665hw5

Ty Darnell

Problem 1

```
interval=c(0,1,2,3)
cex=c(3,5,14,35)
wdex=c(3,2,7,12)
atex=c(118,112,105,84)
noex=atex-cex-wdex
```

```
cep=c(2,3,9,17)
wdep=c(2,1,9,17)
atep=c(118,114,110,92)
noep=atep-cep-wdep
treatv=c(rep("exper",4),rep("exist",4))
q1=tibble("treat"=treatv,"interval"=rep(interval,2),"nocaries"=c(noep,noex),"caries"=c(cep,cex),"withdra
exper=q1%>%filter(treat=="exper")%>%select(-treat)
exist=q1%>%filter(treat=="exist")%>%select(-treat)
```

part a

Construct the life table format for the data separately for each treatment group.

interval: 0-6=0, 6-12=1, 12-18=2, 18-24=3

Table 1: existing

interval	nocaries	caries	withdraw	atrisk
0	112	3	3	118
1	105	5	2	112
2	84	14	7	105
3	37	35	12	84

Table 2: experimental

interval	nocaries	caries	withdraw	atrisk
0	114	2	2	118
1	110	3	1	114
2	92	9	9	110
3	58	17	17	92

part b

Provide life table estimates for the cumulative probabilities (and corresponding standard errors) for no occurrence of dental caries by the end of each of the four periods for each treatment group. State the assumptions for these estimates, and assume withdrawals may be treated as not having dental caries at the time of withdrawal.

Assumptions:

- The first observation for each interval/stratification level includes subjects who withdrew and the second level will include subjects who have the event
- Patients who experienced dental caries are not censored
- Withdrawal is independent of condition being studied
- Multiple withdrawals occur uniformly throughout the interval

Existing					Life Table Survival Estimates				
Interval									
[Lower,	Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error
0	0.5	3	3	116.5	0.0258	0.0147	1.0000	0	0
0.5	1	0	0	112.0	0	0	0.9742	0.0258	0.0147
1	1.5	5	2	111.0	0.0450	0.0197	0.9742	0.0258	0.0147
1.5	2	0	0	105.0	0	0	0.9304	0.0696	0.0238
2	2.5	14	7	101.5	0.1379	0.0342	0.9304	0.0696	0.0238
2.5	3	0	0	84.0	0	0	0.8020	0.1980	0.0379
3	3.5	35	12	78.0	0.4487	0.0563	0.8020	0.1980	0.0379
3.5	4	0	0	37.0	0	0	0.4421	0.5579	0.0498
4	4.5	0	37	18.5	0	0	0.4421	0.5579	0.0498
4.5	.	0	0	0.0	0	0	0.4421	0.5579	0.0498

Experimental					Life Table Survival Estimates				
Interval									
[Lower,	Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error
0	0.5	2	2	117.0	0.0171	0.0120	1.0000	0	0
0.5	1	0	0	114.0	0	0	0.9829	0.0171	0.0120
1	1.5	3	1	113.5	0.0264	0.0151	0.9829	0.0171	0.0120
1.5	2	0	0	110.0	0	0	0.9569	0.0431	0.0188
2	2.5	9	9	105.5	0.0853	0.0272	0.9569	0.0431	0.0188
2.5	3	0	0	92.0	0	0	0.8753	0.1247	0.0312
3	3.5	17	17	83.5	0.2036	0.0441	0.8753	0.1247	0.0312
3.5	4	0	0	58.0	0	0	0.6971	0.3029	0.0459
4	4.5	0	58	29.0	0	0	0.6971	0.3029	0.0459
4.5	.	0	0	0.0	0	0	0.6971	0.3029	0.0459

Table 3: Existing

Interval	Survival Probability	Standard Error
0-6	0.974	0.015
6-12	0.930	0.024
12-18	0.802	0.038
18-24	0.442	0.050

Table 4: Experimental

Interval	Survival Probability	Standard Error
0-6	0.983	0.012
6-12	0.957	0.019
12-18	0.875	0.031
18-24	0.697	0.046

Problem 2

part a

Mathematically specify the structure of the model that includes main effects for treatment and time, as well as their interaction. Mathematically define all variables used in the model

```
nonex=c(112,105,84,37)
failex=c(3,5,14,35)
wdex=c(3,2,7,12)
ptex=6*(nonex+.5*failex+.5*wdex)
ptex
```

```
## [1] 690 651 567 363
```

```
nonep=c(114,110,92,58)
failep=c(2,3,9,17)
wdep=c(2,1,9,17)
ptep=6*(nonep+.5*failep+.5*wdep)
ptep
```

```
## [1] 696 672 606 450
```

	Treatment= Existing			
Interval	No Caries	Dental Caries	Withdrawal	Total person month exposure
0-6	112	3	3	690
6-12	105	5	2	651
12-18	84	14	7	567
18-24	37	35	12	363

	Treatment= Experimental			
Interval	No Caries	Dental Caries	Withdrawal	Total person month exposure
0-6	114	2	2	696
6-12	110	3	1	672
12-18	92	9	9	606
18-24	58	17	17	450

Assumptions:

- Withdrawals are uniformly distributed during time intervals in which they occur and are unrelated to treatment failures
- Within-interval probabilities of treatment failures are small. Time-to-failure events have independent exponential distributions

$logit(\lambda/n) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7$

where λ is the number of caries and n is the person-months of exposure

Analysis of Maximum Likelihood Estimates							
Parameter			DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept			1	-5.4336	0.5786	88.1954	<.0001
time	12-18		1	1.7573	0.6387	7.5692	0.0059
time	18-24		1	3.1960	0.6053	27.8787	<.0001
time	6-12		1	0.5723	0.7323	0.6107	0.4345
treatment	exper		1	-0.4156	0.9144	0.2065	0.6495
treatment*time	exper	12-18	1	-0.1028	1.0110	0.0103	0.9190
treatment*time	exper	18-24	1	-0.5843	0.9638	0.3675	0.5444
treatment*time	exper	6-12	1	-0.1302	1.1715	0.0124	0.9115

Parameter	Value	Definition
α	-5.434	Intercept: existing treatment and 0-6 month follow up
β_1	.572	6-12 month follow up
β_2	1.757	12-18 month follow up
β_3	3.196	18-24 month follow up
β_4	-.415	Experimental treatment
β_5	-.130	Interaction for time 6-12 and experimental treatment
β_6	-.103	Interaction for time 12-18 and experimental treatment
β_7	-.584	Interaction for time 18-24 and experimental treatment

part b

Fit the model you specified in (a) and assess the effect of interaction between treatment and time.

Joint Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
time	3	59.4568	<.0001
treatment	1	0.2065	0.6495
treatment*time	3	1.1497	0.7651

$H_0 : \beta_5 = \beta_6 = \beta_7$ The interaction between treatment and time is nonsignificant

Wald $\chi^2 = 1.15$ df=3 p-value= $.765 > .05$ Fail to reject H_0 and conclude that the interaction between treatment and time is not significant.

part c

Regardless of your conclusion from part (b), fit the main effects model (i.e., the model without interaction terms) for treatment and follow-up period.

i)

Interpret the model parameters.

$logit(\lambda/n) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-5.2992	0.4539	136.3282	<.0001
time	12-18	1	1.7224	0.4953	12.0935	0.0005
time	18-24	1	2.9866	0.4710	40.2138	<.0001
time	6-12	1	0.5237	0.5716	0.8395	0.3595
treatment	exper	1	-0.7927	0.2283	12.0603	0.0005

Parameter	Definition	Value	Interpretation
α	Intercept: existing treatment and 0-6 month follow up	-5.299	Log incidence density for existent treatment and 0-6 month follow up
β_1	6-12 month follow up	.524	Increment for 6-12 month follow up
β_2	12-18 month follow up	1.722	Increment for 12-18 month follow up
β_3	18-24 month follow up	2.987	Increment for 18-24 month follow up
β_4	Experimental treatment	-.793	Increment for treatment experimental

ii)

Provide a 95% confidence interval for the model parameter corresponding to the treatment variable.

```
c(-.7917-1.96*.2283,-.7917+1.96*.2283)
```

```
## [1] -1.239168 -0.344232
```

$\hat{\beta_4} = -.793$ $se = .2283$

95% CI: $(-1.239, -.344)$

part d

For each treatment group, provide model-predicted values using the model in part (c) for cumulative probabilities of no occurrence of dental caries by 6 months, 12 months, 18 months, and 24 months, respectively.

Cumulative probabilities of no occurrence of a cavity

Time,Treat	Failure rate formula	Failure within interval failure	Estimated Survival
6 mo Exist	e^{α}	$e^{-5.299} = .005$	$1 * e^{-.005*6} = .970$
12 mo Exist	$e^{\alpha+\beta_1}$	$e^{-5.299+.524} = .008$	$.974 * e^{-.008*6} = .928$
18 mo Exist	$e^{\alpha+\beta_2}$	$e^{-5.299+1.722} = .028$	$.957 * e^{-.028*6} = .809$
24 mo Exist	$e^{\alpha+\beta_3}$	$e^{-5.299+2.987} = .099$	$.875 * e^{-.099*6} = .483$
6 mo Exper	$e^{\alpha+\beta_4}$	$e^{-5.299+-.793} = .002$	$1 * e^{-.002*6} = .988$
12 mo Exper	$e^{\alpha+\beta_4+\beta_1}$	$e^{-5.299+-.793+.524} = .004$	$.983 * e^{-.004*6} = .957$
18 mo Exper	$e^{\alpha+\beta_4+\beta_2}$	$e^{-5.299+-.793+1.722} = .013$	$.957 * e^{-.013*6} = .885$
24 mo Exper	$e^{\alpha+\beta_4+\beta_3}$	$e^{-5.299+-.793+2.987} = .045$	$.875 * e^{-.045*6} = .668$

Problem 3

part a

Fit a GEE repeated measures logistic regression model across all study follow-up visits (not including at the time of randomization) to describe the marginal relationship of participants' self-rating of good health to the main effects of randomized group, visit (as a class variable), health self-rating at time of randomization, and age group as explanatory variables together with group by visit interaction. Use 15-24 year old women in the control group with a good health assessment at randomization as your reference group, and use 3 months post- randomization as your reference visit. Specify an exchangeable working correlation structure.

y_{ij} denotes the response from subject i at time j

Assuming an exhangeable working correlation structure:

$R_i(\alpha)$ is the working correlation matrix, which may depend on a vector of unknown parameters α which is the same for all subjects.

$$Corr(Y_{ij}, Y_{ij'}) = \begin{cases} 1 & j = j' \\ \alpha & j \neq j' \end{cases}$$

i)

Assess goodness of fit through consideration of other pairwise interactions between explanatory variables in the model. Also assess whether group by visit needs to remain in the model.

Score Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
drug	1	6.04	0.0140
Time	2	0.12	0.9414
base	1	3.79	0.0515
age_group	2	2.65	0.2660
Time*drug	2	1.59	0.4522
base*drug	1	2.36	0.1242
age_group*drug	2	4.05	0.1319
Time*base	2	0.05	0.9772
Time*age_group	4	6.64	0.1560
base*age_group	2	2.91	0.2332

Assessing goodness of fit through consideration of other pairwise interactions between explanatory variables

Model= drug time base agegroup *drug*∗time *drug*∗base *drug*∗agegroup *time*∗base *time*∗agegroup base∗agegroup

The Score statistics for Type 3 GEE Analysis Tests show that none of the interaction terms are significant

$\alpha = .05$

time*drug

$\chi^2_2 = 1.59$ p-value $= .452 > \alpha$ Reject H_0 $time * drug$ interaction term is not significant

base*drug

$\chi^2_1 = 2.36$ p-value $= .124 > \alpha$ Reject H_0 $base * drug$ interaction term is not significant

age_group*drug

$\chi^2_2 = 4.05$ p-value $= .132 > \alpha$ Reject H_0 $agegroup * drug$ interaction term is not significant

time*base

$\chi^2_2 = .05$ p-value $= .977 > \alpha$ Reject H_0 $time * base$ interaction term is not significant

$time * agegroup$ $\chi^2_4 = 6.64$ p-value $= .156 > \alpha$ Reject H_0 $time * agegroup$ interaction term is not significant

$base * agegroup$ $\chi^2_2 = 2.91$ p-value $= .233 > \alpha$ Reject H_0 $base * agegroup$ interaction term is not significant

Conclusion: no significant interaction terms

Assessing whether time*agegroup interaction should be in the model

Model=drug time base agegroup time*agegroup

Score Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
drug	1	18.07	<.0001
Time	2	0.14	0.9342
base	1	14.00	0.0002
age_group	2	5.49	0.0642
Time*age_group	4	6.68	0.1540

Score statistics for Type 3 GEE Analysis Test

time*agegroup

$\chi^2_4 = 6.68$ p-value $= .154 > \alpha$ Reject H_0 time*agegroup interaction term is not significant

Based on this result we will not include time*agegroup in the model

Our final model is:

model = drug time base agegroup

Analysis Of GEE Parameter Estimates							
Empirical Standard Error Estimates							
Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		0.3809	0.4101	-0.4228	1.1846	0.93	0.3530
drug	0	1.9493	0.4662	1.0356	2.8629	4.18	<.0001
drug	1	0.0000	0.0000	0.0000	0.0000	.	.
Time	3	-0.0756	0.3090	-0.6812	0.5299	-0.24	0.8066
Time	4	-0.2234	0.3199	-0.8504	0.4036	-0.70	0.4849
Time	2	0.0000	0.0000	0.0000	0.0000	.	.
base	Poor	-1.6875	0.4475	-2.5646	-0.8104	-3.77	0.0002
base	Good	0.0000	0.0000	0.0000	0.0000	.	.
age_group	25-34	0.9629	0.4400	0.1006	1.8252	2.19	0.0286
age_group	35+	1.0800	0.6632	-0.2198	2.3798	1.63	0.1034
age_group	15-24	0.0000	0.0000	0.0000	0.0000	.	.

ii)

Do the results of i) affect your choice for the final model? Explain in 1-2 sentences.

The interaction terms were non significant thus we have no interaction terms in our final model. Therefore the results from the first part affect the choice for the final model.

iii)

For the final model based on ii), present the table of parameter estimates, along with a standard error, test statistic, and p-value for each.

Parameter	Estimate	Standard Error	ZScore Statistics	p-value
Intercept	.381	.410	.93	.353
Drug 0	1.949	.466	4.18	<.0001
Time 3	-.076	.309	0.24	.807
Time 4	-.223	.320	-.70	.485
Base Poor	-1.688	.448	-3.77	.0002
Age 25-34	.963	.440	2.19	.029
Age 35+	1.080	.663	1.63	.103

reference group: agegroup=15-24, drug=1 (placebo), time=2, base=good

part b

Using your final recommended model from part (a), separately for each follow up visit, provide the odds ratio and corresponding 95% confidence interval that pertains to the intervention effect as estimated by the model for that visit.

α=intercept

β_1 = treatment drug

β_2 = time 3

β_3 = time 4

Odds Ratio comparing treatment effect for time 2

$\exp(\alpha_1 + \beta_1)/\exp(\alpha_1) = \exp(\beta_1) = \exp(1.949) = 7.022$

Odds Ratio comparing treatment effect for time 3

$\exp(\alpha_1 + \beta_1 + \beta_2)/\exp(\alpha_1 + \beta_2) = \exp(\beta_1) = \exp(1.949) = 7.022$

Odds Ratio comparing treatment effect for time 4

$\exp(\alpha_1 + \beta_1 + \beta_3)/\exp(\alpha_1 + \beta_3) = \exp(\beta_1) = \exp(1.949) = 7.022$

The odds ratio is the same at each time. OR= 7.022 95% CI = $(\exp(1.036), \exp(2.863)) = (2.818, 17.512)$

At each visit, the odds of self-rating of good health compared to poor for the treatment group is approximately 7 times that of the control group. The interval does not include the null value of 1 thus the results are statistically significant.

part c

Briefly summarize your findings related to the intervention effect(s) (adjusted for baseline status and age group, and addressing time in a way that is consistent with your final model) in 2-3 sentences.

Score Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
drug	1	18.43	<.0001
Time	2	0.52	0.7692
base	1	14.23	0.0002
age_group	2	5.52	0.0633

Looking at the score statistics for type 3 GEE analysis, the p-value for the treatment drug is highly significant. Based on this result and the odds ratio results from part b, the model indicates that the treatment does have an effect on self-rating of good health, adjusted for baseline status and age group.