

LDA-HW2-al3998

AimingLiu UNI:al3998

Question 1

1. Consider a marginal model for the log odds of moderate or severe onycholysis. Using GEE, set up a suitable model assuming linear trends. Use month as the time variable. Assume “exchangeable” correlation for the association among the repeated binary responses.

Model setup:

$E[Y_{ij}] = \mu_{ij}$ Link function: $\log \mu_{ij} = \eta_{ij} = \sum_{k=1}^p X_{ij} \beta_k$ Under poisson assumption: $Var(Y_{ij}) = \phi V(\mu_{ij}) = \phi \mu_{ij}$

The model I built is

$$\eta_{ij} = \beta_0 + \beta_1 \text{Treatment}_i + \beta_2 \text{Month}_{ij} + \beta_3 \text{Month}_{ij}^2 + \beta_4 (\text{Treatment}_i * \text{Month}_{ij}) + \beta_5 (\text{Treatment}_i * \text{Month}_{ij}^2)$$

Call:

```
geeglm(formula = Response ~ Treatment * (Month + I(Month^2)),
        family = binomial(link = "logit"), data = toenail1, id = Subject_ID,
        corstr = "exchangeable")
```

Coefficients:

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	-0.378812	0.176363	4.614	0.03172 *
Treatment1	-0.053047	0.251016	0.045	0.83263
Month	-0.308201	0.053739	32.892	9.74e-09 ***
I(Month^2)	0.012364	0.004076	9.202	0.00242 **
Treatment1:Month	-0.029879	0.081520	0.134	0.71398
Treatment1:I(Month^2)	-0.003161	0.006998	0.204	0.65145

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = exchangeable

Estimated Scale Parameters:

	Estimate	Std.err
(Intercept)	0.9988	0.2733

Link = identity

Estimated Correlation Parameters:

	Estimate	Std.err
alpha	0.4391	0.1405

Number of clusters: 294 Maximum cluster size: 7

Test if the treatment interaction term is required The null hypothesis is: $\beta_4 = \beta_5 = 0$

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	0	0	0	0	1	0
[2,]	0	0	0	0	0	1

X2.stat	DF	Pr(> X^2)
1.885	2	0.3896

As we can see from the result, p-value is greater than 0.05 so that we fail to reject the null and we can remove the interaction terms.

I fit the model again as

$$\eta_{ij} = \beta_0 + \beta_1 \text{Treatment}_i + \beta_2 \text{Month}_{ij} + \beta_3 \text{Month}_{ij}^2$$

Call:

```
geeglm(formula = Response ~ Treatment + (Month + I(Month^2)),
  family = binomial(link = "logit"), data = toenail1, id = Subject_ID,
  corstr = "exchangeable")
```

Coefficients:

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	-0.39889	0.17545	5.17	0.02300 *
Treatment1	-0.00653	0.25168	0.00	0.97929
Month	-0.32603	0.04039	65.17	6.7e-16 ***
I(Month^2)	0.01151	0.00326	12.43	0.00042 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = exchangeable

Estimated Scale Parameters:

	Estimate	Std.err
(Intercept)	0.992	0.205

Link = identity

Estimated Correlation Parameters:

	Estimate	Std.err
alpha	0.442	0.113

Number of clusters: 294 Maximum cluster size: 7

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	-0.399	0.175	5.169	0.023
Treatment1	-0.007	0.252	0.001	0.979
Month	-0.326	0.040	65.168	0.000
I(Month^2)	0.012	0.003	12.435	0.000

2. Provide Interpretations for the coefficients in your model.

Interpretation:

$$\eta_{ij} = -0.399 - 0.007\text{Treatment} - 0.326\text{Month} + 0.012\text{Month}^2$$

- β_0 :Log odds of moderate or severe onycholysis for oral treatment B at baseline month is -0.399
- β_1 :Log odds ratio of moderate or severe onycholysis comparing treatment A to treatment B with month fixed is -0.007
- β_2 :Log odds ratio of moderate or severe onycholysis for one unit increase in month among treatment B is -0.326
- β_3 :This is a non-linear term and it is difficult to interpret

3. From the results of your analysis what conclusions do you draw about the effect of treatment on changes in the severity of onycholysis over time? Provide results that support your conclusions.

From the result above, we can recall the test of interaction terms for treatment and month, at 0.05 significant level, the p-value is greater than 0.05 so that we fail to reject the null and can conclude that the effect of treatment in the severity of onycholysis will not change over time.

4. Try Different correlation structures. Is the analysis and inference sensitive to this choice?

Correlation structure:unstructured

Call:

```
geeglm(formula = Response ~ Treatment * (Month + I(Month^2)),
        family = binomial(link = "logit"), data = toenail1, id = Subject_ID,
        corstr = "unstructured")
```

Coefficients:

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	-8.58e+15	1.63e+15	27.8	1.3e-07 ***
Treatment1	-1.64e+16	1.66e+15	97.9	< 2e-16 ***
Month	3.04e+15	1.17e+14	675.5	< 2e-16 ***
I(Month^2)	-2.06e+14	3.20e+13	41.5	1.1e-10 ***
Treatment1:Month	2.63e+15	1.82e+14	207.6	< 2e-16 ***
Treatment1:I(Month^2)	-1.23e+14	3.56e+13	12.0	0.00053 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = unstructured

Estimated Scale Parameters:

	Estimate	Std.err
(Intercept)	1.44e+15	3.79e+37

Link = identity

Estimated Correlation Parameters:

	Estimate	Std.err
alpha.1:2	1.0099	2.65e+22
alpha.1:3	0.8120	2.14e+22
alpha.1:4	0.5399	1.44e+22
alpha.1:5	-0.3219	8.45e+21
alpha.1:6	-0.3102	8.14e+21
alpha.1:7	0.0981	2.58e+21
alpha.2:3	0.8339	2.20e+22
alpha.2:4	0.5624	1.50e+22
alpha.2:5	-0.2874	7.54e+21
alpha.2:6	-0.2625	6.89e+21
alpha.2:7	0.0981	2.58e+21
alpha.3:4	0.6074	1.62e+22
alpha.3:5	-0.2414	6.34e+21
alpha.3:6	-0.2148	5.64e+21
alpha.3:7	0.0981	2.58e+21
alpha.4:5	-0.0460	1.21e+21
alpha.4:6	-0.0835	2.19e+21
alpha.4:7	0.0981	2.58e+21
alpha.5:6	1.1694	3.07e+22
alpha.5:7	0.0280	7.37e+20
alpha.6:7	0.0420	1.11e+21

Number of clusters: 294 Maximum cluster size: 7

Test if the treatment interaction term is required The null hypothesis is: $\beta_4=\beta_5=0$

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	0	0	0	0	1	0
[2,]	0	0	0	0	0	1

X2.stat	DF	Pr(> X^2)
283	2	0

As we can see from the result, the p-value is less than 0.05 so that when the correlation structure is unstructured, the analysis and inference are sensitive to this choice.

Correlation structure:independence

Call:

```
geeglm(formula = Response ~ Treatment * (Month + I(Month^2)),
  family = binomial(link = "logit"), data = toenail1, id = Subject_ID,
  corstr = "independence")
```

Coefficients:

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	-0.38412	0.17662	4.73	0.0296 *
Treatment1	-0.04804	0.25059	0.04	0.8480
Month	-0.31836	0.05714	31.04	2.5e-08 ***
I(Month^2)	0.01314	0.00444	8.76	0.0031 **

```
Treatment1:Month      -0.03699  0.08470  0.19  0.6623
Treatment1:I(Month^2) -0.00174  0.00688  0.06  0.8006
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Correlation structure = independence
Estimated Scale Parameters:

```
              Estimate Std.err
(Intercept)    0.997    0.226
Number of clusters: 294 Maximum cluster size: 7
```

Test if the treatment interaction term is required The null hypothesis is: $\beta_4=\beta_5=0$

```
      [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    0    0    0    0    1    0
[2,]    0    0    0    0    0    1
```

X2.stat	DF	Pr(> X^2)
1.53	2	0.465

As we can see from the result, the p-value is greater than 0.05 so that when the correlation structure is independence, the analysis and inference are not sensitive to this choice.

Correlation structure:AR1

```
Call:
geeglm(formula = Response ~ Treatment * (Month + I(Month^2)),
       family = binomial(link = "logit"), data = toenail1, id = Subject_ID,
       corstr = "ar1")
```

```
Coefficients:
              Estimate Std.err Wald Pr(>|W|)
(Intercept)   -0.47915  0.17163  7.79  0.0052 **
Treatment1      0.00670  0.25100  0.00  0.9787
Month         -0.30016  0.05411 30.77 2.9e-08 ***
I(Month^2)     0.01340  0.00412 10.58  0.0011 **
Treatment1:Month -0.04251  0.08196  0.27  0.6040
Treatment1:I(Month^2) -0.00232  0.00656  0.13  0.7235
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Correlation structure = ar1
Estimated Scale Parameters:

```
              Estimate Std.err
(Intercept)    0.984    0.189
Link = identity
```

Estimated Correlation Parameters:

```

      Estimate Std.err
alpha    0.698  0.0833
Number of clusters:   294  Maximum cluster size: 7

```

Test if the treatment interaction term is required The null hypothesis is: $\beta_4=\beta_5=0$

```

      [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    0    0    0    0    1    0
[2,]    0    0    0    0    0    1

```

X2.stat	DF	Pr(> X ²)
2.55	2	0.28

As we can see from the result, the p-value is greater than 0.05 so that when the correlation structure is independence, the analysis and inference are not sensitive to this choice.

Question 2

1. Set up a suitable GEE model for rate of skin cancers with Treatment and Year as covariates.

The model will be fitted as:

model set up

$Y_{ij} \in [0,1]; E[Y_{ij}] = \mu_{ij}$ Mean response model: $\log(\mu_{ij}) = \eta_{ij}$ Under binomial assumption: $V(\mu_{ij}) = \mu_{ij}(1 - \mu_{ij})$

$$\eta_{ij} = \beta_0 + \beta_1 \text{Treatment} + \beta_2 \text{Year} + \beta_3 \text{Treatment} * \text{Year}$$

Call:

```

geeglm(formula = Y ~ Treatment * Year, family = poisson(link = "log"),
       data = skin, id = ID, corstr = "unstructured")

```

Coefficients:

```

      Estimate Std.err Wald Pr(>|W|)
(Intercept) -1.35953  0.12009 128.16 <2e-16 ***
Treatment1    0.04891  0.16353   0.09   0.76
Year         -0.00379  0.03232   0.01   0.91
Treatment1:Year 0.03271  0.04917   0.44   0.51
---

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = unstructured

Estimated Scale Parameters:

```

      Estimate Std.err
(Intercept)    2.67   0.394

```

Link = identity

Estimated Correlation Parameters:

	Estimate	Std.err
alpha.1:2	0.294	0.0862
alpha.1:3	0.328	0.1113
alpha.1:4	0.367	0.1284
alpha.1:5	0.419	0.2276
alpha.2:3	0.243	0.0565
alpha.2:4	0.232	0.0635
alpha.2:5	0.241	0.1113
alpha.3:4	0.741	0.4089
alpha.3:5	0.517	0.2039
alpha.4:5	0.507	0.2262

Number of clusters: 1683 Maximum cluster size: 5

Test if the interaction term is required: The null hypothesis is: $\beta_3=0$

	X2.stat	DF	Pr(> X^2)
1	0.443	1	0.506

We can see that the p-value is $0.506 > 0.05$, so that we fail to reject the null and we can remove the interaction term from the model.

Then I fit the model again as

$$\eta_{ij} = \beta_0 + \beta_1 Treatment + \beta_2 Year$$

Call:

```
geeglm(formula = Y ~ Treatment + Year, family = poisson(link = "log"),
       data = skin, id = ID, corstr = "unstructured")
```

Coefficients:

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	-1.4020	0.1069	172.03	<2e-16 ***
Treatment1	0.1297	0.1052	1.52	0.22
Year	0.0134	0.0250	0.29	0.59

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = unstructured

Estimated Scale Parameters:

	Estimate	Std.err
(Intercept)	2.68	0.401

Link = identity

Estimated Correlation Parameters:

	Estimate	Std.err
alpha.1:2	0.295	0.0871
alpha.1:3	0.329	0.1120
alpha.1:4	0.365	0.1282
alpha.1:5	0.415	0.2248

```

alpha.2:3    0.242  0.0565
alpha.2:4    0.231  0.0630
alpha.2:5    0.237  0.1091
alpha.3:4    0.743  0.4147
alpha.3:5    0.513  0.2052
alpha.4:5    0.504  0.2297
Number of clusters: 1683 Maximum cluster size: 5

```

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	-1.402	0.107	172.031	0.000
Treatment1	0.130	0.105	1.521	0.218
Year	0.013	0.025	0.286	0.593

2. Provide Interpretations for the coefficients in your model.

- β_0 : the log rate of having non-melanoma skin cancers in placebo group at baseline is -1.402.
- β_1 : the log rate ratio of having non-melanoma skin cancer between carotene group and placebo group with year fixed is 0.130.
- β_2 : the log rate ratio of having non-melanoma skin cancer with one unit increase in years of follow-up for placebo group will be 0.013.

3. From the results of your analysis what conclusions do you draw about the effect of beta carotene on the rate of skin cancers? Provide results that support your conclusions.

At 0.05 significant level, the treatment is not significant when adjusted year, we conclude that effect of beta carotene on the rate of skin cancers is not sufficient.

4. Repeat the above analysis adjusting for skin type, age, and the count of the number of previous skin cancers. What conclusions do you draw about the effect of beta carotene on the adjusted rate of skin cancers?

```

Call:
geeglm(formula = Y ~ Treatment + Year + Age + Skin + Exposure,
       family = poisson(link = "log"), data = skin, id = ID, corstr = "unstructured")

```

Coefficients:

```

      Estimate Std.err Wald Pr(>|W|)
(Intercept) -3.06545  0.32970 86.45  <2e-16 ***
Treatment1   0.11595  0.09772  1.41  0.2354
Year         0.01637  0.02469  0.44  0.5072
Age          0.01527  0.00513  8.88  0.0029 **
Skin1        0.18398  0.10808  2.90  0.0887 .
Exposure     0.13806  0.01016 184.49 <2e-16 ***
---

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = unstructured

Estimated Scale Parameters:

```

              Estimate Std.err
(Intercept)    1.64  0.0776
Link = identity

```

Estimated Correlation Parameters:

```

              Estimate Std.err
alpha.1:2    0.164  0.0353
alpha.1:3    0.178  0.0365
alpha.1:4    0.199  0.0572
alpha.1:5    0.186  0.0513
alpha.2:3    0.197  0.0479
alpha.2:4    0.181  0.0436
alpha.2:5    0.150  0.0457
alpha.3:4    0.317  0.0884
alpha.3:5    0.312  0.0773
alpha.4:5    0.245  0.0686
Number of clusters: 1683 Maximum cluster size: 5

```

I fit the model as:

$$\eta_{ij} = \beta_0 + \beta_1 \textit{Treatment} + \beta_2 \textit{Year} + \beta_3 \textit{Age} + \beta_4 \textit{Skin}_1 + \beta_5 \textit{Exposure}$$

null hypothesis: $\beta_1=0$

estimate	std.error	statistic	p.value	beta0	df	lwr	upr
0.116	0.098	1.41	0.235	0	1	-0.076	0.307

The p-value equals to 0.235 which is greater than 0.05,so that we fail to reject the null and conclude that there is no difference in log rate ratios between carotene gorup and placebo group when adjusting for skin type, age, and the count of the number of previous skin cancers.

5.Try Different correlation structures.Is the analysis and inference sensitive to this choice?

For model in part 1) Correlation structure:AR1

Call:

```

geeglm(formula = Y ~ Treatment + Year, family = poisson(link = "log"),
       data = skin, id = ID, corstr = "ar1")

```

Coefficients:

```

              Estimate Std.err Wald Pr(>|W|)
(Intercept) -1.37639  0.11040 155.43  <2e-16 ***
Treatment1   0.15365  0.11098   1.92    0.17
Year         0.00605  0.02453   0.06    0.81

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = ar1

Estimated Scale Parameters:

```
      Estimate Std.err
(Intercept)   2.63   0.384
Link = identity
```

Estimated Correlation Parameters:

```
      Estimate Std.err
alpha    0.545   0.111
```

Number of clusters: 1683 Maximum cluster size: 5

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	-1.376	0.110	155.430	0.000
Treatment1	0.154	0.111	1.917	0.166
Year	0.006	0.025	0.061	0.805

Correlation structure:exchangeable

Call:

```
geeglm(formula = Y ~ Treatment + Year, family = poisson(link = "log"),
       data = skin, id = ID, corstr = "exchangeable")
```

Coefficients:

```
      Estimate Std.err   Wald Pr(>|W|)
(Intercept) -1.4123  0.1080 171.10  <2e-16 ***
Treatment1   0.1478  0.1094   1.83    0.18
Year         0.0173  0.0247   0.49    0.48
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = exchangeable ar1 unstructured userdefined fixed

Estimated Scale Parameters:

```
      Estimate Std.err
(Intercept)   2.65   0.374
Link = identity
```

Estimated Correlation Parameters:

```
      Estimate Std.err
alpha    0.378   0.111
```

Number of clusters: 1683 Maximum cluster size: 5

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	-1.412	0.108	171.095	0.000
Treatment1	0.148	0.109	1.825	0.177
Year	0.017	0.025	0.492	0.483

I change the correlation structures of the model in part 1) and we can see that the coefficients are similar with the former model. So that the analysis and inference is not sensitive to this choice.

For model in part 4) Correlation structure:AR1

Call:

```
geeglm(formula = Y ~ Treatment + Year + Age + Skin + Exposure,
       family = poisson(link = "log"), data = skin, id = ID, corstr = "ar1")
```

Coefficients:

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	-3.02093	0.32857	84.53	<2e-16 ***
Treatment1	0.12808	0.10083	1.61	0.2040
Year	0.01056	0.02508	0.18	0.6737
Age	0.01494	0.00511	8.53	0.0035 **
Skin1	0.15284	0.11232	1.85	0.1736
Exposure	0.13915	0.01065	170.79	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = ar1

Estimated Scale Parameters:

	Estimate	Std.err
(Intercept)	1.64	0.0788

Link = identity

Estimated Correlation Parameters:

	Estimate	Std.err
alpha	0.294	0.0328

Number of clusters: 1683 Maximum cluster size: 5

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	-3.021	0.329	84.532	0.000
Treatment1	0.128	0.101	1.613	0.204
Year	0.011	0.025	0.177	0.674
Age	0.015	0.005	8.534	0.003
Skin1	0.153	0.112	1.852	0.174
Exposure	0.139	0.011	170.786	0.000

estimate	std.error	statistic	p.value	beta0	df	lwr	upr
0.128	0.101	1.61	0.204	0	1	-0.07	0.326

Correlation structure:exchangeable

Call:

```
geeglm(formula = Y ~ Treatment + Year + Age + Skin + Exposure,
       family = poisson(link = "log"), data = skin, id = ID, corstr = "exchangeable")
```

```

Coefficients:
              Estimate Std.err   Wald Pr(>|W|)
(Intercept) -3.04458   0.33263  83.78  <2e-16 ***
Treatment1    0.12357   0.09941   1.55   0.2139
Year          0.01759   0.02521   0.49   0.4854
Age           0.01496   0.00525   8.12   0.0044 **
Skin1         0.16191   0.11079   2.14   0.1439
Exposure      0.13899   0.01055 173.42  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = exchangeable ar1 unstructured userdefined fixed
Estimated Scale Parameters:

              Estimate Std.err
(Intercept)    1.64   0.0769
Link = identity

Estimated Correlation Parameters:
              Estimate Std.err
alpha         0.209   0.0262
Number of clusters: 1683 Maximum cluster size: 5

```

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	-3.045	0.333	83.777	0.000
Treatment1	0.124	0.099	1.545	0.214
Year	0.018	0.025	0.487	0.485
Age	0.015	0.005	8.122	0.004
Skin1	0.162	0.111	2.136	0.144
Exposure	0.139	0.011	173.419	0.000

estimate	std.error	statistic	p.value	beta0	df	lwr	upr
0.124	0.099	1.54	0.214	0	1	-0.071	0.318

I change the correlation structures of the model in part 4) and we can see that the p-value is still larger than 0.05 and the coefficients are similar with the former model. So that the analysis and inference is not sensitive to this choice.

5. Do you need to account for overdispersion. Comment

The scale parameter for the models are : for model in part 1) the scale parameter is `summary(gee_q2)$dispersion[1]` and the for the model in part 4) the scale parameter is `summary(gee_q2_4)$dispersion[1]`. Both of them are greater than 1, so that we need to account for overdispersion.

Appendix

```
options(tinytex.verbose = TRUE)
knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE, comment = "")
library(tidyverse)
library(gee) # for gee()
library(lme4)
library(data.table)
library(geepack)
library(doBy)
## Question 1
# load original data
toenail1 <- read.table("toenail.txt", header = T) %>%
as_tibble() %>%
mutate(Treatment = as.factor(Treatment))
# GEE Model 1:
# - quadratic trend with Month
# - interaction with Treatment
gee1 = geeglm(Response ~ Treatment * (Month + I(Month^2)), data = toenail1, family = "binomial"(link = "logit"))
summary(gee1)
L <- matrix(0,ncol=6,nrow=2) # ncol = number of coefficients in the model, nrow = number of tests
L[1,5] <- -1
L[2,6] <- -1
L
q1 = esticon(gee1,L=L, joint.test = TRUE)
q1 %>%
knitr::kable()
# GEE Model 2:
# - quadratic trend with Month
# - interaction with Treatment
gee2 = geeglm(Response ~ Treatment + (Month + I(Month^2)), data = toenail1, family = "binomial"(link = "logit"))
summary(gee2)
sum_1 = summary(gee2)
sum_1$coefficients %>%
  knitr::kable()
gee3 = geeglm(Response ~ Treatment*(Month + I(Month^2)), data = toenail1, family = "binomial"(link = "logit"))
summary(gee3)
L <- matrix(0,ncol=6,nrow=2) # ncol = number of coefficients in the model, nrow = number of tests
L[1,5] <- -1
L[2,6] <- -1
L
q1_un = esticon(gee3,L=L, joint.test = TRUE)
q1_un %>%
knitr::kable()
gee4 = geeglm(Response ~ Treatment*(Month + I(Month^2)), data = toenail1, family = "binomial"(link = "logit"))
summary(gee4)
L <- matrix(0,ncol=6,nrow=2) # ncol = number of coefficients in the model, nrow = number of tests
L[1,5] <- -1
L[2,6] <- -1
L
q1_in = esticon(gee4,L=L, joint.test = TRUE)
q1_in %>%
  knitr::kable()
```

```

gee5 = geeglm(Response ~ Treatment*(Month + I(Month^2)), data = toenail1, family = "binomial"(link = "logit"))
summary(gee5)
L <- matrix(0, ncol=6, nrow=2) # ncol = number of coefficients in the model, nrow = number of tests
L[1,5] <- -1
L[2,6] <- -1
L
q1_ar = esticon(gee5, L=L, joint.test = TRUE)
q1_ar %>%
knitr::kable()
## Question 2
# load data
skin <- read.table("skin.txt", header = F)
colnames(skin) = c("ID", "Center", "Age", "Skin", "Gender", "Exposure", "Y", "Treatment", "Year")
skin = skin %>%
  as_tibble() %>%
  mutate(Skin = as.factor(Skin),
         Gender = as.factor(Gender),
         Treatment = as.factor(Treatment))
gee_q2 = geeglm(Y ~ Treatment*Year, data = skin, family = "poisson"(link = "log"), id = ID, corstr = "unrestricted")
summary(gee_q2)
L_1 = matrix(0, ncol = 4, nrow = 1)
L_1[1,4] <- 1
esticon(gee_q2, L = L_1, joint.test = T)
gee_q2_final = geeglm(Y ~ Treatment+Year, data = skin, family = "poisson"(link = "log"), id = ID, corstr = "unrestricted")
summary(gee_q2_final)
# Coefficients
gee_q2_sum = summary(gee_q2_final)
gee_q2_sum$coefficients %>% knitr::kable()
gee_q2_4 = geeglm(Y ~ Treatment + Year + Age + Skin + Exposure, data = skin, family = "poisson"(link = "log"))
summary(gee_q2_4)
L3 = c(0,1,0,0,0,0)
q2_4 = esticon(gee_q2_4, L=L3)
q2_4 %>%
knitr::kable()
gee_q2_ar = geeglm(Y ~ Treatment + Year, data = skin, family = "poisson"(link = "log"), id = ID, corstr = "unrestricted")
summary(gee_q2_ar)
gee_q2_sum1 = summary(gee_q2_ar)
gee_q2_sum1$coefficients %>% knitr::kable()
gee_q2_ex = geeglm(Y ~ Treatment + Year, data = skin, family = "poisson"(link = "log"), id = ID, corstr = "unrestricted")
summary(gee_q2_ex)
gee_q2_sum2 = summary(gee_q2_ex)
gee_q2_sum2$coefficients %>% knitr::kable()
gee_q2_4_ar = geeglm(Y ~ Treatment+ Year + Age + Skin + Exposure, data = skin, family = "poisson"(link = "log"))
summary(gee_q2_4_ar)
L6 = c(0,1,0,0,0,0)
q2_6 = esticon(gee_q2_4_ar, L=L6)
gee_q2_sum3 = summary(gee_q2_4_ar)
gee_q2_sum3$coefficients %>% knitr::kable()
q2_6 %>%
  knitr::kable()
gee_q2_4_ex = geeglm(Y ~ Treatment+ Year + Age + Skin + Exposure, data = skin, family = "poisson"(link = "log"))
summary(gee_q2_4_ex)
L7 = c(0,1,0,0,0,0)

```

```
q2_7 = esticon(gee_q2_4_ex,L=L7)
gee_q2_sum4 = summary(gee_q2_4_ex)
gee_q2_sum4$coefficients %>% knitr::kable()
q2_7 %>%
  knitr::kable()
```