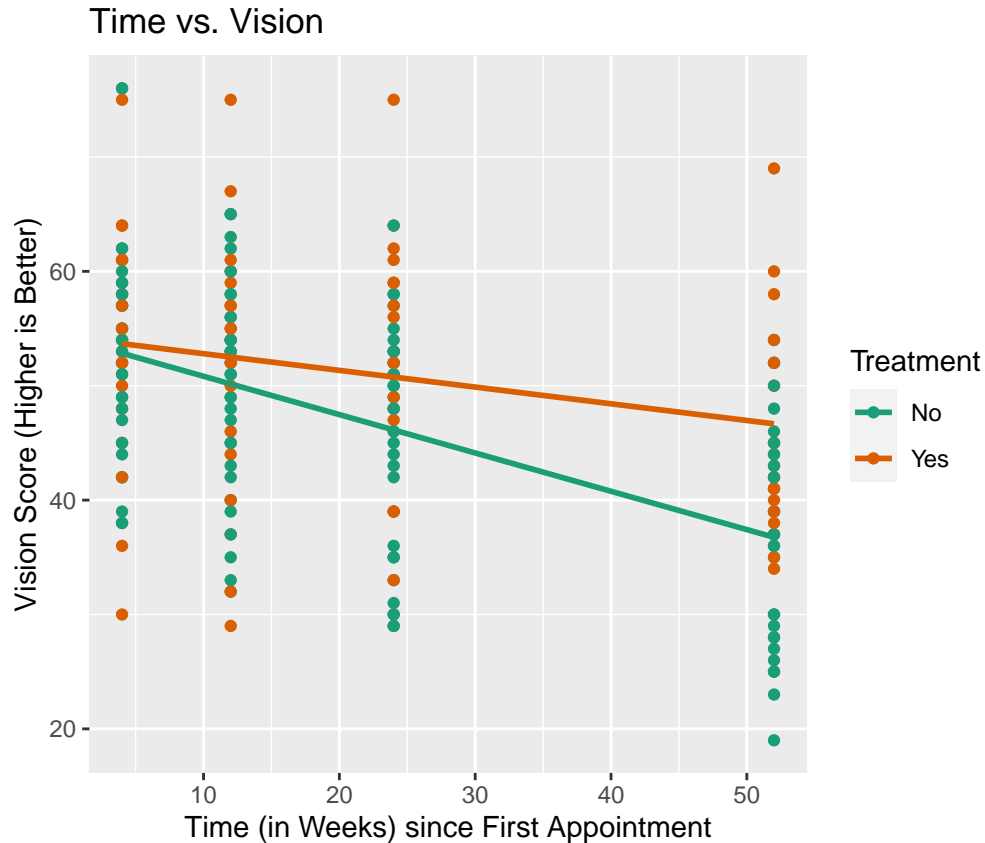


## Analysis with an MLR

1. Verify that the residuals of an independent MLR with a baseline effect and an interaction between `Time` and `Trt` are indeed correlated by calculating the 4 by 4 correlation matrix $\hat{\mathbf{R}}$ between residuals of the same person but between visits (see page 14 of the slides). **Hint:** the easiest way to do this is to organize the residuals into a $50 \times 4$ matrix where the $i^{th}$ row is the residuals from the $i^{th}$ patient then use `cor()`.

```
##            [,1]      [,2]      [,3]      [,4]
## [1,] 1.0000000 0.5157923 0.1508502 0.1358271
## [2,] 0.5157923 1.0000000 0.5242362 0.3328835
## [3,] 0.1508502 0.5242362 1.0000000 0.4416136
## [4,] 0.1358271 0.3328835 0.4416136 1.0000000
```

## Longitudinal MLR Model Fitting and Iterative Optimization

1. Fit a linear regression model for `Vision` using a linear effect for `Baseline`, `Time` and `Trt` as well as the interaction of `Time` and `Trt`. In your model use a block diagonal general symmetric correlation structure within `Subject` but independent between `Subject` (i.e. `form=~1|Subject` or `form=~1:4|Subject`). Identify the constrained estimates of the general correlation structure along with any $\hat{\beta}$ coefficients and the estimate of the variance parameter $\hat{\sigma}^2$.

```
## Generalized least squares fit by maximum likelihood
##   Model: Vision ~ Baseline + Time + Trt + Time:Trt
```

```
##    Data: data
##    Log-likelihood: -642.3705
##
## Coefficients:
## (Intercept)    Baseline        Time         Trt1    Time:Trt1
##   16.8398861   0.6984037   -0.3292298   -0.1565878    0.1614645
##
## Correlation Structure: General
##  Formula: ~1:4 | Subject
##  Parameter estimate(s):
##  Correlation:
##    1     2     3
## 2 0.660
## 3 0.118 0.436
## 4 0.100 0.255 0.376
## Degrees of freedom: 200 total; 195 residual
## Residual standard error: 6.81838
```

```
## [1] 0.6600738 0.1175724 0.1003644 0.4364513 0.2548244 0.3756646
```

2. In Stat 340, you found that the maximum likelihood estimate of $\mu$ when $y_i \overset{iid}{\sim} N(\mu, 5)$ is $\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$. Simulate 100 values of $y_i$ from `rnorm(100, mean=17, sd=sqrt(5))`, write a function in R that evaluates the log-likelihood of $\mu$ and then maximize this function using `optim()` from a starting value (initial guess) of $\hat{\mu} = 13$. Verify that the iterative optimization routine returns a value near $\bar{y}$.

   **Hint:** The `dnorm` function returns the density of the normal distribution where you can specify the option `log=TRUE` to return the log density. The log-likelihood is given by

   $$\mathcal{L}(\mu) = log(\prod_{i=1}^{n} f_Y(y_i | \mu, \sigma^2 = 5))$$

   $$= \sum_{i=1}^{n} log(f_Y(y_i | \mu, \sigma^2 = 5))$$

   $$= \sum_{i=1}^{n} dnorm(y_i, mean = \mu, sd = sqrt(5), log = TRUE)$$

   where $f_Y(y_i | \mu, \sigma^2)$ is the Gaussian density with mean $\mu$ and variance $\sigma^2$ evaluated at $y_i$.

```
## $par
## [1] 16.73308
##
## $value
## [1] -222.5511
##
## $counts
## function gradient
##        5        3
##
## $convergence
## [1] 0
##
## $message
## NULL
```

3. We learned in this class that in an independent MLR model, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ where $\mathbf{X}$ is the design matrix and $\mathbf{y}$ is the vector of response variables. Using the ARMD dataset and corresponding design matrix, use `optim()` to find the maximum likelihood estimates $\hat{\boldsymbol{\beta}}$ and verify that $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is indeed the maximum. Use an initial guess of $\hat{\boldsymbol{\beta}} = (\bar{y}, 0, 0, 0, 0)$ and fix $\sigma^2 = 6.8072$. **Hint:** We found $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ by maximum the log-likelihood of the multivariate normal density. You can use the `dmvnorm` function from the `mvtnorm` library in R to set this up (or write your own function). Basic syntax for the function to be maximized would be something like:

```
## $par
## [1] 16.3219724  0.7091499 -0.3350262 -0.9820549  0.1889976
##
## $value
## [1] -664.8908
##
## $counts
## function gradient
##       52       14
##
## $convergence
## [1] 0
##
## $message
## NULL
```

```
## (Intercept)    Baseline        Time        Trt1    Time:Trt1
##   16.3144478   0.7093761  -0.3351586  -0.9909924   0.1891784
```

4. Using `myFunction` below (which is bimodal), use `optim()` to find the maximum with starting values of -4 and 4. Show that when you start at -4, you get stuck in a local mode and fail to find the global maximum.

```
## $par
## [1] -1.995912
##
## $value
## [1] 0.09983676
##
## $counts
## function gradient
##       37       32
##
## $convergence
## [1] 0
##
## $message
## NULL
```

```
## $par
## [1] 1.999552
##
## $value
## [1] 0.2992402
##
```

4

```
## $counts
## function gradient
##       20       14
##
## $convergence
## [1] 0
##
## $message
## NULL
```

**Validating Longitudinal MLR Model Assumptions**

1. Check the assumption of linearity using added-variable plots from an independent model (note: correlation doesn't change the linearity at all so you can just fit an independent model and look at the added-variable plots).

## Added−Variable Plots
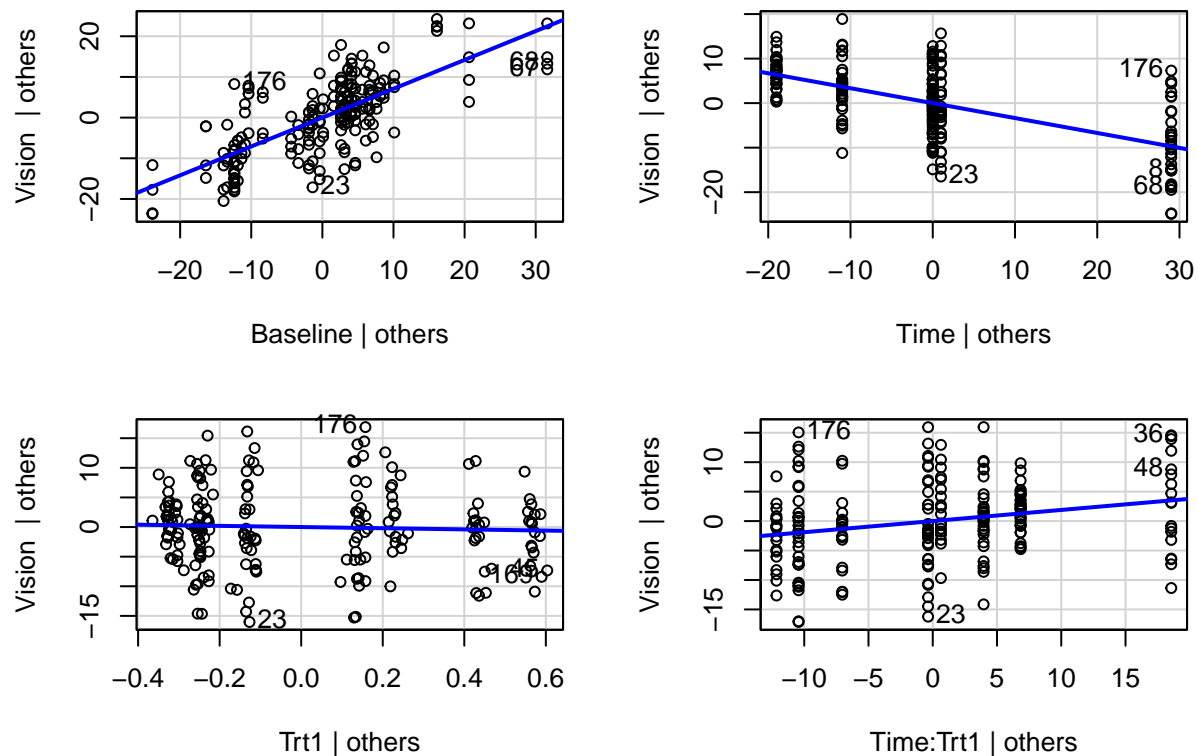


2. Check the assumption of independence by decorrelating residuals and looking at the sample correlation matrix of within-Subject residuals. Refer to Question #1 under the "Analysis with an MLR" section above for a hint on how to do this.

```
##              [,1]        [,2]        [,3]        [,4]
## [1,] 1.00000000 0.08626570 0.03956925 0.05525491
## [2,] 0.08626570 1.00000000 0.08337487 0.08255935
## [3,] 0.03956925 0.08337487 1.00000000 0.09132445
## [4,] 0.05525491 0.08255935 0.09132445 1.00000000
```

5

3. Check the assumption of normality by drawing a histogram of the decorrelated residuals.

## Normality Assumption Check
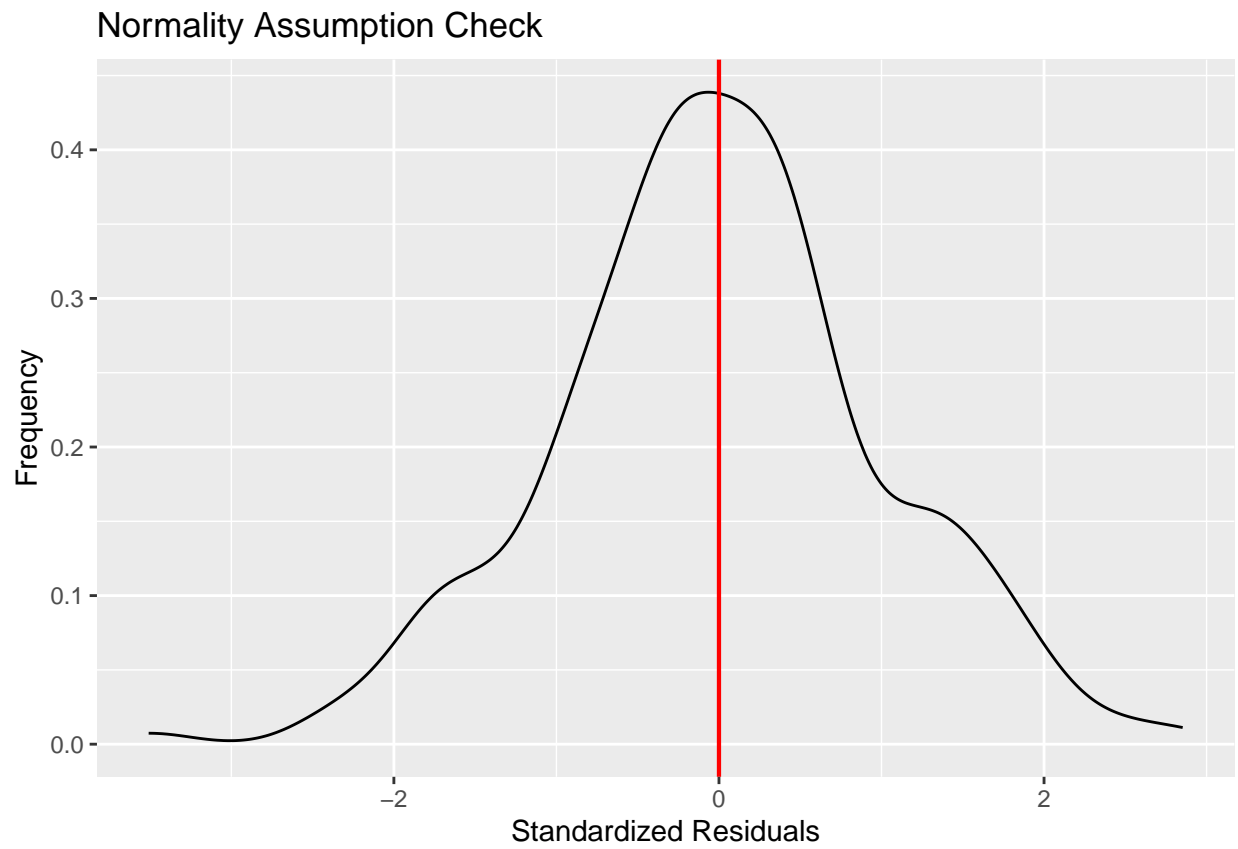


4. Check the assumption of equal variance by plotting the decorrelated residuals vs. the fitted values.

## Equal Variance Assumption Check:
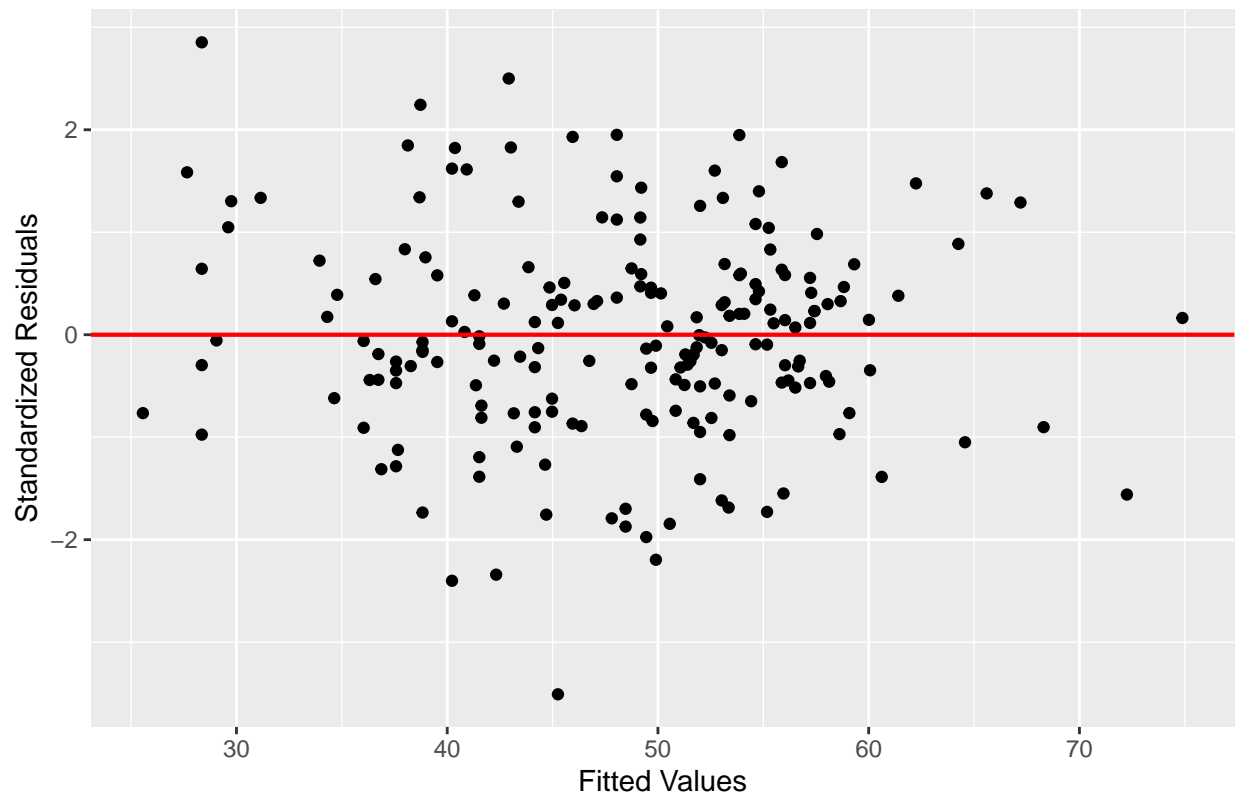


## Statistical Inference

1. Use a general linear hypothesis test on your longitudinal MLR to test if patients on the treatment have a significantly higher vision score at 52 than those not on the treatment.

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: gls(model = Vision ~ Baseline + Time + Trt + Time:Trt, data = data,
##     correlation = corSymm(form = ~1:4 | Subject), method = "ML")
##
## Linear Hypotheses:
##          Estimate Std. Error z value    Pr(>z)
## 1 <= 0     8.240      2.024    4.07 0.0000235 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

2. Use a general linear hypothesis test on your longitudinal MLR to test if the treatment stops vision loss over time.

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
```

```
## Fit: gls(model = Vision ~ Baseline + Time + Trt + Time:Trt, data = data,
##     correlation = corSymm(form = ~1:4 | Subject), method = "ML")
##
## Linear Hypotheses:
##          Estimate Std. Error z value     Pr(<z)
## 1 >= 0 -0.16777    0.04346   -3.86 0.0000567 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

3. Use a general linear hypothesis test on your longitudinal MLR to test if patients on/off the drug with a baseline of 29 are at risk of going legally blind (20).

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: gls(model = Vision ~ Baseline + Time + Trt + Time:Trt, data = data,
##     correlation = corSymm(form = ~1:4 | Subject), method = "ML")
##
## Linear Hypotheses:
##          Estimate Std. Error z value Pr(<z)
## 1 >= 20   19.974     2.043   -0.013  0.495
## (Adjusted p values reported -- single-step method)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: gls(model = Vision ~ Baseline + Time + Trt + Time:Trt, data = data,
##     correlation = corSymm(form = ~1:4 | Subject), method = "ML")
##
## Linear Hypotheses:
##          Estimate Std. Error z value Pr(<z)
## 1 >= 20   28.21      2.38    3.451     1
## (Adjusted p values reported -- single-step method)
```