

Analysing Hospital Episode Statistics (HES)

Practical Session 2 - Looking at diagnoses

Available to download from: www.robin-beaumont.co.uk/virtualclassroom/hes

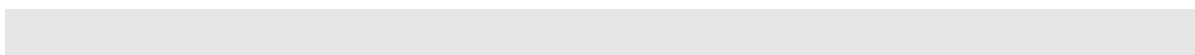
Friday, 24 August 2007

Written by: Robin Beaumont robin@organplayers.co.uk

Status: V1.0

Contents

1.	Introduction.....	2
2.	What are, and how do I calculate bed days?	3
2.1	How do I create new fields derived from values in other fields	3
2.1.1	How do I obtain a chart adding all the bed days together for each consultant?	4
2.2	How do I save the total number of bed days for each consultant to another Data file?.....	5
3.	Investigating diagnoses.....	7
3.1.1	How do I find the number of episodes for a particular diagnosis for each age.....	9
3.1.2	How do I select ranges of diagnoses?	10



1. Introduction

Note:

If you are continuing work after session one without a break you can ignore the exercise below and move onto the next page.

During this session we will carry on working on the Trauma and Orthopaedic dataset.

The data file is called Tando1.sav You should now load this into SPSS.

During the last session we created a filter to remove certain records from the analysis. SPSS does not store the filter so you need to re-create it when you re-open the file if you want to carry on working on the same subset of data.

Exercises:

The aim of this exercise is to re-create the filter you ended of with in the last session:

If you're not too sure how to create a filter look back at the section 'How do I exclude certain cases from an analysis?' in the last sessions handout .

In the edit window in the '*Select Cases: If*' dialogue box edit it so that it contains the following text:

startage<=700 and not (classpat = 2) and not sysmis(epiend)

In English this reads select records which have a valid EPIEND date (i.e. finished episode) as well as not being day cases (i.e. CLASSPAT=2) and over one year old (i.e. startage<=700).

The 'sysmis()' is a SPSS function that returns true if the record for the field specified in the brackets has a missing value. EPIEND in this example. In other words we only want records with valid EPIEND values.

2. Carry out a frequency count on the new filter. You should now have 1463 cases available for analysis.

2. What are, and how do I calculate bed days?

The idea of a bed day is the length of time a bed is occupied. In the data set there were a number of people with a episode duration of '0' that is they were admitted and discharged within a 24 period however it could be argued that they still used a bed for approximately a day. To calculate the number of days a particular bed is used it is necessary to add one onto each episode duration (EPIDUR). This is the purpose of the next exercise.

2.1 How do I create new fields derived from values in other fields

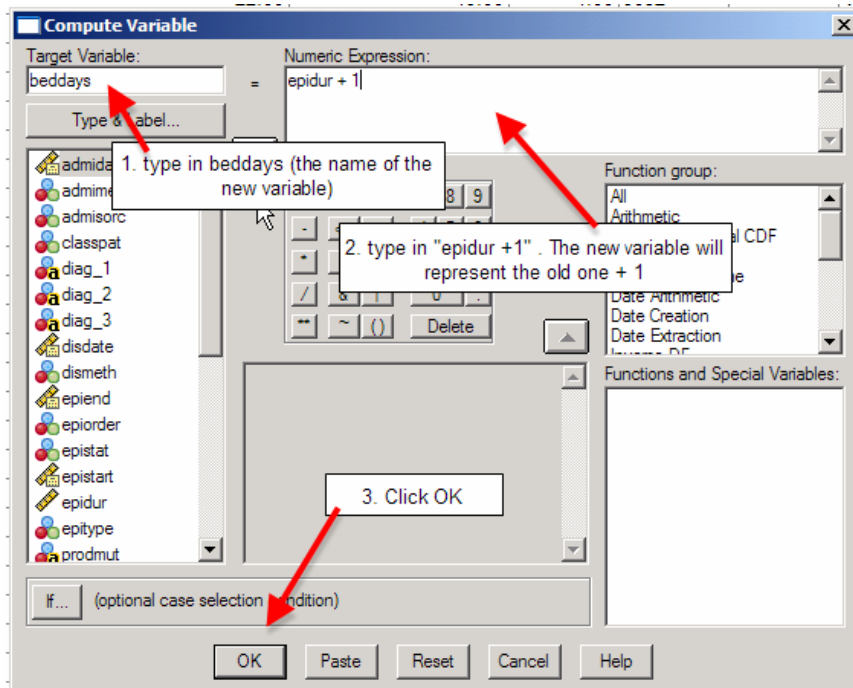
There are several ways to do this in SPSS, we will choose the most common method.

Exercise:

The aim of this exercise is to create a new field (variable) called beddays based upon the length of a episode (EPIDUR) + 1.

Choose the menu option *Transform -> Compute*

You will be presented with the 'Compute Variables' dialogue box shown below.



Modify The dialogue box by changing the following:

Target Variable: Beddays

Numeric Expression: epidur + 1

Click the 'OK' button to run the computation. The last field in the data window (scroll if necessary) should now be a new field called Beddays, containing values which are one more than that in EPIDUR.

We now need to add up all the bed days for each consultant. In Excel we would just use a 'Sum' command or highlighted the column. In SPSS it is slightly more complex. We will achieve the result two different ways to gain experience. The first technique is really a work around using the charts available in SPSS.

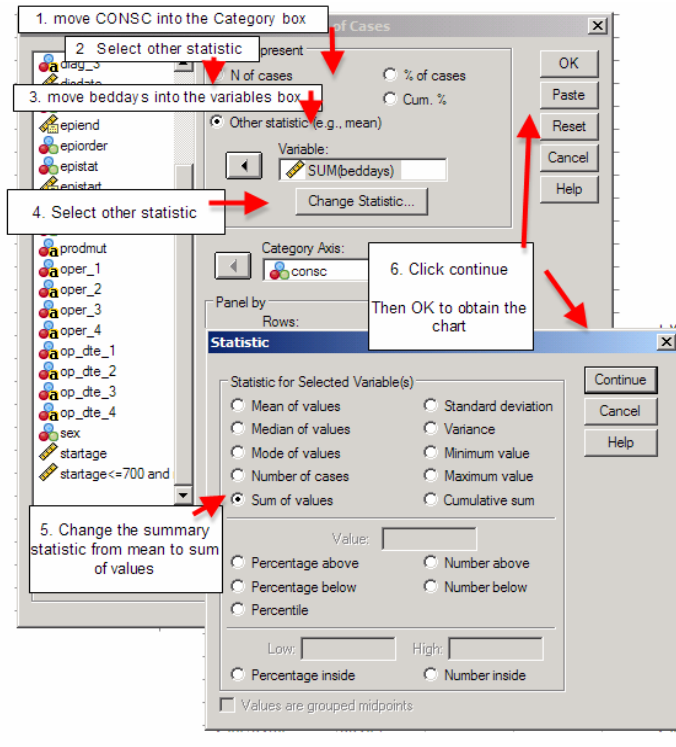
2.1.1 How do I obtain a chart adding all the bed days together for each consultant?

Exercise:

Choose the menu option *Graphs -> Bar*

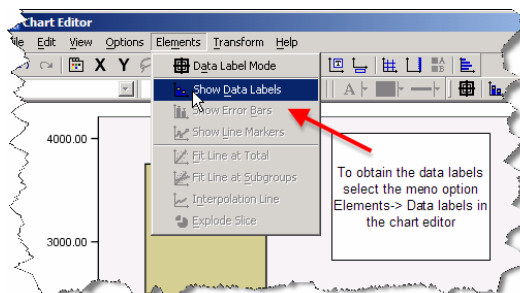
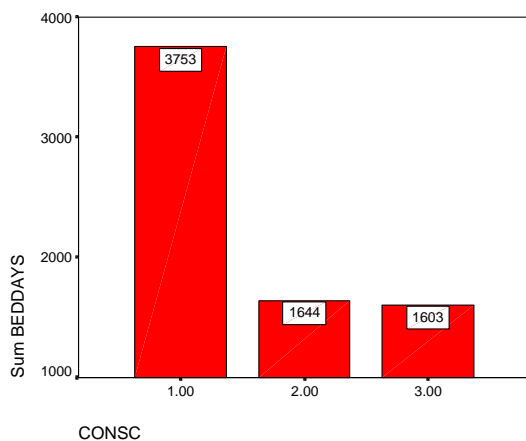
The 'Bar charts' dialogue box will appear. Select the **Simple** chart type option.

Click the **Define** button, you will then be presented with the following dialogue boxes:



Follow the instructions above (step 4 should say select Change statistic)

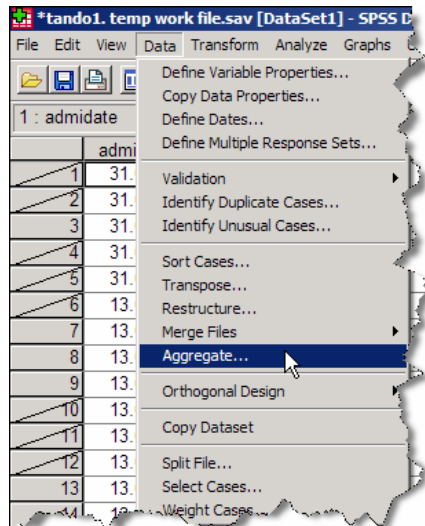
You will obtain a bar chart similar to the one below. To display the actual values you need to edit the chart in SPSS and choose the menu option *Elements -> Show data labels*.



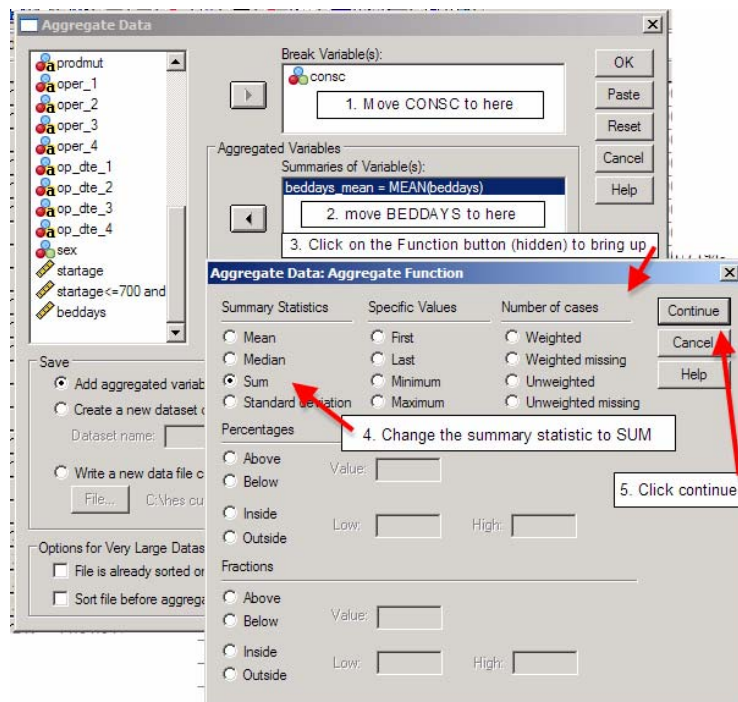
2.2 How do I save the total number of bed days for each consultant to another Data file?

SPSS is designed to guide the user in a particular way to prevent them from making errors. SPSS assumes the dataset consists of a set of cases. It is therefore very easy to add a set of field values together for a particular case or create a new field as shown in the previous section creating the beddays field. In contrast there is no method of adding a set of values in a field across cases and showing the result at the bottom of the datasheet, simply because this would be meaningless in terms of the row = case philosophy. What SPSS provides is a method of aggregating cases to produce another dataset. This prevents users from producing a morass of different types of data all in one place which frequently occurs in spreadsheets. The following exercise should make it clearer.

Exercise



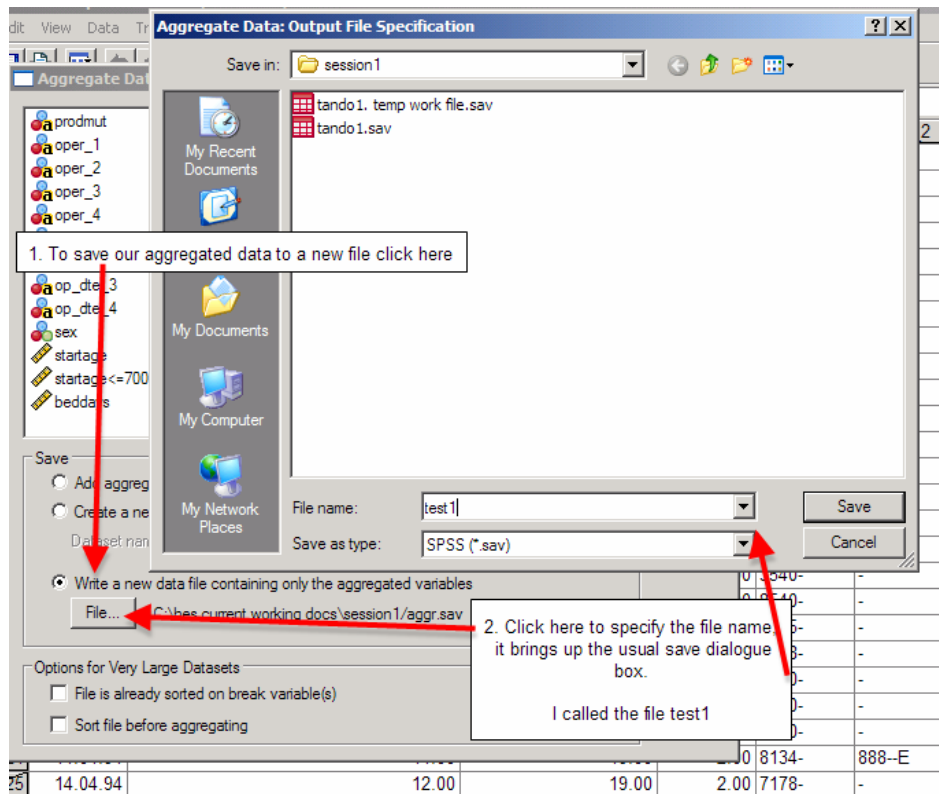
Choose the menu option *Data -> Aggregate* to obtain the following dialogue box shown on the left hand side:



Select the CONCS* field from the list of variables in the left hand window and move it into the 'Break Variable(s)' box by clicking on the button with the → on it . This

indicates that you want the data broken into values for each different consultant id. We need also to choose the Variable(s) we intend to aggregate on. Select BEDDAYS* from the list of variables and move it over into the 'Aggregate Variable(s)' window. As in the previous exercise the wrong function appears with the BEDDAY* field. Click on the 'Function...' button to modify it. You will now be presented with the Aggregate Function dialogue box shown also above. Select the 'Sum' option and then click the continue button to close the dialogue box.

Most of the above should have appeared very similar to the previous exercise. However in contrast to that exercise we now wish to specify where we want the aggregated results (data file) to be saved to. To do this see the next page:



Click on the 'File' button. Note: you will only be able to do this if the 'Create new data file' option is chosen to the left of the button. You will then be presented with the familiar windows save file dialogue box. Save the file as test1 into a suitable folder. Click the 'save' button to return to the previous dialogue box.

Now click on the 'OK' button in the aggregate data dialogue box. The file will be created. You can now create the aggregated data file by. To see the new file you need to select the menu option File -> Open -> data and select the file from the list, the result is shown below.

	consc	bedday_1
1	1.00	3753.00
2	2.00	1644.00
3	3.00	1603.00
4		

It would appear now that consultant one is even doing better, She / he has approximately twice as many episodes as the other consultants and well over twice as many bed days. Does this indicate that she / he has a far greater proportion of the orthopaedic beds? At this point in the analysis it would be sensible to question someone in the organisation directly.

We will continue with the other factors that might explain the difference in numbers of episode for each consultant. The last two factors that were considered to be important were diagnosis and Procedures.

Exercise:

Re open the Tando1.sav data file

Note: There is no need to re-create a filter this time

3. Investigating diagnoses

HES has several diagnosis fields called DIAG_1 to DIAG_7. The fictitious data set we are working with contains the first three. Up until April 1995, the period in which the fictitious data was collected, the 9th Revision of The International Statistical Classification of Diseases, Injuries and Causes of Death (ICD-9) was used since that date ICD-10 has been used. We will start looking at the main diagnosis field DIAG_1.

As usual it is a good idea to start by inspecting the field by looking at the relative frequency of each code. This is left as an exercise for you.

Exercise:

Carry out a frequency count for DIAG_1 (menu option *Analyze -> Descriptive Statistics -> Frequencies*)

Set the format option to display the result as 'Descending counts'

Part of the result is given below. Remember this includes all the records.

(DIAG_1)	Code	Frequency	Percent	Valid Percent	Cumulative Percent
	blank	282	10.6	10.6	10.6
	7151-	123	4.6	4.6	15.3
	7194-	122	4.6	4.6	19.9
	8540-	99	3.7	3.7	23.6
	8134-	94	3.5	3.5	27.1
	7999-	77	2.9	2.9	30.0
	3540-	73	2.8	2.8	32.8
	8200-	60	2.3	2.3	35.1
	7153-	51	1.9	1.9	37.0
	7274-	44	1.7	1.7	38.6
	7242-	43	1.6	1.6	40.3
	8202-	42	1.6	1.6	41.8
	7159-	41	1.5	1.5	43.4
	7225-	40	1.5	1.5	44.9
	8130-	38	1.4	1.4	46.3

	V664	1	.0	.0	99.9
	V712-	1	.0	.0	100.0
	V729-	1	.0	.0	100.0
	Total	2653	100.0	100.0	
Total	2653				

Exercise:

Before looking up the various ICD9 codes write down what you think might be the ten most common Trauma and Orthopaedic diagnoses.

You can look up the most common codes from the handout provided. It is important to realise that the codes form a type of hierarchy for example the ICD9 codes 800 to 829 specify different types of fractures while those for broken lower limbs represent the subset 820-829.

From the above the ten most common diagnoses are:

diag	number of records	Short title	Long title
7151	123	Local.primary osteoarthritis	Localised, primary osteoarthritis
7194	122	Pain in joint - arthralgia	
8540	99	Intracran.inj.NOS+no open i/c	Intracranial injury NOS no open intracranial wound
8134	94	Wrist fracture - closed	
7999	77	[D]Other morbid/mortality NOS	[D]Other and unknown causes of morbidity or mortality NOS
3540	73	Thenar atrophy - partial	
8200	60	Cls # prox femur,transcerv	Closed fracture proximal femur, transcervical
7153	51	Localised OA unspecified	Localised osteoarthritis, unspecified
7274	44	Ganglion and synovial cyst	Ganglion and cyst of synovium, tendon and bursa
7242	43	Low back pain	

	Statistic	Std. Error
STARTAGE Mean	47.0419	.4817
95% Confidence Interval for Mean	Lower Bound 46.0973 Upper Bound 47.9864	
5% Trimmed Mean	47.0849	
Median	48.0000	
Variance	609.313	
Std. Deviation	24.6843	
Minimum	1.00	
Maximum	102.00	
Range	101.00	
Interquartile Range	41.0000	
Skewness	-.056	.048
Kurtosis	-1.094	.096

I wonder if it was what you expected? Applying the less than one year old age exclusion criteria to the data set it would appear that the average age is 47 years:

There is likely to be a great deal of variety between different age groups. For example the teenage age group would show a very different distribution of illness. We will now investigate them.

Exercises:

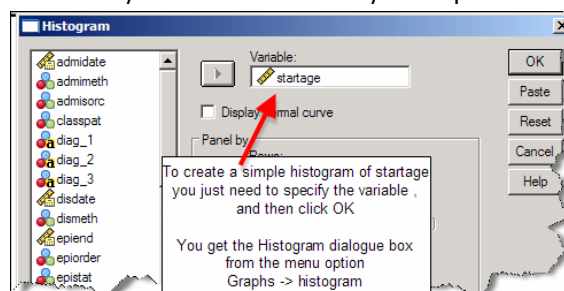
The aim of this exercise is to create and check a filter to allow analysis of diagnoses of 16 to 20 year olds.

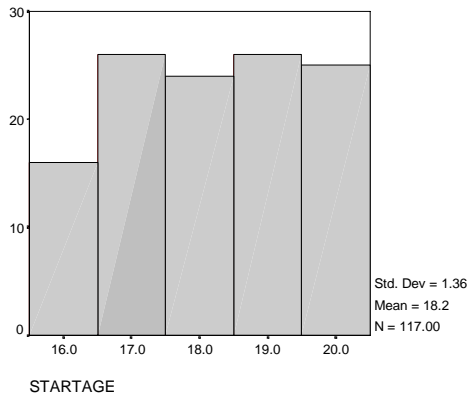
1. Create a filter using the following criteria:

startage<21 and startage > 15

If you're not too sure how to create a filter look back at the section 'How do I exclude certain cases from an analysis?' in the last sessions handout .

2. Create a histogram of the startage variable using the menu option *Graphs -> Histogram* or alternatively using the *Analyze -> Descriptive Statistics -> Explore* option. The result is given on the next page.





Out of the 2626 episodes it would appear that only 117 are for the 16 to 20 age group.

Exercise:

The aim of this exercise is to find the incidence of T & O diagnoses in the 16 to 20 age group.

Carry out a frequency count for DIAG_1

Set the format option to display the result as 'Descending counts'

Part of the result is given below. Fill in the

descriptions from the handout for each of the codes.

DIAG_1	Description	Frequency	Percent	Cumulative Percent
blank		11	9.4	9.4
7194-		7	6.0	15.4
7030-		6	5.1	20.5
8540-		6	5.1	25.6
7183-		4	3.4	29.1
7350-		4	3.4	32.5
8134-		3	2.6	35.0
9058-		3	2.6	37.6
V540-		3	2.6	40.2
3050-		2	1.7	41.9
7177-		2	1.7	43.6
	
Totals		117	100	100

Three of the diagnoses occur in both sets, 194, 8540 and 8134. Strangely the second most common diagnosis is Unguis incarnatus (in-growing toenail) representing 7 out of the 25 occurrences in the data file. One particular diagnosis - late effect tendon injury (ICD9 code 9058) occurs 3 times out of the four in the whole data set obviously a problem more common to 16-20 year olds..

3.1.1 How do I find the number of episodes for a particular diagnosis for each age

Most things you can do several many different ways in SPSS. For example if you were interested in finding out the relative incidence of admissions for one of the commonest diagnoses - Non open Intracranial injury (ICD9 code 8540) - you could create a filter that only allowed you to look at those episodes or create an aggregated file with just that data in it.

Exercise:

Create a filter using the following criteria:

startage < 700 and diag_1 = "8540-"

That is select all episodes with start age of less than 700 (i.e. more than a year old) and which have a diagnosis of 8540. Notice the quotes marks (") indicating that this is a string type field and the '-' as this is how the codes are represented in the data set.

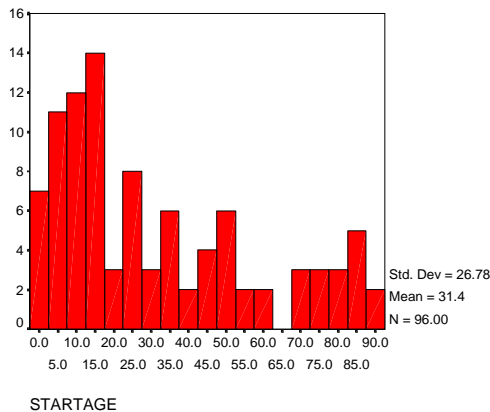
To display the results create a histogram.

Exercise:

Choose the Menu option Graphs -> Histogram

Place the STARTAGE variable in the 'Variable' box.

Click the 'OK' button to run the analysis. The result is given below.



There appears to be a more frequent head injury diagnosis in the early years. What would be more interesting is to find out the proportion of head injures, to all admissions for each age. This will be the topic of the next session.

It should be noted that the choice of a histogram over than of a line chart or bar chart was deliberate. A histogram treats the x axis as a scale measure therefore where there are no observations gaps are left. If a line / bar chart had been chosen the x axis would have been uneven.

Exercise:

Repeat the above steps to look at three more diagnoses of your choice. Use the handout listing the 150 most frequent diagnoses, along with the top 10 listed previously, to guide your choice.

3.1.2 How do I select ranges of diagnoses?

Looking back at the whole data set there are other codes which relate to head injury most notably, 27 episodes of concussion (ICD9 code 850 coded in the data set as both 850-- and 8500-). If you look in the ICD9 handbook you will find that codes 8500 to 8540 relate to head injury. It is quite easy to select ranges of diagnoses in SPSS by using the Range function.

For example `Range(diag_1,"850--", "8540-")` would provide all the head injuries. If your not sure what the range might include sort the data file on `diag_1` (menu option *Data -> Sort Cases*) and scroll down to the relevant cases.

If you wish to select several different diagnosis you can use a filter similar to the following:

`(diag_1 = "8540-" or diag_1 = "7151-") and startage < 700`

or you can use the any function. For example:

`any(diag_1,"8540-", "7151-") and startage < 700`

See the appendix entry at the end of the session three HES handout for an example and production of a chart showing two diagnoses.

End of session