Chapter 14

Stuff for Students

14.1 Tips for Doing Research

As a student or new researcher, you will probably encounter researchers who think that their method of doing research is the only correct way of doing research, but there are dozens of methods that have proven effective.

Familiarity with the literature is important since your research should be original. The field of high breakdown (HB) robust statistics has perhaps produced more literature in the past 40 years than any other field in statistics.

This text presents the author's applied research in the fields of high breakdown robust statistics and regression graphics from 1990–2008, and a summary of the ideas that most influenced the development of this text follows. Important contributions in the location model include detecting outliers with dot plots and other graphs, the sample median and the sample median absolute deviation. Stigler (1973a) and Tukey and McLaughlin (1963) (and others) developed inference for the trimmed mean. Gnanadesikan and Kettenring (1972) suggested an algorithm similar to concentration and suggested that robust covariance estimators could be formed by estimating the elements of the covariance matrix with robust scale estimators. Hampel (1975) introduced the least median of squares estimator. The LTS and LTA estimators were interesting extensions. Devlin, Gnanadesikan and Kettenring (1975, 1981) introduced the concentration technique. Siegel (1982) suggested using elemental sets to find robust regression estimators. Rousseeuw (1984) popularized LMS and extended the LTS/MCD location estimator to the MCD estimator of multivariate location and dispersion. Ruppert (1992) used con-

centration for HB multiple linear regression. Cook and Nachtsheim (1994) showed that robust Mahalanobis distances could be used to reduce the bias of 1D regression estimators. Rousseeuw and Van Driessen (1999) introduced the DD plot. Important references from the regression graphics literature include Stoker (1986), Li and Duan (1989), Cook (1998a), Cook and Ni (2005), Cook and Weisberg (1999a), Li (2000) and Xia, Tong, Li, and Zhu (2002).

Much of the HB literature is not applied or consists of ad hoc methods. In far too many papers, the estimator actually used is an ad hoc inconsistent zero breakdown approximation of an estimator for which there is theory. The MCD, LTS, LMS, LTA, depth and MVE estimators are impractical to compute. The S estimators and projection estimators are currently impossible to compute. Unless there is a computational breakthrough, these estimators can rarely be used in practical problems. Similarly, two stage estimators need a good initial HB estimator, but no good initial HB estimator was available until Olive (2004a) and Olive and Hawkins (2007b, 2008).

There are hundreds of papers on outlier detection. Most of these compare their method with an existing method on one or two outlier configurations where their method does better. However, the new method rarely outperforms the existing method (such as `lmsreg` or `cov.mcd`) if a broad class of outlier configurations is examined. In such a paper, check whether the new estimator is consistent and if the author has shown types of outlier configurations where the method fails. **Try to figure out how the method would perform for the cases of one and two predictors**.

Dozens of papers suggest that a classical method can be made robust by replacing a classical estimator with a robust estimator. Again inconsistent robust estimators are usually used. These methods can be very useful, but rely on perfect classification of the data into outliers and clean cases. Check whether these methods can find outliers that can not be found by the response plot, FCH DD plot and FMCD DD plot.

For example consider making a robust Hotelling's t–test. If the paper uses the FMCD `cov.mcd` algorithm, then the procedure is relying on the perfect classification paradigm. On the other hand, Srivastava and Mudholkar (2001) present an estimator that has large sample theory.

Beginners can have a hard time determining whether a robust algorithm estimator is consistent or not. As a rule of thumb, assume that the approximations (including those for depth, LTA, LMS, LTS, MCD, MVE, S, projection estimators and two stage estimators) are inconsistent unless the authors show that they understand Hawkins and Olive (2002) and Olive

and Hawkins (2007b, 2008). In particular, the elemental or basic resampling algorithms, concentration algorithms and algorithms based on random projections should be considered inconsistent until you can prove otherwise.

After finding a research topic, **paper trailing** is an important technique for finding related literature. To use this technique, find a paper on the topic, go to the bibliography of the paper, find one or more related papers and repeat. Often your university's library will have useful internet resources for finding literature. Usually a research university will subscribe to either *The Web of Knowledge* with a link to ISI Web of Science or to the *Current Index to Statistics.* Both of these resources allow you to search for literature by author, eg Olive, or by topic, eg robust statistics. Both of these methods search for recent papers. With Web of Knowledge, find an article with *General Search*, click on the article and then click on the *Find Related Articles* icon to get a list of related articles. For papers before 1997, use the free *Current Index to Statistics* website (http://query.statindex.org/CIS/OldRecords/queryOld).

The search engines (www.google.com), (www.ask.com), (www.msn.com), (www.yahoo.com), (www.info.com) and (www.scirus.com) are also useful. The google search engine also has a useful link to "Google Scholar." When searching, enter a topic and the word *robust* or *outliers.* For example, enter the keywords *robust factor analysis* or *factor analysis and outliers.* The keywords *sliced inverse regression, dimension reduction* and *single index models* are useful for finding regression graphics literature.

The STATLIB site (http://lib.stat.cmu.edu/) is useful for finding statistics departments, data sets and software. Statistical journals often have websites that make abstracts and preprints available. Two useful websites are given below.

```
(www.stat.ucla.edu/journals/ProbStatJournals/)
(www.statsci.org/jourlist.html)
```

Websites for researchers or research groups can be very useful. Below are websites for Dr. Rousseeuw's group, Dr. He, Dr. Rocke, Dr. Croux, Dr. Hubert's group and for the University of Minnesota.

```
(www.agoras.ua.ac.be/)
(www.stat.uiuc.edu/~he/index.html)
```

```
(http://handel.cipic.ucdavis.edu/~dmrocke/preprints.html)
(www.econ.kuleuven.ac.be/public/NDBAE06/)
(http://wis.kuleuven.be/stat/robust.html)
(www.stat.umn.edu)
```

The latter website has useful links to software. *Arc* and *R* can be downloaded from these links. **Familiarity with a high level programming language** such as FORTRAN or *R/Splus* is essential. A very useful *R* link is (www.r-project.org/#doc).

Finally, a Ph.D. student needs an advisor or **mentor** and most researchers will find collaboration valuable. Attending conferences and making your research available over the internet can lead to contacts.

Some references on research, including technical writing and presentations, include American Society of Civil Engineers (1950), Becker and Keller-McNulty (1996), Ehrenberg (1982), Freeman, Gonzalez, Hoaglin and Kilss (1983), Hamada and Sitter (2004), Rubin (2004) and Smith (1997).

14.2   R/Splus and Arc

R is the free version of *Splus*. The website (www.stat.umn.edu) has useful links for *Arc* which is the software developed by Cook and Weisberg (1999a). The website (www.stat.umn.edu) also has a link to **Cran** which gives *R* support. As of June 2008, the author's personal computer has Version 2.4.1 (December 18, 2006) of *R*, Splus–2000 (see Mathsoft 1999ab) and Version 1.03 (August 2000) of *Arc*. Many of the text *R/Splus* functions and figures were made in the middle 1990's using *Splus* on a workstation.

Downloading the book's R/Splus functions *rpack.txt* into *R* or *Splus*:

Many of the homework problems use *R/Splus* functions contained in the book's website (www.math.siu.edu/olive/ol-bookp.htm) under the file name *rpack.txt*. Suppose that you download *rpack.txt* onto a disk. Enter *R* and wait for the cursor to appear. Then go to the *File* menu and drag down *Source R Code*. A window should appear. Navigate the *Look in* box until it says *3 1/2 Floppy(A:)*. In the *Files of type* box choose *All files(*.*)* and then select *rpack.txt*. The following line should appear in the main *R* window.

```
> source("A:/rpack.txt")
```

Type *ls()*. About 90 *R/Splus* functions from *rpack.txt* should appear.

When you finish your *R/Splus* session, enter the command *q()*. A window asking "*Save workspace image?*" will appear. Click on *No* if you do not want to save the programs in *R*. (If you do want to save the programs then click on *Yes*.)

If you use *Splus*, the command

```
> source("A:/rpack.txt")
```

will enter the functions into *Splus*. Creating a special workspace for the functions may be useful.

This section gives tips on using *R/Splus*, but is no replacement for books such as Becker, Chambers, and Wilks (1988), Braun and Murdoch (2007), Chambers (1998), Crawley (2005), Fox (2002) or Venables and Ripley (2003). Also see Mathsoft (1999ab) and use the website (www.google.com) to search for useful websites. For example enter the search words *R documentation*.

The command *q()* gets you out of *R* or *Splus*.

Least squares regression is done with the function *lsfit*.

The commands *help(fn)* and *args(fn)* give information about the function fn, eg if fn = lsfit.

Type the following commands.

```
x <- matrix(rnorm(300),nrow=100,ncol=3)
y <- x%*%1:3 + rnorm(100)
out<- lsfit(x,y)
out$coef
ls.print(out)
```

The first line makes a 100 by 3 matrix x with N(0,1) entries. The second line makes $y[i] = 0 + 1*x[i,1] + 2*x[i,2] + 3*x[i,2] + e$ where e is N(0,1). The term 1:3 creates the vector $(1, 2, 3)^T$ and the matrix multiplication operator is $\%*\%$. The function `lsfit` will automatically add the constant to the model. Typing "out" will give you a lot of irrelevant information, but *out$coef* and *out$resid* give the OLS coefficients and residuals respectively.

To make a residual plot, type the following commands.

```
fit <- y - out$resid
plot(fit,out$resid)
title("residual plot")
```

The first term in the plot command is always the horizontal axis while the second is on the vertical axis.

To put a graph in *Word,* hold down the *Ctrl* and *c* buttons simultaneously. Then select "paste" from the *Word* Edit menu.

To enter data, open a data set in *Notepad* or *Word.* You need to know the number of rows and the number of columns. Assume that each case is entered in a row. For example, assuming that the file *cyp.lsp* has been saved on your disk from the webpage for this book, open *cyp.lsp* in *Word.* It has 76 rows and 8 columns. In *R* or *Splus,* write the following command.

```
cyp <- matrix(scan(),nrow=76,ncol=8,byrow=T)
```

Then copy the data lines from *Word* and paste them in *R/Splus.* If a cursor does not appear, hit *enter.* The command *dim(cyp)* will show if you have entered the data correctly.

Enter the following commands

```
cypy <- cyp[,2]
cypx<- cyp[,-c(1,2)]
lsfit(cypx,cypy)$coef
```

to produce the output below.

```
   Intercept            X1           X2           X3           X4
205.40825985    0.94653718   0.17514405   0.23415181   0.75927197
 X5            X6
 -0.05318671  -0.30944144
```

To check that the data is entered correctly, fit LS in *Arc* with the response variable *height* and the predictors *sternal height, finger to ground, head length, nasal length, bigonal breadth,* and *cephalic index* (entered in that order). You should get the same coefficients given by *R* or *Splus.*

Making functions in R and Splus is easy.

For example, type the following commands.

484

```
mysquare <- function(x){
# this function squares x
r <- x^2
r }
```

The second line in the function shows how to put comments into functions.

Modifying your function is easy.

Use the fix command.
    fix(mysquare)
This will open an editor such as *Notepad* and allow you to make changes.
    In *Splus*, the command *Edit(mysquare)* may also be used to modify the function *mysquare*.

To save data or a function in R, when you exit, click on *Yes* when the "*Save worksheet image?*" window appears. When you reenter R, type *ls()*. This will show you what is saved. You should rarely need to save anything for the material in the first thirteen chapters of this book. In *Splus*, data and functions are automatically saved. To remove unwanted items from the worksheet, eg x, type *rm(x)*,
pairs(x) makes a scatterplot matrix of the columns of x,
hist(y) makes a histogram of y,
boxplot(y) makes a boxplot of y,
stem(y) makes a stem and leaf plot of y,
scan(), source(), and *sink()* are useful on a *Unix* workstation.
To type a simple list, use $y <- c(1,2,3.5)$.
The commands *mean(y), median(y), var(y)* are self explanatory.

    The following commands are useful for a scatterplot created by the command *plot(x,y)*.
lines(x,y), lines(lowess(x,y,f=.2))
identify(x,y)
abline(out$coef), abline(0,1)

    The usual arithmetic operators are $2 + 4$, $3 - 7$, $8 * 4$, $8/4$, and

 `2^{10}`.

    The ith element of vector y is $y[i]$ while the ij element of matrix x is

$x[i, j]$. The second row of x is $x[2,]$ while the 4th column of x is $x[, 4]$. The transpose of x is $t(x)$.

The command $apply(x,1,fn)$ will compute the row means if fn = mean. The command $apply(x,2,fn)$ will compute the column variances if fn = var. The commands *cbind* and *rbind* combine column vectors or row vectors with an existing matrix or vector of the appropriate dimension.

Downloading the book's R/Splus data sets *robdata.txt* into R or *Splus* is done in the same way for downloading *rpack.txt*. Use the following command.

```
> source("A:/robdata.txt")
```

For example the command

```
> lsfit(belx,bely)
```

will perform the least squares regression for the Belgian telephone data.

Transferring Data to and from *Arc* and R or *Splus*.
For example, suppose that the Belgium telephone data (Rousseeuw and Leroy 1987, p. 26) has the predictor *year* stored in x and the response *number of calls* stored in y in R or *Splus*. Combine the data into a matrix z and then use the *write.table* command to display the data set as shown below. The

```
 sep=' '
```

separates the columns by two spaces.

```
> z <- cbind(x,y)
> write.table(data.frame(z),sep='   ')
row.names   z.1   y
 1    50    0.44
 2    51    0.47
 3    52    0.47
 4    53    0.59
 5    54    0.66
 6    55    0.73
 7    56    0.81
 8    57    0.88
 9    58    1.06
```

486

```
10   59   1.2
11   60   1.35
12   61   1.49
13   62   1.61
14   63   2.12
15   64   11.9
16   65   12.4
17   66   14.2
18   67   15.9
19   68   18.2
20   69   21.2
21   70   4.3
22   71   2.4
23   72   2.7073
24   73   2.9
```

To enter a data set into *Arc*, use the following template *new.lsp*.

```
dataset=new
begin description
Artificial data.
Contributed by David Olive.
end description
begin variables
col 0 = x1
col 1 = x2
col 2 = x3
col 3 = y
end variables
begin data
```

Next open *new.lsp* in *Notepad*. (Or use the *vi* editor in Unix. Sophisticated editors like *Word* will often work, but they sometimes add things like page breaks that do not allow the statistics software to use the file.) Then copy the data lines from *R/Splus* and paste them below *new.lsp*. Then modify the file *new.lsp* and save it on a disk as the file *belg.lsp*. (Or save it in *mdata* where *mdata* is a data folder added within the *Arc data* folder.) The header of the new file *belg.lsp* is shown on the next page.

```
dataset=belgium
begin description
Belgium telephone data from
Rousseeuw and Leroy (1987, p. 26)
end description
begin variables
col 0 = case
col 1 = x = year
col 2 = y = number of calls in tens of millions
end variables
begin data
1 50 0.44
.   .   .
.   .   .
.   .   .
24 73 2.9
```

The file above also shows the first and last lines of data. The header file needs a data set name, description, variable list and a *begin data* command. Often the description can be copied and pasted from source of the data, eg from the STATLIB website. Note that the first variable starts with *Col 0*.

To transfer a data set from Arc to R or Splus, select the item "Display data" from the dataset's menu. Select the variables you want to save, and then push the button for "Save in R/Splus format." You will be prompted to give a file name. If you select *bodfat*, then two files *bodfat.txt* and *bodfat.Rd* will be created. The file *bodfat.txt* can be read into either *R* or *Splus* using the *read.table* command. The file *bodfat.Rd* saves the documentation about the data set in a standard format for *R*.

As an example, the following command was used to enter the body fat data into *Splus*. (The *mdata folder* does not come with *Arc*. The folder needs to be created and filled with files from the book's website. Then the file *bodfat.txt* can be stored in the mdata folder.)

```
bodfat <- read.table("C:\\ARC\\DATA\\MDATA\\BODFAT.TXT",header=T)
bodfat[,16] <- bodfat[,16]+1
```

The last column of the body fat data consists of the case numbers which start with 0 in *Arc*. The second line adds one to each case number.

As another example, use the menu commands
"File>Load>Data>Arcg>forbes.lsp" to activate the forbes data set. From
the *Forbes* menu, select *Display Data*. A window will appear. Double click
on *Temp* and *Pressure*. Click on *Save Data in R/Splus Format* and save as
forbes.txt in the folder *mdata*.

Enter *Splus* and type the following command.

```
forbes<-read.table("C:\\ARC\\DATA\\ARCG\\FORBES.TXT",header=T)
```

The command *forbes* will display the data set.

Getting information about a library in R

In *R*, a *library* is an add–on package of *R* code. The command *library()*
lists all available libraries, and information about a specific library, such as
`MASS` for robust estimators like `cov.mcd` or `ts` for time series estimation, can
be found, eg, with the command *library(help=MASS)*.

Downloading a library into R

Many researchers have contributed a *library* of *R* code that can be down-
loaded for use. To see what is available, go to the website
(http://cran.us.r-project.org/) and click on the Packages icon. Suppose you
are interested the Weisberg (2002) dimension reduction library *dr*. Scroll
down the screen and click on *dr*. Then click on the file corresponding to your
type of computer, eg *dr 2.0.0.zip* for *Windows*. My unzipped files are stored
in my directory

```
C:\unzipped.
```

The file

```
C:\unzipped\dr
```

contains a folder *dr* which is the *R library*. Cut this folder and paste it into
the *R* library folder. (On my computer, I store the folder *rw1011* in the file

```
C:\R-Gui.
```

The folder

```
C:\R-Gui\rw1011\library
```

contains the library packages that came with R.) Open R and type the following command.

    library(dr)

Next type *help(dr)* to make sure that the library is available for use.

Warning: R is free but not fool proof. If you have an old version of R and want to download a library, you may need to update your version of R. The libraries for robust statistics may be useful for outlier detection, but the methods have not been shown to be consistent or high breakdown. All software has some bugs. For example, Version 1.1.1 (August 15, 2000) of R had a random generator for the Poisson distribution that produced variates with too small of a mean θ for $\theta \geq 10$. Hence simulated 95% confidence intervals might contain θ 0% of the time. This bug seems to have been fixed in Version 2.4.1.

14.3   Projects

Straightforward Projects

- Compare the response transformation method illustrated in Example 1.5 with the method given in Section 5.1. Use simulations and real data.

- Investigate the approximations for MED(Y) and MAD(Y) for Gamma data. See Table 2.3.

- Application 2.2 suggests using $U_n = n - L_n$ where $L_n = \lfloor n/2 \rfloor - \lceil \sqrt{n/4} \rceil$ and

$$SE(\text{MED}(n)) = 0.5(Y_{(U_n)} - Y_{(L_n+1)}).$$

  Then use the t_p approximation with $p = U_n - L_n - 1 \approx \lceil \sqrt{n} \rceil$.

  Run a simulation to compare a 95% CI with this interval and a 95% CI that uses
$$SE(\text{MED}(n)) = 0.5(Y_{(U_n)} - Y_{(L_n)})$$

  with $z_{1-\alpha/2}$ instead of $t_{p,1-\alpha/2}$.

- Find a useful technique in Chambers, Cleveland, Kleiner and Tukey (1983), Cook (1998a) or Cook and Weisberg (1999a) that was not presented in this course. Analyze a real data set with the technique.

- Read Stigler (1977). This paper suggests a method for comparing new estimators. Use this method with the two stage estimators $T_{S,n}$ and $T_{A,n}$ described in Section 2.6.

- Read Anscombe (1961) and Anscombe and Tukey (1963). These papers suggest graphical methods for checking multiple linear regression and experimental design methods that were the "state of the art" at the time. What graphical procedures did they use and what are the most important procedures that were not suggested?

- Read Bentler and Yuan (1998) and Cattell (1966). These papers use scree plots to determine how many eigenvalues of the covariance matrix are nonzero. This topic is very important for dimension reduction methods such as principal components.

- The simulation study in Section 4.6 suggests that $T_{S,n}$ does not work well on exponential data. Find a coarse grid so that $T_{S,n}$ works well normal and exponential data. Illustrate with a simulation study.

- Examine via simulation how the graphical method for assessing variable selection complements numerical methods. Find at least two data sets where deleting one case changes the model selected by a numerical variable selection method such as C_p.

- Are numerical diagnostics such as Cook's distance needed? Examine whether Cook's distance can detect influential points that can not be found using the OLS response plot. Are there data sets where the response plot is more effective?

- Are robust estimators needed for multiple linear regression? Examine whether using the OLS response plot is as effective as robust methods for detecting outliers.

- Find some benchmark multiple linear regression outlier data sets. Fit OLS, L_1 and M-estimators from *R/Splus*. Are any of the M-estimators as good as L_1? (Note: `l1fit` is in *Splus* but not in *R*.)

- Compare `lmsreg` and the MBA regression estimator on real and simulated multiple linear regression data.

- Find some benchmark multiple linear regression outlier data sets. Fit robust estimators such as `ltsreg` from *R/Splus*, but do not use `lmsreg`. Are any of the robust estimators as good as the MBA estimator?

- Make a graphical version of the Durbin-Watson test for dependent errors in multiple linear regression.

- There are several papers that give tests or diagnostics for linearity. Find a data set from such a paper and find the fitted values from some nonparametric method. Plot these fitted values versus the fitted values from a multiple linear regression such as OLS. What should this plot look like? How can the response plot and trimmed views be used as a diagnostic for linearity? See Hawkins and Olive (2002, p. 158).

- *R/Splus* provides several regression functions for examining data when the multiple linear regression model is not appropriate such as projection pursuit regression and methods for time series. Examine the FY plot of Section 6.4 for such methods. Generate outlier data and check whether the outliers can be found with the FY plot. Run the `rpack` function `fysim` and include the output and last plot in *Word*.

- Remark 10.3 estimates the percentage of outliers that the FMCD algorithm can tolerate. At the end of Section 10.7, data is generated such that the FMCD estimator works well for $p = 4$ but fails for $p = 8$. Generate similar data sets for $p = 8, 9, 10, 12, 15, 20, 25, 30, 35, 40, 45,$ and 50. For each value of p find the smallest integer valued percentage of outliers needed to cause the FMCD and FCH estimators to fail. Use the `rpack` function `concsim`. If `concsim` is too slow for large p, use `covsim2` which will only give counts for the fast FCH estimator. As a criterion, a count ≥ 16 is good. Compare these observed FMCD percentages with Remark 10.3 (use the `gamper2` function). Do not forget the *library(MASS)* command if you use *R*.

- DD plots: compare classical–FCH vs classical–cov.mcd DD plots on real and simulated data. Do problems 10.14, 11.2 and 11.3 but with a wider variety of data sets, n, p and gamma.

- Many papers substitute the latest MCD (or LTS) algorithm for the classical estimator and have titles like "Fast and Robust Factor Anal-

ysis." Find such a paper (see Section 11.4) that analyzes a data set on

i) factor analysis,

ii) discriminant analysis,

iii) principal components,

iv) canonical correlation analysis,

v) Hotelling's t test, or

vi) principal component regression.

For the data, make a scatterplot matrix of the classical, FCH and FMCD Mahalanobis distances. Delete any outliers and run the classical procedure on the undeleted data. Did the paper's procedure perform as well as this procedure?

- Examine the DD plot as a diagnostic for multivariate normality and elliptically contoured distributions. Use real and simulated data.

- Resistant regression: modify `tvreg` by using OLS–covfch instead of OLS–cov.mcd. (L_1–cov.mcd and L_1–covfch are also interesting.) Compare your function with `tvreg`. The `tvreg` and `covfch` functions are in *rpack.txt*.

- *Using ESP to Search for the Missing Link*: Compare `trimmed views` which uses OLS and `cov.mcd` with another regression–MLD combo. There are 8 possible projects: i) OLS–FCH, ii) OLS–Classical (use `ctrviews`), iii) SIR–cov.mcd (`sirviews`), iv) SIR–FCH, v) SIR–classical, vi) lmsreg–cov.mcd (`lmsviews`), vii) lmsreg–FCH, and viii) lmsreg –classical. Do Problem 12.7ac (but just copy and paste the best view instead of using the essp(nx,ncuby,M=40) command) with both your estimator and `trimmed views`. Try to see what types of functions work for both estimators, when `trimmed views` is better and when the procedure i)–viii) in better. If you can invent interesting 1D functions, do so.

- Many 1D regression models where Y_i is independent of $\boldsymbol{x}_i$ given the sufficient predictor $\boldsymbol{x}_i^T \boldsymbol{\beta}$ can be made resistant by making EY plots of the estimated sufficient predictor $\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}$ versus Y_i for the 10 trimming

proportions. Since 1D regression is the study of the conditional distribution of Y_i given $\boldsymbol{x}_i^T \boldsymbol{\beta}$, the EY plot is used to visualize this distribution and needs to be made anyway. See how well trimmed views work when outliers are present.

- Investigate using trimmed views to make various procedures such as sliced inverse regression resistant against the presence of nonlinearities. The functions *sirviews*, *drsim5*, *drsim6* and *drsim7* in *rpack.txt* may be useful.

- Examine the method of variable selection for 1D regression models suggested in Section 12.4.

- The DGK estimator with 66% coverage should be able to tolerate a cluster of about 30% extremely distant outliers. Compare the DGK estimators with 50% and 66% coverage for various outlier configurations.

Harder Projects

- Which estimator is better FCH, RFCH, CMBA or RCMBA?

- For large data sets, make the DD plot of the DGK estimator vs MB estimator and the DD plot of the classical estimator versus the MB estimator. Which DD plot is more useful? Does your answer depend on n and p? These two plots are among the fastest outlier diagnostics for multivariate data.

- Resampling algorithms such as the bootstrap, jackknife and permutation tests draw B_n random samples from the set of all bootstrap samples, all jackknife samples or all permutations. A statistic T_n is computed from each sample resulting in B_n statistics. If H_n is the cdf of the statistic T_n computed from all possible samples, then the sample of B_n statistics is often used to estimate the α_1 and α_2 percentiles ξ_{α_i} of H_n where $P_{H_n}(T_n \leq \xi_{\alpha_i}) = \alpha_i$ and $\alpha_1 + 1 - \alpha_2 = \alpha$. Use $\alpha = 0.05$ and the SHORTH estimator on the B_n values of T_n to estimate ξ_{α_i} in the same way that Olive (2007) used the SHORTH estimator to estimate percentiles for prediction intervals.

- Olive (2007) gives a technique for finding asymptotically optimal $100(1-\alpha)\%$ prediction intervals for regression models of the form $Y = m(\boldsymbol{x}, \boldsymbol{\beta}) + e$ where the errors are iid with zero mean and constant variance. The intervals tend to be too short for finite n. Try to get good simulated coverage for moderate n by using an asymptotically conservative $100(1 - \alpha/2)\%$ PI in place of the $100(1 - \alpha)\%$ PI. So use a 95% PI if a 90% PI is desired and use a 97.5% PI if a 95% PI is desired.

- *The Super Duper Outlier Scooper for MLR:* Consider the MLR algorithm from Theorem 8.8 that uses LTS concentration steps to create attractors as well. OLS and a high breakdown estimator are also used as attractors. The attractors can be screened with either the LMS or the LTS criterion. Which criterion results in a better estimator? Write *R/Splus* functions to compute the two estimators. Compare these estimators with `lmsreg` and `ltsreg` on real and simulated data.

- *The Super Duper Outlier Scooper for Multivariate Location and Dispersion:* Consider the modified MBA estimator for multivariate location and dispersion given in Problem 10.18. This MBA estimator uses 8 starts using 0%, 50%, 60%, 70%, 80%, 90%, 95% and 98% trimming of the cases closest to the coordinatewise median in Euclidean distance. The estimator is $\sqrt{n}$ consistent on elliptically contoured distributions with nonsingular covariance matrix. For small data sets the *cmba2* function can fail because the covariance estimator applied to the closest 2% cases to the coordinatewise median is singular. Modify the function so that it works well on small data sets. Then consider the following proposal that may make the estimator asymptotically equivalent to the classical estimator when the data are from a multivariate normal (MVN) distribution. The attractor corresponding to 0% trimming is the DGK estimator $(\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0)$. Let $(\hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\Sigma}}_T) = (\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0)$ if $det(\hat{\boldsymbol{\Sigma}}_0) \leq det(\hat{\boldsymbol{\Sigma}}_M)$ and $(\hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\Sigma}}_T) = (\hat{\boldsymbol{\mu}}_M, \hat{\boldsymbol{\Sigma}}_M)$ otherwise where $(\hat{\boldsymbol{\mu}}_M, \hat{\boldsymbol{\Sigma}}_M)$ is the attractor corresponding to $M\%$ trimming. Then make the DD plot of the classical Mahalanobis distances versus the distances corresponding to $(\hat{\mu}_T, \hat{\boldsymbol{\Sigma}}_T)$ for $M = 50, 60, 70, 80, 90, 95$ and 98. If all seven DD plots "look good" then use the classical estimator. The resulting estimator will be asymptotically equivalent to the classical estimator if P(all seven DD plots "look good") goes to one as $n \rightarrow \infty$. We conjecture that all seven plots will look good because if n is large

and the trimmed attractor "beats" the DGK estimator, then the plot will look good. Also if the data is MVN but not spherical, then the DGK estimator will almost always "beat" the trimmed estimator, so all 7 plots will be identical.

- The TV estimator for MLR has a good combination of resistance and theory. Consider the following modification to make the method asymptotically equivalent to OLS when the Gaussian model holds: if each trimmed view "looks good," use OLS. The method is asymptotically equivalent to OLS if the probability P(all 10 trimmed views look good) goes to one as $n \to \infty$. Rousseeuw and Leroy (1987, p. 128) shows that if the predictors are bounded, then the ith residual r_i converges in probability to the ith error e_i for $i = 1, ..., n$. Hence all 10 trimmed views will look like the OLS view with high probability if n is large.

- Modify the trimmed views estimator for resistant logistic regression. Make an ESS plot for each of the trimming proportions with the logistic curve and step function of observed proportions added to the plot. The `rpack` function `lressp` may be useful.

- Modify the trimmed views estimator for resistant Poisson regression. Make an EY plot for each of the trimming proportions with the exponential curve and lowess curve added to the plot. The `rpack` function `llressp` may be useful.

- Try to robustify the discriminant function estimators for binary regression given in Definition 13.4 by replacing the classical estimator of multivariate location and dispersion by the FCH or FMCD estimator.

- Modify the minimum chi–square estimator to make a resistant Poisson regression estimator by replacing OLS by a resistant regression estimator such as `tvreg`, `mbareg` or `lmsreg`. The `rpack` function `llrwtfrp` may be useful.

- For nonlinear regression models of the form $y_i = m(\boldsymbol{x}_i, \boldsymbol{\beta}) + e_i$, the fitted values are $\hat{y}_i = m(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}})$ and the residuals are $r_i = y_i - \hat{y}_i$. The points in the FY plot of the fitted values versus the response should follow the identity line. The TV estimator would make FY and residual plots for each of the trimming proportions. The MBA estimator with the

median squared residual criterion can also be used for many of these models.

- A useful plot for 1D binary regression is the binary response plot of the first SIR direction versus the second SIR direction. Cases with $y = 0$ are plotted with an open circle while cases with $y = 1$ are plotted with a cross. If the 1D binary regression model holds and if the first SIR direction is a useful estimated sufficient predictor, then the symbol density in any narrow vertical strip is approximately constant. See Cook (1998a, ch. 5), Cook and Lee (1999) and Cook and Weisberg (1999a, section 22.2). In analogy with trimmed views, use trimming to make ten binary response plots.

- Econometrics project: Suppose that the MLR model holds but $\mathrm{Var}(\boldsymbol{e}) = \sigma^2 \boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{U}'$ where $\boldsymbol{U}$ is known and nonsingular. Show that $\boldsymbol{U}^{-1}\boldsymbol{Y} = \boldsymbol{U}^{-1}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U}^{-1}\boldsymbol{e}$, and the TV and MBA estimators can be applied to $\tilde{\boldsymbol{Y}} = \boldsymbol{U}^{-1}\boldsymbol{Y}$ and $\tilde{\boldsymbol{X}} = \boldsymbol{U}^{-1}\boldsymbol{X}$ provided that OLS is fit without an intercept.

- Econometrics project: Modify the MBA estimator for time series by choosing cases that are close together in time. For example, if the time series is $y_1, y_2, ..., y_{1000}$ and if y_{100} is a center, then the 10% of cases closest to y_{100} in time are (roughly) $y_{50}, ..., y_{150}$.

- Agresti (2002, p. 109) states that a confidence interval for $\mu_1 - \mu_2$ based on single sample estimates $\hat{\mu}_i$ and confidence intervals (L_i, U_i) for $i = 1, 2$ is

$$\left(\hat{d} - \sqrt{(\hat{\mu}_1 - L_1)^2 + (U_2 - \hat{\mu}_2)^2}, \ \ \hat{d} + \sqrt{(U_1 - \hat{\mu}_1)^2 + (\hat{\mu}_2 - L_2)^2} \right)$$

where $\hat{d} = \hat{\mu}_1 - \hat{\mu}_2$. This method is used when μ_i is a proportion or odds ratio. Try the method when μ_i is a mean and compare this method to Welch intervals given by Remark 2.2.

- Compare outliers and missing values, especially missing and outlying at random. See Little and Rubin (2002).

- Suppose that the data set contains missing values. Code the missing value as $\pm 99999 +$ rnorm(1). Run a robust procedure on the data. The idea is that the case with the missing value will be given weight zero if

497

the variable is important, and the variable will be given weight zero if the case is important. See Hawkins and Olive (1999b).

- Econometrics project: Let $w_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$ be the fitted values for the L_1 estimator. Apply regression quantiles (see Koenker, 2005) to the response and w_i and plot the result. When is this technique competitive with the usual regression quantiles method?

- Read Stefanski and Boos (2002). One of the most promising uses of M-estimators is as generalized estimating equations.

- Download the `dr` function for R, (contributed by Sanford Weisberg), and make PHD and SAVE trimmed views.

- Example 1.4 illustrates a robust prediction interval for multiple linear regression. Run a simulation study to compare the simulated coverage proportion with the nominal coverage.

- Robust sequential procedures do not seem to work very well. Try using analogs of the two stage trimmed means. An ad hoc procedure that has worked very well is to clean the data using the median and mad at each sample size. Then apply the classical sequential method and stopping rule to the cleaned data. This procedure is rather expensive since the median and mad need to be recomputed with each additional observation until the stopping rule ends data collection. Another idea is to examine similar methods in the quality control literature.

- Try to make nonparametric prediction intervals for multiple linear regression by finding ordering the residuals and taking the "shortest interval" containing 90% of the residuals where shortest is in the sense of LMS, LTS or LTA. See Di Bucchianico, Einmahl and Mushkudiani (2001) and Olive (2007). The functions `piplot` and `pisim` in *rpack.txt* may be useful.

- See if swapping with elemental sets is a good technique.

- Apply the Cook and Olive (2001) graphical procedure for response transformations described in Section 5.1 with the power family replaced by the Yeo and Johnson (2000) family of transformations.

Research Ideas that have Confounded the Author

- If the attractor of a randomly selected elemental start is (in)consistent, then FMCD and FLTS are (in)consistent. Hawkins and Olive (2002) showed that the attractor is inconsistent if k concentration steps are used. Suppose K elemental starts are used for an LTS or MCD concentration estimator and that the starts are iterated until convergence instead of for k steps. Prove or disprove the conjecture that the resulting estimator is inconsistent. (Intuitively, the elemental starts are inconsistent and hence are tilted away from the parameter of interest. Concentration may reduce but probably does not eliminate the tilt.)

- Prove that applying an LTA concentration step results in an estimator with the same rate as the start.

- Prove Conjecture 7.1: the LTA estimator is consistent and $O_p(n^{-1/2})$.

- Do elemental set and concentration algorithms for MLR give consistent estimators if the number of starts increases to ∞ with the sample size n? For example, prove or disprove Conjecture 8.1. (Algorithms that use a fixed number of elemental sets along with the classical estimator and a biased but easily computed high breakdown estimator will be easier to compute and have better statistical properties. See Theorem 8.8 and Olive and Hawkins, 2007bc.)

- Prove or disprove Conjecture 11.1. Do elemental set and concentration algorithms for multivariate location and dispersion (MLD) give consistent estimators if the number of starts increases to ∞ with the sample size n? (Algorithms that use a fixed number of elemental sets along with the classical estimator and a biased but easily computed high breakdown estimator will be easier to compute and have better statistical properties. See Theorem 10.15 and Olive and Hawkins, 2007b, 2008.)

It is easy to create consistent algorithm estimators that use $O(n)$ randomly chosen elemental sets. He and Wang (1997) show that the all elemental subset approximation to S estimators for MLD is consistent for $(\boldsymbol{\mu}, c\boldsymbol{\Sigma})$. Hence an algorithm that randomly draws $g(n)$ elemental sets and searches all $C(g(n), p+1)$ elemental sets is also consistent if $g(n) \to \infty$ as $n \to \infty$. For example, $O(n)$ elemental sets are used if $g(n) \propto n^{1/(p+1)}$.

When a fixed number of K elemental starts are used, the best attractor is inconsistent but gets close to $(\boldsymbol{\mu}, c_{MCD}\boldsymbol{\Sigma})$ if the data distribution is EC. (The estimator may be unbiased but the variability of the component estimators does not go to 0 as $n \to \infty$.) If $K \to \infty$, then the best attractor should approximate the highest density region arbitrarily closely and the algorithm should be consistent. However, the time for the algorithm greatly increases, the convergence rate is very poor (possibly between $K^{1/2p}$ and $K^{1/p}$), and the elemental concentration algorithm can not guarantee that the determinant is bounded when outliers are present.

- A promising two stage estimator is the "cross checking estimator" that uses a standard consistent estimator and an alternative consistent estimator with desirable properties such as a high breakdown value. The final estimator uses the standard estimator if it is "close" to the alternative estimator, and hence is asymptotically equivalent to the standard estimator for clean data. One important area of research for robust statistics is finding good computable consistent robust estimators to be used in plots and in the cross checking algorithm. The estimators given in Theorems 10.14 and 10.15 (see Olive 2004a and Olive and Hawkins 2007b, 2008) finally make the cross checking estimator practical, but better estimators are surely possible. He and Wang (1996) suggested the cross checking idea for multivariate location and dispersion, and additional applications are given in He and Fung (1999).

  For MLR, cross checking is not needed since Theorem 8.8 and Remark 8.7 provide a better way for making a HB MLR estimator asymptotically equivalent to an efficient MLR estimator.

14.4 Hints for Selected Problems

Chapter 1

1.1 $\|r_{i,1} - r_{i,2}\| = \|Y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_1 - (Y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_2)\| = \|\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_2 - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_1\| = \|\hat{Y}_{2,i} - \hat{Y}_{1,i}\| = \|\hat{Y}_{1,i} - \hat{Y}_{2,i}\|$.

1.2 The plot should be similar to Figure 1.6, but since the data is simulated, may not be as smooth.

1.3 c) The histograms should become more like a normal distribution as n increases from 1 to 200. In particular, when $n = 1$ the histogram should be right skewed while for $n = 200$ the histogram should be nearly symmetric. Also the scale on the horizontal axis should decrease as n increases.

d) Now $\overline{Y} \sim N(0, 1/n)$. Hence the histograms should all be roughly symmetric, but the scale on the horizontal axis should be from about $-3/\sqrt{n}$ to $3/\sqrt{n}$.

1.4 e) The plot should be strongly nonlinear, having "V" shape.

1.5 You could save the data set from the text's website on a disk, and then open the data in *Arc* from the disk.

c) Most students should delete cases 5, 47, 75, 95, 168, 181, and 199.

f) The forward response plot looks like a line while the residual plot looks like a curve. A residual plot emphasizes lack of fit while the forward response plot emphasizes goodness of fit.

h) The quadratic model looks good.

Chapter 2

2.2. $F_W(w) = P(W \le w) = P(Y \le w - \mu) = F_Y(w - \mu)$. So $f_W(w) = \frac{d}{dw}F_Y(w - \mu) = f_Y(w - \mu)$.

2.3. $F_W(w) = P(W \le w) = P(Y \le w/\sigma) = F_Y(w/\sigma)$. So $f_W(w) = \frac{d}{dw}F_Y(w/\sigma) = f_Y(w/\sigma)\frac{1}{\sigma}$.

2.4. $F_W(w) = P(W \le w) = P(\sigma Y \le w - \mu) = F_Y(\frac{w-\mu}{\sigma})$. So $f_W(w) = \frac{d}{dw}F_Y(\frac{w-\mu}{\sigma}) = f_Y(\frac{w-\mu}{\sigma})\frac{1}{\sigma}$.

2.5 $N(0, \sigma_M^2)$

2.9 a) $8.25 \pm 0.7007 = (6.020, 10.480)$

b) $8.75 \pm 1.1645 = (7.586, 9.914)$.

2.10 a) $\overline{Y} = 24/5 = 4.8$.

b)
$$S^2 = \frac{138 - 5(4.8)^2}{4} = 5.7$$

so $S = \sqrt{5.7} = 2.3875$.

c) The ordered data are 2,3,5,6,8 and $\text{MED}(n) = 5$.

d) The ordered $|Y_i - \text{MED}(n)|$ are 0,1,2,2,3 and $\text{MAD}(n) = 2$.

2.11 a) $\overline{Y} = 15.8/10 = 1.58$.

b)
$$S^2 = \frac{38.58 - 10(1.58)^2}{9} = 1.5129$$

so $S = \sqrt{1.5129} = 1.230$.

c) The ordered data set is 0.0,0.8,1.0,1.2,1.3,1.3,1.4,1.8,2.4,4.6 and $\text{MED}(n) = 1.3$.

d) The ordered $|Y_i - \text{MED}(n)|$ are 0,0,0.1,0.1,0.3,0.5,0.5,1.1,1.3,3.3 and $\text{MAD}(n) = 0.4$.

e) 4.6 is unusually large.

2.12 a) $S/\sqrt{n} = 3.2150$.

b) $n - 1 = 9$.

c) 94.0

d) $L_n = \lfloor n/2 \rfloor - \lceil \sqrt{n/4} \rceil = \lfloor 10/2 \rfloor - \lceil \sqrt{10/4} \rceil = 5 - 2 = 3$.

e) $U_n = n - L_n = 10 - 3 = 7$.

f) $p = U_n - L_n - 1 = 7 - 3 - 1 = 3$.

g) $\text{SE}(\text{MED}(n)) = (Y_{(U_n)} - Y_{(L_n+1)})/2 = (95 - 90.0)/2 = 2.5$.

2.13 a) $L_n = \lfloor n/4 \rfloor = \lfloor 2.5 \rfloor = 2$.

b) $U_n = n - L_n = 10 - 2 = 8$.

c) $p = U_n - L_n - 1 = 8 - 2 - 1 = 5$.

d) $(89.7 + 90.0 + \cdots + 95.3)/6 = 558/6 = 93.0$.

e) 89.7 89.7 89.7 90.0 94.0 94.0 95.0 95.3 95.3 95.3

f) $(\sum d_i)/n = 928/10 = 92.8$.

g) $(\sum d_i^2 - n(\overline{d})^2)/(n-1) = (86181.54 - 10(92.8)^2)/9 = 63.14/9 = 7.0156$.

e)
$$V_{SW} = \frac{S_n^2(d_1, ..., d_n)}{([U_n - L_n]/n)^2} = \frac{7.0156}{(\frac{8-2}{10})^2} = 19.4877,$$

so
$$SE(T_n) = \sqrt{V_{SW}/n} = \sqrt{19.4877/10} = 1.3960.$$

2.14 a) $L_n = \lfloor n/2 \rfloor - \lceil \sqrt{n/4} \rceil = \lfloor 5/2 \rfloor - \lceil \sqrt{5/4} \rceil = 2 - 2 = 0$.

$U_n = n - L_n = 5 - 0 = 5.$

$p = U_n - L_n - 1 = 5 - 0 - 1 = 4.$

$SE(MED(n)) = (Y_{(U_n)} - Y_{(L_n+1)})/2 = (8-2)/2 = 3.$

b) $L_n = \lfloor n/4 \rfloor = \lfloor 1 \rfloor = 1.$

$U_n = n - L_n = 5 - 1 = 4.$

$p = U_n - L_n - 1 = 4 - 1 - 1 = 2.$

$T_n = (3 + 5 + 6)/3 = 4.6667.$

The $d's$ are 3 3 5 6 6.

$(\sum d_i)/n = 4.6$

$(\sum d_i^2 - n(\overline{d})^2)/(n-1) = (115 - 5(4.6)^2)/4 = 9.2/4 = 2.3.$

$$V_{SW} = \frac{S_n^2(d_1, ..., d_n)}{([U_n - L_n]/n)^2} = \frac{2.3}{(\frac{4-1}{5})^2} = 6.3889,$$

so

$$SE(T_n) = \sqrt{V_{SW}/n} = \sqrt{6.3889/5} = 1.1304.$$

The *R/Splus* functions for Problems 2.15–2.29 are available from the text's website file *rpack.txt* and should have been entered into the computer using the *source("A:/rpack.txt")* as described on p. 482-483.

2.16 Simulated data: a) about 0.669   b) about 0.486.

2.17 Simulated data: a) about 0.0   b) $\overline{Y} \approx 1.00$ and $T_n \approx 0.74.$

2.21 Simulated data gives about (1514,1684).

2.22 Simulated data gives about (1676,1715).

2.23 Simulated data gives about (1679,1712).

Chapter 3

3.2 a) $F(y) = 1 - \exp(-y/\lambda)$ for $y \geq 0$. Let $M = MED(Y) = \log(2)\lambda.$ Then $F(M) = 1 - \exp(-\log(2)\lambda/\lambda) = 1 - \exp(-\log(2)) = 1 - \exp(\log(1/2)) = 1 - 1/2 = 1/2.$

b) $F(y) = \Phi([\log(y) - \mu]/\sigma)$ for $y > 0$. Let $M = MED(Y) = \exp(\mu).$ Then $F(M) = \Phi([\log(\exp(\mu)) - \mu]/\sigma) = \Phi(0) = 1/2.$

3.3 a) $M = \mu$ by symmetry. Since $F(U) = 3/4$ and $F(y) = 1/2 + (1/\pi)\arctan([y - \mu]/\sigma)$, want $\arctan([U - \mu]/\sigma) = \pi/4$ or $(U - \mu)/\sigma = 1.$ Hence $U = \mu + \sigma$ and $MAD(Y) = D = U - M = \mu + \sigma - \mu = \sigma.$

b) $M = \theta$ by symmetry. Since $F(U) = 3/4$ and $F(y) = 1 - 0.5 \exp(-[y - \theta]/\lambda)$ for $y \geq 0$, want $0.5 \exp(-[U - \theta]/\lambda) = 0.25$ or $\exp(-[U - \theta]/\lambda) = 1/2$. So $-(U - \theta)/\lambda = \log(1/2)$ or $U = \theta - \lambda \log(1/2) = \theta - \lambda(-\log(2)) = \theta + \lambda \log(2)$. Hence $\mathrm{MAD}(Y) = D = U - M = U - \theta = \lambda \log(2)$.

3.4. f) $E(Y^r) = E(e^{rX}) = m_X(r) = \exp(r\mu + r^2\sigma^2/2)$ where $m_X(t)$ is the mgf of a $N(\mu, \sigma^2)$ random variable. Use $r = 1$.

k) Use the fact that $E(Y^r) = E[(Y^\phi)^{r/\phi}] = E(W^{r/\phi})$ where $W \sim EXP(\lambda)$. Take $r = 1$.

3.5. f) $E(Y^r) = E(e^{rX}) = m_X(r) = \exp(r\mu + r^2\sigma^2/2)$ where $m_X(t)$ is the mgf of a $N(\mu, \sigma^2)$ random variable. Use $r = 1, 2$.

k) Use the fact that $E(Y^r) = E[(Y^\phi)^{r/\phi}] = E(W^{r/\phi})$ where $W \sim EXP(\lambda)$. Use $r = 1, 2$.

3.9 a) $\mathrm{MED}(W) = \sqrt{\lambda \log(2)}$.

3.10 a) $\mathrm{MED}(W) = \theta - \sigma \log(\log(2))$.

b) $\mathrm{MAD}(W) \approx 0.767049\sigma$.

c) Let $W_i = \log(X_i)$ for $i = 1, ..., n$. Then $\hat{\sigma} = \mathrm{MAD}(W_1, ..., W_n)/0.767049$ and $\hat{\theta} = \mathrm{MED}(W_1, ..., W_n) - \hat{\sigma} \log(\log(2))$. So take $\hat{\phi} = 1/\hat{\sigma}$ and $\hat{\lambda} = \exp(\hat{\theta}/\hat{\sigma})$.

3.11 a) $\mathrm{MED}(Y) = \mu$.

b) $\mathrm{MAD}(Y) = 1.17741\sigma$.

3.12 a) $\mathrm{MED}(Y) = \mu + \sigma$.

b) $\mathrm{MAD}(Y) = 0.73205\sigma$.

3.13 Let $\hat{\mu} = \mathrm{MED}(W_1, ..., W_n)$ and $\hat{\sigma} = \mathrm{MAD}(W_1, ..., W_n)$.

3.14 $\mu + \log(3)\sigma$

3.15 a) $\mathrm{MED}(Y) = 1/\phi$
b) $\hat{\tau} = \log(3)/\mathrm{MAD}(W_1, ..., W_n)$ and $\hat{\phi} = 1/\mathrm{MED}(Y_1, ..., Y_n)$.

3.16 $\mathrm{MED}(Y) \approx (p - 2/3)/p \approx 1$ if p is large.

3.21.
$$\mathrm{MED}(Y) = \frac{\sigma}{[\Phi^{-1}(3/4)]^2}.$$

3.22. Let MED(n) and MAD(n) be computed using $W_1, ..., W_n$. Use $-\log(\hat{\tau}) = \text{MED}(n) - 1.440\text{MAD}(n) \equiv A$, so $\hat{\tau} = e^{-A}$. Also $\hat{\lambda} = 2.0781\text{MAD}(n)$.

Chapter 4

4.1 a) 200

b) $0.9(10) + 0.1(200) = 29$

4.2 a) $400(1) = 400$

b) $0.9(10) + 0.1(400) = 49$

The *R/Splus* functions for Problems 4.10–4.14 are available from the text's website file *rpack.txt* and should have been entered into the computer using the *source("A:/rpack.txt")* as described on p. 482-483.

4.13b i) Coverages should be near 0.95. The lengths should be about 4.3 for $n = 10$, 4.0 for $n = 50$ and 3.96 for $n = 100$.

ii) Coverage should be near 0.78 for $n = 10$ and 0 for $n = 50, 100$. The lengths should be about 187 for $n = 10$, 173 for $n = 50$ and 171 for $n = 100$. (It can be shown that the expected length for large n is 169.786.)

Chapter 5

5.1 a) $7 + \beta X_i$
b) $b = \sum (Y_i - 7) X_i / \sum X_i^2$
c) The second derivative $= 2 \sum X_i^2 > 0$.

5.4 $F_o = 0.904$, p–value > 0.1, fail to reject Ho, so the reduced model is good

5.5 a) 25.970

b) $F_o = 0.600$, p–value > 0.5, fail to reject Ho, so the reduced model is good

5.6 a) $b_3 = \sum X_{3i}(Y_i - 10 - 2X_{2i}) / \sum X_{3i}^2$. The second partial derivative $= \sum X_{3i}^2 > 0$.

5.9 a) (1.229, 3.345)
b) (1.0825, 3.4919)

5.11 c) $F_o = 265.96$, pvalue $= 0.0$, reject Ho, there is a MLR relationship between the response variable height and the predictors sternal height and finger to ground.

5.13 No, the relationship should be linear.

5.14 No, since 0 is in the CI. X could be a very useful predictor for Y, eg if $Y = X^2$.

5.16 The model using constant, finger to ground and sternal height is a good candidate. So is the model using constant and sternal height. (You can tell what the variable are by looking at which variables are deleted.)

5.17 Use L3. L1 and L2 have more predictors and higher C_p than L3 while L4 does not satisfy the $C_p \leq 2k$ screen.

5.18 Use L3. L1 has too many predictors. L2 has almost the same summary statistics as L3 but has one more predictor while L4 does not satisfy the $C_p \leq 2k$ screen.

5.19 Use a constant, A, B and C since this is the only model that satisfies the $C_p \leq 2k$ screen.

b) Use the model with a constant and B since it has the smallest C_p and the smallest k such that the $C_p \leq 2k$ screen is satisfied.

5.20 d) The plot should have $\log(Y)$ on the horizontal axis.

e) Since randomly generated data is used, answers vary slightly, but $\widehat{\log(Y)} \approx 4 + X_1 + X_2 + X_3$.

5.22 a) The plot looks roughly like the SW corner of a square.

b) No, the plot is nonlinear.

c) Want to spread small values of y, so make λ smaller. Hence use $y^{(0)} = \log(y)$.

5.23 d) The first assumption to check would be the constant variance assumption.

5.24 Several of the marginal relationships are nonlinear, including $E(M|H)$.

5.25 This problem has the student reproduce Example 5.1. Hence $\log(Y)$ is the appropriate response transformation.

5.26 Plots b), c) and e) suggest that $\log(ht)$ is needed while plots d), f) and g) suggest that $\log(ht)$ is not needed. Plots c) and d) show that the residuals from both models are quite small compared to the fitted values. Plot d) suggests that the two models produce approximately the same fitted values. Hence if the goal is prediction, the expensive $\log(ht)$ measurement does not seem to be needed.

5.27 h) The submodel is ok, but the forward response and residual plots found in f) for the submodel do not look as good as those for the full model found in d). Since the submodel residuals do not look good, more terms are probably needed in the model.

5.30 b) Forward selection gives constant, $(size)^{1/3}$, age, sex, breadth and cause.

c) Backward elimination gives constant, age, cause, cephalic, headht, length and sex.

d) Forward selection is better because it has fewer terms and a smaller C_p.

e) The variables are highly correlated. Hence backward elimination quickly eliminates the single best predictor $(size)^{1/3}$ and can not get a good model that only has a few terms.

f) Although the model in c) could be used, a better model uses constant, age, sex and $(size)^{1/3}$.

j) The FF and RR plots are good and so are the forward response and residual plots if you ignore the good leverage points corresponding to the 5 babies.

Chapter 6

6.1 b) Masking since 3 outliers are good cases with respect to Cook's distances.

c) and d) usually the MBA residuals will be large in magnitude, but for some students MBA, ALMS and ALTS will be highly correlated.

6.4 a) The AR(2) model has the highest correlation with the response and is the simplest model. The top row of the scatterplot matrix gives the FY plots for the 5 different estimators.

b) The AR(11) and AR(12) fits are highly correlated as are the SE-TAR(2,7,2) and SETAR(2,5,2) fits.

6.6 The response Y with a constant and X_3, X_7, X_{13} and X_{14} as predictors is a good submodel. (A competitor would delete X_{13} but then the residual plot is not as good.)

6.8 The response Y with a constant, X_2 and X_5 as predictors is a good submodel. One outlier is visible in the residual plot. (A competitor would also use X_3.)

6.9 The submodel using a constant and X_1 is ok although the residual plot does not look very good.

6.13 The model using $\log(X_3)$, $\log(X_4)$, $\log(X_6)$, $\log(X_{11})$, $\log(X_{13})$ and $\log(X_{14})$ plus a constant has a good FF plot but more variables may be needed to get a good RR plot.

6.14 There are many good models including the submodel that uses $Y = \log(BigMac)$ and a constant, $\log$(BusFare) $\log$(EngSal), $\log$(Service), $\log$(TeachSal) and $\log$(TeachTax) as predictors.

6.16 e) R^2 went from 0.978 with outliers to $R^2 = 0.165$ without the outliers. (The low value of R^2 suggests that the MLR relationship is weak, not that the MLR model is bad.)

Chapter 7

7.4 b) The line should go through the left and right cluster but not through the middle cluster of outliers.

c) The identity line should NOT PASS through the cluster of outliers with Y near 0 and the residuals corresponding to these outliers should be large in magnitude.

7.5 e) Usually the MBA esitmator based on the median squared residual will pass through the outliers with the MBA LATA estimator gives zero weight to the outliers (so that the outliers are large in magnitude).

Chapter 8

8.1 Approximately 2 n^δ $f(0)$ cases have small errors.

Chapter 9

9.3 Adding $\mathbf{1}$ to $\mathbf{Y}$ is equivalent to using $\boldsymbol{u} = (1, 0, ..., 0)^T$ in Equation

(9.7), and the result follows.

Chapter 10

10.1 a) $X_2 \sim N(100, 6)$.

b)
$$\left(\begin{array}{c} X_1 \\ X_3 \end{array} \right) \sim N_2 \left(\left(\begin{array}{c} 49 \\ 17 \end{array} \right), \left(\begin{array}{cc} 3 & -1 \\ -1 & 4 \end{array} \right) \right).$$

c) $X_1 \perp\!\!\!\perp X_4$ and $X_3 \perp\!\!\!\perp X_4$.

d)
$$\rho(X_1, X_2) = \frac{Cov(X_1, X_3)}{\sqrt{\text{VAR}(X_1)\text{VAR}(X_3)}} = \frac{-1}{\sqrt{3}\sqrt{4}} = -0.2887.$$

10.2 a) $Y|X \sim N(49, 16)$ since $Y \perp\!\!\!\perp X$. (Or use $E(Y|X) = \mu_Y + \Sigma_{12}\Sigma_{22}^{-1}(X - \mu_x) = 49 + 0(1/25)(X - 100) = 49$ and $\text{VAR}(Y|X) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = 16 - 0(1/25)0 = 16.$)

b) $E(Y|X) = \mu_Y + \Sigma_{12}\Sigma_{22}^{-1}(X - \mu_x) = 49 + 10(1/25)(X - 100) = 9 + 0.4X.$

c) $\text{VAR}(Y|X) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = 16 - 10(1/25)10 = 16 - 4 = 12.$

10.4 The proof is identical to that given in Example 10.2. (In addition, it is fairly simple to show that $M_1 = M_2 \equiv M$. That is, M depends on Σ but not on c or g.)

10.6 a) Sort each column, then find the median of each column. Then $\text{MED}(\boldsymbol{W}) = (1430, 180, 120)^T$.

b) The sample mean of $(X_1, X_2, X_3)^T$ is found by finding the sample mean of each column. Hence $\overline{\boldsymbol{x}} = (1232.8571, 168.00, 112.00)^T$.

10.11 $\Sigma\boldsymbol{B} = E[E(\boldsymbol{X}|\boldsymbol{B}^T\boldsymbol{X})\boldsymbol{X}^T\boldsymbol{B})] = E(\boldsymbol{M}_B\boldsymbol{B}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{B}) = \boldsymbol{M}_B\boldsymbol{B}^T\Sigma\boldsymbol{B}.$ Hence $\boldsymbol{M}_B = \Sigma\boldsymbol{B}(\boldsymbol{B}^T\Sigma\boldsymbol{B})^{-1}$.

10.15 The 4 plots should look nearly identical with the five cases 61–65 appearing as outliers.

10.16 Not only should none of the outliers be highlighted, but the highlighted cases should be ellipsoidal.

10.17 Answers will vary since this is simulated data, but should get gam near 0.4, 0.3, 0.2 and 0.1 as p increases from 2 to 20.

Chapter 11

11.2 b Ideally the answer to this problem and Problem 11.3b would be nearly the same, but students seem to want correlations to be very high and use n too high. Values of n around 60, 120 and 120 for $p = 2, 3$ and 4 should be enough.

11.3 b Values of n should be near 60, 120 and 120 for $p = 2, 3$ and 4.

11.4 This is simulated data, but for most plots the slope is near 2.

11.8 The identity line should NOT PASS through the cluster of outliers with Y near 0. The amount of trimming seems to vary some with the computer (which should not happen unless there is a bug in the `tvreg2` function or if the computers are using different versions of `cov.mcd`), but most students liked 70% or 80% trimming.

Chapter 12

12.1
a) $\hat{e}_i = Y_i - T(Y)$.
b) $\hat{e}_i = Y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}$.
c)

$$\hat{e}_i = \frac{Y_i}{\hat{\beta}_1 \exp[\hat{\beta}_2 (x_i - \bar{x})]}.$$

d) $\hat{e}_i = \sqrt{w_i}(Y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}})$.

12.2
a) Since Y is a (random) scalar and $E(\boldsymbol{w}) = \boldsymbol{0}$, $\Sigma_{\boldsymbol{x},Y} = E[(\boldsymbol{x} - E(\boldsymbol{x}))(Y - E(Y))^T] = E[\boldsymbol{w}(Y - E(Y))] = E(\boldsymbol{w}Y) - E(\boldsymbol{w})E(Y) = E(\boldsymbol{w}Y)$.

b) Using the definition of z and $\boldsymbol{r}$, note that $Y = m(z) + e$ and $\boldsymbol{w} = \boldsymbol{r} + (\Sigma_{\boldsymbol{x}}\boldsymbol{\beta})\boldsymbol{\beta}^T \boldsymbol{w}$. Hence $E(\boldsymbol{w}Y) = E[(\boldsymbol{r} + (\Sigma_{\boldsymbol{x}}\boldsymbol{\beta})\boldsymbol{\beta}^T \boldsymbol{w})(m(z) + e)] = E[(\boldsymbol{r} + (\Sigma_{\boldsymbol{x}}\boldsymbol{\beta})\boldsymbol{\beta}^T \boldsymbol{w})m(z)] + E[\boldsymbol{r} + (\Sigma_{\boldsymbol{x}}\boldsymbol{\beta})\boldsymbol{\beta}^T \boldsymbol{w}]E(e)$ since e is independent of $\boldsymbol{x}$. Since $E(e) = 0$, the latter term drops out. Since $m(z)$ and $\boldsymbol{\beta}^T \boldsymbol{w}m(z)$ are (random) scalars, $E(\boldsymbol{w}Y) = E[m(z)\boldsymbol{r}] + E[\boldsymbol{\beta}^T \boldsymbol{w} \ m(z)]\Sigma_{\boldsymbol{x}}\boldsymbol{\beta}$.

c) Using result b), $\Sigma_{\boldsymbol{x}}^{-1}\Sigma_{\boldsymbol{x},Y} = \Sigma_{\boldsymbol{x}}^{-1}E[m(z)\boldsymbol{r}] + \Sigma_{\boldsymbol{x}}^{-1}E[\boldsymbol{\beta}^T \boldsymbol{w} \ m(z)]\Sigma_{\boldsymbol{x}}\boldsymbol{\beta} = E[\boldsymbol{\beta}^T \boldsymbol{w} \ m(z)]\Sigma_{\boldsymbol{x}}^{-1}\Sigma_{\boldsymbol{x}}\boldsymbol{\beta} + \Sigma_{\boldsymbol{x}}^{-1}E[m(z)\boldsymbol{r}] = E[\boldsymbol{\beta}^T \boldsymbol{w} \ m(z)]\boldsymbol{\beta} + \Sigma_{\boldsymbol{x}}^{-1}E[m(z)\boldsymbol{r}]$ and the result follows.

d) $E(\boldsymbol{w}z) = E[(\boldsymbol{x} - E(\boldsymbol{x}))\boldsymbol{x}^T\boldsymbol{\beta}] = E[(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{x}^T - E(\boldsymbol{x}^T) + E(\boldsymbol{x}^T))\boldsymbol{\beta}]$
$= E[(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{x}^T - E(\boldsymbol{x}^T))]\boldsymbol{\beta} + E[\boldsymbol{x} - E(\boldsymbol{x})]E(\boldsymbol{x}^T)\boldsymbol{\beta} = \Sigma_{\boldsymbol{x}}\boldsymbol{\beta}$.

e) If $m(z) = z$, then $c(\boldsymbol{x}) = E(\boldsymbol{\beta}^T\boldsymbol{w}z) = \boldsymbol{\beta}^T E(\boldsymbol{w}z) = \boldsymbol{\beta}^T\Sigma_{\boldsymbol{x}}\boldsymbol{\beta} = 1$ by result d).

f) Since z is a (random) scalar, $E(z\boldsymbol{r}) = E(\boldsymbol{r}z) = E[(\boldsymbol{w} - (\Sigma_{\boldsymbol{x}}\boldsymbol{\beta})\boldsymbol{\beta}^T\boldsymbol{w})z]$
$= E(\boldsymbol{w}z) - (\Sigma_{\boldsymbol{x}}\boldsymbol{\beta})\boldsymbol{\beta}^T E(\boldsymbol{w}z)$. Using result d), $E(\boldsymbol{r}z) = \Sigma_{\boldsymbol{x}}\boldsymbol{\beta} - \Sigma_{\boldsymbol{x}}\boldsymbol{\beta}\boldsymbol{\beta}^T\Sigma_{\boldsymbol{x}}\boldsymbol{\beta} = \Sigma_{\boldsymbol{x}}\boldsymbol{\beta} - \Sigma_{\boldsymbol{x}}\boldsymbol{\beta} = \boldsymbol{0}$.

g) Since z and $\boldsymbol{r}$ are linear combinations of $\boldsymbol{x}$, the joint distribution of z and $\boldsymbol{r}$ is multivariate normal. Since $E(\boldsymbol{r}) = \boldsymbol{0}$, z and $\boldsymbol{r}$ are uncorrelated and thus independent. Hence $m(z)$ and $\boldsymbol{r}$ are independent and $\boldsymbol{u}(\boldsymbol{x}) = \Sigma_{\boldsymbol{x}}^{-1} E[m(z)\boldsymbol{r}] = \Sigma_{\boldsymbol{x}}^{-1} E[m(z)]E(\boldsymbol{r}) = \boldsymbol{0}$.

12.4 The submodel I that uses a constant and A, C, E, F, H looks best since it is the minimum $C_p(I)$ model and I has the smallest value of k such that $C_p(I) \leq 2k$.

12.6 a) No strong nonlinearities for MVN data but there should be some nonlinearities present for the non–EC data.

b) The plot should look like a cubic function.

c) The plot should use 0% trimming and resemble the plot in b), but may not be as smooth.

d) The plot should be linear and for many students some of the trimmed views should be better than the OLS view.

e) The EY plot should look like a cubic with trimming greater than 0%.

f) The plot should be linear.

12.7 b) and c) It is possible that none of the trimmed views look much like the sinc(ESP) = sin(ESP)/ESP function.

d) Now at least one of the trimmed views should be good.

e) More lms trimmed views should be good than the views from the other 2 methods, but since simulated data is used, one of the plots from b) or c) could be as good or even better than the plot in d).

Chapter 13

13.2 a) ESP = 1.11108, exp(ESP) = 3.0376 and $\hat{\rho}$ = exp(ESP)/(1 +

511

$\exp(ESP)) = 3.0376/(1 + 3.0376) = 0.7523.$

13.3 $G^2(O|F) = 62.7188 - 13.5325 = 49.1863$, df = 3, p–value = 0.00, reject Ho, there is a LR relationship between ape and the predictors lower jaw, upper jaw and face length.

13.4 $G^2(R|F) = 17.1855 - 13.5325 = 3.653$, df = 1, $0.05 <$ p–value < 0.1, fail to reject Ho, the reduced model is good.

13.5a ESP = 0.2812465 and $\hat{\mu} = \exp(ESP) = 1.3248.$

13.6 $G^2(O|F) = 187.490 - 138.685 = 48.805$, df = 2, p–value = 0.00, reject Ho, there is a LLR relationship between possums and the predictors habitat and stags.

13.8 a) B4

b) EE plot

c) B3 is best. B1 has too many predictors with large Wald p–values, B2 still has too many predictors (want $\leq 300/10 = 30$ predictors) while B4 has too small of a p–value for the change in deviance test.

13.12 a) A good submodel uses a constant, Bar, Habitat and Stags as predictors.

d) The EY and EE plots are good as are the Wald p–values. Also AIC(full) = 141.506 while AIC(sub) = 139.644.

13.14 b) Use the log rule: (max age)/(min age) = 1400 > 10.

e) The slice means track the logistic curve very well if 8 slices are used.

i) The EE plot is linear.

j) The slice means track the logistic curve very well if 8 slices are used.

n) The slice form -0.5 to 0.5 is bad since the symbol density is not approximately constant from the top to the bottom of the slice.

13.15 a) Should have 200 cases, df = 178 and deviance = 112.168.

b) The ESS plot with 12 slices suggests that the full model is good.

c) The submodel I_1 that uses a constant, AGE, CAN, SYS, TYP and FLOC and the submodel I_2 that is the same as I_1 but also uses FRACE seem to be competitors. If the factor FRACE is not used, then the EY plot

follows 3 lines, one for each race. The Wald p–values suggest that FRACE is not needed.

13.16 b) The ESS plot (eg with 4 slices) is bad, so the LR model is bad.

d) Now the ESS plot (eg with 12 slices) is good in that slice smooth and the logistic curve are close where there is data (also the LR model is good at classifying 0's and 1's).

f) The MLE does not exist since there is perfect classification (and the logistic curve can get close to but never equal a discontinuous step function). Hence Wald p–values tend to have little meaning; however, the change in deviance test tends to correctly suggest that there is an LR relationship when there is perfect classification.

For this problem, $G^2(O|F) = 62.7188 - 0.00419862 = 62.7146$, df $= 1$, p–value $= 0.00$, so reject Ho and conclude that there is an LR relationship between ape and the predictor x_3.

13.18 k) The deleted point is certainly influential. Without this case, there does not seem to be a LLR relationship between the predictors and the response.

m) The weighted residual plot suggests that something is wrong with the model since the plotted points scatter about a line with positive slope rather than a line with 0 slope. The deviance residual plot does not suggest that anything is wrong with the model.

13.19 The ESS plot should look ok, but the function uses a default number of slices rather than allowing the user to select the number of slices using a "slider bar" (a useful feature of *Arc*).

13.20 a) Since this is simulated LLR data, the EY plot should look ok, but the function uses a default lowess smoothing parameter rather than allowing the user to select smoothing parameter using a "slider bar" (a useful feature of *Arc*).

b) The data should the identity line in the weighted forward response plots. In about 1 in 20 plots there will be a very large count that looks like an outlier. The weighted residual plot based on the MLE usually looks better than the plot based on the minimum chi-square estimator (the MLE plot tend to have less of a "left opening megaphone shape").

13.21 a)

```
Number in Model  Rsquare C(p)   Variables in model
      6           0.2316  7.0947 X3 X4 X6 X7 X9 X10
```

c) The slice means follow the logistic curve fairly well with 8 slices.

e) The EE plot is linear.

f) The slice means follow the logistic curve fairly well with 8 slices.

14.5 Tables

Tabled values are F(0.95,k,d) where $P(F < F(0.95, k, d)) = 0.95$.
00 stands for ∞. Entries produced with the `qf(.95,k,d)` command in R.
The numerator degrees of freedom are k while the denominator degrees of freedom are d.

k d	1	2	3	4	5	6	7	8	9	00
1	161	200	216	225	230	234	237	239	241	254
2	18.5	19.0	19.2	19.3	19.3	19.3	19.4	19.4	19.4	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.41
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	1.84
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	1.62
00	3.84	3.00	2.61	2.37	2.21	2.10	2.01	1.94	1.88	1.00

Tabled values are $t_{\alpha,d}$ where $P(t < t_{\alpha,d}) = \alpha$ where t has a t distribution with d degrees of freedom. If $d > 30$ use the $N(0,1)$ cutoffs given in the second to last line with $d = Z = \infty$.

alpha	0.95	0.975	0.995
d			
1	6.314	12.706	63.657
2	2.920	4.303	9.925
3	2.353	3.182	5.841
4	2.132	2.776	4.604
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.895	2.365	3.499
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
11	1.796	2.201	3.106
12	1.782	2.179	3.055
13	1.771	2.160	3.012
14	1.761	2.145	2.977
15	1.753	2.131	2.947
16	1.746	2.120	2.921
17	1.740	2.110	2.898
18	1.734	2.101	2.878
19	1.729	2.093	2.861
20	1.725	2.086	2.845
21	1.721	2.080	2.831
22	1.717	2.074	2.819
23	1.714	2.069	2.807
24	1.711	2.064	2.797
25	1.708	2.060	2.787
26	1.706	2.056	2.779
27	1.703	2.052	2.771
28	1.701	2.048	2.763
29	1.699	2.045	2.756
30	1.697	2.042	2.750
Z	1.645	1.960	2.576
CI	90%	95%	99%