Biostatistics III: Survival analysis for epidemiologists

Department of Medical Epidemiology and Biostatistics
Karolinska Institutet
Stockholm, Sweden

http://www.biostat3.net/

Extra lecture on modelling

- This lecture covers general theory of modelling and effect parameterisation (not specific to survival analysis).

- Binary, categorical and continuous effects

- The main effects model

- Interaction models

- Parameterisation of effects to obtain the effect (group comparison) of interest

- As an illustrative example, we will use the diet data and use Poisson regression to model the rate, assuming constant (average) rates over time. So we will ignore time for now.

Binary and categorical exposure variables

- The energy intake variable `hieng` has two levels.

- The variable `eng3`, created below, has 3 levels. Here are the crude rates and IRRs.

```
. egen eng3=cut(energy),at(1500,2500,3000,4500)
```

```
     energy    eng3        Rate
-------------------------------
   1500-2499   1500  |   16.90
   2500-2999   2500  |   10.91    IRR (2500 vs 1500) = 10.91/16.90 = 0.65
   3000-4500   3000  |    4.88    IRR (3000 vs 1500) =  4.88/16.90 = 0.29
-------------------------------
```

- To include `eng3` in regression commands we need to use indicator variables for the 3 levels.

eng3	$X1$	$X2$	$X3$
1500	1	0	0
2500	0	1	0
3000	0	0	1

- We create dummy variables X1, X2, X3 for exposure eng3.

```
. tabulate eng3, generate(X)
```

- We set exposure level 1500 as reference by omitting X1 from the model.

```
. poisson chd X2 X3, e(y) irr

  chd |      IRR  [95% Conf. Interval]
------+--------------------------------
   X2 |    .6452  .3388815    1.228561
   X3 |    .2886  .1235342    .6744495
_cons |    .0169  .0103547    .0275892
```

- The variable ($X1$) that indicates the category with the lowest energy intake (1500) is omitted, meaning this is the reference category.

- In terms of the parameters

$$\ln(\lambda) = \beta_0 + \beta_2 X_2 + \beta_3 X_3$$

energy	X	ln(rate)	rate
1500	X_1	β_0	$\exp(\beta_0)$
2500	X_2	$\beta_0 + \beta_2$	$\exp(\beta_0 + \beta_2) = \exp(\beta_0)\exp(\beta_2)$
3000	X_3	$\beta_0 + \beta_3$	$\exp(\beta_0 + \beta_3) = \exp(\beta_0)\exp(\beta_3)$

$$\text{IRR}_{X_2/X_1} = \exp(\beta_2) = \frac{\exp(\beta_0)\exp(\beta_2)}{\exp(\beta_0)}$$

$$\text{IRR}_{X_3/X_1} = \exp(\beta_3) = \frac{\exp(\beta_0)\exp(\beta_3)}{\exp(\beta_0)}$$

Automatic generation of indicators using factor variables

```
. poisson chd i.eng3, e(y) irr
-------------------------------------------
     chd |       IRR    [95% Conf. Interval]
---------+---------------------------------
    eng3 |
  2500   |   .6452416  .3388815    1.228561
  3000   |   .2886479  .1235342    .6744495
```

- i. tells Stata that eng3 should be treated as a categorical variable
- The baseline is, by default, the first level (1500-2499), but this can be changed to (say) the third level (3000–4500) with

```
. poisson chd ib3000.eng3, e(y) irr
```

Metric (continuous) exposure variables

- The effect of `energy` on the CHD rate, when energy is measured as a continuous variable

```
. poisson chd energy , e(y) irr
-----------------------------------------
   chd |         IRR  [95% Conf. Interval]
-------+---------------------------------
energy |      .99885  .9981367    .9995637
```

- For each 1 unit increase in energy intake, the CHD rate is reduced by 0.1%. The units of energy are kcals per day.

- However, a 1 unit increase may not be clinically relevant. We may want to rescale the continuous variable.

```
. summarize energy
Variable |  Obs      Mean    Std. Dev.      Min       Max
---------+--------------------------------------------------
  energy |  337    2828.9    441.8       1748.4   4395.8
```

- To get the IRR for an increase of, say, 100 units

```
. gen energy100=energy/100
. poisson chd energy100, e(y) irr
----------------------------------------
      chd |        IRR  [95% Conf. Interval]
----------+-----------------------------
energy100 |   .8913034  .8298593    .9572968
```

- The estimated IRR is $0.99885^{100} = 0.8913$. That is, for each 100 unit increase in energy intake, we estimate that the CHD rate is reduced by 11%.

The main effects model — constant effect over strata

- If the true effect of exposure does not vary across strata of another variable we can use a main effects model.

- For example, if the effect of high energy intake is the same in all occupations, we can estimate an energy effect that is the same for all occupations.

- Thus, if the estimates of high energy differ only randomly over occupation level, we can consider a model in which the true effect of high energy is constant over occupation level, i.e. no interaction.

- This allows us to combine the information from different strata to yield a single estimate of exposure effect.

- This combined estimate of the effect we call *the main effect*, which is then *controlled for* the stratifying (confounding) variable(s).

- The effect of exposure is assumed to be the same in all levels of other variables in a main effects model.

- Statistical tests for the presence of confounding are not available, although there are statistical tests for effect modification.

Main effects model using Poisson regression

```
. poisson chd i.hieng i.job, e(y) irr
--------------------------------------------------
         chd |        IRR   [95% Conf. Interval]
-------------+------------------------------------
     1.hieng |   .5247666   .290225    .9488499
             |
         job |
          2  |   1.358442   .6282879    2.937133
          3  |   .8843023   .4322823    1.808981
             |
       _cons |   .0132337   .0071823    .0243837
```

- Occupation is measured in job variable (1=driver, 2=conductor, 3=banker).
- The Stata Poisson regression command makes no distinction between the exposure variable and the confounding variable.

- The first number reported is the effect of `hieng` adjusted for `job`, and the next two are the effects of `job` adjusted for `hieng`.

Models and parameters in Poisson regression

- In the Poisson regression model we estimated 4 parameters. One parameter (the intercept) is a log rate and the other three are log incidence rate ratios.

- The model is $\ln(\lambda) = \beta_0 + \beta_1 \text{hieng} + \beta_2 \text{cond} + \beta_3 \text{bank}$

- The parameters are

$$\begin{aligned}
\exp(\beta_0) &= \text{rate at reference level of all covariates, i.e. driver with low energy} \\
\exp(\beta_1) &= \text{rate ratio (comparing high vs low energy)} \\
\exp(\beta_2) &= \text{rate ratio (comparing conductors vs drivers)} \\
\exp(\beta_3) &= \text{rate ratio (comparing bank vs drivers)}
\end{aligned}$$

Parameters estimates with and without the `irr` option

```
. poisson chd i.hieng i.job, e(y)
         chd |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.hieng |  -.6448017   .3021979    -2.13   0.033    -1.237099   -.0525046
         job |
         2   |   .3063385   .3934232     0.78   0.436    -.4647568    1.077434
         3   |  -.1229563     .36517    -0.34   0.736    -.8386764    .5927639
       _cons |  -4.324988   .3118157   -13.87   0.000    -4.936136   -3.713841

. poisson chd i.hieng i.job, e(y) irr
         chd |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.hieng |   .5247666   .1585834    -2.13   0.033      .290225    .9488499
         job |
         2   |   1.358442   .5344426     0.78   0.436     .6282879    2.937133
         3   |   .8843023   .3229207    -0.34   0.736     .4322823    1.808981
       _cons |   .0132337   .0041265   -13.87   0.000     .0071823    .0243837
```

- The irr option exponentiates the coefficients.

- From the model, the estimated rate for each combination of explanatory variables can be formulated as a function of the baseline rate λ and the three incidence rate ratios. The baseline is the reference group of all variables (drivers with low energy).

- The model is $\ln(\lambda) = \beta_0 + \beta_1\text{hieng} + \beta_2\text{cond} + \beta_3\text{bank}$

- Which on the rate scale is $\lambda = \exp(\beta_0 + \beta_1\text{hieng} + \beta_2\text{cond} + \beta_3\text{bank})$

- These are the rates, λ:

job	hieng=0	hieng=1
1=driv	$\exp(\beta_0)$	$\exp(\beta_0 + \beta_1)$
2=cond	$\exp(\beta_0 + \beta_2)$	$\exp(\beta_0 + \beta_1 + \beta_2)$
3=bank	$\exp(\beta_0 + \beta_3)$	$\exp(\beta_0 + \beta_1 + \beta_3)$

- The estimated incidence rate for drivers with a low energy intake (ref group) is $\exp(-4.325) = 0.0132$ events/person-year.

- The estimated incidence rate for conductors with a high energy intake is $\exp(\beta_0 + \beta_1 + \beta_2) = \exp(\beta_0) \times \exp(\beta_1) \times \exp(\beta_2) = 0.0132 \times 1.358 \times 0.525 = 0.0094$ events/person-year.

- To get the IRRs (with reference group 'drivers with low energy intake') we simply divide all cells with the baseline rate $\lambda = \exp(\beta_0)$

job	hieng=0	hieng=1
1=driv	$\exp(\beta_0)/\exp(\beta_0) = 1$	$\exp(\beta_0 + \beta_1)/\exp(\beta_0) = \exp(\beta_1)$
2=cond	$\exp(\beta_0 + \beta_2)/\exp(\beta_0) = \exp(\beta_2)$	$\exp(\beta_0 + \beta_1 + \beta_2)/\exp(\beta_0) = \exp(\beta_1 + \beta_2)$
3=bank	$\exp(\beta_0 + \beta_3)/\exp(\beta_0) = \exp(\beta_3)$	$\exp(\beta_0 + \beta_1 + \beta_3)/\exp(\beta_0) = \exp(\beta_1 + \beta_3)$

- If we put the estimates from the Stata output into our table of parameters, we get the IRRs

job	hieng=0	hieng=1
driv	1.0	0.52
cond	1.36	$0.52 \times 1.36 = 0.71$
bank	0.88	$0.52 \times 0.88 = 0.46$

- Compared to drivers with low energy intake, the conductors with high energy have a 29% lower incidence of CHD (IRR=0.71).

- Effect of being high vs low energy is 0.52 (48% lower).

- Effect of being conductor vs driver is 1.36 (36% higher).

- Effect of being banker vs driver is 0.88 (12% lower).

0.0132 x 1.36 $\exp(\beta_0+\beta_2)$

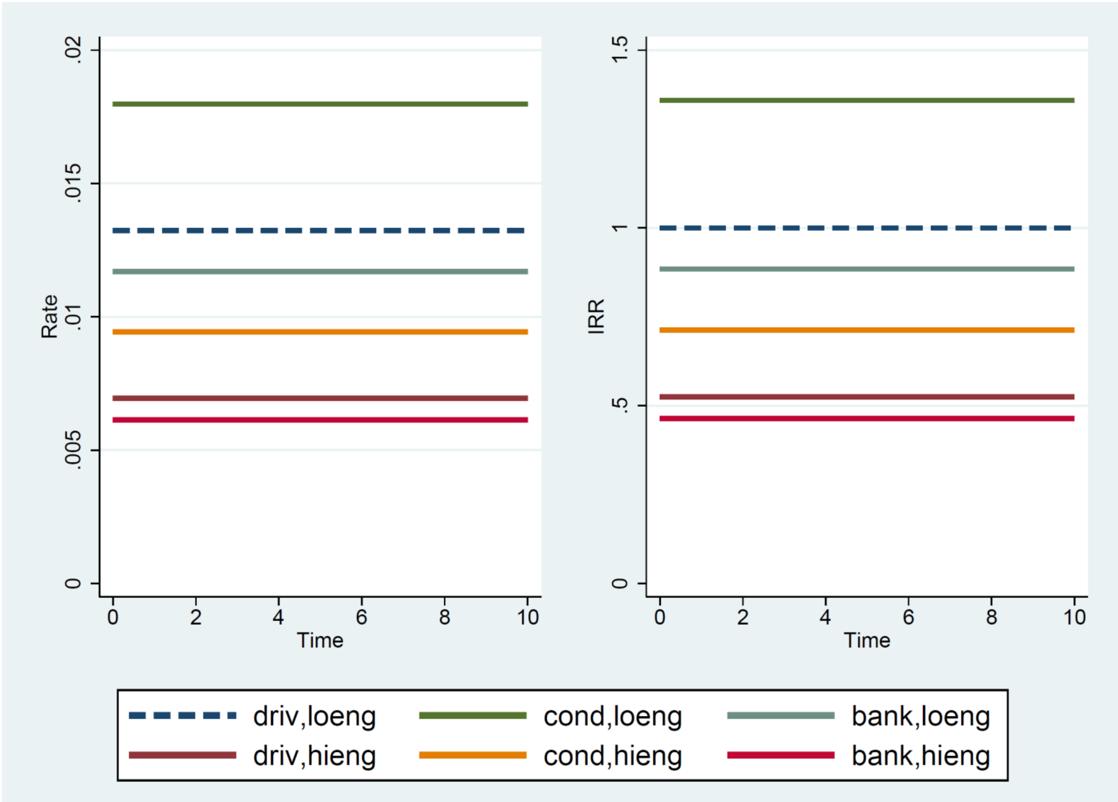
0.0132 $\exp(\beta_0)$

0.0132 x 0.88 $\exp(\beta_0+\beta_3)$

0.0132 x 0.52 x 1.36 $\exp(\beta_0+\beta_1+\beta_2)$

0.0132 x 0.52 $\exp(\beta_0+\beta_1)$

0.0132 x 0.52 x 0.88 $\exp(\beta_0+\beta_1+\beta_3)$



$\exp(\beta_2)$ **1.36**

1.0 (ref)

$\exp(\beta_3)$ **0.88**

$\exp(\beta_1+\beta_2)$ **0.52 x 1.36=0.71**

$\exp(\beta_1)$ **0.52**

$\exp(\beta_1+\beta_3)$ **0.52 x 0.88=0.46**

- If we want to get the effect of energy separately in all levels of occupation, we simply change the reference rate to be low energy in all levels.

job	hieng=0	hieng=1
1=driv	$\exp(\beta_0)/\exp(\beta_0) = 1$	$\exp(\beta_0 + \beta_1)/\exp(\beta_0) = \exp(\beta_1)$
2=cond	$\exp(\beta_0 + \beta_2)/\exp(\beta_0 + \beta_2) = 1$	$\exp(\beta_0 + \beta_1 + \beta_2)/\exp(\beta_0 + \beta_2) = \exp(\beta_1)$
3=bank	$\exp(\beta_0 + \beta_3)/\exp(\beta_0 + \beta_3) = 1$	$\exp(\beta_0 + \beta_1 + \beta_3)/\exp(\beta_0 + \beta_3) = \exp(\beta_1)$

- We see that the main effect models give the same effect of energy, $\exp(\beta_1)$, regardless of job level.

job	hieng=0	hieng=1
1=driv	1.0	$\exp(\beta_1)$
2=cond	1.0	$\exp(\beta_1)$
3=bank	1.0	$\exp(\beta_1)$

- This is what a main effect model assumes, i.e. the effect of exposure is the same in all levels of another variable.

- Similarly, if we want to get the effect of job in all levels of energy intake, we change the reference rate to the driver in both levels of energy intake.

job	hieng=0	hieng=1
1=driv	$\exp(\beta_0)/\exp(\beta_0) = 1$	$\exp(\beta_0 + \beta_1)/\exp(\beta_0 + \beta_1) = 1$
2=cond	$\exp(\beta_0 + \beta_2)/\exp(\beta_0) = \exp(\beta_2)$	$\exp(\beta_0 + \beta_1 + \beta_2)/\exp(\beta_0 + \beta_1) = \exp(\beta_2)$
3=bank	$\exp(\beta_0 + \beta_3)/\exp(\beta_0) = \exp(\beta_3)$	$\exp(\beta_0 + \beta_1 + \beta_3)/\exp(\beta_0 + \beta_1) = \exp(\beta_3)$

- Similarly, the main effects of job are the same, $\exp(\beta_2)$ and $\exp(\beta_3)$, regardless of energy level.

job	hieng=0	hieng=1
1=driv	1.0	1.0
2=cond	$\exp(\beta_2)$	$\exp(\beta_2)$
3=bank	$\exp(\beta_3)$	$\exp(\beta_3)$

Effect modification

- If the true effect of exposure is not the same, but varies across strata of another variable there is said to be 'effect modification' (interaction) — the effect of exposure cannot then be represented by one IRR for all levels.

- For example, the effect of high energy may depend on occupation, so that drivers have a greater effect of high energy than conductors.

- Then we say that the effect is *modified by occupation*. There is an interaction between energy intake and occupation.

- Does `job` modify the effect of `hieng`? If we estimate the IRR of high vs low energy separately in all three job groups we get

```
Job        Effect of hieng
driver        0.41
conductor     0.66
bank          0.52
```

- The numbers represent the incidence rate ratios (comparing high to low energy intake) within each job category.
- If the effect of high energy is not modified by job then we would expect these to be similar.
- In the previous main effect model, we assumed (and estimated) the effect to be 0.52 in all job levels.

Interaction model using Poisson regression

- If we want to estimate separate effects of hieng in job levels, then four parameters are not enough.

- We have six combinations of energy and job level, hence we need six parameters to estimate six different rates.

- This is easily done by including interaction terms in the Poisson model.

- The model is
$\ln(\lambda) = \beta_0 + \beta_1 \text{hieng} + \beta_2 \text{cond} + \beta_3 \text{bank} + \beta_4 \text{hieng x cond} + \beta_5 \text{hieng x bank}$

- The parameters are

$$\begin{aligned}
\exp(\beta_0) &= \text{rate at reference level of all covariates, i.e. driver with low energy} \\
\exp(\beta_1) &= \text{rate ratio (comparing high vs low energy among drivers)} \\
\exp(\beta_2) &= \text{rate ratio (comparing conductors vs drivers among low energy)} \\
\exp(\beta_3) &= \text{rate ratio (comparing bank vs drivers among low energy)} \\
\exp(\beta_4) &= \text{interaction term (excess rate for cond vs driv among high energy)} \\
\exp(\beta_5) &= \text{interaction term (excess rate for bank vs driv among high energy)}
\end{aligned}$$

```
. poisson chd i.hieng##i.job, e(y) irr
         chd |        IRR   [95% Conf  Int]
-------------+------------------------
     1.hieng |   .4102648   .1235412    1.362438
             |
         job |
          2  |   1.136857   .4266828    3.029051
          3  |    .813427   .3325064    1.989927
             |
   hieng#job |
        1 2  |   1.596755   .3222813    7.911183
        1 3  |   1.261973   .2824452    5.638532
             |
       _cons |   .0144648   .0072338     .028924
```

- 0.41 is the effect of `hieng` when `job` is at its first level.
- 1.14 and 0.81 are the effects of `job` when `hieng` is at its first level.
- 1.60 and 1.26 are the interactions between `hieng` and `job`.

- 0.014 is the baseline rate in the reference level of both hieng and job.

Parameters for the interaction model

- The *fitted rates* (λ) from the interaction model can be expressed in terms of the model parameters as follows

job	hieng=0	hieng=1
1=driv	$\exp(\beta_0)$	$\exp(\beta_0 + \beta_1)$
2=cond	$\exp(\beta_0 + \beta_2)$	$\exp(\beta_0 + \beta_1 + \beta_2 + \beta_4)$
3=bank	$\exp(\beta_0 + \beta_3)$	$\exp(\beta_0 + \beta_1 + \beta_3 + \beta_5)$

- $\exp(\beta_4)$ and $\exp(\beta_5)$ are the interaction parameters. They measure deviations from the hypothesis of main effect of `hieng` in all `job` categories.

- The model is
$\ln(\lambda) = \beta_0 + \beta_1$hieng $+ \beta_2$cond $+ \beta_3$bank $+ \beta_4$hieng x cond $+ \beta_5$hieng x bank

- If we wish to tabulate the IRR, using the drivers with low energy intake as reference group, then we simply divided all cells with the reference rate.

job	hieng=0	hieng=1
1=driv	$\exp(\beta_0)/\exp(\beta_0) = 1$	$\exp(\beta_0+\beta_1)/\exp(\beta_0) = \exp(\beta_1)$
2=cond	$\exp(\beta_0+\beta_2)/\exp(\beta_0) = \exp(\beta_2)$	$\exp(\beta_0+\beta_1+\beta_2+\beta_4)/\exp(\beta_0) = \exp(\beta_1+\beta_2+\beta_4)$
3=bank	$\exp(\beta_0+\beta_3)/\exp(\beta_0) = \exp(\beta_3)$	$\exp(\beta_0+\beta_1+\beta_3+\beta_5)/\exp(\beta_0) = \exp(\beta_1+\beta_3+\beta_5)$

- If we add the Stata output into the table of IRRs we get

job	hieng=0	hieng=1
driv	1.0	0.41
cond	1.14	$0.41 \times 1.14 \times 1.60 = 0.75$
bank	0.81	$0.41 \times 0.81 \times 1.26 = 0.42$

- Compare to the estimates from main effects model

job	hieng=0	hieng=1
driv	1.0	0.52
cond	1.36	$0.52 \times 1.36 = 0.71$
bank	0.88	$0.52 \times 0.88 = 0.46$

$\exp(\beta_2)$ **1.36**

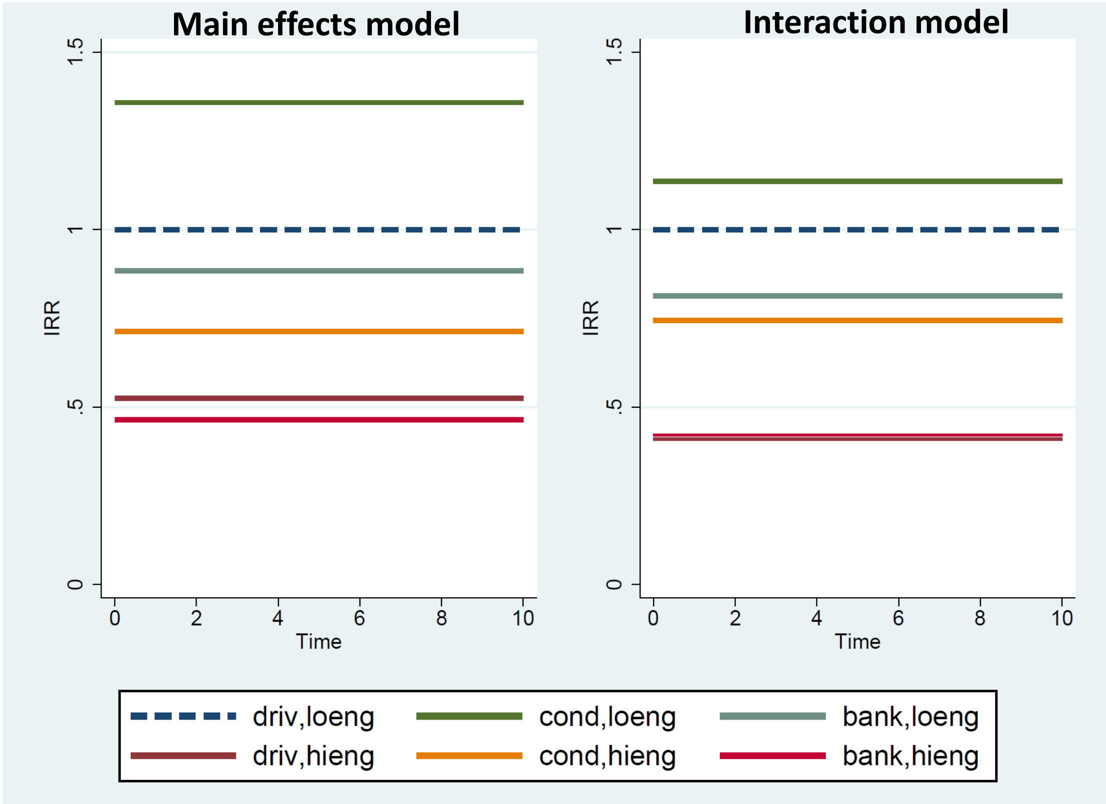
1.0 (ref)

$\exp(\beta_3)$ **0.88**

$\exp(\beta_1+\beta_2)$ **0.52 x 1.36=0.71**

$\exp(\beta_1)$ **0.52**

$\exp(\beta_1+\beta_3)$ **0.52 x 0.88=0.46**



$\exp(\beta_2)$ **1.14**

1.0 (ref)

$\exp(\beta_3)$ **0.81**

$\exp(\beta_1+\beta_2+\beta_4)$ **0.41 x 1.14 x 1.60=0.75**

$\exp(\beta_1+\beta_3+\beta_5)$ **0.41 x 0.81 x 1.26=0.42**

$\exp(\beta_1)$ **0.41**

Testing for interaction using Stata

- A test of interaction is simply to test if the excess (1.26 and 1.60) is equal to 1 (or 0 on the log scale).

```
. poisson chd i.hieng i.job i.hieng##i.job, e(y) irr
-------------------------------------------------
         chd |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+-----------------------------------
     1.hieng |   .4102648   .2512349    -1.45   0.146     .1235412    1.362438
             |
         job |
          2  |   1.136857   .5684285     0.26   0.798     .4266828    3.029051
          3  |    .813427   .3712769    -0.45   0.651     .3325064    1.989927
             |
   hieng#job |
        1 2  |   1.596755   1.303745     0.57   0.567     .3222813    7.911183
        1 3  |   1.261973   .9638479     0.30   0.761     .2824452    5.638532
             |
       _cons |   .0144648   .0051141   -11.98   0.000     .0072338     .028924
```

```
. testparm 1.hieng#2.job 1.hieng#3.job
  chi2(  2) =    0.33
Prob > chi2 =    0.8475
```

- No evidence of a statistically significant interaction.
- This is a so-called Wald test, which approximates the likelihood ratio test. We could also use a likelihood ratio test, where we compare the log-likelihoods from the main effects model and the interaction model.

Reparameterising the model to directly estimate the effect of exposure in each stratum

- We are often interested in the effect of the exposure (comparison between high and low energy intake) for each level of the modifier (job).

- We can divide the rates with different reference rates to obtain the effect of high energy in each level of job.

job	hieng=0	hieng=1
1=driv	$\exp(\beta_0)/\exp(\beta_0) = 1$	$\exp(\beta_0 + \beta_1)/\exp(\beta_0) = \exp(\beta_1)$
2=cond	$\exp(\beta_0 + \beta_2)/\exp(\beta_0 + \beta_2) = 1$	$\exp(\beta_0 + \beta_1 + \beta_2 + \beta_4)/\exp(\beta_0 + \beta_2) = \exp(\beta_1 + \beta_4)$
3=bank	$\exp(\beta_0 + \beta_3)/\exp(\beta_0 + \beta_3) = 1$	$\exp(\beta_0 + \beta_1 + \beta_3 + \beta_5)/\exp(\beta_0 + \beta_3) = \exp(\beta_1 + \beta_5)$

- This yields the following IRRs:

job	hieng=0	hieng=1
driv	1.0	$\exp(\beta_1)$
cond	1.0	$\exp(\beta_1 + \beta_4)$
bank	1.0	$\exp(\beta_1 + \beta_5)$

- We can reparameterise the model in Stata to directly estimate parameters of high energy, one for each job level.

How to make Stata produce stratified effects

- Instead of just one baseline rate, you need three baseline rates, λ (one for each job level). In addition, for each job level, you need an IRR for the energy effect (three rate ratios).

```
. poisson chd ibn.job i.hieng#i.job, e(y) irr nocons
         chd |        IRR   [95% Conf. Interval]
-------------+--------------------------------
         job |
          1  |   .0144648   .0072338    .028924
          2  |   .0164445   .0082238   .0328825
          3  |   .0117661   .0066821   .0207183
             |
   hieng#job |
        1 1  |   .4102648   .1235412   1.362438
        1 2  |   .6550924   .2273009   1.888008
        1 3  |   .5177431    .211639   1.266581
```

- Note that this is the same model; there are still 6 parameters and the fitted rates are identical. It's just that the 6 parameters in this model have a different interpretation.

- The log-likelihood for this model is the same as the previous interaction model. This is because we are fitting the exact same model, but with different parameterisation.

- Technical note: The ibn. factor-variable operator specifies that a categorical variable should be treated as if it has no base, or, in other words, that all levels of the categorical variable are to be included in the model.

Effects of exposure within each stratum of the modifier

- If we insert the Stata output in our previous table, we get

job	hieng=0	hieng=1
driv	1.0	0.41
cond	1.0	0.66
bank	1.0	0.52

- The stratum-specific IRRs are similar, there is no evidence of interaction.
- Effect of being high vs low energy is 0.41 for drivers, 0.41x1.60=0.66 for conductors, and 0.41x1.26=0.52 for bankers.
- Compare this to the main effects model: 0.52 for all occupation levels.

- Similarly, we can reparameterise the models to show effect of job in levels of energy.

- This yields the following IRRs

job	hieng=0	hieng=1
driv	1.0	1.0
cond	$\exp(\beta_2)$	$\exp(\beta_2 + \beta_4)$
bank	$\exp(\beta_3)$	$\exp(\beta_3 + \beta_5)$

```
. poisson chd ibn.hieng i.job#i.hieng, e(y) irr nocons
         chd |        IRR   [95% Conf. Interval]
-------------+------------------------------------
       hieng |
          0  |   .0144648   .0072338    .028924
          1  |   .0059344   .0022273    .0158117
             |
   job#hieng |
        2 0  |   1.136857   .4266828    3.029051
        2 1  |   1.815282   .5122665    6.432687
        3 0  |   .813427    .3325064    1.989927
        3 1  |   1.026523   .3091123    3.408953
```

- If we insert the Stata output in our previous table, we get

job	hieng=0	hieng=1
driv	1.0	1.0
cond	1.14	1.82
bank	0.81	1.03

- Effect of being conductor vs driver is 1.14 for the low energy intake group, whereas it is 1.82 for the high energy group.

- Compare this to the main effects model: 1.36 for all energy intake levels.

- Effect of being banker vs driver is 0.81 for the low energy intake group, whereas it is 1.03 for the high energy group.

- Compare this to the main effects model: 0.88 for all energy intake levels.

Linear combinations of parameters

- As an alternative to reparameterising the interaction model we can use the Stata `lincom()` command to estimate the effect of exposure within each level of the modifier together with confidence intervals. Here again is the interaction model with the default parameterisation.

```
.poisson chd i.hieng##i.job, exp(y) irr
         chd |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.hieng |   .4102648   .2512349    -1.45   0.146     .1235412    1.362438
             |
         job |
          2  |   1.136857   .5684285     0.26   0.798     .4266828    3.029051
          3  |    .813427   .3712769    -0.45   0.651     .3325064    1.989927
             |
   hieng#job |
        1 2  |   1.596755   1.303745     0.57   0.567     .3222813    7.911183
        1 3  |   1.261973   .9638479     0.30   0.761     .2824452    5.638532
             |
       _cons |   .0144648   .0051141   -11.98   0.000     .0072338     .028924
------------------------------------------------------------------------------
```

- The effect of `hieng` for job=1 is 0.41. We now estimate the effect of `hieng` for the other two categories of job.

```
. lincom 1.hieng + 1.hieng#2.job, irr
------------------------------------------------------------------------
chd |         IRR   Std. Err.       z    P>|z|     [95% Conf. Interval]
----+-------------------------------------------------------------------
(1) |    .6550924   .3537903    -0.78   0.434     .2273009    1.888008
------------------------------------------------------------------------

. lincom 1.hieng + 1.hieng#3.job, irr
------------------------------------------------------------------------
chd |         IRR   Std. Err.       z    P>|z|     [95% Conf. Interval]
----+-------------------------------------------------------------------
(1) |    .5177431   .2363163    -1.44   0.149      .211639    1.266581
------------------------------------------------------------------------
```

- The calculation $0.410 \times 1.596 = 0.655$ isn't difficult but calculating the standard error and CI is non-trivial (a combination of variances and covariances). Lincom is useful for obtaining CI's.

Review: Parameterisations of the interaction model

- Interaction models can be parameterised in different ways to show effects using different reference groups.
- Interaction models measure deviations from the hypothesis of main effect of `hieng` in all `job` categories.
- Interactions are symmetrical, meaning that we can choose either variable job or hieng as the effect modifier of the other. The test for interaction will be the same regardless of parameterisation.
- If we wish to change reference group, simply divide the cells with corresponding reference group.
- E.g. if we wish to tabulate the IRR, using the drivers with low energy intake as reference group, then we simply divided all cells with the rate for drivers with low energy intake.

How the two parameterisations are related

job	hieng=0	hieng=1
1=driv	$\exp(\beta_0) = \exp(\gamma_0)$	$\exp(\beta_0 + \beta_1) = \exp(\gamma_0 + \gamma_3)$
2=cond	$\exp(\beta_0 + \beta_2) = \exp(\gamma_1)$	$\exp(\beta_0 + \beta_1 + \beta_2 + \beta_4) = \exp(\gamma_0 + \gamma_4)$
3=bank	$\exp(\beta_0 + \beta_3) = \exp(\gamma_2)$	$\exp(\beta_0 + \beta_1 + \beta_3 + \beta_5) = \exp(\gamma_0 + \gamma_5)$

- The model is $\ln(\lambda) =$
  $\gamma_0$driv $+ \gamma_1$cond $+ \gamma_2$bank $+ \gamma_3$hieng x driv $+ \gamma_4$hieng x cond $+ \gamma_5$hieng x bank