Exercise 1.2

1. The following table presents the association between an exposure and outcome in a cohort of 10,000 individuals.

	Outcome	
	Yes	No
Exposed	230	4885
Unexposed	106	4779

Using the on-line calculator at Openepi.com or direct/immediate commands in statistical software (such as the immediate commands in Stata), verify the following estimates and 95% confidence intervals:

$$RR = 2.07 \ (1.65, 2.60)$$
$$OR = 2.12 \ ((1.68, 2.68)$$

2. (i) Using the Openepi.com online calculator, and the following three tables describing the 3505 individuals older than 41 in the Framingham teaching cohort data, find the crude OR, the three separate ORs, and the Mantel-Haenszel OR.

Age 42–51	Diabetes			Age 52–61	Diabetes			Age 62–71	Diabetes		
	Yes	No			Yes	No			Yes	No	
Yes CHD	1	32	33		4	74	78		7	69	76
No CHD	28	1553	1581		47	1221	1268		29	440	469
	29	1585	1614		51	1295	1346		36	509	545

(ii) Is there evidence of heterogeneity in the ORs in the three age categories? In other words, is it reasonable to assume they are homogenous, and report the Mantel-Haenszel adjusted OR? Why/why not?

3. The following table is taken from Table 4 of "Chlordecone exposure, length of gestation and risk of preterm birth" by Kadhel et al, *American Journal of Epidemiology*, 2014 Mar 1;179(5):536-44. doi: 10.1093/aje/kwt313.PMID: 24401561

Chlordecone level µg/L	Total births	Preterm births
< 0.14	163	16
0.14-0.28	165	21
0.29-0.51	162	23
0.52-0.97	165	30
>=0.98	163	25

These summary data are provided in the Excel sheet **Ex 1.2 chlordecone.xls**, which has 4 columns (category: 1,2,3,4,5; level as above; preterm 1 (preterm) ,0 (non-preterm; count). The summary data can be used by any command that allows you to specify a frequency option (e.g. *fweights* in Stata) or you can modify the data to have a record for each individual (e.g. using the "expand" command in Stata)

(i)   Conduct a Chi-squared test of the association between chlordecone level and preterm birth.

(ii)   Test for evidence of a trend for increasing risk with increasing exposure, using the category numbers (1 to 5) as scores. The test for trend may provide the separate ORs in the categories, in which case you will note that all except one are non-significant, yet the trend you found is borderline significant.

(iii)   Explain and compare the conclusion in part (i) and (ii) after obtaining the separate ORs if necessary.

4. The following table presents the cross-tabulation of alcohol intake and lung cancer from *Zang E and Wynder E. Preventive Med 3, 359-370,2001:*

	Lung Cancer	
Alcohol intake	Yes	No
< 1	1090	3976
1-3.9	806	2244
4-6.9	378	783
7+	679	1166

(i)   Test for a trend in risk with alcohol level using simple categories and/or scores for alcohol intake.

(ii)   The following tables present the association between the binary exposure (heavy/light drinker) and cancer, stratified by smoking status:

	Smoker Lung Cancer			Non-smoker Lung Cancer		
	Yes	No		Yes	No	
Heavy Drinker	786	665	1451	271	1284	1555
Light Drinker	702	591	1293	1194	5629	6823
All	1488	1256	2744	1465	6913	8378

Use the appropriate four numbers from this information in OpenEpi or a statistical package to calculate the OR for the association between smoking and lung cancer. Would you expect this association to be confounded by drinking? Why/why not?

(iii)   These data are available as **"alcohol_lungcancer_smoking.dta"**. Open the dataset, display the table in (i) above and request the test of trend.

(iv)   For the same data, find the OR for the association between smoking and lung cancer. Check if it is confounded by alcohol and compare to your expectation in (ii)

HINTS

Stata immediate commands csi (cross-sectional data) and cci (case-control data)

Test for trend
OpenEpi: Dose-Response command
Stata: tabodds command