

## Reading Compressed Data with SAS and SPSS on a UNIX<sup>1</sup>

SAS and SPSS can read raw data that has been compressed with either the gzip or compress utility. It saves disk space because the data do not need to be uncompressed. The trade-off is that jobs may take longer to run, because the data will be uncompressed as they are read (but not stored in the uncompressed format). Thus, two processes will run when you submit the job - the SAS or SPSS process, and the uncompressing processes.

### SAS

If reading a standard ASCII file of raw data with SAS, you could use a filename statement to tell SAS where the data are stored - such as that shown below.

```
filename in1 '/full/path/raw.data' lrecl = 80 ;
```

If you want SAS to read a compressed raw data set, the filename statement is modified in the following way.

\* for a data set compressed with the “compress” utility;

```
filename in1 PIPE 'zcat /full/path/raw.data.Z' lrecl = 80 ;
```

\* for a data set compressed with the “gzip” utility;

```
filename in1 PIPE 'zcat /full/path/raw.data.gz' lrecl = 80 ;
```

Note the items in blue/bold. The words PIPE and zcat are necessary. And, remember that UNIX is case sensitive (upper and lower case matter). Files compressed with the “compress” utility will have the extension of “.Z”, and those compressed with “gzip” will have the extension of “.gz”. These extensions must be included to fully identify the location of the data. And, while it is sometimes necessary to provide information about LRECL on a filename reading uncompressed ASCII data, it is always required when reading compressed data.

### SPSS

If reading a standard ASCII file of raw data with SPSS, you could use a file handle statement to tell SPSS where the data are stored - such as that shown below.

```
file handle indd1 / name = '/full/path/raw.data' LRECL=1629.
```

If you want SPSS to read a compressed raw data set, the filename statement is modified in the following way.

\* for a data set compressed with the “compress” utility .

```
file handle in1 / inpipe = 'zcat /full/path/raw.data.Z' LRECL=1629 .
```

\* for a data set compressed with the “gzip” utility .

```
file handle in1 / inpipe = 'zcat /full/path/raw.data.gz' LRECL=1629 .
```

Note the items in blue/bold. The words “inpipe” and “zcat” are necessary. And, remember that UNIX is case sensitive (upper and lower case matter). Files compressed with the “compress” utility will have the extension of “.Z”, and those compressed with “gzip” will have the extension of “.gz”. These extensions must be included to fully identify the location of the data. And, while it is sometimes necessary to provide information about LRECL on a file handle when reading uncompressed ASCII data, it is always required when reading compressed data.

---

<sup>1</sup>Prepared by Patty Glynn, University of Washington, November 30, 2000 C:\all\help\helpnew\compdata.wpd