A close-up of a math problem

Description automatically generated

$Q

[,1] [,2] [,3] [,4]

[1,] 0.7071068 0.4082483 0.2886751 0.2236068

[2,] -0.7071068 0.4082483 0.2886751 0.2236068

[3,] 0.0000000 -0.8164966 0.2886751 0.2236068

[4,] 0.0000000 0.0000000 -0.8660254 0.2236068

[5,] 0.0000000 0.0000000 0.0000000 -0.8944272

The resulting matrix Q is the orthonormal matrix formed from the columns of the original matrix, representing the Helmert contrasts. Helmert contrasts are used to compare the mean of a current level with the means of all preceding levels. The first contrast (first column) compares the first level with the second. The second contrast (second column) compares the average of the first two levels with the third level. The third contrast (third column) compares the average of the first three levels with the fourth level, and so on.

2. This problem uses a data set from **a study of strength training** with human subjects using weights. Participants in the study were followed for seven weeks, with a measurement of strength taken each week. The measurements were equally spaced over time. The study had **three treatment groups**: a group with weights increasing over time (WI), a group with repetitions increasing over time (RI), and a control group (CONT). Participants were randomly allocated to treatment groups, and subject are considered to be a sample from a larger population. The data are contained in the file **strength.dat** on Canvas. The data are in wide format, where each column represents successive measurements over the seven weeks of the study. The purpose of the study was to determine if either WI or RI produced larger increase in strength over time. You may use SAS, R, python, or Matlab to obtain answers to the questions below.

(a) Obtain a **profile plot** for strength over weeks for the three treatment groups. What does the plot suggest about polynomial trends in strength over time? Why?

A graph showing the strength of a patient

Description automatically generated

* WI (Weights Increasing over time): The trend appears to be mostly linear but could have a slight quadratic component due to the accelerating increase in later weeks.
* RI (Repetitions Increasing over time): A quadratic trend is likely, with an increase followed by a leveling off of strength.
* CONT (Control): A flat or slightly negative linear trend, indicating no major polynomial trend.

(b) Write down the **assumptions** for the **split plot model** of expression 3.1 in Chapter 3 of the Marie Davidian course notes as applied to this study.

**(3.1) Yhℓj​=μ+τℓ​+bhℓ​+γj​+(τγ)ℓj​+ehℓj​**

* Yhℓj​: The observed value of the dependent variable (in this case, strength) for subject h, treatment group ℓ, and time point j.
* μ: The overall mean of the response variable.
* τℓ ​: The **fixed effect** of the treatment group ℓ (e.g., WI, RI, or CONT).
* bhℓ​: The **random effect** of the individual subject h nested within treatment group ℓ (accounts for variability between subjects).
* γj ​: The **fixed effect** of time point j (e.g., week 1, week 2, etc.).
* (τγ)ℓj ​: The **interaction** effect between treatment group ℓ and time point j, indicating how the treatment effect varies across time.
* ehℓj​: The residual error term, representing unexplained variability at the subject-by-time level.

**Random Effects bhℓ ​**:

* **bhℓ ∼ N (0, σb2​):** The random deviations bhℓ​, representing among-individual variation, are assumed to be normally distributed with a mean of 0 and variance σb2​.
* The random effects **bhℓ​ are independent** for all individuals h and treatment groups ℓ. This implies that the individual variation is the same across all treatment groups, indicating that individuals’ deviations are not affected by the specific treatment.
* The variance component σb2​​ represents the **among-individual variation** and is the same for all populations (**treatment groups)**. (constant variation)

**Error Term ehℓj ​**:

* **ehℓj ∼ N (0, σe2​)**: The residual errors σe2​, representing within-individual (repeat measure) variation, are assumed to be normally distributed.
* The errors **ehℓj are independent** for all individuals h, treatment groups ℓ, and time points j, meaning that there is no correlation between errors within or across individuals and time points.
* The variance component σe2​​ ​ represents the **within-individual variation** (across time), and it is assumed to be the same for all individuals and treatment groups.

**Independence of Random Effects and Errors**:

* The random effects **bhℓ**​ and errors **ehℓj** are assumed to be **mutually independent** for all h, ℓ, and j. This means that individual variation (between subjects) is independent of within-individual variation (over time).

**Independence Across Time**:

* Model assumes **no correlation within individuals across time**, meaning no correlation between measurements taken at different time points for the same individual. This implies that the strength measurements at different time points for the same individual are independent of each other: **Compound Symmetry Structure**

(c) Fit the split plot model of expression 3.1 in Chapter 3 of the course notes. Assume participants (i.e., subjects) to be a random effect, with bhl ~ N (0, σ b 2). Assume ehlj to be distributed normal. Write down the **analysis of variance** table.

**1) Fit split plot model**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* Group: Effect is marginally significant (p=0.0548)
* Week: Effect is on strength is significant (p = 0.000544).
* Interaction (time \* group): Effect is significant: the changes in strength over time differ between the groups (p = 0.000544). 🡺 **At least one of (τγ)ℓj ​≠ 0**

**2) ANOVA**

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

* Time effect: that strength changes significantly over the weeks.
* Group effect: marginal

F-value for Groupf is 209.718/68.420 = 3.065

* Interaction (Time \* Group): the effect of treatment varies over time. 🡺 **At least one of (τγ)ℓj ​≠ 0**

A screenshot of a computer program

Description automatically generatedA screenshot of a computer code

Description automatically generated

A close-up of a white background

Description automatically generated

(d) Test whether the time and treatment group factors interact; use α = 0.0025 State the alternatives, decision rule, and conclusion. What is the p-value for the test?

**H0: (τγ)ℓj = 0 for all ℓ**

**H1: (τγ)ℓj ​≠ 0 for at least one ℓ**

F (interaction) = 2.99 (df: 12, 324)

p-value = 0.000544 < α = 0.0025, thus we **reject H0**.

Thus, there is enough evidence suggesting interaction effect present: the changes in strength over time significantly differ between the groups (p = 0.000544).

(e) Test separately whether or not time and treatment group main effects are present; use α = 0.025 for each test. State the alternatives, decision rule and conclusion for each test. What is the p-value for each test?

To test group:

**H0: (τ)ℓ = 0 for all ℓ**

**H1: (τ)ℓ ​≠ 0 for at least one ℓ**

F (Groupf) = 209.718/68.420 = 3.065 (df: 2, 54)

DF (group) = 3 – 1 = 2

DF (within) = 57 - 3 = 54

A red border with black text

Description automatically generated

p=0.548 > α = 0.0025, thus we **DON’T reject H0**.

Thus, no main effect of treatment.

To test time (week):

**H0: (γ)j = 0 for all j**

**H1: (γ)j ​≠ 0 for at least one j**

F (week) = 7.3702 (df: 6, 324)

P = 2.1e-7 < α = 0.0025, thus we **reject H0**.

Thus, there is enough evidence suggesting time effect exists.

(f) Use orthogonal polynomial contrasts to test for polynomial trends across time and interaction of polynomial trend by treatment group. Is there a Signiant difference in polynomials trends across time for the treatment groups. If so, what is the degree of polynomial trends that differ across groups.

A screenshot of a computer

Description automatically generated

Contrast of Linear trend CONT to RI or WI

* (Weekf.L:GroupfRI): Significant difference in the linear trend across time between the CONT and RI groups (p < 0.0001).
* (Weekf.L:GroupfWI): Significant difference in the linear trend across time between the CONT and WI groups (p < 0.0001).

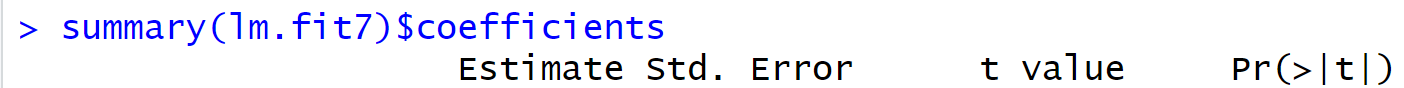
Contrast of Quadratic and higher order polynomial trend:

* No significant differences in quadratic or higher-order polynomial trends across time between the treatment groups.

The trend for CONT appears to be flat. The trend for RI may exhibit a parabolic shape, while the trend for WI shows a more pronounced linear increase within the scope of the model. The differences between the treatment groups are primarily in their linear trends, with WI demonstrating a stronger linear trend.

Thus, **the degree of polynomial trend that differs across the groups is primarily linear.**

3. Extra Credit: Obtain an estimate of the quadratic trend parameters for the RI and WI treatment groups. Also obtain standard errors for these estimates. Are the population average growth curves for WI and RI a section of a parabola? Why?



A screenshot of a computer

Description automatically generated

The estimates and standard errors for the estimates are shown in the table above.

Based on the results, no, the quadratic trend parameters for both the RI and WI are NOT significantly different. Because the p-value is too big (p > 0.05). Therefore, **the population average growth curves for WI and RI are likely not sections of a parabola**.

A math equations on a white background

Description automatically generated

Since P is orthonormal, P’P = I, identical matrix.

Each column of P sum to 0.

A math equations on a piece of paper

Description automatically generated

R code for HW2

Jiseon Yang

2024-10-01

## 1.

# Define the matrix  
A <- matrix(c(1, 0, 0, 0,  
 -1, 1, 0, 0,  
 0, -1, 1, 0,  
 0, 0, -1, 1,  
 0, 0, 0, -1),   
 nrow = 5, ncol = 4, byrow = TRUE)  
  
library(pracma)  
  
# Gram-Schmidt process  
Q <- gramSchmidt(A)$Q  
  
# Output the orthonormal matrix  
gramSchmidt(Q)

## $Q  
## [,1] [,2] [,3] [,4]  
## [1,] 0.7071068 0.4082483 0.2886751 0.2236068  
## [2,] -0.7071068 0.4082483 0.2886751 0.2236068  
## [3,] 0.0000000 -0.8164966 0.2886751 0.2236068  
## [4,] 0.0000000 0.0000000 -0.8660254 0.2236068  
## [5,] 0.0000000 0.0000000 0.0000000 -0.8944272  
##   
## $R  
## [,1] [,2] [,3] [,4]  
## [1,] 1 -1.110223e-16 -1.110223e-16 -1.110223e-16  
## [2,] 0 1.000000e+00 1.110223e-16 1.110223e-16  
## [3,] 0 0.000000e+00 1.000000e+00 8.326673e-17  
## [4,] 0 0.000000e+00 0.000000e+00 1.000000e+00

## 2 (a) Profile plot

• WI (Weights Increasing): The trend appears to be mostly linear but could have a slight quadratic component due to the accelerating increase in later weeks. • RI (Repetitions Increasing): A quadratic trend is likely, with an increase followed by a leveling off of strength. • CONT (Control): A flat or slightly negative linear trend, indicating no major polynomial trend.

# Load necessary libraries  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)  
#install.packages("tidyr")  
library(tidyr)  
library(reshape2)

##   
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':  
##   
## smiths

# Load the data  
strength\_wide <- read.table("C:/Users/jyang/OneDrive - Arizona State University/10 Classes\_OneDrive/2024 8F\_STP598\_Logitudinal/zz HW/Longitudinal\_HW2/strength.dat", header = FALSE)  
  
# Add proper column names for better understanding  
names(strength\_wide) <- c("ID", "Group", "Week1", "Week2", "Week3", "Week4", "Week5", "Week6", "Week7")  
  
# Convert data from wide format to long format  
strength\_long <- melt(strength\_wide, id.vars = c("ID", "Group"),  
 variable.name = "Week", value.name = "Strength")  
  
# Convert 'Week' to a numeric variable (removing 'Week' prefix)  
strength\_long$Week <- as.numeric(gsub("Week", "", strength\_long$Week))  
  
length(unique(strength\_long$ID))

## [1] 21

table(strength\_wide$Group)

##   
## CONT RI WI   
## 20 16 21

# Generate the plot  
strength\_long %>%   
 group\_by(Group, Week) %>%   
 summarise(Strength = mean(Strength, na.rm = TRUE), .groups = "drop") %>%   
 ggplot(aes(x = Week, y = Strength, color = as.factor(Group))) +   
 geom\_point() +  
 geom\_line() +  
 ylab("Strength Means") +   
 scale\_color\_discrete(name = "Group") +  
 ggtitle("Profile Plot of Strength Over Time by Treatment Group") +  
 xlab("Week")

A graph with a line graph

Description automatically generated with medium confidence

## 2 (c) Fit the Split-Plot Model: ANOVA approach -. use aov() function -. best suit for banlanced model (not for unbalanced model) -. contrast function

1. aov.fit1: Split-plot model with a random effect for ID handled by the Error(IDf) term. Best suited for balanced designs.
2. aov.fit2: ANOVA model treating ID as a fixed effect. Overfits when there are many participants and treats individual variability as a fixed difference rather than random variability.
3. lm.fit3: Similar to aov.fit1, it treats ID as a fixed effect using a standard linear regression model. Again, this leads to overfitting when there are many participants.
4. lmer.fit4: Mixed-effects model that treats ID as a random effect. This is the best choice for repeated measures data or hierarchical data, and it can handle unbalanced data. It provides a flexible structure to account for both fixed and random effects, and it generalizes better than models treating ID as a fixed effect.

# Load the data  
strength\_wide <- read.table("C:/Users/jyang/OneDrive - Arizona State University/10 Classes\_OneDrive/2024 8F\_STP598\_Logitudinal/zz HW/Longitudinal\_HW2/strength.dat", header = FALSE)  
  
# Add proper column names for better understanding  
names(strength\_wide) <- c("ID", "Group", "Week1", "Week2", "Week3", "Week4", "Week5", "Week6", "Week7")  
  
# Provide a Unique ID for each subject by combining ID and Group so that the same ID# is not considered to be the same subject accross different groups  
strength\_wide$UniqueID <- interaction(strength\_wide$ID, strength\_wide$Group)  
  
# Convert data from wide format to long format (reshaping only once)  
strength\_long <- melt(strength\_wide, id.vars = c("UniqueID", "Group"),  
 variable.name = "Week", value.name = "Strength")  
  
# Remove 'Week' prefix and convert Week to a numeric variable  
strength\_long$Week <- as.numeric(gsub("Week", "", strength\_long$Week))

## Warning: NAs introduced by coercion

# Convert Week, UniqueID, and Group to factors  
strength\_long$Weekf <- as.factor(strength\_long$Week)  
strength\_long$UniqueIDf <- as.factor(strength\_long$UniqueID)  
strength\_long$Groupf <- as.factor(strength\_long$Group)  
  
# Contrast  
levels(strength\_long$Weekf) # Check levels of Weekf

## [1] "1" "2" "3" "4" "5" "6" "7"

contrasts(strength\_long$Weekf) <- contr.poly(7)  
  
# 1) aov.fit1: Split-plot model with a random effect for ID handled by the Error(IDf) term. Best suited for balanced designs.  
# aov.fit1 <- aov(Strength ~ Weekf + Groupf + Weekf\*Group + Error(UniqueIDf), data = strength\_long)  
# aov.fit1  
# summary(aov.fit1)  
  
# 2) aov.fit2: ANOVA model treating ID as a fixed effect.   
aov.fit2 <- aov(Strength ~ Weekf + Groupf + Weekf\*Group + UniqueIDf, data = strength\_long)  
summary(aov.fit2)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Weekf 6 53 8.82 7.370 2.1e-07 \*\*\*  
## Groupf 2 419 209.72 175.221 < 2e-16 \*\*\*  
## UniqueIDf 54 3695 68.42 57.166 < 2e-16 \*\*\*  
## Weekf:Group 12 43 3.58 2.994 0.000544 \*\*\*  
## Residuals 324 388 1.20   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## 57 observations deleted due to missingness

# 3) lm.fit3: Similar to aov.fit1, it treats ID as a fixed effect using a standard linear regression model.   
lm.fit3 <- lm(Strength ~ Weekf + Groupf + Weekf\*Groupf + UniqueIDf, data = strength\_long)  
summary(lm.fit3)

##   
## Call:  
## lm(formula = Strength ~ Weekf + Groupf + Weekf \* Groupf + UniqueIDf,   
## data = strength\_long)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.3500 -0.6929 0.0143 0.6170 3.5179   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 85.85714 0.41350 207.635 < 2e-16 \*\*\*  
## Weekf.L -0.25513 0.24463 -1.043 0.297769   
## Weekf.Q -0.31096 0.24463 -1.271 0.204588   
## Weekf.C 0.16330 0.24463 0.668 0.504904   
## Weekf^4 0.10476 0.24463 0.428 0.668770   
## Weekf^5 0.02728 0.24463 0.112 0.911286   
## Weekf^6 -0.06415 0.24463 -0.262 0.793307   
## GroupfRI -1.71429 0.58478 -2.932 0.003613 \*\*   
## GroupfWI -4.71429 0.58478 -8.062 1.47e-14 \*\*\*  
## UniqueIDf2.CONT -7.14286 0.58478 -12.215 < 2e-16 \*\*\*  
## UniqueIDf3.CONT -9.00000 0.58478 -15.390 < 2e-16 \*\*\*  
## UniqueIDf4.CONT -1.71429 0.58478 -2.932 0.003613 \*\*   
## UniqueIDf5.CONT -6.00000 0.58478 -10.260 < 2e-16 \*\*\*  
## UniqueIDf6.CONT -9.00000 0.58478 -15.390 < 2e-16 \*\*\*  
## UniqueIDf7.CONT -6.28571 0.58478 -10.749 < 2e-16 \*\*\*  
## UniqueIDf8.CONT -10.71429 0.58478 -18.322 < 2e-16 \*\*\*  
## UniqueIDf9.CONT -6.71429 0.58478 -11.482 < 2e-16 \*\*\*  
## UniqueIDf10.CONT -7.28571 0.58478 -12.459 < 2e-16 \*\*\*  
## UniqueIDf11.CONT -5.14286 0.58478 -8.795 < 2e-16 \*\*\*  
## UniqueIDf12.CONT -8.85714 0.58478 -15.146 < 2e-16 \*\*\*  
## UniqueIDf13.CONT -2.85714 0.58478 -4.886 1.62e-06 \*\*\*  
## UniqueIDf14.CONT -4.28571 0.58478 -7.329 1.86e-12 \*\*\*  
## UniqueIDf15.CONT -4.85714 0.58478 -8.306 2.73e-15 \*\*\*  
## UniqueIDf16.CONT -7.85714 0.58478 -13.436 < 2e-16 \*\*\*  
## UniqueIDf17.CONT -2.71429 0.58478 -4.642 5.03e-06 \*\*\*  
## UniqueIDf18.CONT -7.85714 0.58478 -13.436 < 2e-16 \*\*\*  
## UniqueIDf19.CONT -6.14286 0.58478 -10.505 < 2e-16 \*\*\*  
## UniqueIDf20.CONT -6.28571 0.58478 -10.749 < 2e-16 \*\*\*  
## UniqueIDf1.RI -4.85714 0.58478 -8.306 2.73e-15 \*\*\*  
## UniqueIDf2.RI 1.00000 0.58478 1.710 0.088213 .   
## UniqueIDf3.RI -1.85714 0.58478 -3.176 0.001638 \*\*   
## UniqueIDf4.RI -2.57143 0.58478 -4.397 1.49e-05 \*\*\*  
## UniqueIDf5.RI -1.71429 0.58478 -2.932 0.003613 \*\*   
## UniqueIDf6.RI -8.28571 0.58478 -14.169 < 2e-16 \*\*\*  
## UniqueIDf7.RI -0.42857 0.58478 -0.733 0.464161   
## UniqueIDf8.RI -4.57143 0.58478 -7.817 7.63e-14 \*\*\*  
## UniqueIDf9.RI 2.14286 0.58478 3.664 0.000290 \*\*\*  
## UniqueIDf10.RI -7.00000 0.58478 -11.970 < 2e-16 \*\*\*  
## UniqueIDf11.RI -7.57143 0.58478 -12.948 < 2e-16 \*\*\*  
## UniqueIDf12.RI 1.14286 0.58478 1.954 0.051520 .   
## UniqueIDf13.RI -3.85714 0.58478 -6.596 1.72e-10 \*\*\*  
## UniqueIDf14.RI -7.71429 0.58478 -13.192 < 2e-16 \*\*\*  
## UniqueIDf15.RI -7.42857 0.58478 -12.703 < 2e-16 \*\*\*  
## UniqueIDf16.RI NA NA NA NA   
## UniqueIDf1.WI 2.57143 0.58478 4.397 1.49e-05 \*\*\*  
## UniqueIDf2.WI -5.85714 0.58478 -10.016 < 2e-16 \*\*\*  
## UniqueIDf3.WI 1.42857 0.58478 2.443 0.015102 \*   
## UniqueIDf4.WI 5.57143 0.58478 9.527 < 2e-16 \*\*\*  
## UniqueIDf5.WI 3.00000 0.58478 5.130 5.00e-07 \*\*\*  
## UniqueIDf6.WI -1.71429 0.58478 -2.932 0.003613 \*\*   
## UniqueIDf7.WI -0.42857 0.58478 -0.733 0.464161   
## UniqueIDf8.WI 9.14286 0.58478 15.635 < 2e-16 \*\*\*  
## UniqueIDf9.WI 0.71429 0.58478 1.221 0.222797   
## UniqueIDf10.WI 2.85714 0.58478 4.886 1.62e-06 \*\*\*  
## UniqueIDf11.WI -0.85714 0.58478 -1.466 0.143683   
## UniqueIDf12.WI 0.14286 0.58478 0.244 0.807158   
## UniqueIDf13.WI 2.42857 0.58478 4.153 4.20e-05 \*\*\*  
## UniqueIDf14.WI 1.42857 0.58478 2.443 0.015102 \*   
## UniqueIDf15.WI -2.57143 0.58478 -4.397 1.49e-05 \*\*\*  
## UniqueIDf16.WI 2.00000 0.58478 3.420 0.000706 \*\*\*  
## UniqueIDf17.WI -1.28571 0.58478 -2.199 0.028610 \*   
## UniqueIDf18.WI 0.14286 0.58478 0.244 0.807158   
## UniqueIDf19.WI 4.85714 0.58478 8.306 2.73e-15 \*\*\*  
## UniqueIDf20.WI -0.85714 0.58478 -1.466 0.143683   
## UniqueIDf21.WI NA NA NA NA   
## Weekf.L:GroupfRI 1.36540 0.36694 3.721 0.000234 \*\*\*  
## Weekf.Q:GroupfRI -0.35733 0.36694 -0.974 0.330880   
## Weekf.C:GroupfRI 0.01531 0.36694 0.042 0.966747   
## Weekf^4:GroupfRI -0.21052 0.36694 -0.574 0.566559   
## Weekf^5:GroupfRI 0.12275 0.36694 0.335 0.738209   
## Weekf^6:GroupfRI 0.11761 0.36694 0.321 0.748789   
## Weekf.L:GroupfWI 1.94697 0.34182 5.696 2.75e-08 \*\*\*  
## Weekf.Q:GroupfWI -0.01117 0.34182 -0.033 0.973950   
## Weekf.C:GroupfWI -0.04666 0.34182 -0.136 0.891513   
## Weekf^4:GroupfWI 0.01036 0.34182 0.030 0.975838   
## Weekf^5:GroupfWI 0.12859 0.34182 0.376 0.707010   
## Weekf^6:GroupfWI -0.09250 0.34182 -0.271 0.786849   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.094 on 324 degrees of freedom  
## (57 observations deleted due to missingness)  
## Multiple R-squared: 0.9157, Adjusted R-squared: 0.8964   
## F-statistic: 47.53 on 74 and 324 DF, p-value: < 2.2e-16

# ANOVA  
#anova(aov.fit1)  
anova(aov.fit2)

## Analysis of Variance Table  
##   
## Response: Strength  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Weekf 6 52.9 8.821 7.3702 2.102e-07 \*\*\*  
## Groupf 2 419.4 209.718 175.2213 < 2.2e-16 \*\*\*  
## UniqueIDf 54 3694.7 68.420 57.1658 < 2.2e-16 \*\*\*  
## Weekf:Group 12 43.0 3.583 2.9939 0.000544 \*\*\*  
## Residuals 324 387.8 1.197   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(lm.fit3)

## Analysis of Variance Table  
##   
## Response: Strength  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Weekf 6 52.9 8.821 7.3702 2.102e-07 \*\*\*  
## Groupf 2 419.4 209.718 175.2213 < 2.2e-16 \*\*\*  
## UniqueIDf 54 3694.7 68.420 57.1658 < 2.2e-16 \*\*\*  
## Weekf:Groupf 12 43.0 3.583 2.9939 0.000544 \*\*\*  
## Residuals 324 387.8 1.197   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(lm.fit3)

## Analysis of Variance Table  
##   
## Response: Strength  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Weekf 6 52.9 8.821 7.3702 2.102e-07 \*\*\*  
## Groupf 2 419.4 209.718 175.2213 < 2.2e-16 \*\*\*  
## UniqueIDf 54 3694.7 68.420 57.1658 < 2.2e-16 \*\*\*  
## Weekf:Groupf 12 43.0 3.583 2.9939 0.000544 \*\*\*  
## Residuals 324 387.8 1.197   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## R Markdown

## **2(a). Using New data table**

Because of missing data, I recreated datasheet to provide each different subject ID. The new data file is saved as csv.

# Load necessary libraries  
library(dplyr)  
library(ggplot2)  
#install.packages("tidyr")  
library(tidyr)  
library(reshape2)  
  
# Load the data  
strength\_wide1 <- read.csv("C:/Users/jyang/OneDrive - Arizona State University/10 Classes\_OneDrive/2024 8F\_STP598\_Logitudinal/zz HW/Longitudinal\_HW2/strength.csv", header = FALSE)  
  
str(strength\_wide)

## 'data.frame': 57 obs. of 10 variables:  
## $ ID : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ Group : chr "CONT" "CONT" "CONT" "CONT" ...  
## $ Week1 : int 85 80 78 84 80 76 79 76 77 79 ...  
## $ Week2 : int 85 79 77 84 81 78 79 76 78 79 ...  
## $ Week3 : int 86 79 77 85 80 77 80 76 78 79 ...  
## $ Week4 : int 85 78 77 84 80 78 79 75 80 79 ...  
## $ Week5 : int 87 78 76 83 79 78 80 75 80 77 ...  
## $ Week6 : int 86 79 76 84 79 77 79 74 81 78 ...  
## $ Week7 : int 87 78 77 85 80 74 81 74 80 79 ...  
## $ UniqueID: Factor w/ 63 levels "1.CONT","2.CONT",..: 1 2 3 4 5 6 7 8 9 10 ...

# Add proper column names for better understanding  
names(strength\_wide1) <- c("ID", "Group", "Week1", "Week2", "Week3", "Week4", "Week5", "Week6", "Week7")  
  
head(strength\_wide1)

## ID Group Week1 Week2 Week3 Week4 Week5 Week6 Week7  
## 1 1 CONT 85 85 86 85 87 86 87  
## 2 2 CONT 80 79 79 78 78 79 78  
## 3 3 CONT 78 77 77 77 76 76 77  
## 4 4 CONT 84 84 85 84 83 84 85  
## 5 5 CONT 80 81 80 80 79 79 80  
## 6 6 CONT 76 78 77 78 78 77 74

# Convert data from wide format to long format  
strength\_long1 <- melt(strength\_wide1, id.vars = c("ID", "Group"),  
 variable.name = "Week", value.name = "Strength")  
  
# Convert 'Week' to a numeric variable (removing 'Week' prefix)  
strength\_long1$Week <- as.numeric(gsub("Week", "", strength\_long1$Week))  
  
length(unique(strength\_long1$ID))

## [1] 57

table(strength\_wide1$Group)

##   
## CONT RI WI   
## 20 16 21

# Generate the plot  
strength\_long1 %>%   
 group\_by(Group, Week) %>%   
 summarise(Strength = mean(Strength, na.rm = TRUE), .groups = "drop") %>%   
 ggplot(aes(x = Week, y = Strength, color = as.factor(Group))) +   
 geom\_point() +  
 geom\_line() +  
 ylab("Strength Means") +   
 scale\_color\_discrete(name = "Group") +  
 ggtitle("Profile Plot of Strength Over Time by Treatment Group-1") +  
 xlab("Week")

A graph with a line graph

Description automatically generated with medium confidence

## 2(c) Using New data table

# Convert data from wide format to long format (reshaping only once)  
strength\_long1 <- melt(strength\_wide1, id.vars = c("ID", "Group"),  
 variable.name = "Week", value.name = "Strength")  
  
# Remove 'Week' prefix and convert Week to a numeric variable  
strength\_long1$Week <- as.numeric(gsub("Week", "", strength\_long1$Week))  
  
# Convert Week, ID, and Group to factors  
strength\_long1$Weekf <- as.factor(strength\_long1$Week)  
strength\_long1$IDf <- as.factor(strength\_long1$ID)  
strength\_long1$Groupf <- as.factor(strength\_long1$Group)  
  
# Contrast  
levels(strength\_long1$Weekf) # Check levels of Weekf

## [1] "1" "2" "3" "4" "5" "6" "7"

contrasts(strength\_long1$Weekf) <- contr.poly(7)  
contrasts(strength\_long1$Weekf)

## .L .Q .C ^4 ^5  
## 1 -5.669467e-01 5.455447e-01 -4.082483e-01 0.2417469 -1.091089e-01  
## 2 -3.779645e-01 9.690821e-17 4.082483e-01 -0.5640761 4.364358e-01  
## 3 -1.889822e-01 -3.273268e-01 4.082483e-01 0.0805823 -5.455447e-01  
## 4 2.098124e-17 -4.364358e-01 4.532467e-17 0.4834938 5.342065e-16  
## 5 1.889822e-01 -3.273268e-01 -4.082483e-01 0.0805823 5.455447e-01  
## 6 3.779645e-01 0.000000e+00 -4.082483e-01 -0.5640761 -4.364358e-01  
## 7 5.669467e-01 5.455447e-01 4.082483e-01 0.2417469 1.091089e-01  
## ^6  
## 1 0.03289758  
## 2 -0.19738551  
## 3 0.49346377  
## 4 -0.65795169  
## 5 0.49346377  
## 6 -0.19738551  
## 7 0.03289758

# 1) aov.fit5: Split-plot model with a random effect for ID handled by the Error(IDf) term. Best suited for balanced designs.  
aov.fit5 <- aov(Strength ~ Weekf + Groupf + Weekf\*Group + Error(IDf), data = strength\_long1)  
summary(aov.fit5)

##   
## Error: IDf  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Groupf 2 419 209.72 3.065 0.0548 .  
## Residuals 54 3695 68.42   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Error: Within  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Weekf 6 52.9 8.821 7.370 2.1e-07 \*\*\*  
## Weekf:Group 12 43.0 3.583 2.994 0.000544 \*\*\*  
## Residuals 324 387.8 1.197   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# 2) aov.fit6: ANOVA model treating ID as a fixed effect.   
aov.fit6 <- aov(Strength ~ Weekf + Groupf + Weekf\*Group + IDf, data = strength\_long1)  
summary(aov.fit6)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Weekf 6 53 8.82 7.370 2.1e-07 \*\*\*  
## Groupf 2 419 209.72 175.221 < 2e-16 \*\*\*  
## IDf 54 3695 68.42 57.166 < 2e-16 \*\*\*  
## Weekf:Group 12 43 3.58 2.994 0.000544 \*\*\*  
## Residuals 324 388 1.20   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# 3) lm.fit7: Similar to aov.fit1, it treats ID as a fixed effect using a standard linear regression model.   
lm.fit7 <- lm(Strength ~ Weekf + Groupf + Weekf\*Groupf + IDf, data = strength\_long1)  
summary(lm.fit7)

##   
## Call:  
## lm(formula = Strength ~ Weekf + Groupf + Weekf \* Groupf + IDf,   
## data = strength\_long1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.3500 -0.6929 0.0143 0.6170 3.5179   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 85.85714 0.41350 207.635 < 2e-16 \*\*\*  
## Weekf.L -0.25513 0.24463 -1.043 0.297769   
## Weekf.Q -0.31096 0.24463 -1.271 0.204588   
## Weekf.C 0.16330 0.24463 0.668 0.504904   
## Weekf^4 0.10476 0.24463 0.428 0.668770   
## Weekf^5 0.02728 0.24463 0.112 0.911286   
## Weekf^6 -0.06415 0.24463 -0.262 0.793307   
## GroupfRI -1.71429 0.58478 -2.932 0.003613 \*\*   
## GroupfWI -4.71429 0.58478 -8.062 1.47e-14 \*\*\*  
## IDf2 -7.14286 0.58478 -12.215 < 2e-16 \*\*\*  
## IDf3 -9.00000 0.58478 -15.390 < 2e-16 \*\*\*  
## IDf4 -1.71429 0.58478 -2.932 0.003613 \*\*   
## IDf5 -6.00000 0.58478 -10.260 < 2e-16 \*\*\*  
## IDf6 -9.00000 0.58478 -15.390 < 2e-16 \*\*\*  
## IDf7 -6.28571 0.58478 -10.749 < 2e-16 \*\*\*  
## IDf8 -10.71429 0.58478 -18.322 < 2e-16 \*\*\*  
## IDf9 -6.71429 0.58478 -11.482 < 2e-16 \*\*\*  
## IDf10 -7.28571 0.58478 -12.459 < 2e-16 \*\*\*  
## IDf11 -5.14286 0.58478 -8.795 < 2e-16 \*\*\*  
## IDf12 -8.85714 0.58478 -15.146 < 2e-16 \*\*\*  
## IDf13 -2.85714 0.58478 -4.886 1.62e-06 \*\*\*  
## IDf14 -4.28571 0.58478 -7.329 1.86e-12 \*\*\*  
## IDf15 -4.85714 0.58478 -8.306 2.73e-15 \*\*\*  
## IDf16 -7.85714 0.58478 -13.436 < 2e-16 \*\*\*  
## IDf17 -2.71429 0.58478 -4.642 5.03e-06 \*\*\*  
## IDf18 -7.85714 0.58478 -13.436 < 2e-16 \*\*\*  
## IDf19 -6.14286 0.58478 -10.505 < 2e-16 \*\*\*  
## IDf20 -6.28571 0.58478 -10.749 < 2e-16 \*\*\*  
## IDf21 -4.85714 0.58478 -8.306 2.73e-15 \*\*\*  
## IDf22 1.00000 0.58478 1.710 0.088213 .   
## IDf23 -1.85714 0.58478 -3.176 0.001638 \*\*   
## IDf24 -2.57143 0.58478 -4.397 1.49e-05 \*\*\*  
## IDf25 -1.71429 0.58478 -2.932 0.003613 \*\*   
## IDf26 -8.28571 0.58478 -14.169 < 2e-16 \*\*\*  
## IDf27 -0.42857 0.58478 -0.733 0.464161   
## IDf28 -4.57143 0.58478 -7.817 7.63e-14 \*\*\*  
## IDf29 2.14286 0.58478 3.664 0.000290 \*\*\*  
## IDf30 -7.00000 0.58478 -11.970 < 2e-16 \*\*\*  
## IDf31 -7.57143 0.58478 -12.948 < 2e-16 \*\*\*  
## IDf32 1.14286 0.58478 1.954 0.051520 .   
## IDf33 -3.85714 0.58478 -6.596 1.72e-10 \*\*\*  
## IDf34 -7.71429 0.58478 -13.192 < 2e-16 \*\*\*  
## IDf35 -7.42857 0.58478 -12.703 < 2e-16 \*\*\*  
## IDf36 NA NA NA NA   
## IDf37 2.57143 0.58478 4.397 1.49e-05 \*\*\*  
## IDf38 -5.85714 0.58478 -10.016 < 2e-16 \*\*\*  
## IDf39 1.42857 0.58478 2.443 0.015102 \*   
## IDf40 5.57143 0.58478 9.527 < 2e-16 \*\*\*  
## IDf41 3.00000 0.58478 5.130 5.00e-07 \*\*\*  
## IDf42 -1.71429 0.58478 -2.932 0.003613 \*\*   
## IDf43 -0.42857 0.58478 -0.733 0.464161   
## IDf44 9.14286 0.58478 15.635 < 2e-16 \*\*\*  
## IDf45 0.71429 0.58478 1.221 0.222797   
## IDf46 2.85714 0.58478 4.886 1.62e-06 \*\*\*  
## IDf47 -0.85714 0.58478 -1.466 0.143683   
## IDf48 0.14286 0.58478 0.244 0.807158   
## IDf49 2.42857 0.58478 4.153 4.20e-05 \*\*\*  
## IDf50 1.42857 0.58478 2.443 0.015102 \*   
## IDf51 -2.57143 0.58478 -4.397 1.49e-05 \*\*\*  
## IDf52 2.00000 0.58478 3.420 0.000706 \*\*\*  
## IDf53 -1.28571 0.58478 -2.199 0.028610 \*   
## IDf54 0.14286 0.58478 0.244 0.807158   
## IDf55 4.85714 0.58478 8.306 2.73e-15 \*\*\*  
## IDf56 -0.85714 0.58478 -1.466 0.143683   
## IDf57 NA NA NA NA   
## Weekf.L:GroupfRI 1.36540 0.36694 3.721 0.000234 \*\*\*  
## Weekf.Q:GroupfRI -0.35733 0.36694 -0.974 0.330880   
## Weekf.C:GroupfRI 0.01531 0.36694 0.042 0.966747   
## Weekf^4:GroupfRI -0.21052 0.36694 -0.574 0.566559   
## Weekf^5:GroupfRI 0.12275 0.36694 0.335 0.738209   
## Weekf^6:GroupfRI 0.11761 0.36694 0.321 0.748789   
## Weekf.L:GroupfWI 1.94697 0.34182 5.696 2.75e-08 \*\*\*  
## Weekf.Q:GroupfWI -0.01117 0.34182 -0.033 0.973950   
## Weekf.C:GroupfWI -0.04666 0.34182 -0.136 0.891513   
## Weekf^4:GroupfWI 0.01036 0.34182 0.030 0.975838   
## Weekf^5:GroupfWI 0.12859 0.34182 0.376 0.707010   
## Weekf^6:GroupfWI -0.09250 0.34182 -0.271 0.786849   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.094 on 324 degrees of freedom  
## Multiple R-squared: 0.9157, Adjusted R-squared: 0.8964   
## F-statistic: 47.53 on 74 and 324 DF, p-value: < 2.2e-16

# 4) lmer.fit8: Mixed-effects model that treats ID as a random effect. Best recommended for repeated measures data. Also, it can handle unbalanced data. It provides a flexible structure to account for both fixed and random effects, and it generalizes better than models treating ID as a fixed effect.  
# lmer.fit8 <- lmer(Strength ~ Weekf + Group + Weekf\*Group + (1 | IDf), data = strength\_long1)  
# summary(lmer.fit8)  
  
# ANOVA  
#anova(aov.fit5)  
anova(aov.fit6)

## Analysis of Variance Table  
##   
## Response: Strength  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Weekf 6 52.9 8.821 7.3702 2.102e-07 \*\*\*  
## Groupf 2 419.4 209.718 175.2213 < 2.2e-16 \*\*\*  
## IDf 54 3694.7 68.420 57.1658 < 2.2e-16 \*\*\*  
## Weekf:Group 12 43.0 3.583 2.9939 0.000544 \*\*\*  
## Residuals 324 387.8 1.197   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(lm.fit7)

## Analysis of Variance Table  
##   
## Response: Strength  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Weekf 6 52.9 8.821 7.3702 2.102e-07 \*\*\*  
## Groupf 2 419.4 209.718 175.2213 < 2.2e-16 \*\*\*  
## IDf 54 3694.7 68.420 57.1658 < 2.2e-16 \*\*\*  
## Weekf:Groupf 12 43.0 3.583 2.9939 0.000544 \*\*\*  
## Residuals 324 387.8 1.197   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# anova(lmer.fit8)

## 3. Extra credit

summary(lm.fit7)$coefficients

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 85.85714286 0.4134995 207.63544009 0.000000e+00  
## Weekf.L -0.25512602 0.2446296 -1.04290748 2.977687e-01  
## Weekf.Q -0.31096049 0.2446296 -1.27114837 2.045877e-01  
## Weekf.C 0.16329932 0.2446296 0.66753708 5.049043e-01  
## Weekf^4 0.10475699 0.2446296 0.42822697 6.687704e-01  
## Weekf^5 0.02727724 0.2446296 0.11150424 9.112855e-01  
## Weekf^6 -0.06415029 0.2446296 -0.26223440 7.933074e-01  
## GroupfRI -1.71428571 0.5847765 -2.93152268 3.612939e-03  
## GroupfWI -4.71428571 0.5847765 -8.06168738 1.467538e-14  
## IDf2 -7.14285714 0.5847765 -12.21467785 1.759602e-28  
## IDf3 -9.00000000 0.5847765 -15.39049409 1.677480e-40  
## IDf4 -1.71428571 0.5847765 -2.93152268 3.612939e-03  
## IDf5 -6.00000000 0.5847765 -10.26032939 1.421545e-21  
## IDf6 -9.00000000 0.5847765 -15.39049409 1.677480e-40  
## IDf7 -6.28571429 0.5847765 -10.74891650 2.988716e-23  
## IDf8 -10.71428571 0.5847765 -18.32201677 5.838627e-52  
## IDf9 -6.71428571 0.5847765 -11.48179717 7.842848e-26  
## IDf10 -7.28571429 0.5847765 -12.45897140 2.234285e-29  
## IDf11 -5.14285714 0.5847765 -8.79456805 8.648941e-17  
## IDf12 -8.85714286 0.5847765 -15.14620053 1.481489e-39  
## IDf13 -2.85714286 0.5847765 -4.88587114 1.621736e-06  
## IDf14 -4.28571429 0.5847765 -7.32880671 1.864395e-12  
## IDf15 -4.85714286 0.5847765 -8.30598093 2.733351e-15  
## IDf16 -7.85714286 0.5847765 -13.43614563 5.101987e-33  
## IDf17 -2.71428571 0.5847765 -4.64157758 5.028993e-06  
## IDf18 -7.85714286 0.5847765 -13.43614563 5.101987e-33  
## IDf19 -6.14285714 0.5847765 -10.50462295 2.083300e-22  
## IDf20 -6.28571429 0.5847765 -10.74891650 2.988716e-23  
## IDf21 -4.85714286 0.5847765 -8.30598093 2.733351e-15  
## IDf22 1.00000000 0.5847765 1.71005490 8.821306e-02  
## IDf23 -1.85714286 0.5847765 -3.17581624 1.637826e-03  
## IDf24 -2.57142857 0.5847765 -4.39728402 1.488454e-05  
## IDf25 -1.71428571 0.5847765 -2.93152268 3.612939e-03  
## IDf26 -8.28571429 0.5847765 -14.16902630 8.459185e-36  
## IDf27 -0.42857143 0.5847765 -0.73288067 4.641611e-01  
## IDf28 -4.57142857 0.5847765 -7.81739382 7.629912e-14  
## IDf29 2.14285714 0.5847765 3.66440335 2.895463e-04  
## IDf30 -7.00000000 0.5847765 -11.97038429 1.365593e-27  
## IDf31 -7.57142857 0.5847765 -12.94755852 3.459217e-31  
## IDf32 1.14285714 0.5847765 1.95434846 5.151990e-02  
## IDf33 -3.85714286 0.5847765 -6.59592604 1.720614e-10  
## IDf34 -7.71428571 0.5847765 -13.19185207 4.224915e-32  
## IDf35 -7.42857143 0.5847765 -12.70326496 2.798148e-30  
## IDf37 2.57142857 0.5847765 4.39728402 1.488454e-05  
## IDf38 -5.85714286 0.5847765 -10.01603583 9.485607e-21  
## IDf39 1.42857143 0.5847765 2.44293557 1.510198e-02  
## IDf40 5.57142857 0.5847765 9.52744872 3.933190e-19  
## IDf41 3.00000000 0.5847765 5.13016470 4.995735e-07  
## IDf42 -1.71428571 0.5847765 -2.93152268 3.612939e-03  
## IDf43 -0.42857143 0.5847765 -0.73288067 4.641611e-01  
## IDf44 9.14285714 0.5847765 15.63478764 1.889746e-41  
## IDf45 0.71428571 0.5847765 1.22146778 2.227970e-01  
## IDf46 2.85714286 0.5847765 4.88587114 1.621736e-06  
## IDf47 -0.85714286 0.5847765 -1.46576134 1.436831e-01  
## IDf48 0.14285714 0.5847765 0.24429356 8.071581e-01  
## IDf49 2.42857143 0.5847765 4.15299047 4.201457e-05  
## IDf50 1.42857143 0.5847765 2.44293557 1.510198e-02  
## IDf51 -2.57142857 0.5847765 -4.39728402 1.488454e-05  
## IDf52 2.00000000 0.5847765 3.42010980 7.060376e-04  
## IDf53 -1.28571429 0.5847765 -2.19864201 2.860997e-02  
## IDf54 0.14285714 0.5847765 0.24429356 8.071581e-01  
## IDf55 4.85714286 0.5847765 8.30598093 2.733351e-15  
## IDf56 -0.85714286 0.5847765 -1.46576134 1.436831e-01  
## Weekf.L:GroupfRI 1.36539666 0.3669444 3.72099088 2.338730e-04  
## Weekf.Q:GroupfRI -0.35733180 0.3669444 -0.97380372 3.308805e-01  
## Weekf.C:GroupfRI 0.01530931 0.3669444 0.04172107 9.667468e-01  
## Weekf^4:GroupfRI -0.21052125 0.3669444 -0.57371434 5.665591e-01  
## Weekf^5:GroupfRI 0.12274756 0.3669444 0.33451273 7.382091e-01  
## Weekf^6:GroupfRI 0.11760887 0.3669444 0.32050871 7.487893e-01  
## Weekf.L:GroupfWI 1.94696699 0.3418151 5.69596532 2.754474e-08  
## Weekf.Q:GroupfWI -0.01117068 0.3418151 -0.03268047 9.739495e-01  
## Weekf.C:GroupfWI -0.04665695 0.3418151 -0.13649762 8.915127e-01  
## Weekf^4:GroupfWI 0.01036058 0.3418151 0.03031048 9.758381e-01  
## Weekf^5:GroupfWI 0.12859269 0.3418151 0.37620539 7.070104e-01  
## Weekf^6:GroupfWI -0.09250488 0.3418151 -0.27062840 7.868492e-01