Homework Number 2
student 2931

These tables summarize my results. I attempt to answer the questions below.

Associations for Death Within Four Years (Response Variable "cvddeath4")

Predictor: estrogen

Measure	Value	95% CI		Stat. Significance
RD	-.026	-.0378	-.0134	0.011 (chi-sq)
OR	.250	.07865	.7945	0.019 (p)
RR	.257	.08164	.8065	0.020 (p)

Predictors: estrogen, prevdis:NO (stratified)

Measure	Value	95% CI		Stat. Significance
RD	-.0116	-.0223	-.0010	chi2 = 0.1345
OR	.3524	.0845	1.4695	0.152 p
RR	.3566	.0864	1.4725	0.154 p

Predictors: estrogen, prevdis:YES (stratified)

Measure	Value	95% CI		Stat. Significance
RD	-.0659	-.1351	.0034	chi2 = 0.2328
OR	.3130	.0418	2.347	0.258 p
RR	.3359	.0480	2.353	0.272 p

Predictors: estrogen, prevdis

Measure	Value	95% CI		Stat. Significance
OR estrogen	.3382	.1055	1.084	0.068
OR prevdis	5.956	3.882	9.137	0.000
RR estrogen	.3412	.0779	1.495	0.154 p
RR prevdis	5.4792	3.6459	8.234	0.000 p

Predictors: estrogen, prevdis, estrogen*prevdis

Measure	Value	95% CI		Stat. Significance
OR estrogen	.3524	.0845	1.4695	0.152
OR prevdis	5.978	3.8687	9.2371	0.000
OR estroxprev	.888	.0752	10.4928	0.925
RR estrogen	.3566	.0864	1.4724	0.154
RR prevdis	5.484	3.6317	8.2810	0.000
RR estroxprev	.9423	.0849	10.4545	0.961

Predictors: estrogen, prevdis, age

Measure	Value	95% CI		Stat. Significance
OR estrogen	.4272	.1323	1.3788	0.155
OR prevdis	5.061	3.2751	7.8211	0.000
OR age	1.097	1.0617	1.1338	0.000

```
 RR estrogen   .3046       .0838    1.1073    0.071
 RR  prevdis  3.749       2.0714    6.7844    0.000
 RR      age  1.055       1.0088    1.1037    0.019


   confounding auxiliary data
Predictors: prevdis for cvddeath4
   OR  6.29      4.106    9.635  0.000
   RR  5.78      3.859    8.669  0.000

Predictors: prevdis for estrogen
   OR  .3850     .26146   .5670    0.000
   RR  .4189     .29140   .6023    0.000

Predictors: age for cvddeath4
  OR  1.120      1.085    1.155   0.000
  RR  1.102      1.060    1.146   0.000

Predictors: age for prevdis
   OR  1.063      1.046    1.0804     0.000
   RR  1.046      1.034    1.0573     0.000

Predictors: age for estrogen
   OR   .9142     .8914    .9376     0.000
   RR   .9293     .9107    .9484     0.000
```

We are interested in measuring any association between estrogen use at any time prior to study enrollment (estrogen==1) and CVD death within 4 years using three different measures of association. For each measure, perform these analyses:

a. Provide complete statistical inference regarding such an association. (Include point estimates, confidence intervals, and a p value, along with a full interpretation of those quantities.)

b. Is there evidence in the dataset that any such effect is modified by a history of prior CVD (as measured by variable prevdis)? Provide results of a statistical analysis in support of your answer.

c. Suppose we just want to ignore any such effect modification. Is there evidence in the dataset that any estrogen-CVD mortality association is confounded by a history of prior CVD? Provide results of a statistical analysis in support of your answer.

d. Provide complete statistical inference regarding an association between estrogen and CVD mortality after adjustment for a prior history of CVD.

e. Is there evidence in the dataset that the prior disease adjusted analysis of an association between estrogen-CVD mortality is further confounded by age? Provide results of a statistical analysis in support of your answer.

f. Provide complete statistical inference regarding an association between estrogen and CVD mortality after adjustment for age and any prior history of CVD.

1. Measure: Risk difference (RD).
Unfortunately, I was only able to obtain numerical RD's from tabular data. For the results I got, it appears that prior estrogen use was protective for cvddeath4. No further time to investigate.

2. Measure: odds ratio (OR)
a. Prior estrogen use provides a statistically significant protective effect, with an odds ratio of .250 [.07865 .7945]

b. Prior CVD disease does not appear to modify the effect. The interaction term "estrogen*prevdis" is not statistically significant (p=.925), so we would accept the null hypothesis that the coefficient of this term is 0, and a modification effect does not occur.

c. Prior CVD disease also does appear to be a confounder for the estrogen-cvddeath4 association. The values of the regression coefficient (OR) in the model have changed from .250 to .338, and both cvddeath4 and prior estrogen use appear to be associated with prevdis (OR's 6.29 and .385). The .385 value is not a strong association for estrogen, and scientifically it seems unlikely that prevdis would prompt estrogen use, so perhaps prevdis is a precision variable for the estrogen-cvddeath4 association.

d. When adjusted for prevdis, the estrogen-cvddeath4 association seems to break down, as the association between prevdis-cvddeath4 (OR=5.95) overwhelms it in the regression model. Indeed the coefficient OR for estrogen-cvddeath4 (.338) is no longer statistically significant (p=.068), so we would not reject the null hypothesis that this coefficient in the regression is 0.

e. Age does appear to be a confounding variable in the adjusted model associating cvddeath4 with estrogen use and prevdis. Including age in the model does appear to change the coefficients in the model (.352->.427 and 5.95->5.06). Age also appears to be associated with both estrogen use, prevdis, and cvddeath4, as shown in the auxiliary confounding data.

f. When adjusted for age and prior cvd disease, the association between estrogen and cvddeath4 is not very strong (OR=.42) and with a p value of .155, we would likely accept the null hypothesis that there is not such an association.

3. Measure: risk ratio (RR)

For all the questions above, the RR measure of association leads to the same conclusions. I am out of time to elaborate further, but the table illustrates this argument.

4. Of the three measures of association used above, how similar were the conclusions? What are the relative advantages and disadvantages of the three?
I was not able to make much progress on using the RD, and so I must conclude it is not as useful as the OR or RR. Both of these latter associations are quite consistent with each other, and either could be used successfully. I believe overall the OR is the more versitile measure, since it can be generated buy logistic regression, which appears to be more robust than the glm techniques.