

# Models In Epidemiology And Biostatistics

## Gordon Hilton Fick

### Stratified Analysis with R

I have developed an R function called strat. This function was constructed starting with the R function mhor which is in the epiDisplay package.

If you have a dichotomous outcome : disease [0 : no dis, 1 : dis], a dichotomous exposure : exposure [0 : not exp, 1 : exp] and strata with levels: 0, 1, 2, 3, ....

strat(disease, exposure, strata) gives a stratified analysis for the OR, the RR and the RD.

If you have more than one stratification variable, say age [0 : Y, 1 : O] and gender [0 : F, 1 : M], you can create a single variable using the R function interaction.

a\_g <- interaction(age, gender) creates a factor a\_g with levels 0.0, 0.1, 1.0 and 1.1 corresponding to YF, YM, OF and OM respectively.

then you can use strat(disease, exposure, a\_g) which gives the 3 stratified analyses with the 4 strata.

Here are 2 examples :

Using the kalbfleisch data;

```
> strat(suc,tr,surg)
      tr exposed not exposed
surg suc
1  cases      100        5
   controls    900       95
2  cases       95      500
   controls     5      500
```

```
surg 1
      exposed not exposed odds est
cases      100.0      5.00 20.000000
controls   900.0     95.00  9.473684
risk est    0.1      0.05      NA
```

```
surg 2
      exposed not exposed odds est
cases       95.00     500.0    0.19
controls     5.00     500.0    0.01
risk est     0.95       0.5      NA
```

```
Stratified analysis for OR by surg
      OR est      lower      upper      p-value
surg 1    2.10998    0.84347    6.80312    1.5029e-01
surg 2   18.97140    7.75038   60.25999    2.9193e-21
crude     0.25404    0.20775    0.30983    1.4540e-46
adjusted   8.14286    4.34278   15.26814    1.7661e-16
OR Homogeneity test, chi-squared 1 df = 11.56 , p-value = 0.000672723
```

```
Stratified analysis for RR by surg
      RR est      lower      upper      p-value
surg 1    2.00000    0.83428    4.7945    1.5029e-01
surg 2    1.90000    1.75994    2.0512    2.9193e-21
crude     0.38614    0.33484    0.4453    1.4540e-46
adjusted   1.90909    1.71543    2.1246    1.7661e-16
```

RR Homogeneity test, chi-squared 1 df = 0.03 , p-value = 0.8723836

Stratified analysis for RD by surg

	RD est	lower	upper	p-value
surg 1	0.05000	0.0034122	0.096588	1.5029e-01
surg 2	0.45000	0.3972264	0.502774	2.9193e-21
crude	-0.28182	-0.3189200	-0.244716	1.4540e-46
adjusted	0.25000	0.2137274	0.286273	1.7661e-16

RD Homogeneity test, chi-squared 1 df = 125.97 , p-value = 3.123441e-29

Using the dis expo age gen data ;

```
> a_g<-interaction(age,gender)
```

```
> strat(dis,expo,a_g)
```

	expo	exposed	not exposed
a_g dis			
0.0 cases		103	15
controls		897	285
1.0 cases		157	16
controls		843	284
0.1 cases		93	251
controls		47	249
1.1 cases		57	249
controls		17	251

a\_g 0.0

	exposed	not exposed	odds est
cases	103.000	15.00	6.866667
controls	897.000	285.00	3.147368
risk est	0.103	0.05	NA

a\_g 1.0

	exposed	not exposed	odds est
cases	157.000	16.00000000	9.81250
controls	843.000	284.00000000	2.96831
risk est	0.157	0.05333333	NA

a\_g 0.1

	exposed	not exposed	odds est
cases	93.0000000	251.000	0.3705179
controls	47.0000000	249.000	0.1887550
risk est	0.6642857	0.502	NA

a\_g 1.1

	exposed	not exposed	odds est
cases	57.0000000	249.000	0.22891566
controls	17.0000000	251.000	0.06772908
risk est	0.7702703	0.498	NA

Stratified analysis for OR by a\_g

	OR est	lower	upper	p-value
a_g 0.0	2.18063	1.23749	4.10422	4.0526e-03
a_g 1.0	3.30343	1.92967	6.02645	9.0533e-07
a_g 0.1	1.96094	1.30532	2.97419	7.5981e-04
a_g 1.1	3.37321	1.87203	6.36610	9.4538e-06
crude	0.45764	0.39288	0.53277	7.3142e-25
adjusted	2.53557	1.98162	3.24437	3.6252e-14

OR Homogeneity test, chi-squared 3 df = 3.85 , p-value = 0.2783922

Stratified analysis for RR by a\_g

	RR est	lower	upper	p-value
a_g 0.0	2.0600	1.21730	3.48606	4.0526e-03
a_g 1.0	2.9437	1.78921	4.84328	9.0533e-07
a_g 0.1	1.3233	1.14285	1.53220	7.5981e-04
a_g 1.1	1.5467	1.32808	1.80137	9.4538e-06

```

crude      0.5580    0.49904  0.62391  7.3142e-25
adjusted   1.6446    1.45324  1.86112  3.6252e-14
RR Homogeneity test, chi-squared 3 df = 15.03 , p-value = 0.001794812

```

```

Stratified analysis for RD by a_g
      RD est      lower      upper      p-value
a_g 0.0      0.05300    0.021965   0.084035   4.0526e-03
a_g 1.0      0.10367    0.069682   0.137651   9.0533e-07
a_g 0.1      0.16229    0.072620   0.251951   7.5981e-04
a_g 1.1      0.27227    0.166882   0.377658   9.4538e-06
crude      -0.14669   -0.174871  -0.118509   7.3142e-25
adjusted    0.11246    0.087493   0.137428   3.6252e-14
RD Homogeneity test, chi-squared 3 df = 24.38 , p-value = 2.082897e-05

```

```

> strat(dis,expo,age)
      expo exposed not exposed

```

```

age dis
0  cases      196      266
   controls    944     534
1  cases      214     265
   controls    860     535

```

```

age 0
      exposed not exposed odds est
cases    196.0000000    266.0000 0.7368421
controls 944.0000000    534.0000 1.7677903
risk est  0.1719298      0.3325    NA

```

```

age 1
      exposed not exposed odds est
cases    214.0000000    265.0000 0.8075472
controls 860.0000000    535.0000 1.6074766
risk est  0.1992551      0.33125    NA

```

```

Stratified analysis for OR by age
      OR est      lower      upper      p-value
age 0      0.41701    0.33490   0.51860   6.8369e-16
age 1      0.50255    0.40491   0.62320   1.1615e-10
crude      0.45764    0.39288   0.53277   7.3142e-25
adjusted   0.45826    0.39466   0.53211   4.1204e-25
OR Homogeneity test, chi-squared 1 df = 1.5 , p-value = 0.2212831

```

```

Stratified analysis for RR by age
      RR est      lower      upper      p-value
age 0      0.51708    0.44026   0.60731   6.8369e-16
age 1      0.60152    0.51508   0.70247   1.1615e-10
crude      0.55800    0.49904   0.62391   7.3142e-25
adjusted   0.55870    0.49970   0.62465   4.1204e-25
RR Homogeneity test, chi-squared 1 df = 1.76 , p-value = 0.1845693

```

```

Stratified analysis for RD by age
      RD est      lower      upper      p-value
age 0     -0.16057   -0.19988  -0.121258   6.8369e-16
age 1     -0.13199   -0.17242  -0.091567   1.1615e-10
crude     -0.14669   -0.17487  -0.118509   7.3142e-25
adjusted  -0.14646   -0.17465  -0.118271   4.1204e-25
RD Homogeneity test, chi-squared 1 df = 0.99 , p-value = 0.3205582

```

```

> strat(dis,expo,gender)
      expo exposed not exposed

```

```

gender dis
0  cases      260      31
   controls   1740     569
1  cases      150     500
   controls     64     500

```

gender 0				
	exposed	not exposed	odds est	
cases	260.00	31.00000000	8.387097	
controls	1740.00	569.00000000	3.057996	
risk est	0.13	0.05166667		NA

gender 1				
	exposed	not exposed	odds est	
cases	150.0000000	500.0	0.300	
controls	64.0000000	500.0	0.128	
risk est	0.7009346	0.5		NA

Stratified analysis for OR by gender				
	OR est	lower	upper	p-value
gender 0	2.74183	1.85907	4.16854	1.4423e-08
gender 1	2.34216	1.68951	3.27441	6.8953e-08
crude	0.45764	0.39288	0.53277	7.3142e-25
adjusted	2.51945	1.97161	3.21950	4.2236e-14

OR Homogeneity test, chi-squared 1 df = 0.39 , p-value = 0.5330144

Stratified analysis for RR by gender				
	RR est	lower	upper	p-value
gender 0	2.5161	1.75357	3.61029	1.4423e-08
gender 1	1.4019	1.25931	1.56056	6.8953e-08
crude	0.5580	0.49904	0.62391	7.3142e-25
adjusted	1.6391	1.44921	1.85397	4.2236e-14

RR Homogeneity test, chi-squared 1 df = 13.58 , p-value = 0.0002287772

Stratified analysis for RD by gender				
	RD est	lower	upper	p-value
gender 0	0.078333	0.055291	0.10138	1.4423e-08
gender 1	0.200935	0.132208	0.26966	6.8953e-08
crude	-0.146690	-0.174871	-0.11851	7.3142e-25
adjusted	0.112217	0.087219	0.13722	4.2236e-14

RD Homogeneity test, chi-squared 1 df = 14.71 , p-value = 0.0001254938

Some of the background

For the crude and stratum specific tables, the p-values presented here are based on Fisher's Exact Test [FET].

For the crude and stratum specific tables, the estimates of the OR are based on the conditional likelihood and so they will be consistent with FET.

The p-values for the 'adjusted' estimates are based based on approximate methods.

The adjusted estimates are based on weighted averages of the stratum specific estimates. The adjusted OR estimates are quite well established and can be compared to the estimates obtained from maximum likelihood with logistic regression. The adjusted RR estimates and adjusted RD estimates are perhaps less accepted and both are the subject of current methodological research.