Answer Key

> attach(fram)

**Question #1**

Are smokers at higher risk of death over follow-up than non-smokers?  Test this with a chi-square test, reporting the proportion of smokers and non-smokers who have died over 20 years of follow-up, the risk ratio, the value of the chi-square statistic, degrees of freedom, and p-value.  Summarize your conclusions.

> table(CURSMOKE,ANYDEATH)
 ANYDEATH
CURSMOKE   0    1
 0 178  57
 1 197  68
> prop.table(table(CURSMOKE,ANYDEATH),1)
 ANYDEATH
CURSMOKE         0         1
 0 0.7574468 0.2425532
 1 0.7433962 0.2566038

> chisq.test(table(CURSMOKE,ANYDEATH),correct = FALSE)

  Pearson's Chi-squared test
data:  table(CURSMOKE, ANYDEATH)
X-squared = 0.1311, df = 1, p-value = 0.7173

**Question #2**

Are smokers at higher risk of death over follow-up than non-smokers?  Find the risk ratio of death for smokers vs. non-smokers, and the 95% confidence interval for this risk ratio (remember that the orientation of the table matters.)

> RRtable<-matrix(c(178,197,57,68),nrow=2,ncol=2)
> RRtable
 [,1] [,2]
[1,]  178   57
[2,]  197   68

> riskratio.wald(RRtable)

$data
 Outcome
 Predictor  Disease1 Disease2 Total
 Exposed1       178       57   235
 Exposed2       197       68   265
 Total          375      125   500

$measure
 risk ratio with 95% C.I.
 Predictor  estimate     lower    upper
 Exposed1 1.000000        NA
 Exposed2 1.057928 0.7797999 1.435255

$p.value
 two-sided
 Predictor  midp.exact fisher.exact chi.square
 Exposed1           NA
 Exposed2  0.7194548     0.756695  0.7172542

$correction
[1] FALSE

attr(,"method")
[1] "Unconditional MLE & normal approximation (Wald) CI

**Comment:** 25.7% of Current smokers at baseline died over the 20 year follow up compared to 24.3% in those who were not current smokers at baseline. The RR=1.06 with 95% confidence interval 0.78 t0 1.44. This is not significant because it includes the null value of RR=1. The chi-square statistic = 0.13 with 1 degree of freedom, and the p-value=0.72 (also indicating that the difference was not significant.

**Question 3**

Is there an association between coffee consumption and death over the 20-year follow-up? Test this with a chi-square statistic, reporting the proportion who have died in each category of coffee consumption, the value of the test statistic, degrees of freedom, and p-value. Summarize your conclusions.

> table(COFFEE,ANYDEATH)
 ANYDEATH
COFFEE  0  1
 0 22  9
 1 50 30
 2 91 32
 3 90 26
 4 51 10
 5 20  5
 6 51 13

> prop.table(table(COFFEE,ANYDEATH),1)

 ANYDEATH
COFFEE         0         1
 0 0.7096774 0.2903226
 1 0.6250000 0.3750000
 2 0.7398374 0.2601626
 3 0.7758621 0.2241379
 4 0.8360656 0.1639344
 5 0.8000000 0.2000000
 6 0.7968750 0.2031250

> chisq.test(table(COFFEE,ANYDEATH),correct=FALSE)

 Pearson's Chi-squared test
data:  table(COFFEE, ANYDEATH)
X-squared = 10.9102, df = 6, p-value = 0.09119

There were small differences in ANYDEATH among those who drank different amounts of coffee, but the differences did not achieve statistical significance.

**Question 4**

Do those who develop heart disease have a higher risk of death over follow-up? What percent of those with and without heart disease die over follow-up? Test through a chi-square statistic, reporting the value of the test statistic, degrees of freedom, and p-value. Summarize your conclusions.

> table(HEARTDIS,ANYDEATH)
 ANYDEATH
HEARTDIS  0   1
 0 344  87
 1 31  38

> prop.table(table(HEARTDIS,ANYDEATH),1)

 ANYDEATH
HEARTDIS         0         1
 0 0.7981439 0.2018561
 1 0.4492754 0.5507246

> chisq.test(table(HEARTDIS,ANYDEATH),correct=FALSE)

 Pearson's Chi-squared test
data:  table(HEARTDIS, ANYDEATH)
X-squared = 38.6081, df = 1, p-value = 5.18e-10

**Question 5**

Do those who develop heart disease have a higher risk of death over follow-up? Find the risk ratio of death for with vs. without heart disease, and the 95% confidence interval for this relative risk (remember that the orientation of the table matters when finding a RR). Given this confidence interval, do those who develop heart disease have significantly higher risk of death? Explain

> RRtableHRT<-matrix(c(344,31,87,38),nrow = 2,ncol = 2)
> RRtableHRT

 [,1] [,2]
[1,]  344   87
[2,]   31   38

> riskratio.wald(RRtableHRT)

$data
 Outcome
 Predictor  Disease1 Disease2 Total
 Exposed1      344       87   431
 Exposed2       31       38    69
 Total         375      125   500

$measure
 risk ratio with 95% C.I.
 Predictor  estimate    lower    upper
 Exposed1 1.000000        NA
 Exposed2 2.728303 2.053762 3.624389

$p.value
two-sided
Predictor    midp.exact fisher.exact   chi.square
Exposed1             NA
Exposed2 7.485284e-09 6.667311e-09 5.180315e-10

$correction
[1] FALSE

attr(,"method")
[1] "Unconditional MLE & normal approximation (Wald) CI"

***Comment:*** The risk ratio is 2.7 and the 95% confidence interval is (2.05, 3.62). The 95% confidence interval does not contain the null value of RR=1, so there is a statistically significant increased risk of ANYDEATH in those with HEARTDIS. Those with HEARTDIS had 2.7 times the risk of ANYDEATH during the 20 year follow up period.