HW #3
Biost 515

Scott Emerson
1/27/13

1. The relationship between serum LDL and 5 year all-cause mortality was assessed using logistic regression. The odds of death within 5 years was compared between serum LDL groups. Serum LDL level was dichotomized. Subjects were grouped according to high serum LDL (LDL≥ 160 mg/dL) and low serum LDL (LDL<160 mg/dL). Observation time was also dichotomized to differentiate subjects who died within 5 years and those that did not. Odds of death and probability of death for each LDL group was calculated. The odds ratio was also calculated to compare odds of death between groups. Wald based confidence intervals were also calculated for the odds ratio. The null hypothesis that there is no association between serum LDL and 5 year all-cause mortality, or equivalently, that the odds ratio is equal to 1, was tested. An alpha level of 0.05 was used to calculate confidence intervals and make inference on test statistics. Ten observations were dropped and not included in the analysis due to missing serum LDL measurements.

	1. This model is saturated. We are modeling the odds of death in two groups, subjects with high serum LDL levels and subjects with low serum LDL levels. There are two parameters in the model, serum LDL and survival status at 5 years. The number of parameters and the number of groups are equal, so this model is saturated.
	2. The odds of dying within 5 years for subjects who have low LDL is 0.205. The estimated probability of dying within 5 years for subjects who have low LDL is 0.17. The observed proportion of subjects with low LDL who died within 5 years is the same as the estimated probability of dying, 0.17.
	3. The odds of dying within 5 years for subjects who have high LDL is 0.151. The estimated probability of dying within 5 years for subjects who have high LDL is 0.131. The observed proportion of subjects with high LDL who died within 5 years is also 0.131.
	4. The estimated odds of dying within 5 years for subjects who have high or low LDL are reported above. The estimated odds of dying within 5 years for subjects who have high LDL is 0.735 times the odds of dying within 5 years for subjects who have low LDL. It would not be unusual to observe a ratio comparing the odds of death for high LDL patients to low LDL patients between 0.404 and 1.340. Note this interval contains 1, so it would not be unusual if both LDL groups had the same odds of dying within 5 years. The p-value yielded from the Wald test is 0.315. Therefore, we cannot reject the null hypothesis that there is an association between 5-year all-cause mortality and LDL.
	This is the same conclusion that was reached in HW #1 (parts 5 and 6). The p-value in reached in HW #1 part 6, is slightly larger than the one reported above. The difference is due to the test used. The Wald test was used in this analysis, while Fisher’s exact test was used in HW #1. The confidence intervals are also slightly different from those that were found in HW #1. This is because the confidence intervals were calculating Wald based confidence intervals in the above analysis, while Woolf’s formula was used to calculate the confidence interval in homework 1.
	5. If the predictor of interest was an indicator for low LDL the estimates for the odds of dying in 5 years for each group would not change, as this would just be a reparamterization of the model. The estimated odds ratio and corresponding confidence intervals would change. The estimated odds ratio would be the inverse of the reported value above. If the response variable was and indicator of survival for at least 5 years the estimated odds of surviving for each group would change.

	If we made the reparamterization that the response was the odds of survival for at least 5 years the estimates reported would changes. The odds of surviving for at least 5 years for each group would be the inverse of the odds reported in (b) and (c). The odds ratio would also be inverted. In both cases the model would still not be saturated.
	6. If we changed the response variable to be serum LDL groups and the predictor of interest to be the survival status at 5 years the odds ratio reported in part (d) would not change. However, the point estimates for the odds of having high LDL for subjects who died within 5 years and the odds of having high LDL for subjects who survived for more than 5 years would not be the same as our estimates in (b) and (c). The model would still be saturated.
2. The relationship between serum LDL and 5 year all-cause mortality was assessed using classical linear regression. The probability of death within 5 years was compared between LDL groups. Serum LDL level was dichotomized. Subjects were grouped according to high serum LDL (LDL≥ 160 mg/dL) and low serum LDL (LDL<160 mg/dL). Observation time was also dichotomized to differentiate subjects who died within 5 years and those that did not. The risk of death for each LDL group was calculated. The risk difference was also calculated to compare probability of death between LDL groups. Confidence intervals were also calculated for the risk difference. The null hypothesis that there is no association between serum LDL and 5 year all-cause mortality, or equivalently, that the risk difference is equal to 0, was tested using a Chi-squared test. Ten observations were dropped and not included in the analysis due to missing serum LDL measurements.

	1. This model is saturated. There are two parameters in the model, serum LDL and survival status at 5 years. We are modeling the risk of death in two groups, subjects with high LDL and subjects with normal LDL. The number of groups is equal to the number of parameters so this is a saturated regression model.
	2. The estimated probability of dying within 5 years for subjects with low LDL is 0.17. The estimated odds of dying within 5 years for subjects with low LDL is 0.205. These estimates are the same as the observed proportion of subjects with low LDL dying within 5 years and the observed odds of dying within 5 years.
	3. The estimated probability of dying with 5 years for subjects with high LDL is 0.131. The estimated odds of dying within 5 years for subjects with high LDL is 0.151. These estimates are the same as the observed proportion of subjects with high LDL dying within 5 years and the observed odds of dying within 5 years.
	4. The estimated risk of dying for each LDL group is reported above. The estimated difference in probability of dying within 5 years is -0.039. That is, the estimated probability of dying within 5 years for subjects with high LDL is 0.039 less than the estimated probability of dying within 5 years for subjects with low LDL. Comparing high LDL subjects to low LDL subjects, it would not be unusual to observe a difference in probability of death within 5 years between -0.115 and .037. Note this interval contains 0, so it would not be unusual to observe the same risk of death for both groups. The p-value yielded from the test of the null hypothesis is 0.315. Therefore, we cannot reject the null hypothesis that there is an association between 5-year all-cause mortality and LDL. This conclusion in supported by the inference made from the confidence intervals.

	The p-values found in this analysis and in homework 1 are the same. The confidence intervals found in this analysis are slightly different than those found in homework 1. Wald based confidence intervals were used in homework 1 while a normal approximation was used to calculate confidence in this homework
	5. If we changed the predictor of interest to be an indicator of low LDL none of the estimates reported in (b) or (c) would change, as it would be just a re-parameterization of the model. The model would also still be saturated since the number of groups and parameters would not change.

	If we changed the response variable to be an indicator of survival for at least 5 years the estimates reported in (b) or (c) would change. The estimated probability of survival would be 1 minus the estimated probability of death reported in parts (b) and (c). The model would still be saturated since the number of groups and parameters would not change.
	6. If we switch the predictor of interest and the response variable in our model the estimated probabilities reported above would change. The probability of having high LDL for subjects who survived longer than 5 years would be larger than the probability of having high LDL for subjects who died within 5 years, but these would be different probabilities than reported above. The risk difference would also change.
3. The relationship between serum LDL and 5 year all-cause mortality was assessed using Poisson regression. The ratio of risk of death within 5 years was used compared LDL groups. Serum LDL level was dichotomized. Subjects were grouped according to high serum LDL (LDL≥ 160 mg/dL) and low serum LDL (LDL<160 mg/dL). Observation time was also dichotomized to differentiate subjects who died within 5 years and those that did not. The risk of death for each LDL group was calculated. The relative risk was also calculated to compare risk of death between LDL groups. Wald-based confidence intervals were also calculated for the relative risk. The null hypothesis that there is no association between serum LDL and 5 year all-cause mortality, or equivalently, that the relative risk is equal to 1, was tested. Ten observations were dropped and not included in the analysis due to missing serum LDL measurements.

	1. This model is saturated. There are two parameters in the model, serum LDL and survival status at 5 years. We are modeling the risk of death in two groups, subjects with high LDL and subjects with normal LDL. The number of groups is equal to the number of parameters so this is a saturated regression model.
	2. The estimated probability of dying within 5 years for subjects with low LDL is 0.17. The estimated odds of dying within 5 years for subjects with low LDL is 0.205. These estimates are the same as the observed proportion of subjects with low LDL dying within 5 years and the observed odds of dying within 5 years.
	3. The estimated probability of dying within 5 years for subjects with high LDL is 0.131. The estimated odds of dying within 5 years for subjects with high LDL is 0.151. These estimates are the same as the observed proportion of subjects with high LDL dying within 5 years and the observed odds of dying within 5 years.
	4. The estimated probability and odds of dying within 5 years by LDL group is reported above. The estimated ratio of the probability of death for the high LDL group compared to the low LDL group is 0.77. It would not be unusual to observe a ratio comparing the probability of death for the high LDL to low LDL group between 0.441 and 1.36. That is, it would not be unusual to observe a probability of death for the high LDL group that is 44.1% to 136% of the low LDL group. Note this interval contains 1 (100%), so it would not be unusual to observe the same probability of death for both high and low LDL groups. The p-value yielded from testing the hypothesis that that the ratio of risk is same between each LDL group is 0.359. We, therefore, fail reject the null hypothesis, and cannot claim there is an association between serum LDL and 5-year all-cause mortality.

	This is the same conclusion as reached in homework 1. The p-values reported in homework 1 differ the p-value reported in the above analysis. The difference is due to the type of test that is used to test the null hypothesis. A Wald based test statistic was used in this analysis, while a Chi-squared based test statistic was used in homework 1. The confidence interval reported in homework 1 is also different then the confidence interval reported in the above analysis.
	5. If the model were re-parameterized such that the predictor of interest is an indicator of low LDL the estimates in (b) and (c) would not change. The model would also still be saturated since the number of groups and parameters would not change.

	If we changed the response variable to be an indicator of survival for at least 5 years the estimates reported in (b) or (c) would change. The estimated probability of survival would be 1 minus the estimated probability of death reported in parts (b) and (c). The model would still be saturated since the number of groups and parameters would not change.
	6. If we switch the predictor of interest and the response variable in our model the estimated probabilities reported above would change. We would instead be modeling the probability of having high LDL give a subject’s survival group. Because the risk ratio is the ratio of these two probabilities it would also change compared to what was reported above. The model would still be saturated.
	7. The relationship between serum LDL and 5 year all-cause mortality was assessed using robust linear regression. The difference in probability of death was calculated for subjects who had a 1 mg/dL difference in serum LDL levels. Confidence intervals were also calculated for the difference in probabilities. The null hypothesis that there is no association between serum LDL and 5 year all cause mortality (or equivalently there is no difference in probability of dying within 5 years for different LDL levels) was tested. All confidence intervals and tests were evaluated at the 0.05 alpha level.

	The estimated difference in probability of dying for subjects who have a 1 mg/dL difference in LDL levels is -0.00103. That is, for a 1 mg/dL increase in serum LDL levels the probability of dying within 5 years decreases by 0.00103. It would not be unusual to observe a risk difference between -0.00188 and -.000185. Note this interval does not contain 1, so it would be unusual if the probability of death were the same between subjects with different LDL levels. The p-value yielded in the test of the null hypothesis is 0.017. We reject the null hypothesis and find that there is a statistical association between 5 year all-cause mortality and serum LDL.
	8. The relationship between serum LDL and 5 year all-cause mortality was assessed using Poisson regression. The ratio of probabilities of death was calculated for subjects who had a 1 mg/dL difference in serum LDL levels. Wald confidence intervals were also calculated for the ratio of probabilities. The null hypothesis that there is no association between serum LDL and 5 year all cause mortality (or equivalently there is ratio of probabilities of dying within 5 years for different LDL levels is 1) was tested. All confidence intervals and tests were evaluated at the 0.05 alpha level.

	The estimated ratio of probabilities of dying within 5 groups comparing subjects with a 1 mg/dL difference in LDL levels is 0.994. That is, a subject with a 1mg/dL higher LDL level will have 99.4% of the probability of dying within 5 years that a subject with a with a 1mg/dL lower LDL level has. It would not be unusual to observe a ratio of probabilities of death within 5 years between 0.988 and 0.999. Note this interval does not contain 1, so it would be unusual to observe the same probability of death for subjects with different serum LDL levels. The p-value yielded in the test of the null hypothesis is 0.021. We reject the null hypothesis and find that there is a statistical association between 5 year all-cause mortality and serum LDL. However, these estimated ratio and confidence interval are very close to 1, so the scientific association may not be meaningful.
	9. The relationship between serum LDL and 5 year all-cause mortality was assessed using logistic regression. The odds of death within 5 years were calculated for subjects who had a 1 mg/dL difference in serum LDL levels. Wald confidence intervals were also calculated for the odds of death. The null hypothesis that there is no association between serum LDL and 5 year all cause mortality (or equivalently the ratio of odds of dying within 5 years for different LDL levels is 1) was tested. All confidence intervals and tests were evaluated at the 0.05 alpha level.

	The estimated ratio of odds of dying within 5 years for subjects who have a 1 mg/dL difference in serum LDL levels is 0.992. That is, the odds of death for a subject that has a 1 mg/dL LDL level higher compared to another subject has 0.992 times the second subjects odds of dying within 5 years. It would not be unusual to observe a ratio of odds of dying within 5 years between 0.986 and 0.998. Note this interval does not contain 1, so it would be unusual to observe the same probability of death for subjects with different serum LDL levels. The p-valued yielded from the test of the null hypothesis is 0.012. We reject the null hypothesis and find that there is a statistical association between the 5 year all-cause mortality and serum LDL levels. However, these estimated ratio and confidence interval are very close to 1, so the scientific association may not be meaningful.
	10. The conclusions reached in homework 2 and the conclusions reached above are different. In questions 1-3, when we considered death within 5 years to be the outcome of interest, we did not find an association between serum LDL and 5 year all-cause mortality. In homework 2, when we considered serum LDL as the response variable, we did find a significant association between serum LDL and 5 year all-cause mortality. A priori, I would be prefer the model we used in part (1). The odds ratio was shown to be invariant regardless of the parameterization of the model. Regardless of what we choose to be our response and predictor of interest our estimate of the odds ratio will not change. That gives us more flexibility a priori.