```
{\rtf1\ansi\ansicpg1252\cocoartf1038\cocoasubrtf360
{\fonttbl\f0\fmodern\fcharset0 Courier;}
{\colortbl;\red255\green255\blue255;}
\margl1440\margr1440\vieww12960\viewh12780\viewkind0
\deftab720
\pard\pardeftab720\ql\qnatural

\f0\fs28 \cf0 Regression\
R scripts from Chpt 7 of\
Modern Statistical Methods for Astronomy with R Applications\
Feigelson & Babu 2012\
R scripts at http://astrostatistics.psu.edu/MSMA\
\
\
# Linear regression with heteroscedastic measurement errors\
# Construct and plot dataset of SDSS quasars with \
qso1 <- read.table('http://astrostatistics.psu.edu/MSMA/datasets/SDSS_QSO.dat',
h=T)  \
dim(qso1) ; summary(qso1) ; attach(qso1)\
sig_z_mag[sig_z_mag<0.01] <- 0.01\
\
dev.new(2)\
plot(i_mag, z_mag, pch=20, cex=0.1, col=grey(0.5), xlim=c(18,21.5), \
  ylim=c(17.5,22), xlab="SDSS i (mag)", ylab="SDSS z (mag)")\
for(i in 50:150) \{\
  lines(c(i_mag[i],i_mag[i]),c((z_mag[i]+sig_z_mag[i]),\
    (z_mag[i]-sig_z_mag[i])))\
  lines(c((i_mag[i]+sig_i_mag[i]),(i_mag[i]-sig_i_mag[i])),\
    c(z_mag[i],z_mag[i]))   \}\
\
# Ordinary least squares fit\
\
fit_ols <- lm(z_mag~i_mag)\
summary(fit_ols) \
confint(fit_ols, level=0.997)\
dev.set(2) ;  abline(fit_ols$coef, lty=1 ,lwd=2)\
\
dev.new(3) ; par(mfrow=c(2,2))\
plot(fit_ols, which=c(2:5), caption='', sub.caption='' ,pch=20, cex=0.3, \
  cex.lab=1.3, cex.axis=1.3)\
par(mfrow=c(1,1))\
\
# Weighted least squares fit\
\
fit_wt <- lm(z_mag~i_mag, x=T, weights=1/(sig_z_mag*sig_z_mag))\
summary(fit_wt)\
```

```
dev.set(2) ; abline(fit_wt$coef,lty=2,lwd=2)\
\
# Robust M-estimator\
\
library(MASS)\
fit_M <- rlm(z_mag~i_mag, method='M', weights=1/(sig_z_mag*sig_z_mag), \
  wt.method='inv.var')\
summary(fit_M) \
length(which(fit_M$w<1.0))  \
aM <- fit_M$coef[[1]] ; bM <- fit_M$coef[[2]]\
dev.set(2) ; lines(c(18,22), c(aM+bM*18, aM+bM*22), lty=3, lwd=3)\
\
# Linear quantile regression\
\
install.packages('quantreg') ; library(quantreg)\
fit_rq <- rq(z_mag~i_mag, tau=c(0.10,0.50,0.90))\
print.summary.rq(fit_rq) \
par(mfrow=c(1,2))\
plot(i_mag, z_mag, pch=20, cex=0.1, col=grey(0.5), xlim=c(18,22),\
  ylim=c(17.5,22), xlab="SDSS i (mag)", ylab="SDSS z (mag)")\
for(j in 0:2) \{ \
 aquant=fit_rq$coef[[2*j+1]] ; bquant=fit_rq$coef[[2*j+2]]\
 lines(c(18,22),c((aquant+bquant*18),(aquant+bquant*22)),lwd=2) \}\
\
# Nonlinear quantile regression\
\
fit_rqss.1 <- rqss(z_mag~qss(i_mag), data=qso1, tau=0.10)\
fit_rqss.5 <- rqss(z_mag~qss(i_mag), data=qso1, tau=0.50)\
fit_rqss.9 <- rqss(z_mag~qss(i_mag), data=qso1, tau=0.90)\
plot.rqss(fit_rqss.1, rug=F, ylim=c(17.5,22), titles='')\
points(i_mag, z_mag, cex=0.1, pch=20, col=grey(0.5))\
plot.rqss(fit_rqss.1, shade=F, rug=F, add=T, titles='', lwd=2)\
plot.rqss(fit_rqss.5, shade=F, rug=F, add=T, titles='', lwd=2)\
plot.rqss(fit_rqss.9, shade=F, rug=F, add=T, titles='', lwd=2)\
par(mfrow=c(1,1))\
\
# Fit Sersic function to NGC 4472 elliptical galaxy surface brightness profile\
\
NGC4472 <- \
  read.table("http://astrostatistics.psu.edu/MSMA/datasets/NGC4472_profile.dat", \
  header=T)\
attach(NGC4472)\
NGC4472.fit <-  nls(surf_mag ~ -2.5*log10(I.e * 10^(-(0.868*n-0.142)*\
  ((radius/r.e)^\{1/n\}-1))) + 26, data=NGC4472, start=list(I.e=20.,\
  r.e=120.,n=4.), model=T, trace=T)\
summary(NGC4472.fit)\
```

```
deviance(NGC4472.fit)\
logLik(NGC4472.fit)\
\
# Plot NGC 4472 data and best-fit model\
\
plot(NGC4472.fit$model$radius, NGC4472.fit$model$surf_mag, pch=20, \
  xlab="r  (arcsec)", ylab=expression(mu ~~ (mag/sq.arcsec)), ylim=c(16,28), \
  cex.lab=1.5, cex.axis=1.5)\
lines(NGC4472.fit$model$radius, fitted(NGC4472.fit))\
\
# Fit and plot radial profiles of NGC 4406 and NGC 4451\
\
NGC4406 <- \
  read.table("http://astrostatistics.psu.edu/MSMA/datasets/NGC4406_profile.dat", \
  header=T)\
attach(NGC4406)\
NGC4406.fit <- nls(surf_mag ~ -2.5*log10(I.e * 10^(-(0.868*n-0.142)*\
  ((radius/r.e)^\{1/n\}-1))) + 32, data=NGC4406, start=list(I.e=20.,\
  r.e=120.,n=4.), model=T, trace=T)\
summary(NGC4406.fit)\
points(NGC4406.fit$model$radius, NGC4406.fit$model$surf_mag, pch=3)\
lines(NGC4406.fit$model$radius, fitted(NGC4406_fit))\
\
NGC4551 <- \
  read.table("http://astrostatistics.psu.edu/MSMA/datasets/NGC4551_profile.dat", \
  header=T)\
attach(NGC4551)\
NGC4551.fit <-  nls(surf_mag ~ -2.5*log10(I.e * 10^(-(0.868*n-0.142)*\
  ((radius/r.e)^\{1/n\}-1))) + 26, data=NGC4551, start=list(I.e=20.,r.e=15.,n=4.), \
  model=T, trace=T)\
summary(NGC4551.fit)\
points(NGC4551.fit$model$radius, NGC4551.fit$model$surf_mag, pch=5)\
lines(NGC4551.fit$model$radius, fitted(NGC4551_fit))\
legend(500, 20, c("NGC 4472","NGC 4406", "NGC 4551"), pch=c(20,3,5))\
\
# NGC 4472 analysis\
# Residual plot\
\
plot(NGC4472.fit$model$radius,residuals(NGC4472.fit), xlab="r (arcsec)", \
  ylab="Residuals", pch=20, cex.lab=1.5, cex.axis=1.5)\
lines(supsmu(NGC4472.fit$model$radius, residuals(NGC4472.fit), span=0.05), \
  lwd=2)\
\
# Test for normality of residuals \
\
qqnorm(residuals(NGC4472.fit) / summary(NGC4472.fit)$sigma) \
```

```
abline(a=0,b=1)\
shapiro.test(residuals(NGC4472.fit) / summary(NGC4472.fit)$sigma)\
\
# Bootstrap parameter estimates\
\
install.packages('nlstools') ; library(nlstools)\
NGC4472.boot <- nlsBoot(NGC4472.fit)\
summary(NGC4472.boot)\
curve(dnorm(x,m=5.95, sd=0.10)*58/5.95, xlim=c(5.6,6.4), ylim=c(0,50))\
hist(NGC4472.boot$coefboot[,3], breaks=50, add=T)  # not shown}
```