Disease Outbreak Data

1) **Carefully describe and interpret the AGE effect. Make sure your estimate of the effect is accompanied by a reliability measure (e.g. a confidence interval).**

```
proc genmod data=disease descending;
   class SES SECT SAVINGS ;
   model Y = AGE SES SECT SAVINGS /dist=binomial link=logit;
   estimate 'Age estimate for 5 years' Age 5 / exp;
   estimate 'Age estimate for 1 years' Age 1 / exp;
run;
```

Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha
Age estimate for 5 years	0.5340	0.5117	0.5562	0.1364	0.0457	0.05
Exp(Age estimate for 5 years)				**1.1461**	0.0523	0.05
Age estimate for 1 year	0.5068	0.5023	0.5113	0.0273	0.0091	0.05
Exp(Age estimate for 1 year)				**1.0277**	0.0094	0.05

Label	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Age estimate for 5 years	0.0469	0.2259	8.92	0.0028
Exp(Age estimate for 5 years)	**1.0480**	**1.2534**		
Age estimate for 1 year	0.0094	0.0452	8.92	0.0028
Exp(Age estimate for 1 year)	**1.0094**	**1.0462**		

Controlling for other covariates, the odds of the disease is estimated to increase by 14.61% (95% CI: [4.80%, 25.34%]) for every 5 year increase in age.

2) **Test the overall goodness of fit of the model "logit=AGE+SES[3]+SECT[2]+SAVINGS[2]."**

```
proc logistic data=disease descending;
   class SES SECT SAVINGS ;
   model Y = AGE SES SECT SAVINGS /lackfit;
run;
```

Partition for the Hosmer and Lemeshow Test

Group	Total	Y = 1 Observed	Y = 1 Expected	Y = 0 Observed	Y = 0 Expected
1	21	1	2.28	20	18.72
2	20	0	2.67	20	17.33
3	22	4	3.53	18	18.47
4	20	8	3.67	12	16.33
5	20	6	4.54	14	15.46

6	20	4	6.05	16	13.95
7	20	7	7.01	13	12.99
8	20	8	8.22	12	11.78
9	20	11	10.26	9	9.74
10	13	8	8.77	5	4.23

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
12.1626	8	**0.1441**

There is no evidence (p=0.1441) for overall lack of fit.

3) Fit the model "logit=AGE+SES[3]+SECT[2]+SAVINGS[2]+SES[3]*AGE." Describe the AGE effect.

```
SES = socio-economic status (1=upper, 2=middle, 3=lower)

proc genmod data=disease descending;
   class SES SECT SAVINGS ;
   model Y = AGE SES SECT SAVINGS SES*AGE/dist=binomial link=logit;
   estimate 'Age effect for SES 1' Age 5 SES*AGE 5 0 0 / exp;
   estimate 'Age effect for SES 2' Age 5 SES*AGE 0 5 0 / exp;
   estimate 'Age effect for SES 3' Age 5 SES*AGE 0 0 5 / exp;

   estimate 'Age effect for SES 1' Age 1 SES*AGE 1 0 0 / exp;
   estimate 'Age effect for SES 2' Age 1 SES*AGE 0 1 0 / exp;
   estimate 'Age effect for SES 3' Age 1 SES*AGE 0 0 1 / exp;

run;
```

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
Age effect for SES 1	0.1051	0.0678	0.05	-0.0279	0.2381	2.40	0.1213
Exp(Age effect for SES 1)	**1.1108**	0.0754	0.05	**0.9725**	**1.2688**		
Age effect for SES 2	0.2801	0.1215	0.05	0.0419	0.5182	5.31	0.0212
Exp(Age effect for SES 2)	**1.3232**	0.1608	0.05	**1.0428**	**1.6790**		
Age effect for SES 3	0.1165	0.0687	0.05	-0.0181	0.2511	2.88	0.0899
Exp(Age effect for SES 3)	**1.1235**	0.0772	0.05	**0.9820**	**1.2854**		

Controlling for other covariates, the partial effect of age depends on the level of socio-economic status (SES):

For upper class, the odds of the disease is estimated to increase by 11.08% (95% CI: [-2.75%, 26.88%]) for every 5 year increase in age.

For middle class, the odds of the disease is estimated to increase by 32.32% (95% CI: [4.28%, 67.90%]) for every 5 year increase in age.

For lower class, the odds of the disease is estimated to increase by 12.35% (95% CI: [-1.80%, 28.54%]) for every 5 year increase in age.

For one year increase in age:

Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha
Age effect for SES 1	0.5053	0.4986	0.5119	0.0210	0.0136	0.05
Exp(Age effect for SES 1)				1.0212	0.0139	0.05
Age effect for SES 2	0.5140	0.5021	0.5259	0.0560	0.0243	0.05
Exp(Age effect for SES 2)				1.0576	0.0257	0.05
Age effect for SES 3	0.5058	0.4991	0.5126	0.0233	0.0137	0.05
Exp(Age effect for SES 3)				1.0236	0.0141	0.05

Contrast Estimate Results

Label	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Age effect for SES 1	-0.0056	0.0476	2.40	0.1213
Exp(Age effect for SES 1)	0.9944	1.0488		
Age effect for SES 2	0.0084	0.1036	5.31	0.0212
Exp(Age effect for SES 2)	1.0084	1.1092		
Age effect for SES 3	-0.0036	0.0502	2.88	0.0899
Exp(Age effect for SES 3)	0.9964	1.0515		

Number of Children Data

```
data children; set children;
Y = avg*n;
logn = log(n);
run;

proc genmod data=children;
  class dur res educ;
  model Y = dur res educ / dist=poisson link=log offset = logn;
  estimate  'logrr(res=rural v suva)' res -1 1 0 /exp;
  estimate  'logrr(res=rural v urban)' res 0 1 -1 /exp;
  estimate  'logrr(res=urban v suva)' res -1 0 1 /exp;
run;
```

1) **Using the model of 1, describe the residence effect on the expected number of births, controlling for the other covariates.**

Label	Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq

logrr(res=rural v suva)	0.1512	0.0283	0.05	0.0957	0.2067	28.50	<.0001
Exp(logrr(res=rural v suva))	**1.1632**	0.0330	0.05	**1.1004**	**1.2297**		
logrr(res=rural v urban)	0.0390	0.0246	0.05	-0.0093	0.0872	2.50	0.1136
Exp(logrr(res=rural v urban))	**1.0397**	0.0256	0.05	**0.9907**	**1.0911**		
logrr(res=urban v suva)	0.1123	0.0325	0.05	0.0486	0.1759	11.94	0.0006
Exp(logrr(res=urban v suva))	**1.1188**	0.0364	0.05	**1.0498**	**1.1924**		

Controlling for other covariates, the expected number of births per woman in rural areas is estimated to be 1.1632 (95% CI: [1.1004, 1.2297]) times higher than that in Suva and 1.0397 (95% CI: [0.9907, 1.0911]) times higher than that in urban areas.

Controlling for other covariates, the expected number of births per woman in urban areas is estimated to be 1.1188 (95% CI: [1.0498, 1.1924]) times higher than that in Suva.

2) Do you suspect overdispersion? NO

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	59	70.6526	**1.1975**
Scaled Deviance	59	70.6526	1.1975
Pearson Chi-Square	59	71.5057	1.2120
Scaled Pearson X2	59	71.5057	1.2120
Log Likelihood		49456.1204	

In this model setting, the dispersion parameter is known and there is generally a large Poisson mean for each observation (out of the 70 data points only 4 observations are less than 10). Therefore, an approximate chi-square distribution is reasonable for testing overall goodness of fit.

Treatment Data

```
Assume that the population-average probabilities can be modeled as "logit
P(Y=1|c) = 1 + Center[2] + Treatment[2] + Sex[2] + Age + Baseline + Visit".
Use GEE to estimate and describe the treatment effect.

proc genmod data=Resp1 descending;
class IDnew Center Treatment Sex;
model Resp = Center Treatment Sex Age Baseline visit /dist=binomial
link=logit;
repeated subject=IDnew/type=exch covb;
estimate 'Active vs Placebo' Treatment 1 -1 /exp;
run;
```

Label	Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
Active vs Placebo	1.2556	0.3463	0.05	0.5768	1.9344	13.14	0.0003
Exp(Active vs Placebo)	**3.5100**	**1.2156**	**0.05**	**1.7804**	**6.9200**		

Controlling for other covariates and averaging over all patients, the odds of feeling well with active treatment is estimated to be 3.51 times higher (95% CI: [1.78, 6.92]) than the odds of feeling well with placebo.

Supplementary analysis (wasn't included in the HW assignment)

Re-do Part 1, but this time incorrectly assume that the four responses per patient are independent. How do your answers in Parts 1 and 2 compare?

```
proc genmod data=Resp1 descending;
class obs Center Treatment Sex;
model Resp = Center Treatment Sex Age Baseline visit /dist=binomial
link=logit;
repeated subject=obs/type=exch covb;
estimate 'Active vs Placebo' Treatment 1 -1 /exp;
run;
```

Label	Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
Active vs Placebo	1.2673	0.2352	0.05	0.8062	1.7284	29.02	<.0001
Exp(Active vs Placebo)	3.5513	0.8354	0.05	2.2394	5.6315		

Controlling for other covariates and averaging over all patients, the odds of feeling well with active treatment is estimated to be 3.55 times higher (95% CI: [2.24, 5.63]) than the odds of feeling well with placebo.

Incorrectly assuming independence results in a smaller confidence interval for the estimated effect. This puts more confidence in the estimate than there really is.