This is an edited version of Chapters 6-8 of

A SHORT INTRODUCTION TO

STATA FOR BIOSTATISTICS

Michael Hills and Bianca L. De Stavola

Chapter 6

More about variables

So far we have allowed a variable to have only one attribute: whether it is coded numerically or as a string. In this chapter we shall introduce attributes which indicate how a variable is to be used in an analysis. These include designating a variable as *response* or *explanatory*, and then further classifying the types of response variable.

6.1 Response and explanatory variables

Most questions in statistical analysis take the form of asking whether the value which one variable takes for a given subject depends on the value taken by another variable. For example, in the `births` data we might be interested in whether the birth weight of a baby depends on whether it is a boy. The variable which is of primary interest is called the *response* variable; the variable on which the response variable may depend is called the *explanatory* variable. In the example just quoted, the response variable is birth weight and the explanatory variable is whether the baby is a boy.

In biostatistics four types of response variables are particularly common:

1. Binary
2. Metric
3. Failure
4. Count

A binary response has just two values which should be coded 0 and 1. A metric response (also called a quantitative response) measures some quantity and usually has many possible values. A failure response indicates whether or not a subject fails at the end of a period of observation, and is used with survival data. Finally, a count response records a number of events, and often arises with aggregated failure data. The type of response determines how it will be summarized. For example, a binary response is usually summarized using the proportion of 1's, and a metric response is usually summarized using its mean or median.

6.2 Declaring the type of response with tabmenu1

To facilitate the preparation of tables for a variety of responses a Stata menu invoked by the command `tabmenu1` invites you to specify the type of response. Consider a question such as whether `bweight` varies with `hyp` and try

```
. tabmenu1, clear
```

The different parts of the menu are used as follows:

- Select the response variable – choose `bweight`.
- Select the type of response – choose `metric`.
- Select follow-up time variable if appropriate. This is needed with survival data. Ignore it for now.
- Select the explanatory variable which will be used for the rows of the table – choose `hyp`.
- Select the explanatory variable which will be used for the columns of the table – ignore it for now.

From here on we shall abbreviate these instructions as follows:

```
. tabmenu1, clear
---> select bweight as response
---> select metric as type
---> select hyp as rows
```

Note that only the first line is typed – all the rest are instructions to follow within the menu. The option `clear` is used with `tabmenu1` to remove any previous selections, so the menu will start off with blanks.

6.3 Producing tables

Clicking on the Tables button in the first menu calls up another menu. Because the response which was chosen in `tabmenu1` is metric you are offered three possibilities for summarizing the response - the mean, the geometric mean, and the median. Select Mean, and press OK. You should see something like this:

```
Response variable is: bweight which is metric
Row variable is: hyp
Number of records used:  500

Summary using means
----------------------
hypertens |    bweight
----------+-----------
        0 |    3198.90
        1 |    2768.21
----------------------
```

The figures in the table are the mean values of the babies' birth weights for mothers who were normal (`hyp=0`) or hypertensive (`hyp=1`). The default is to provide minimum information, because it is often easier to start with this, but to get more you could check (tick) the boxes for frequencies and confidence intervals that appear in the second menu. Try

```
. tabmenu1
---> click on Tables
     ---> select Mean
     ---> check Frequencies
     ---> check Confidence intervals
     ---> click on OK
```

The second level of indentation above refers to the second menu.

With a binary response, such as `lowbw`, the choice of summary is different. Try

```
. tabmenu1, clear
---> select lowbw as response
---> select binary as type
---> select hyp as rows
---> click on Tables
```

Because the response is binary you will be offered a choice between Proportions and Odds. Select Proportions and click on OK. You should see something like this

```
Response variable is: lowbw which is binary
Row variable is: hyp
Number of records used:  500

Summary using proportions per 100
----------------------
hypertens |      lowbw
----------+-----------
        0 |       9.35
        1 |      27.78
----------------------
```

The figures in the table are the percentages of low birth weight babies for mothers who were normal or hypertensive. To produce 90% confidence intervals for these proportions, try

```
. tabmenu1
---> click on Tables
     ---> check Confidence intervals
     ---> enter 90 in the Level of confidence box
     ---> click on OK
```

The command `tabmenu1` works entirely through menus, but there is an equivalent command, `tabmenu2`, which does the same thing without menus. Using the option `display` with `tabmenu1` will show the equivalent `tabmenu2` command, and this can then be cut and pasted into the Command window or a do file. For example, try

```
. tabmenu1, display
```

and repeat the table you have just created. You should now see the line

tabmenu2, res(lowbw) typ(binary) row(hyp) summ(prop) ci level(90)

displayed in the Results window. Cut and paste this into the Commands window using Edit/Copy Text and Edit/Paste, press return, and you will see the same table again. The command can be edited, so small changes can be made without going through the menus again, but the main use of this facility is to include tables produced by the `tabmenu1` command in a do file, without having to fill out the menus. There is more information about `tabmenu1` and `tabmenu2` in the help files.

6.4 A second explanatory variable

We found a strong relationship between birth weight and whether the mother was hypertensive, but is this relationship the same for both male and female babies, i.e. is it *modified* by `sex`? To study this we need to produce a table of mean birth weight by both `hyp` and `sex`. Using

```
. use births, clear
. table hyp sex, contents(freq mean bweight)
```

we find that the birth weight of both male and female babies is lower when the mother is hypertensive than when the mother is normal – about 500 g lower for males babies and about 400 g for female babies.

To produce the same table using `tabmenu1`, try

```
. tabmenu1, clear
---> select bweight as response
---> select type as metric
---> select hyp as rows
---> select sex as columns
---> click on Tables
     ---> select Mean
     ---> click on OK
```

To reverse the rows and columns, check the box marked Reverse variables in the second menu that appears after clicking on Tables.

6.5 Odds

To make a table of the odds of the baby being low birth weight for normal and hypertensive mothers, try

```
. tabmenu1, clear
---> select lowbw as response
---> select binary as type
---> select hyp as rows
---> click on Tables
     ---> select Odds
     ---> click on OK
```

To obtain frequencies and confidence intervals for the odds you can check the appropriate boxes. The Stata command `tabodds` will also tabulate odds, but only by one explanatory variable:

```
. tabodds lowbw hyp
```

Tables with too many rows are not particularly useful, and for this reason an upper limit of 10 values for the row and column variables has been set in `tabmenu1`. This can be increased if required. Try

```
. tabmenu1, clear
---> select lowbw as response
---> select binary as type
---> select matage as rows
---> click on Tables
     ---> select Odds
     ---> click on OK
```

and you will get a message telling you that there are too many values in the row variable (`matage`). Try again, and set the maximum number of values that a row or column variable can have to 25, select Odds as the summary, and you will see a table of the odds of being low birth weight by maternal age. This is not very useful, and it would be better to group the values of `matage` using `egen` with `cut`.

6.6 Survival data and rates

Data involving survival times are often summarized using *rates*, i.e. the number of events per unit time. There are no survival time variables in the `births` data, so we need another dataset to demonstrate how to make tables of rates. Try

```
. use diet, clear
. describe
```

These data refer to a follow–up study of 337 male subjects who were asked to weigh the different components of their diet for a week. They were then followed until

1. They developed, and possibly died from, coronary heart disease (CHD)

2. They died from some other cause, or were withdrawn from the study for some reason, or the study ended.

The time for which each subject is followed is the true survival time in the first case, but in the second case the true survival time has been *censored* by death from another cause, withdrawal, or the end of the study. To record data with true and censored survival times we need two variables: the time spent in the study and a variable which indicates whether the subject developed CHD or not. These are called the *time* and *failure* variables, respectively. In the absence of censoring, the failure variable takes the value 1 for all subjects, and the response variable is time, but when there is censoring, the response is a combination of the time and the failure variables.

In this example the time variable is `y`, and the failure variable is `chd`, coded 1 if the subject developed coronary heart disease (CHD) during the period of the study, and 0 otherwise. For a preliminary analysis the total energy intake per day is converted to a binary variable `hieng` coded 1 if the energy intake is > 2750 kcal, and 0 otherwise. To create a table of rates for `chd` by `hieng`, try

```
. tabmenu1, clear
---> select chd as response
---> select failure as type
---> select y as follow-up time
---> select hieng as rows
---> click on Tables
     ---> select Rates per 1000
     ---> click on OK
```

Chapter 7

Making comparisons

The word *effect* is a general term referring to ways of comparing the values of the response variable at different levels of an explanatory variable. This chapter covers the measurement of effects as differences in means for a metric response; differences in proportions for a binary response; ratios of odds or proportions for a binary response; and ratios of rates for a failure response.

7.1 Comparing means

In Chapter 6 we prepared a table of mean `bweight` by `hyp` using `tabmenu1`. This showed that hypertensive mothers had babies which were on average about 430 g smaller than those delivered to normal mothers. This difference in mean birth weight is called the *effect* of hypertension on birth weight. The command `effmenu1`, like `tabmenu1`, brings up a menu which first invites you to specify the type of response. When you bring up this menu you will see that the term *exposure variable* is used instead of *explanatory variable*. The reason for this change is that when calculating effects it is important to distinguish between the different roles which the explanatory variables might have in an analysis. The explanatory variable whose effects you want to calculate is called the *exposure* variable; *control* and *modifying* variables will be introduced later in this chapter. In this example the response variable is `bweight` and the exposure variable is `hyp`, so try

```
. use births, clear
. effmenu1, clear
---> select bweight as response
---> select metric as type
---> select hyp as exposure
---> click on Effects
     ---> select Difference in means
     ---> click on OK
```

The option `clear` after `effmenu1` removes any selections which might remain from a previous use of this command. You will see that the effect of `hyp` on `bweight`, measured using the difference in the mean response, is −430.7 g. The 95% confidence interval for this effect is from −276 to −586 g. The statistical test is for the null hypothesis that the true effect of hypertension is zero. The P-value is very low so there is strong evidence against the null hypothesis.

Now cut `gestwks` into 4 groups with

```
. egen gest4=cut(gestwks), at(20,35,37,39,45)
. tabulate gest4
```

When comparing the mean birth weight between the four different levels of `gest4` there will be three effects: the effect comparing level 2 with level 1; level 3 with level 1; and level 4 with level 1.

The level with which each of the other levels is compared is called the *baseline* (level 1 in this case). To prepare a table of the three effects of gest4, try

```
. effmenu1
---> select gest4 as exposure
---> click on Effects
     ---> select Difference in means
     ---> click on OK
```

There are three effects, and the statistical test is for the null hypothesis that the true values of these effects are all zero. The second menu allows you to change the baseline from its default value of 1. Try changing the baseline to 3: each of the levels 1, 2, 4 is now compared with level 3.

Like tabmenu1, the command effmenu1, with the option display, will display the equivalent effmenu2 command which works without menus. See the help on effmenu1 and effmenu2 for more information.

7.2 Comparing proportions and odds

When examining the proportion of low birth weight babies according to the length of their gestation, using gestation time in four groups, each level of gest4 can be compared with the baseline level using the difference in proportions, or the ratio of proportions, or the ratio of odds. The preferred method is to use the ratio of odds.

Start by using tabmenu1 to prepare a table of the odds that a baby has low birth weight, by the levels of gest4, and note that the odds are 4.1667 for level 1 of gest4, 0.6842 for level 2, 0.1208 for level 3, and 0.0117 for level 4. Using ratios of odds to measure effects we should get $0.6842/4.1667 = 0.1642$ for level 2 compared with level 1, and so on. Now try

```
. effmenu1, clear
---> select lowbw as response
---> select binary as type
---> select gest4 as exposure
---> click on Effects
     ---> select Odds ratios
     ---> click on OK
```

You should see this table of three effects comparing levels 2, 3, 4 against level 1, using odds ratios:

```
                        Effect    95% Confidence Interval

Level 2 vs level 1      0.1642   [ 0.053 , 0.512 ]
Level 3 vs level 1      0.0290   [ 0.010 , 0.080 ]
Level 4 vs level 1      0.0028   [ 0.001 , 0.012 ]
```

Compared with level 1 of gest4, mothers at level 2 have lower odds of a low birth weight baby by a factor of 0.1642, mothers at level 3 have lower odds by a factor of 0.0290, while mothers at level 4 have lower odds by a factor of 0.0028. The statistical test that appears below the table is for the null hypothesis that the true values of these three effects (odds ratios) are all 1. It takes the form of a chi–squared statistic, because the response is binary, and is on 3 df because there are three effects being tested.

We have chosen to measure the effects of gest4 as odds ratios, but they can also be measured using the ratios of proportions. Try

```
. effmenu1
---> click on Effects
     ---> select Ratio of proportions
     ---> click on OK
```

and you will see a table showing the three effects of gest4 as ratios of proportions instead of odds ratios. You should be aware that when using the ratio or difference of proportions to measure effects, the program may fail to reach an answer for datasets where some of the proportions being compared are close to 0 or 1.

7.3 Comparing rates

Load the diet data and cut energy into 3 groups with

```
. use diet, clear
. egen eng3=cut(energy),at(1500,2500,3000,4500)
. tabulate eng3
```

Use tabmenu1 to prepare a table showing the rates of CHD for different levels of eng3, and note that the rate goes down from 16.90 to 4.88 per 1000, with increasing level of energy. The two effects of eng3 can be measured as rate differences or rate ratios. To prepare a table of effects using rate ratios, try

```
. effmenu1, clear
---> select chd as response
---> select failure as type
---> select y as follow-up time
---> select eng3 as exposure
---> click on Effects
     ---> select Rate ratios
     ---> click on OK
```

You should see something like this:

```
                    Effect   95% Confidence Interval

Level 2 vs level 1  0.6452   [ 0.339 , 1.229 ]
Level 3 vs level 1  0.2886   [ 0.124 , 0.674 ]
```

Compared with level 1 of eng3, subjects at level 2 have a lower rate of CHD by a factor of 0.6452, and subjects at level 3 have a lower rate of CHD by a factor of 0.2886. The statistical test that appears below the table is for the null hypothesis that the true values of both effects of eng3 are 1. The P-value is very small, so there is strong evidence against the null hypothesis.

7.4 Controlling for confounding variables

The effects calculated so far are *marginal* effects, and take no account of the possibility of confounding due to other variables. For example, in the births data, the marginal effect of hyp is −430.7 g, but the sex of the baby is associated with its birth weight, so if sex is also associated with hypertension, then part of the marginal effect could be due to differences in the sex ratio between the babies of normal and hypertensive women. To exclude this possibility we need to *control* the effect of hyp for sex by selecting sex as a control variable:

```
. use births, clear
. effmenu1, clear
---> select bweight as response
---> select metric as type
---> select hyp as exposure
```

```
---> click on Effects
     ---> select Difference in means
     ---> select sex as a control variable
     ---> click on OK
```

The effect of hyp controlled for sex is −448.1 g, but to understand how this is obtained we need to look at the effect of hyp on bweight separately for each sex. This is done by specifying sex as a modifying variable.

7.5 Effect modification

To specify sex as a modifying variable, try

```
. effmenu1
---> select sex as modifier
---> click on Effects
     ---> select Difference in means
     ---> remove sex as a control variable
     ---> click on OK
```

The result should look like this:

```
Level or value
of sex                        Effect   95% Confidence Interval

1                            -496.3513     [ -296.143 , -696.560 ]
2                            -379.7734     [ -141.606 , -617.941 ]
```

The effect of hyp is −496.3 g for boys and −379.8 g for girls. These two separate effects of hyp are really what is meant by "controlled for sex" because all boy babies have the same value for sex, and all girl babies have the same value for sex, so no part of these effects can be due to differences in the sex ratio. However, because there is no evidence that the true values of these two effects differ, reporting the separate effects is unnecessary, and they are combined to give a single effect, −448.1 g. It is this combined effect that is called the effect of hyp controlled for sex. The word *control* (unfortunately) is used to cover both controlling, i.e. finding effects separately for different levels of the confounder, and combining the separate effects when they appear to be the same.

Although many people control for a potential confounder without first looking at the separate effects at different levels of the confounder, it is better to look first, because the separate effects are combined on the assumption that there is no effect modification, i.e. that their true values are the same. The statistical test for no effect modification is used to confirm this assumption, although commonsense also plays a role. When there is strong effect modification the effect of exposure should be reported separately for each level of the modifying variable.

As another example, we shall return to the diet data, and control the effect of hieng on the rate of CHD for job, first looking at the separate effects of hieng in the different jobs. Start by finding the marginal effect of hieng (level 2 vs level 1) on chd as a rate ratio, with

```
. effmenu1, clear
---> select chd as response
---> select failure as type
---> select y as follow-up time
---> select hieng as exposure
---> click on Effects
     ---> select Rate ratios
     ---> click on OK
```

The answer should be 0.5204. Now go back to `effmenu1`, select `job` as the modifying variable, click on Effects, and then OK as before. You will now see three effects of `hieng`, one for each level of `job`:

```
Level or value
of job            Effect   95% Confidence Interval

driver            0.4103   [ 0.124 , 1.362 ]
conductor         0.6551   [ 0.227 , 1.888 ]
bank              0.5177   [ 0.212 , 1.267 ]
```

All three effects are measuring the effect of `hieng`, but in subjects who have different jobs. These three effects are similar in size, so there is no evidence that `job` modifies the size of the effect of `hieng`. This is confirmed by the significance test for no effect modification which appears below the table. The chi–squared statistic is on 2 df because we are making two comparisons: 0.6551 (conductors) with 0.4103 (driver) and 0.5177 (bank) with 0.4103 (driver). The P-value is large, confirming that there is no evidence that `job` modifies the size of the effect of `hieng`.

We can now combine the three separate effects as a single effect, by removing `job` from being a modifying variable in the first menu, and selecting `job` as a control variable in the second menu. Try this now – you should get 0.5248 for the effect of `hieng` controlled for `job`, not very different from the marginal, or uncontrolled effect of 0.5204, showing that the confounding influence of `job`, if any, is minimal.

It is also possible to control for one variable while allowing another to be a modifier. For example, try

```
. effmenu1
---> select hieng as exposure
---> select job as modifier
---> click on Effects
     ---> select month as control
     ---> select Rate ratios
     ---> click on OK
```

The output shows the effect of `hieng`, controlled for month, but separately by job. Any number of control variables can be selected with `effmenu1`, but there can be only one modifier.

Chapter 8

Effects and metric variables

In this chapter we extend the ideas of Chapter 7 to cover metric exposures, metric control variables, and metric modifiers.

8.1 Metric exposure variables

When using `tabmenu1` no distinction is drawn between categorical and metric explanatory variables. If a metric variable such as `gestwks` in the `births` data is selected as explanatory, `tabmenu1` simply refuses to make a table because `gestwks` has too many values. When using `effmenu1` the distinction between categorical and metric is essential because, as we shall now see, it is possible to find the effects of a metric exposure even when it has many values.

We shall illustrate this by finding the effect of `gestwks` on `bweight`. Even though `gestwks` has many values we can still find the effect of one unit increase in `gestwks` on `bweight` provided we can assume that this effect is the same throughout the range, i.e. that the effect of changing from 30 to 31 weeks of gestation is the same as changing from 31 to 32, and so on throughout the range. If this is the case then the relationship between `bweight` and `gestwks` is *linear*, so start by checking this with

```
. use births, clear
. graph bweight gestwks
```

You will see that birth weight tends to go up with gestational age, and that the increase in birth weight per unit increase in gestation (in weeks) is roughly constant throughout its range of values. Now we can find the effect of a unit increase in `gestwks` with

```
. effmenu1, clear
---> select bweight as response
---> select metric as type
---> select gestwks as exposure
---> check metric exposure
---> click on Effects
     ---> select Difference in means
```

The menu now shows that the exposure variable is metric, and offers a choice of effects *per something*. The default is per 1 unit, so go with this and click OK to produce the table of effects. In this case there is only one effect, namely 197.0 g per unit increase in `gestwks`, i.e. per week of gestation.

We can do the same thing to find the effect of `gestwks` on `lowbw`, using odds ratios to measure the effect. The assumption we are now making is that the odds that the baby has low birth weight is reduced by the same factor for a change in gestation from 30 to 31 weeks, 31 to 32 weeks, and so throughout the range. This is the same as saying that the relationship between the log odds and `gestwks` is linear. The easiest way to check this is to cut `gestwks` into, say, 4 equally spaced groups and to look at how the odds change from one group to the next. Now try

```
. effmenu1, clear
---> select lowbw as response
---> select binary as type
---> select gestwks as exposure
---> check metric exposure
---> click on Effects
     ---> select Odds ratios
     ---> click on OK
```

The effect of a unit increase in `gestwks` is to multiply the odds that the baby has low birth weight by 0.52, i.e. a reduction to 52% of its current level for each extra week of gestation. To see the effect per 5 weeks of gestation, which is the interval between the levels of `gest4`, re-run `effmenu1` and select per 5 units in the second menu. The result is 0.0375, a sort of average of the three odds ratios 0, 0.034, and 0.065, which compare each level of `gest4` with the previous one.

We shall now return to the `diet` data and find the effect of `energy` on the rate of CHD, where `energy` is a metric exposure variable, and the effect is measured as a rate ratio per unit of energy. As before we assume that this effect is the same throughout the range, that is the effect of changing from 1700 to 1701 kcal is the same as changing from 1702 to 1703, and so on throughout the range. This is the same as saying that the relationship between the log rate and `energy` is linear. Now try

```
. effmenu1, clear
---> select chd as the response
---> select failure as type
---> select y as follow-up time
---> select energy as exposure
---> check metric exposure
---> click on Effects
     ---> select Rate ratio
     ---> click on OK
```

The effect of a unit increase in `energy` is 0.9988 per unit of energy, i.e. the CHD rate is reduced by a factor of 0.9988 for each increase of 1 kcal in total energy. An increase of 1 kcal is a very small amount of energy, which explains why the effect is so close to 1. It would be better to measure the effect per 100 kcal, or even per 500 kcal. Go back by typing `effmenu1`, click on Effects, and then change the units for computing the rate ratio to per 100 kcal. The effect is now 0.8913 per 100 kcal. Repeat this and change the units to per 500 kcal. The effect is now 0.645 per 500 kcal, a sort of average of the rate ratios 0.93, 0.61, 0.63, and 0, which compare each level of `eng5` with the previous one.

8.2 Controlling the effect of a metric exposure

The effect of a metric exposure variable can be controlled for a confounding variable, in the same way as for a categorical exposure. For example, try

```
. effmenu1
---> select energy as exposure
---> check metric exposure
```

```
---> select job as modifier
---> click on Effects
     ---> enter 100 in per unit box
     ---> select Rate ratios
     ---> click on OK
```

The three effects you will see are the effects of energy per 100 kcal at each of the levels of `job`. They seem similar, and this is confirmed by the test for effect modification. Combine them by moving `job` from being a modifier to being a control variable, to get 0.8919 as the effect of `energy` per 100 kcal, controlled for `job`.

8.3 Metric modifying and control variables

To complete the story we need to discuss metric modifying and control variables. For example, suppose we want to control the effect of `hieng` on `chd` for `height`. On the assumption that the log rate changes linearly with `height`, this is easily done as follows:

```
. effmenu1, clear
---> select chd as response
---> select failure as type
---> select y as follow-up time
---> select hieng as exposure
---> click on Effects
     ---> select Rate ratios
     ---> select height as control
     ---> select height as metric control
     ---> click on OK
```

The effect of `hieng` controlled for `height` (metric) is 0.6132. Of course, before doing this, you should check on the linear relationship between the log rate and `height` by grouping `height`, and making a table of rates by grouped height, as follows:

```
. egen htgrp=cut(height),at(150(10)190)
. tabmenu1, clear
---> select chd as response
---> select failure as type
---> select y as follow-up time
---> select htgrp as rows
---> click on Tables
     ---> select Rates per 1000
     ---> click on OK
```

One man has a height above 190 cm, but since we are only checking the linear assumption, he can be excluded. You will see that the rate is going down at each level, so the assumption of a linear relationship between the log rate and `height` is not unreasonable. You should also check that `height` does not modify the effect of `hieng` by specifying `height` as a modifier and producing a table of effects for different values of `height`. To do this, use `effmenu1` again, specify `height` as a modifier, and check the metric modifier box. You will see (among other things):[1]

[1] If you don't see this, you have probably forgotten to check the metric modifier box.

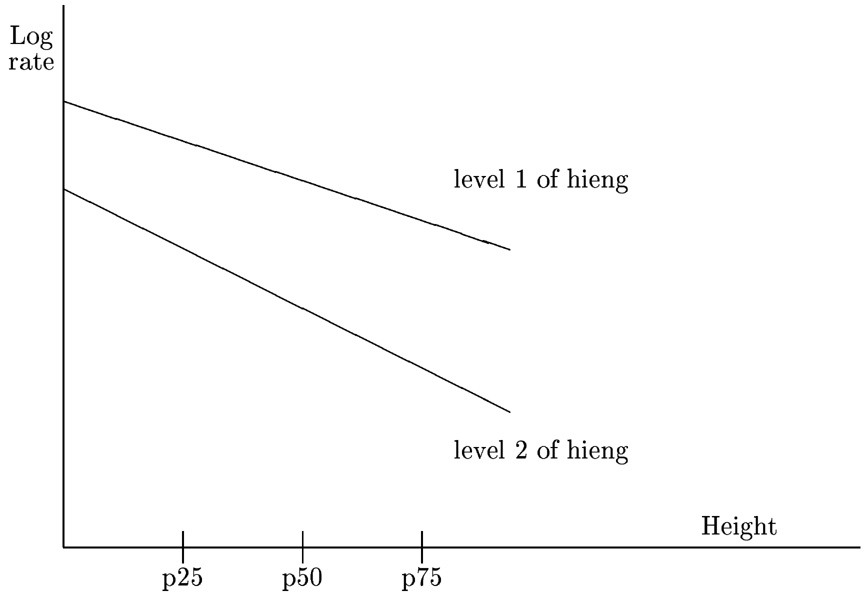

Figure 8.1: Displaying effects when the modifier is metric (hypothetical example)

Level or value of height	Effect	95% Confidence Interval
p25	0.6409	[0.348 , 1.181]
p50	0.5673	[0.295 , 1.091]
p75	0.4916	[0.193 , 1.249]

What the program has done is best explained in terms of log rates. A linear relationship between the log rate and `height` is assumed for each level of `hieng` – these are the two straight lines you see in Figure 8.1. The 25th, 50th, and 75th percentiles of height are marked on the horizontal axis as p25, p50, and p75. Because the log of the ratio of two rates is equal to the difference between the two log-transformed rates, the log of the effect of `hieng` when height is at the 25th percentile is the vertical distance between the two lines at p25. The log of the effect when height is at the 50th percentile is the vertical distance between the two lines at p50, and similarly for height at the 75th percentile. If there is no effect modification these three effects of `hieng` at different levels of `height` will be the same, apart from random variation, i.e. the lines will be parallel. If there is substantial divergence or convergence between them then the effect of `hieng` is modified by `height`.

The `showat` box in the second menu allows you to choose the points of the metric modifier at which the effects of the exposure are calculated. For example, you might prefer to show the effects of `hieng` at heights 160, 165, 170, 175 cm, instead of at the three percentile values shown above.