02\_question2

44

2022-06-01

# Q2 Carry out the steps of the statistical analysis outlined below.

Turn in **commented statistical code** used to obtain the results reported. Use no more than **3 decimal places**.

Lower-limb amputation (LLA) results primarily from complications of severe peripheral artery disease, diabetes mellitus, or trauma.

Sustained exercise has been suggested by previous studies as a promising rehabilitation target to improve the long-term health after LLA.

Using telehealth, a clinical trial has been conducted at the Veterans Affairs Eastern Colorado Regional Amputation Center to test the potential of sustaining walking exercise using **exercise self-management (EXP)** versus an **attention-control education (CTL)** program.

* The **primary outcome** is *step count per day*
* A few other **secondary outcomes** are also of interest including *presence of any adverse events*, as binary outcome.
* Outcomes were collected at **baseline (time=0)**, **6 months (time=6)** and **18 months (time=18)**.
* The primary hypothesis of the trial was that **the intervention would improve daily steps in at least 1000 steps**.
* Because randomization was **stratified by age ( and )** and **level of amputation (below knee and above knee)**, these two variables should be adjusted for in the analyses.
* The longitudinal dataset of the study is in the table data-lla.csv with data dictionary below:
  + treat: treat = EXP for individuals in EXP group, and treat = CTL for individuals in CTL group
  + stepc: Step count
  + amp.above: amp.above=1 for amputation above the knee, amp.above=0 for amputation below the knee
  + time: Month from baseline
  + older: older=1 for individuals 60 or older, and older=0 for individuals younger than 60 years of age
* Use R or SAS to answer the following questions regarding the hypothesis above.

## Part 2a

The primary hypothesis of the study was to test **whether there was a difference of more than 1000 steps in the change of step count/day at 6 months between exposure groups**.

This particular type of trial is often referred to as a **superiority trial by a margin**. Carry out this hypothesis test and interpret your results for a clinical collaborator.

### i.

Write the model formally in mathematical form. Allow for a difference in daily step count at baseline.

There are four different types of model can be used to solve only for 2a(i) - (iv):

* a linear change score model (R code fit\_2\_0)
* a linear mixed model with random intercept (R code fit\_2\_1)
* a generalized linear mixed model for Poisson regression with random intercept (R code fit\_2\_2)
* a generalized estimate equation model for Poisson regression with compound symmetry correlation structure (not included due to only include marginal model)

However if we still need to consider the question for 2b, a linear mixed model or a generalized linear mixed model conditional on each individual are preferred.

There are pros and cons for each model: For linear mixed model, the outcomes are counts as integers (not continuous), so there will be violations on the assumption for normally distributed residuals. however it is easier to calculate expected mean, standard error, and statistics for given groups, as well as, to test the primary hypothesis of the numeric difference between group.

For generalized linear mixed model on Poisson regressor, the model is perfect for count-outcomes. However the model can only provide the risk rate (log risk rate), it will be hard to calculate a meaningful standard error for given groups, so it is hard to test the primary hypothesis. Besides there might be over- or under-dispersion situation to mitigate the performance of the model.

According to figures and tables below in 2a(v): There is nonlinear trend for the trajectories in treatment group, it is reasonable to treat time as a categorical variable with (0, 6, and 18 three levels) other than a continuous variable. Due to the parallel trajectory patterns for subjects from different groups, a random intercept will be applied for both models. The time effects are different between different intervention groups, so a time, intervention interaction term must be included in the model.

I will also recommend to build regression model on risk difference in future studies, however the related materials are not included in the lectures.

#### Linear mixed model with only random intercept

The and is identical for every individual subject . Here we only use to distinguish the individual variance-covariance from a population level varaince-covariance and

#### Poisson generalized linear mixed model

In such case, the interpretation for s are risk ratios and the variance is not meaningful for the variance for numeric difference. Hence we have to use a first order Taylor series approximation, similar to variance stablization.

### ii.

Write down the null and alternative hypothesis of interest in terms of the model parameters specified in (a); keep the terms consistent with (a).

Null hypothesis:

After adjusted for the other variables, there is no difference between LLA patients from the treatment group and control group, on *whether the change of step counts over 1000 at 6 month*. Alternative hypothesis:

After adjustment for baseline step counts and stratification on age and amputation location, treatment as a statistically significant effect on *whether the step counts over 1000 at 6 month* over control group.

For the older age group and amputation location above:

:

:

For the younger age group and amputation location above:

:

:

For the older age group and amputation location below:

:

:

For the younger age group and amputation location below:

:

:

### iii. Calculate the change (from baseline) in expected means and standard error of the step counts at 6 months for each the intervention and the control group.

## import dataset data\_lla  
data\_lla <- read\_csv("data/data-lla.csv",   
 col\_types = cols(  
 ## amputation location as binary variable  
 amp.above = col\_factor(levels = c("0", "1")),  
 ## older age or not as binary variable   
 older = col\_factor(levels = c("0", "1"))))   
  
  
data\_lla\_diff6 <- data\_lla %>%   
 ## change the from long format into wide format  
 pivot\_wider(names\_from = "time",  
 values\_from = "stepc") %>%  
 ## calculated the different from time0 to time6  
 mutate(step\_diff = abs(`6` - `0`)) %>%   
 ## binary outcome of whether the change of step is over 1000  
 mutate(step\_diff\_1k = case\_when(step\_diff >= 1000 ~ 1,  
 step\_diff < 1000 ~ 0),  
 baseline = `0`)  
  
# View(data\_lla\_diff6)

#### using change score model

fit\_2\_0 <- lm(step\_diff ~ 1 + treat + older + amp.above,  
 # control = fit\_control,  
 data = data\_lla\_diff6)  
  
## to get the coefficients  
beta\_2\_0 <- coef(fit\_2\_0) %>% round(digits = 3)  
beta\_2\_0

## (Intercept) treatEXP older1 amp.above1   
## 58.818 1081.156 -0.543 1.223

## to get the variance-covariance  
vcov\_2\_0 <- vcov(fit\_2\_0) %>% round(digits = 3)  
vcov\_2\_0

## (Intercept) treatEXP older1 amp.above1  
## (Intercept) 0.551 -0.349 -0.316 -0.266  
## treatEXP -0.349 0.658 0.036 0.017  
## older1 -0.316 0.036 0.680 0.048  
## amp.above1 -0.266 0.017 0.048 0.778

## [1] [2] [3] [4]   
## coefs (Int) treatEXP old1 amp.above1  
con\_trt\_old\_abv <- c(1, 1, 1, 1)  
con\_trt\_old\_blw <- c(1, 1, 1, 0)  
con\_trt\_yng\_abv <- c(1, 1, 0, 1)  
con\_trt\_yng\_blw <- c(1, 1, 0, 0)  
  
con\_ctl\_old\_abv <- c(1, 0, 1, 1)  
con\_ctl\_old\_blw <- c(1, 0, 1, 0)  
con\_ctl\_yng\_abv <- c(1, 0, 0, 1)  
con\_ctl\_yng\_blw <- c(1, 0, 0, 0)  
  
## this is the contrast of all the   
contr1 <- cbind(con\_trt\_old\_abv,  
 con\_trt\_old\_blw,  
 con\_trt\_yng\_abv,  
 con\_trt\_yng\_blw,  
 con\_ctl\_old\_abv,  
 con\_ctl\_old\_blw,  
 con\_ctl\_yng\_abv,  
 con\_ctl\_yng\_blw)  
  
contr1\_1 <- cbind(con\_trt\_old\_abv - con\_ctl\_old\_abv,  
 con\_trt\_old\_blw - con\_ctl\_old\_blw,  
 con\_trt\_yng\_abv - con\_ctl\_yng\_abv,  
 con\_trt\_yng\_blw - con\_ctl\_yng\_blw)  
  
rownames(contr1) <- rownames(vcov\_2\_0)

mu\_2\_0 <- t(contr1) %\*% beta\_2\_0 %>%  
 as.data.frame() %>%  
 dplyr::select("expected\_mean" = 1)  
## contrast variance covariance matrix  
sd\_2\_0 <- t(contr1) %\*% vcov\_2\_0 %\*% contr1 %>%  
 diag() %>% sqrt()  
cbind(mu\_2\_0, "standard\_error" = sd\_2\_0) %>%  
 round(3)

## expected\_mean standard\_error  
## con\_trt\_old\_abv 1140.654 1.003  
## con\_trt\_old\_blw 1139.431 0.794  
## con\_trt\_yng\_abv 1141.197 0.889  
## con\_trt\_yng\_blw 1139.974 0.715  
## con\_ctl\_old\_abv 59.498 0.970  
## con\_ctl\_old\_blw 58.275 0.774  
## con\_ctl\_yng\_abv 60.041 0.893  
## con\_ctl\_yng\_blw 58.818 0.742

mu\_2\_01 <- t(contr1\_1) %\*% beta\_2\_0 %>%  
 as.data.frame() %>%  
 dplyr::select("expected\_mean" = 1)  
## contrast variance covariance matrix  
sd\_2\_01 <- t(contr1\_1) %\*% vcov\_2\_0 %\*% contr1\_1 %>%  
 diag() %>% sqrt()  
cbind(mu\_2\_01, "standard\_error" = sd\_2\_01) %>%  
 round(3)

## expected\_mean standard\_error  
## 1 1081.156 0.811  
## 2 1081.156 0.811  
## 3 1081.156 0.811  
## 4 1081.156 0.811

qnorm(0.05, 1081.156, 0.811)

## [1] 1079.822

#### using linear mixed model

data\_lla\_tf <- data\_lla %>%  
 mutate(time = as.factor(time))  
  
  
fit\_2\_1 <- lmer(stepc ~ 1 + treat \* time + older + amp.above +   
 + (1|id),  
 # control = fit\_control,  
 data = data\_lla\_tf)  
  
## to get the coefficients  
beta\_2\_1 <- (summary(fit\_2\_1)$coefficients[, 1]) %>% round(digits = 3)  
# beta\_2\_1  
## to get the variance-covariance  
vcov\_2\_1 <- vcov(fit\_2\_1)   
# vcov\_2\_1  
  
## lmm contrast---------------------------------------------------------------------------  
## [0] [1] [2] [3] [4] [5] [6] [7]  
## coefs (Int) treatEXP time6 time18 older1 amp.abv1 trtEXP:time6 trtEXP:time18  
con\_trt\_t0 <- c(1, 1, 0, 0, 1, 1, 0, 0)  
con\_trt\_t6 <- c(1, 1, 1, 0, 1, 1, 1 \* 1, 0)  
con\_trt\_t18 <- c(1, 1, 0, 1, 1, 1, 0, 1 \* 1)  
con\_ctl\_t0 <- c(1, 0, 0, 0, 1, 1, 0, 0)  
con\_ctl\_t6 <- c(1, 0, 1, 0, 1, 1, 0, 0)  
con\_ctl\_t18 <- c(1, 0, 0, 1, 1, 1, 0, 0)  
con\_trt\_t6\_t0 <- con\_trt\_t6 - con\_trt\_t0  
con\_ctl\_t6\_t0 <- con\_ctl\_t6 - con\_ctl\_t0  
con\_trt\_t6\_t0 - con\_ctl\_t6\_t0

## [1] 0 0 0 0 0 0 1 0

contrast0 <- cbind(con\_trt\_t6\_t0, con\_ctl\_t6\_t0)  
mu\_lm <- t(contrast0) %\*% beta\_2\_1   
## standard error for each group  
sd\_lm <- t(contrast0) %\*% vcov\_2\_1 %\*% contrast0 %>%  
 diag() %>% sqrt()  
  
cbind(mu\_lm, sd\_lm) %>%  
 as.data.frame() %>%  
 dplyr::select("expected mean" = 1,  
 "standard\_error" = 2) %>%  
 round(3)

## expected mean standard\_error  
## con\_trt\_t6\_t0 1140.121 1.062  
## con\_ctl\_t6\_t0 -58.962 1.062

t <- (1140.121 - (-58.962))/ 1.062; t

## [1] 1129.08

con\_trt\_t6\_t0 <- con\_trt\_t6 - con\_trt\_t0  
con\_ctl\_t6\_t0 <- con\_ctl\_t6 - con\_ctl\_t0  
  
con\_tvc\_t6\_t0 <- con\_trt\_t6\_t0 - con\_ctl\_t6\_t0  
mu\_2\_3 <- t(con\_tvc\_t6\_t0 ) %\*% beta\_2\_1 %>%  
 as.data.frame() %>%  
 dplyr::select("expected\_mean" = 1)  
## contrast variance covariance matrix  
sd\_2\_3 <- t(con\_tvc\_t6\_t0 ) %\*% vcov\_2\_1 %\*% con\_tvc\_t6\_t0 %>%  
 diag() %>% sqrt()  
  
cbind(mu\_2\_3, "standard\_error" = sd\_2\_3) %>%  
 round(3)

## expected\_mean standard\_error  
## 1 1199.083 1.501

qnorm(0.05, mean = 1199.083, sd = 1.501)

## [1] 1196.614

pnorm(1000, mean = 1199.083, sd = 1.501)

## [1] 0

The 95% confidence interval for *the change of step counts over 1000 at 6 month* is

#### using generalized linear mixed model

fit\_2\_2 <- glmer(stepc ~ 1 + treat \* time + older + amp.above +   
 + (1|id),  
 family = "poisson",  
 # control = fit\_control,  
 data = data\_lla\_tf)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unidentifiable: very large eigenvalue  
## - Rescale variables?

## to get the coefficients  
beta\_2\_2 <- (summary(fit\_2\_2)$coefficients[, 1]) %>% round(digits = 3)  
# beta\_2\_2  
## to get the variance-covariance  
vcov\_2\_2 <- vcov(fit\_2\_2)   
# vcov\_2\_2

## old amputation above ---------------------------------------------------------------------------  
## [0] [1] [2] [3] [4] [5] [6] [7]  
## coefs (Int) treatEXP time6 time18 older1 amp.abv1 trtEXP:time6 trtEXP:time18  
con\_old\_abv\_trt\_t0 <- c(1, 1, 0, 0, 1, 1, 0, 0)  
con\_old\_abv\_trt\_t6 <- c(1, 1, 1, 0, 1, 1, 1 \* 1, 0)  
con\_old\_abv\_trt\_t18 <- c(1, 1, 0, 1, 1, 1, 0, 1 \* 1)  
(con\_old\_abv\_trt\_t6\_t0 <- con\_old\_abv\_trt\_t6 - con\_old\_abv\_trt\_t0)

## [1] 0 0 1 0 0 0 1 0

(con\_old\_abv\_trt\_t18\_t0 <- con\_old\_abv\_trt\_t18 - con\_old\_abv\_trt\_t0)

## [1] 0 0 0 1 0 0 0 1

con\_old\_abv\_ctl\_t0 <- c(1, 0, 0, 0, 1, 1, 0, 0)  
con\_old\_abv\_ctl\_t6 <- c(1, 0, 1, 0, 1, 1, 0, 0)  
con\_old\_abv\_ctl\_t18 <- c(1, 0, 0, 1, 1, 1, 0, 0)  
(con\_old\_abv\_ctl\_t6\_t0 <- con\_old\_abv\_ctl\_t6 - con\_old\_abv\_ctl\_t0)

## [1] 0 0 1 0 0 0 0 0

(con\_old\_abv\_ctl\_t18\_t0 <- con\_old\_abv\_ctl\_t18 - con\_old\_abv\_ctl\_t0)

## [1] 0 0 0 1 0 0 0 0

## young amputation above ---------------------------------------------------------------------------  
## [0] [1] [2] [3] [4] [5] [6] [7]  
## coefs (Int) treatEXP time6 time18 older1 amp.abv1 trtEXP:time6 trtEXP:time18  
con\_yng\_abv\_trt\_t0 <- c(1, 1, 0, 0, 1, 0, 0, 0)  
con\_yng\_abv\_trt\_t6 <- c(1, 1, 1, 0, 1, 0, 1 \* 1, 0)  
con\_yng\_abv\_trt\_t18 <- c(1, 1, 0, 1, 1, 0, 0, 1 \* 1)  
con\_yng\_abv\_trt\_t6\_t0 <- con\_yng\_abv\_trt\_t6 - con\_yng\_abv\_trt\_t0  
con\_yng\_abv\_trt\_t18\_t0 <- con\_yng\_abv\_trt\_t18 - con\_yng\_abv\_trt\_t0  
con\_yng\_abv\_ctl\_t0 <- c(1, 0, 0, 0, 1, 0, 0, 0)  
con\_yng\_abv\_ctl\_t6 <- c(1, 0, 1, 0, 1, 0, 0, 0)  
con\_yng\_abv\_ctl\_t18 <- c(1, 0, 0, 1, 1, 0, 0, 0)  
con\_yng\_abv\_ctl\_t6\_t0 <- con\_yng\_abv\_ctl\_t6 - con\_yng\_abv\_ctl\_t0  
con\_yng\_abv\_ctl\_t18\_t0 <- con\_yng\_abv\_ctl\_t18 - con\_yng\_abv\_ctl\_t0  
## old amputation below ---------------------------------------------------------------------------  
## [0] [1] [2] [3] [4] [5] [6] [7]  
## coefs (Int) treatEXP time6 time18 older1 amp.abv1 trtEXP:time6 trtEXP:time18  
con\_old\_blw\_trt\_t0 <- c(1, 1, 0, 0, 0, 1, 0, 0)  
con\_old\_blw\_trt\_t6 <- c(1, 1, 1, 0, 0, 1, 1 \* 1, 0)  
con\_old\_blw\_trt\_t18 <- c(1, 1, 0, 1, 0, 1, 0, 1 \* 1)  
con\_old\_blw\_trt\_t6\_t0 <- con\_old\_blw\_trt\_t6 - con\_old\_blw\_trt\_t0  
con\_old\_blw\_trt\_t18\_t0 <- con\_old\_blw\_trt\_t18 - con\_old\_blw\_trt\_t0  
con\_old\_blw\_ctl\_t0 <- c(1, 0, 0, 0, 0, 1, 0, 0)  
con\_old\_blw\_ctl\_t6 <- c(1, 0, 1, 0, 0, 1, 0, 0)  
con\_old\_blw\_ctl\_t18 <- c(1, 0, 0, 1, 0, 1, 0, 0)  
con\_old\_blw\_ctl\_t6\_t0 <- con\_old\_blw\_ctl\_t6 - con\_old\_blw\_ctl\_t0  
con\_old\_blw\_ctl\_t18\_t0 <- con\_old\_blw\_ctl\_t18 - con\_old\_blw\_ctl\_t0  
## young amputation below ---------------------------------------------------------------------------  
## [0] [1] [2] [3] [4] [5] [6] [7]  
## coefs (Int) treatEXP time6 time18 ynger1 amp.abv1 trtEXP:time6 trtEXP:time18  
con\_yng\_blw\_trt\_t0 <- c(1, 1, 0, 0, 0, 0, 0, 0)  
con\_yng\_blw\_trt\_t6 <- c(1, 1, 1, 0, 0, 0, 1 \* 1, 0)  
con\_yng\_blw\_trt\_t18 <- c(1, 1, 0, 1, 0, 0, 0, 1 \* 1)  
con\_yng\_blw\_trt\_t18\_t0 <- con\_yng\_blw\_trt\_t18 - con\_yng\_blw\_trt\_t0  
con\_yng\_blw\_trt\_t6\_t0 <- con\_yng\_blw\_trt\_t6 - con\_yng\_blw\_trt\_t0  
con\_yng\_blw\_ctl\_t0 <- c(1, 0, 0, 0, 0, 0, 0, 0)  
con\_yng\_blw\_ctl\_t6 <- c(1, 0, 1, 0, 0, 0, 0, 0)  
con\_yng\_blw\_ctl\_t18 <- c(1, 0, 0, 1, 0, 0, 0, 0)  
con\_yng\_blw\_ctl\_t18\_t0 <- con\_yng\_blw\_ctl\_t18 - con\_yng\_blw\_ctl\_t0  
con\_yng\_blw\_ctl\_t6\_t0 <- con\_yng\_blw\_ctl\_t6 - con\_yng\_blw\_ctl\_t0  
  
# contast\_old\_abv\_t06 <- cbind(con\_old\_abv\_trt\_t6\_t0, con\_old\_abv\_ctl\_t6\_t0)  
# contast\_yng\_abv\_t06 <- cbind(con\_yng\_abv\_trt\_t6\_t0, con\_yng\_abv\_ctl\_t6\_t0)  
# contast\_old\_blw\_t06 <- cbind(con\_old\_blw\_trt\_t6\_t0, con\_old\_blw\_ctl\_t6\_t0)  
# contast\_yng\_blw\_t06 <- cbind(con\_yng\_blw\_trt\_t6\_t0, con\_yng\_blw\_ctl\_t6\_t0)  
  
contrast1 <- cbind(con\_old\_abv\_trt\_t6\_t0,  
 con\_yng\_abv\_trt\_t6\_t0,  
 con\_old\_blw\_trt\_t6\_t0,  
 con\_yng\_blw\_trt\_t6\_t0,  
 con\_old\_abv\_ctl\_t6\_t0,  
 con\_yng\_abv\_ctl\_t6\_t0,  
 con\_old\_blw\_ctl\_t6\_t0,  
 con\_yng\_blw\_ctl\_t6\_t0)  
rownames(contrast1) <- rownames(vcov\_2\_2)

con\_ctl\_t0 <- cbind(con\_old\_abv\_trt\_t0,  
 con\_yng\_abv\_trt\_t0,  
 con\_old\_blw\_trt\_t0,  
 con\_yng\_blw\_trt\_t0,  
 con\_old\_abv\_ctl\_t0,  
 con\_yng\_abv\_ctl\_t0,  
 con\_old\_blw\_ctl\_t0,  
 con\_yng\_blw\_ctl\_t0)  
  
con\_ctl\_t6 <- cbind(con\_old\_abv\_trt\_t6,  
 con\_yng\_abv\_trt\_t6,  
 con\_old\_blw\_trt\_t6,  
 con\_yng\_blw\_trt\_t6,  
 con\_old\_abv\_trt\_t6,  
 con\_yng\_abv\_trt\_t6,  
 con\_old\_blw\_trt\_t6,  
 con\_yng\_blw\_trt\_t6)  
  
mu\_ctl\_t0 <- t(con\_ctl\_t0) %\*% beta\_2\_2 %>%  
 exp() %>%  
 as.data.frame() %>%  
 dplyr::select("t0" = 1)  
sd\_ctl\_t0 <- t(con\_ctl\_t0) %\*% vcov\_2\_2 %\*% con\_ctl\_t0 %>%  
 diag() %>% sqrt()  
mu\_ctl\_t6 <- t(con\_ctl\_t6) %\*% beta\_2\_2 %>%  
 exp() %>%  
 as.data.frame() %>%  
 dplyr::select("t6" = 1)  
sd\_ctl\_t6 <- t(con\_ctl\_t6) %\*% vcov\_2\_2 %\*% con\_ctl\_t6 %>%  
 diag() %>% sqrt()  
  
  
mu\_2\_2 <- t(contrast1) %\*% beta\_2\_2 %>%  
 as.data.frame() %>%  
 dplyr::select("log\_rate" = 1) %>%  
 mutate(rate = exp(log\_rate))   
## contrast variance covariance matrix  
sd\_2\_2 <- t(contrast1) %\*% vcov\_2\_2 %\*% contrast1 %>%  
 diag() %>% sqrt()  
  
result2 <- cbind(mu\_2\_2,  
 sd\_log\_rate = sd\_2\_2,  
 mu\_ctl\_t0,  
 mu\_ctl\_t6) %>%   
 mutate(diff\_t6\_t0 = (rate - 1) \* t0,  
 sd\_t6\_t0 = abs(rate \* sd\_2\_2 \* t0)) %>%  
 round(3)   
result2

## log\_rate rate sd\_log\_rate t0 t6 diff\_t6\_t0  
## con\_old\_abv\_trt\_t6\_t0 0.349 1.418 0.002 2657.125 3766.871 1109.746  
## con\_yng\_abv\_trt\_t6\_t0 0.349 1.418 0.002 2670.444 3785.753 1115.309  
## con\_old\_blw\_trt\_t6\_t0 0.349 1.418 0.002 2581.175 3659.201 1078.026  
## con\_yng\_blw\_trt\_t6\_t0 0.349 1.418 0.002 2594.113 3677.542 1083.429  
## con\_old\_abv\_ctl\_t6\_t0 -0.023 0.977 0.002 2232.774 3766.871 -50.768  
## con\_yng\_abv\_ctl\_t6\_t0 -0.023 0.977 0.002 2243.966 3785.753 -51.022  
## con\_old\_blw\_ctl\_t6\_t0 -0.023 0.977 0.002 2168.953 3659.201 -49.317  
## con\_yng\_blw\_ctl\_t6\_t0 -0.023 0.977 0.002 2179.825 3677.542 -49.564  
## sd\_t6\_t0  
## con\_old\_abv\_trt\_t6\_t0 7.528  
## con\_yng\_abv\_trt\_t6\_t0 7.566  
## con\_old\_blw\_trt\_t6\_t0 7.313  
## con\_yng\_blw\_trt\_t6\_t0 7.350  
## con\_old\_abv\_ctl\_t6\_t0 4.921  
## con\_yng\_abv\_ctl\_t6\_t0 4.945  
## con\_old\_blw\_ctl\_t6\_t0 4.780  
## con\_yng\_blw\_ctl\_t6\_t0 4.804

According to the model, the effect size for older age and amputation location is very small, hence the expected mean values and the standard errors of *the change of step counts over 1000 at 6 month from baseline* are very close to each others among different age group and amputation location groups.

As we can see the difference on linear mixed model and generalized linear mixed model, the difference between t0 to t6 in linear mixed model is larger than generalized linear mixed model. There might be a over-dispersion for generalized linear mixed model.

### iv.

Calculate the difference (EXP - CLT) in change of expected means in the two intervention groups. Report a 1-sided 95% confidence interval for this difference.

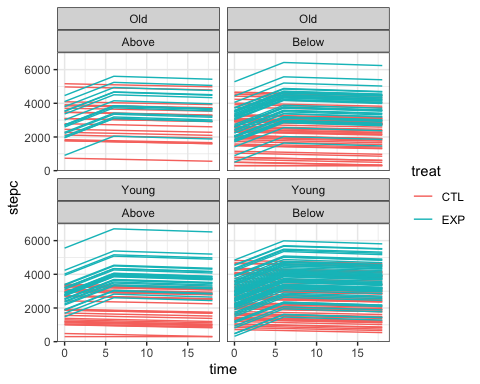
result3a <- result2[1:4, ]  
result3b <- result2[5:8, ]  
  
result4 <- data.frame(diff\_trt\_ctl = result3a$diff\_t6\_t0 - result3b$diff\_t6\_t0,  
 sd\_trt\_ctl = sqrt(result3a$sd\_t6\_t0^2 + result3b$sd\_t6\_t0^2)) %>%  
 mutate(q05 = qnorm(0.05, mean = diff\_trt\_ctl, sd = sd\_trt\_ctl)) %>%  
 round(3)  
  
result4

## diff\_trt\_ctl sd\_trt\_ctl q05  
## 1 1160.514 8.994 1145.721  
## 2 1166.331 9.039 1151.464  
## 3 1127.343 8.737 1112.973  
## 4 1132.993 8.781 1118.550

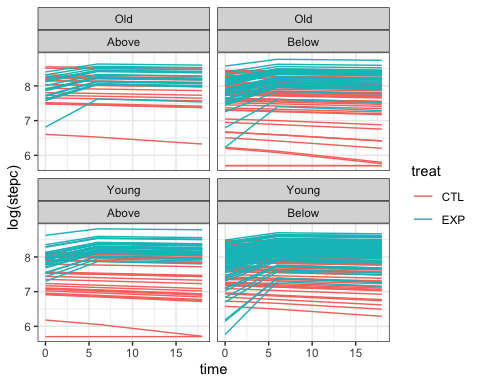
### v.

Make a graph, and include any appropriate descriptive statistics, that you would present as evidence related to the primary hypothesis of interest.

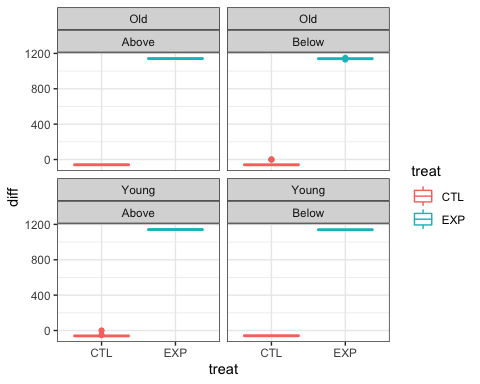
plot\_lla <- data\_lla %>%  
 mutate(older = case\_when(older == 0 ~ "Young",  
 older == 1 ~ "Old"),  
 amp.above = case\_when(amp.above == 1 ~ "Above",  
 amp.above == 0 ~ "Below")) %>%  
 ggplot(aes(x = time,   
 y = stepc,  
 group = id,  
 color = treat)) +  
 geom\_line() +  
 facet\_wrap(older ~ amp.above)  
  
plot\_lla\_log <- data\_lla %>%  
 mutate(older = case\_when(older == 0 ~ "Young",  
 older == 1 ~ "Old"),  
 amp.above = case\_when(amp.above == 1 ~ "Above",  
 amp.above == 0 ~ "Below")) %>%  
 ggplot(aes(x = time,   
 y = log(stepc),  
 group = id,  
 color = treat)) +  
 geom\_line() +  
 facet\_wrap(older ~ amp.above)  
  
plot\_lla + theme\_bw()



plot\_lla\_log + theme\_bw()



plot\_lla2 <- data\_lla %>%  
 mutate(older = case\_when(older == 0 ~ "Young",  
 older == 1 ~ "Old"),  
 amp.above = case\_when(amp.above == 1 ~ "Above",  
 amp.above == 0 ~ "Below")) %>%   
 pivot\_wider(names\_from = time,   
 names\_prefix = "t",   
 values\_from = stepc) %>%  
 mutate(diff = t6 - t0) %>%   
 ggplot(aes(y = diff,   
 x = treat,  
 color = treat)) +  
 geom\_boxplot() +  
 facet\_wrap(older ~ amp.above)  
  
plot\_lla2 + theme\_bw()



### vi.

Write your conclusion of the hypothesis test in (b).

Based on the linear mixed model fit\_2\_1 as well as the generalized linear mixed model fit\_2\_2, conditional on given individual adjusted for age group and operational amputation location, we can reject the null hypothesis and claim there is a very highly statistically significant difference on *the change of step counts over 1000 at 6 month* between two intervention groups (p-value < 0.05).

### vii.

Comment on your assumptions of the correlation structure for the repeated measurements within subject.

The correlation structure is assumed to be compound symmetric structure, with ICC as .

### viii.

Write a summary paragraph interpreting the results for a clinical collaborator.

For linear mixed model, on average in population level, after adjusted for age group and amputation location, we found that *the change of step counts over 1000 at 6 month* in treatment group is 1199 steps more than the control group (sd = 1.5). Hence *the change of step counts over 1000 at 6 month* from different intervention groups is larger than 1000 with both practical and statistical significance. Therefore we can reject the null hypothesis and claim that after adjustment for baseline step counts and stratification on age and amputation location, treatment as a statistically significant effect on *whether the step counts over 1000 at 6 month* over control group.

For generalized linear mixed model with Poisson regression, the conclusion is very similar to linear mixed model.

## Part 2b

A secondary hypothesis of interest was whether or not the change in step counts was sustained from 6 to 18 months, specifically, interest was in testing the difference between **M18-M0** and **M6-M0** in the intervention group.

Carry out the appropriate hypothesis to test whether there was a sustained change and interpret the results for a clinical collaborator.

For linear mixed model:

: There is no difference between *the change of step counts over 1000 at 6 month from baseline* and *the change of step counts over 1000 at 18 month from baseline*

equation form:

: There is a statistically significant difference between *the change of step counts over 1000 at 6 month from baseline* and *the change of step counts over 1000 at 18 month from baseline*

equation form:

## lmm contrast---------------------------------------------------------------------------  
## [0] [1] [2] [3] [4] [5] [6] [7]  
## coefs (Int) treatEXP time6 time18 older1 amp.abv1 trtEXP:time6 trtEXP:time18  
con\_trt\_t0 <- c(1, 1, 0, 0, 1, 1, 0, 0)  
con\_trt\_t6 <- c(1, 1, 1, 0, 1, 1, 1 \* 1, 0)  
con\_trt\_t18 <- c(1, 1, 0, 1, 1, 1, 0, 1 \* 1)  
  
con\_trt\_t6\_t0 <- con\_trt\_t6 - con\_trt\_t0  
con\_trt\_t18\_t0 <- con\_trt\_t18 - con\_trt\_t0  
contrast4 <- con\_trt\_t18\_t0 - con\_trt\_t6\_t0  
contrast4

## [1] 0 0 -1 1 0 0 -1 1

mu\_b <- t(contrast4) %\*% beta\_2\_1   
mu\_b

## [,1]  
## [1,] -180.428

## contrast variance covariance matrix  
cov\_b <- t(contrast4) %\*% vcov\_2\_1 %\*% contrast4 %>% as.matrix()  
  
  
## with both point estimates and standard deviation   
## an anova or pairwise comparison can be performed  
W0 <- mu\_b^2 \* solve(cov\_b)  
pchisq(W0, df = 1, lower.tail = FALSE)

## [,1]  
## [1,] 0

For generalized linear mixed model:

: There is no difference between *the change of step counts over 1000 at 6 month from baseline* and *the change of step counts over 1000 at 18 month from baseline*

equation form:

: There is a statistically significant difference between *the change of step counts over 1000 at 6 month from baseline* and *the change of step counts over 1000 at 18 month from baseline*

equation form:

contrast5 <- con\_old\_abv\_trt\_t18\_t0 - con\_old\_abv\_trt\_t6\_t0  
contrast5

## [1] 0 0 -1 1 0 0 -1 1

mu\_b2 <- t(contrast5) %\*% beta\_2\_2   
exp(mu\_b2)

## [,1]  
## [1,] 0.9521811

## contrast variance covariance matrix  
cov\_b2 <- t(contrast5) %\*% vcov\_2\_2 %\*% contrast5 %>% as.numeric()  
  
W1 <- mu\_b2^2 / cov\_b2  
pchisq(W1, df = 1, lower.tail = FALSE)

## [,1]  
## [1,] 3.323379e-156

As we can see in both linear mixed model and generalized linear mixed model, with Wald tests, the Wald statistics follow distribution. There is a change of -180.428 steps change between *the change of step counts over 1000 at 6 month from baseline* and *the change of step counts over 1000 at 18 month from baseline* in linear mixed model with (p-value < 0.001); *the change of step counts over 1000 at 18 month from baseline* is 4.781887% less than the *the change of step counts over 1000 at 6 month from baseline* in Poisson generalized linear mixed model with (p-value < 0.001). In both cases, the difference is statistically significant, hence we can reject the null hypothesis and claim that *the change in step counts* was not sustained from 6 to 18 months.