Assignment 7

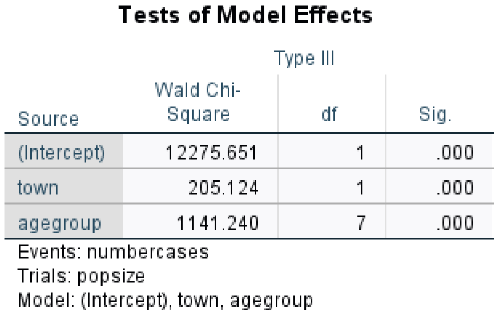
Sainivedhitha Arunajatesan

Exercise 1

Effect of age group and town on number of skin cancer patients

We perform logistic regression of the dependent variable (Number of patients) with age group and town.

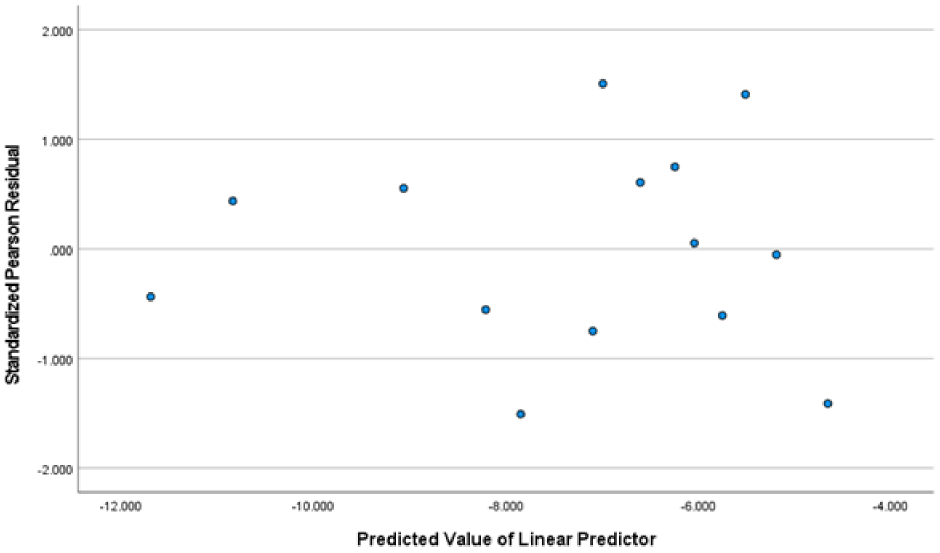
$$
\begin{aligned}
ln(\frac{p}{1-p}) &= \eta \\
&= \beta_0 + \beta_1 TOWN + \beta_2 AGEGROUP(2) + \beta_3 AGEGROUP(3) + \beta_4 AGEGROUP(4) \\
&+ \beta_5 AGEGROUP(5) + \beta_6 AGEGROUP(6) + \beta_7 AGEGROUP(7) + \beta_8 AGEGROUP(8)
\end{aligned}
$$

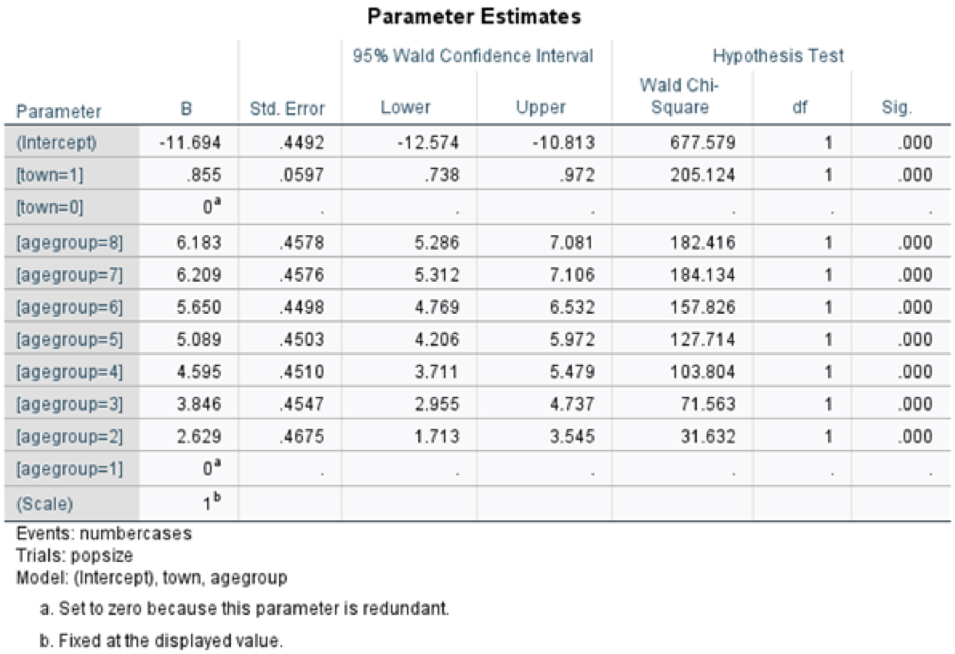


Tests of Model Effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	12275.651	1	.000
town	205.124	1	.000
agegroup	1141.240	7	.000

Events: numbercases
Trials: popsize
Model: (Intercept), town, agegroup





Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-11.694	.4492	-12.574	-10.813	677.579	1	.000
[town=1]	.855	.0597	.738	.972	205.124	1	.000
[town=0]	0[a]	.	.	.	.	.	.
[agegroup=8]	6.183	.4578	5.286	7.081	182.416	1	.000
[agegroup=7]	6.209	.4576	5.312	7.106	184.134	1	.000
[agegroup=6]	5.650	.4498	4.769	6.532	157.826	1	.000
[agegroup=5]	5.089	.4503	4.206	5.972	127.714	1	.000
[agegroup=4]	4.595	.4510	3.711	5.479	103.804	1	.000
[agegroup=3]	3.846	.4547	2.955	4.737	71.563	1	.000
[agegroup=2]	2.629	.4675	1.713	3.545	31.632	1	.000
[agegroup=1]	0[a]	.	.	.	.	.	.
(Scale)	1[b]						

Events: numbercases
Trials: popsize
Model: (Intercept), town, agegroup

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

The significance value of all the variables is 0.000 (or p<0.05). So, we reject the null hypothesis. In this case, the null hypothesis is that both age group and town has no effect on the number of skin cancer patients.

This can also be proven by the value of Wald Chi Square value. For town (205.124), we reject the null hypothesis if Wald $\geq$ 3.841 (df=1 & α=5%) and for age group (1141.240), we reject the null hypothesis if Wald $\geq$ 14.067 (df=4 & α=5%).

Odds Ratio:

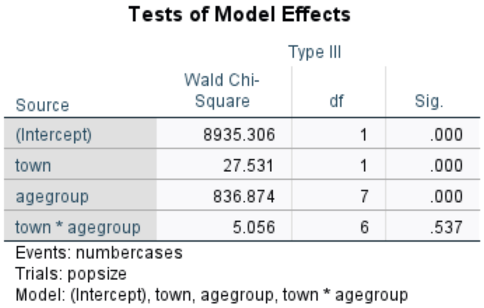
$$\beta_1 TOWN = e^{\beta_1} = e^{0.855} = 2.35$$
$$\beta_2 AGEGROUP(2) = e^{\beta_2} = e^{2.629} = 13.86$$
$$\beta_3 AGEGROUP(3) = e^{\beta_3} = e^{3.846} = 46.81$$
$$\beta_4 AGEGROUP(4) = e^{\beta_4} = e^{4.595} = 98.99$$
$$\beta_5 AGEGROUP(5) = e^{\beta_5} = e^{5.089} = 162.23$$
$$\beta_6 AGEGROUP(6) = e^{\beta_6} = e^{5.650} = 284.29$$
$$\beta_7 AGEGROUP(7) = e^{\beta_7} = e^{6.209} = 497.20$$
$$\beta_8 AGEGROUP(8) = e^{\beta_8} = e^{6.183} = 484.44$$

The confidence interval was set to be 95%. The boundaries for each variable can be found in the parameters table given above.

Now we add a new variable as interaction between age group and town to determine the interaction effect between the independent variables.

The significance value of this interaction is 0.537 (or p$\geq$0.05). So we accept the null hypothesis. In this case, the null hypothesis is that there is no significant interaction between the age group and town value.

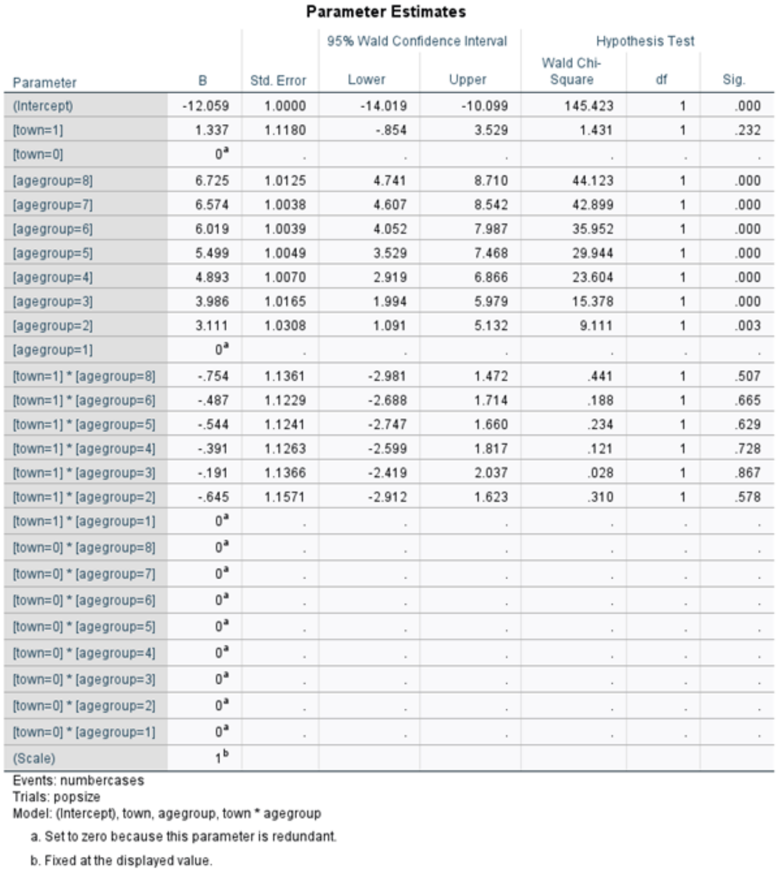
This can also be proven by the value of Wald Chi Square value. For town*age group (5.056), we reject the null hypothesis if Wald ≥ 12.592 (df=6 & α=5%).



Tests of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	8935.306	1	.000
town	27.531	1	.000
agegroup	836.874	7	.000
town * agegroup	5.056	6	.537

Events: numbercases
Trials: popsize
Model: (Intercept), town, agegroup, town * agegroup



Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval Lower	 Upper	Hypothesis Test Wald Chi-Square	 df	 Sig.
(Intercept)	-12.059	1.0000	-14.019	-10.099	145.423	1	.000
[town=1]	1.337	1.1180	-.854	3.529	1.431	1	.232
[town=0]	0[a]	.	.	.	.	.	.
[agegroup=8]	6.725	1.0125	4.741	8.710	44.123	1	.000
[agegroup=7]	6.574	1.0038	4.607	8.542	42.899	1	.000
[agegroup=6]	6.019	1.0039	4.052	7.987	35.952	1	.000
[agegroup=5]	5.499	1.0049	3.529	7.468	29.944	1	.000
[agegroup=4]	4.893	1.0070	2.919	6.866	23.604	1	.000
[agegroup=3]	3.986	1.0165	1.994	5.979	15.378	1	.000
[agegroup=2]	3.111	1.0308	1.091	5.132	9.111	1	.003
[agegroup=1]	0[a]	.	.	.	.	.	.
[town=1] * [agegroup=8]	-.754	1.1361	-2.981	1.472	.441	1	.507
[town=1] * [agegroup=6]	-.487	1.1229	-2.688	1.714	.188	1	.665
[town=1] * [agegroup=5]	-.544	1.1241	-2.747	1.660	.234	1	.629
[town=1] * [agegroup=4]	-.391	1.1263	-2.599	1.817	.121	1	.728
[town=1] * [agegroup=3]	-.191	1.1366	-2.419	2.037	.028	1	.867
[town=1] * [agegroup=2]	-.645	1.1571	-2.912	1.623	.310	1	.578
[town=1] * [agegroup=1]	0[a]	.	.	.	.	.	.
[town=0] * [agegroup=8]	0[a]	.	.	.	.	.	.
[town=0] * [agegroup=7]	0[a]	.	.	.	.	.	.
[town=0] * [agegroup=6]	0[a]	.	.	.	.	.	.
[town=0] * [agegroup=5]	0[a]	.	.	.	.	.	.
[town=0] * [agegroup=4]	0[a]	.	.	.	.	.	.
[town=0] * [agegroup=3]	0[a]	.	.	.	.	.	.
[town=0] * [agegroup=2]	0[a]	.	.	.	.	.	.
[town=0] * [agegroup=1]	0[a]	.	.	.	.	.	.
(Scale)	1[b]						

Events: numbercases
Trials: popsize
Model: (Intercept), town, agegroup, town * agegroup

[a] Set to zero because this parameter is redundant.

[b] Fixed at the displayed value.

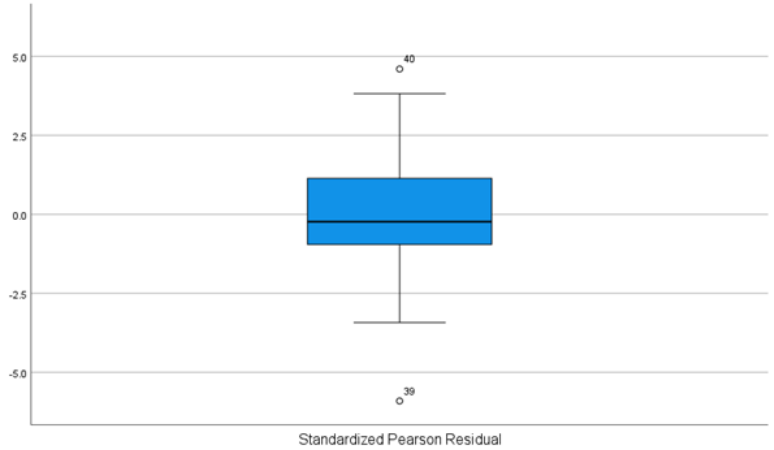
CONCLUSION: The number of women having non-melanoma skin cancer is affected by both the towns (Minneapolis and Dallas) and the age group (all the 8 groups) while the age group and town have no interaction between each other.

Exercise 2

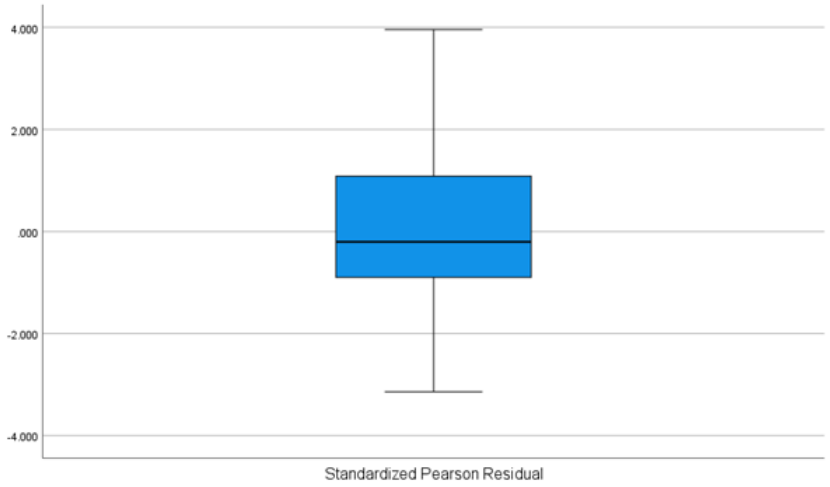
Effect of Area, Age and Sex on the number of deaths

We perform logistic regression of the dependent variable (Number of deaths) with area, age group and sex.

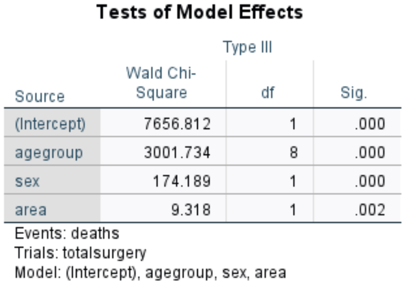
$$\begin{aligned} ln(\frac{p}{1-p}) = \eta = \beta_0 + \beta_1 AREA + \beta_2 AGEGROUP(2) + \beta_3 AGEGROUP(3) + \beta_4 AGEGROUP(4) \\ + \beta_5 AGEGROUP(5) + \beta_6 AGEGROUP(6) + \beta_7 AGEGROUP(7) + \beta_8 AGEGROUP(8) \\ + \beta_9 AGEGROUP(9) + \beta_{10} AGEGROUP(10) + \beta_{11} SEX \end{aligned}$$



Based on the box plot obtained above using the residuals of logistic regression, we find a few outliers in age group 10. It is also evident through the data view that age group 10 has outliers. So, I decided to leave out age group 10 to obtain a better result. The box plot without any outliers after omitting the age group 10 is given below.



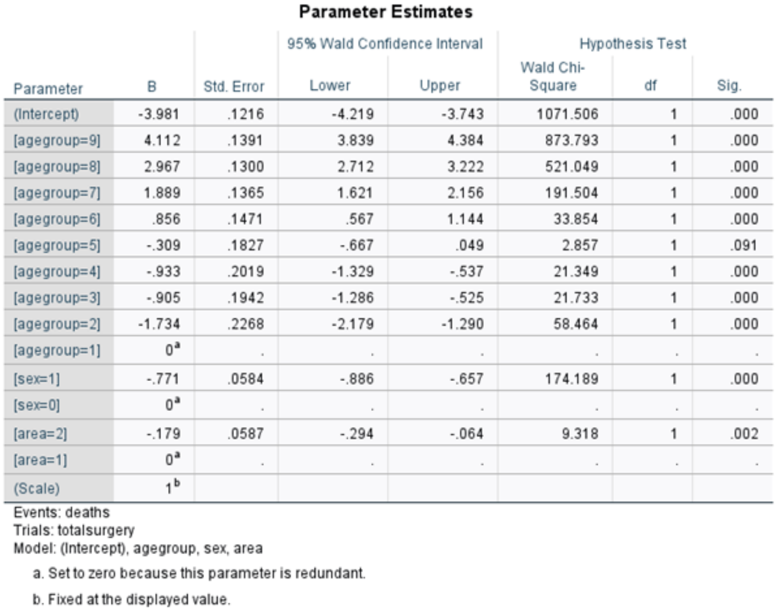
WITHOUT INTERACTION:



Tests of Model Effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	7656.812	1	.000
agegroup	3001.734	8	.000
sex	174.189	1	.000
area	9.318	1	.002

Events: deaths
Trials: totalsurgery
Model: (Intercept), agegroup, sex, area



Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-3.981	.1216	-4.219	-3.743	1071.506	1	.000
[agegroup=9]	4.112	.1391	3.839	4.384	873.793	1	.000
[agegroup=8]	2.967	.1300	2.712	3.222	521.049	1	.000
[agegroup=7]	1.889	.1365	1.621	2.156	191.504	1	.000
[agegroup=6]	.856	.1471	.567	1.144	33.854	1	.000
[agegroup=5]	-.309	.1827	-.667	.049	2.857	1	.091
[agegroup=4]	-.933	.2019	-1.329	-.537	21.349	1	.000
[agegroup=3]	-.905	.1942	-1.286	-.525	21.733	1	.000
[agegroup=2]	-1.734	.2268	-2.179	-1.290	58.464	1	.000
[agegroup=1]	0[a]	.	.	.	.	.	.
[sex=1]	-.771	.0584	-.886	-.657	174.189	1	.000
[sex=0]	0[a]	.	.	.	.	.	.
[area=2]	-.179	.0587	-.294	-.064	9.318	1	.002
[area=1]	0[a]	.	.	.	.	.	.
(Scale)	1[b]						

Events: deaths
Trials: totalsurgery
Model: (Intercept), agegroup, sex, area

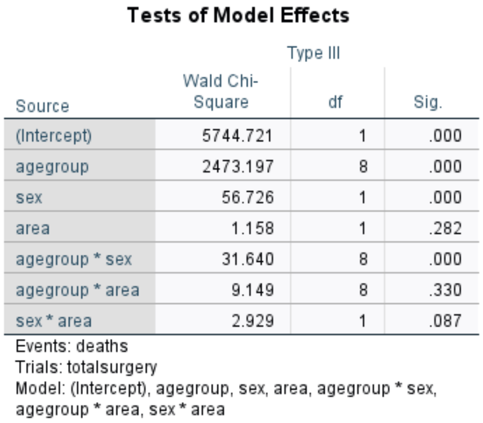
a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

The significance value of all the variables is 0.000 (or $p<0.05$). So, we reject the null hypothesis. In this case, the null hypothesis is that area, age group and sex has no effect on the number of deaths.

This can also be proven by the value of Wald Chi Square value. For age group (3001.734), we reject the null hypothesis if Wald $\geq$ 15.507 (df=8 & α=5%), for area (9.319), we reject the null hypothesis if Wald $\geq$ 3.841 (df=1 & α=5%) and for sex (174.189), we reject the null hypothesis if Wald $\geq$ 3.841 (df=1 & α=5%)

WITH INTERACTION:

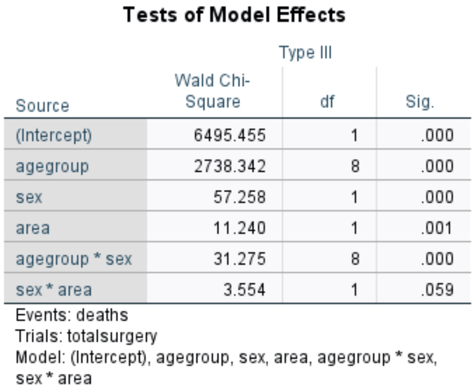


Tests of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	5744.721	1	.000
agegroup	2473.197	8	.000
sex	56.726	1	.000
area	1.158	1	.282
agegroup * sex	31.640	8	.000
agegroup * area	9.149	8	.330
sex * area	2.929	1	.087

Events: deaths
Trials: totalsurgery
Model: (Intercept), agegroup, sex, area, agegroup * sex, agegroup * area, sex * area

The results show that there is no significant interaction between area*age group and area*sex. Therefore, we first omit the interaction with the highest significance value (0.330) and re-run the output results.

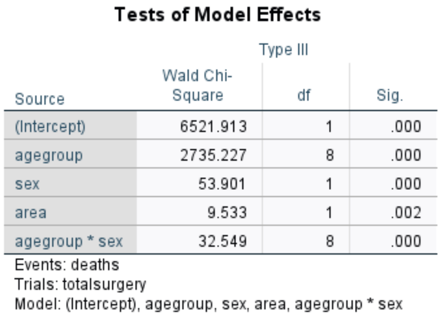


Tests of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	6495.455	1	.000
agegroup	2738.342	8	.000
sex	57.258	1	.000
area	11.240	1	.001
agegroup * sex	31.275	8	.000
sex * area	3.554	1	.059

Events: deaths
Trials: totalsurgery
Model: (Intercept), agegroup, sex, area, agegroup * sex, sex * area

The results show that there is no significant interaction between area*sex. Therefore, we now omit the interaction with the significance value (0.059) and re-run the output results.

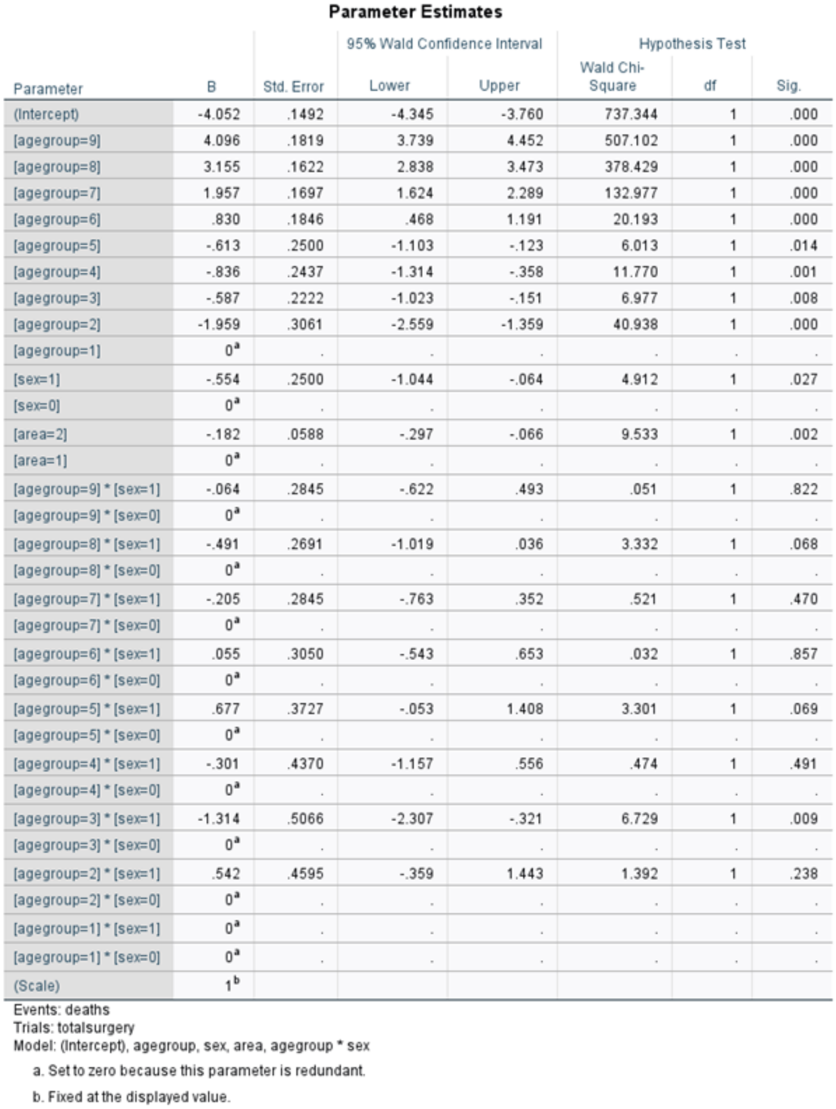


Tests of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	6521.913	1	.000
agegroup	2735.227	8	.000
sex	53.901	1	.000
area	9.533	1	.002
agegroup * sex	32.549	8	.000

Events: deaths
Trials: totalsurgery
Model: (Intercept), agegroup, sex, area, agegroup * sex

The results show the remaining variables having a significant effect on the dependent variable (number of deaths).
The more detailed parameter estimates are displayed below.



Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-4.052	.1492	-4.345	-3.760	737.344	1	.000
[agegroup=9]	4.096	.1819	3.739	4.452	507.102	1	.000
[agegroup=8]	3.155	.1622	2.838	3.473	378.429	1	.000
[agegroup=7]	1.957	.1697	1.624	2.289	132.977	1	.000
[agegroup=6]	.830	.1846	.468	1.191	20.193	1	.000
[agegroup=5]	-.613	.2500	-1.103	-.123	6.013	1	.014
[agegroup=4]	-.836	.2437	-1.314	-.358	11.770	1	.001
[agegroup=3]	-.587	.2222	-1.023	-.151	6.977	1	.008
[agegroup=2]	-1.959	.3061	-2.559	-1.359	40.938	1	.000
[agegroup=1]	0[a]	.	.	.	.	.	.
[sex=1]	-.554	.2500	-1.044	-.064	4.912	1	.027
[sex=0]	0[a]	.	.	.	.	.	.
[area=2]	-.182	.0588	-.297	-.066	9.533	1	.002
[area=1]	0[a]	.	.	.	.	.	.
[agegroup=9] * [sex=1]	-.064	.2845	-.622	.493	.051	1	.822
[agegroup=9] * [sex=0]	0[a]	.	.	.	.	.	.
[agegroup=8] * [sex=1]	-.491	.2691	-1.019	.036	3.332	1	.068
[agegroup=8] * [sex=0]	0[a]	.	.	.	.	.	.
[agegroup=7] * [sex=1]	-.205	.2845	-.763	.352	.521	1	.470
[agegroup=7] * [sex=0]	0[a]	.	.	.	.	.	.
[agegroup=6] * [sex=1]	.055	.3050	-.543	.653	.032	1	.857
[agegroup=6] * [sex=0]	0[a]	.	.	.	.	.	.
[agegroup=5] * [sex=1]	.677	.3727	-.053	1.408	3.301	1	.069
[agegroup=5] * [sex=0]	0[a]	.	.	.	.	.	.
[agegroup=4] * [sex=1]	-.301	.4370	-1.157	.556	.474	1	.491
[agegroup=4] * [sex=0]	0[a]	.	.	.	.	.	.
[agegroup=3] * [sex=1]	-1.314	.5066	-2.307	-.321	6.729	1	.009
[agegroup=3] * [sex=0]	0[a]	.	.	.	.	.	.
[agegroup=2] * [sex=1]	.542	.4595	-.359	1.443	1.392	1	.238
[agegroup=2] * [sex=0]	0[a]	.	.	.	.	.	.
[agegroup=1] * [sex=1]	0[a]	.	.	.	.	.	.
[agegroup=1] * [sex=0]	0[a]	.	.	.	.	.	.
(Scale)	1[b]						

Events: deaths
Trials: totalsurgery
Model: (Intercept), agegroup, sex, area, agegroup * sex

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Odds Ratio:

$$\beta_1 AREA = e^{\beta_1} = e^{-0.182} = 0.8336$$
$$\beta_{11} SEX = e^{\beta_{11}} = e^{-0.554} = 0.5746$$

Similarly, the odds ratio can be calculated for age groups 2 to 9 and for interaction between each age group (from 2 to 9) and female (sex=1).

The confidence interval was set to be 95%. The boundaries for each variable can be found in the parameters table given above.

CONCLUSION: Based on this data we can conclude age groups*sex have a significant effect on the number of deaths. Furthermore, all individual age groups, sex as well as area has a significant effect on the number of deaths.