PSTAT 194CS Final Project
Monte Carlo Bootstrapping

Ostapenko, Vasiliy (vostapenko, 774 970 8)
Collaborated with: Arthur Starodynov, Jake Bentley

Contents

DATA

Load Data

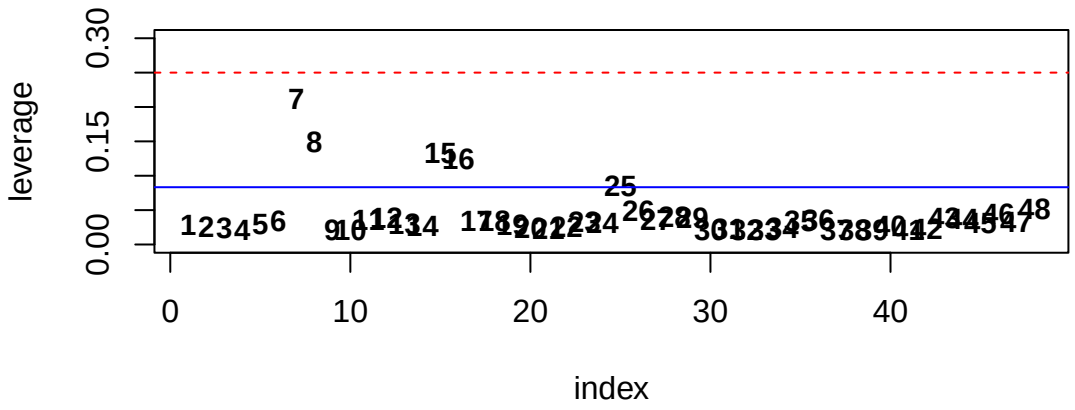
```
df = read.csv("./data/bee_data.csv")
df$Ratio = 100 * df$Ratio
```

Analysis

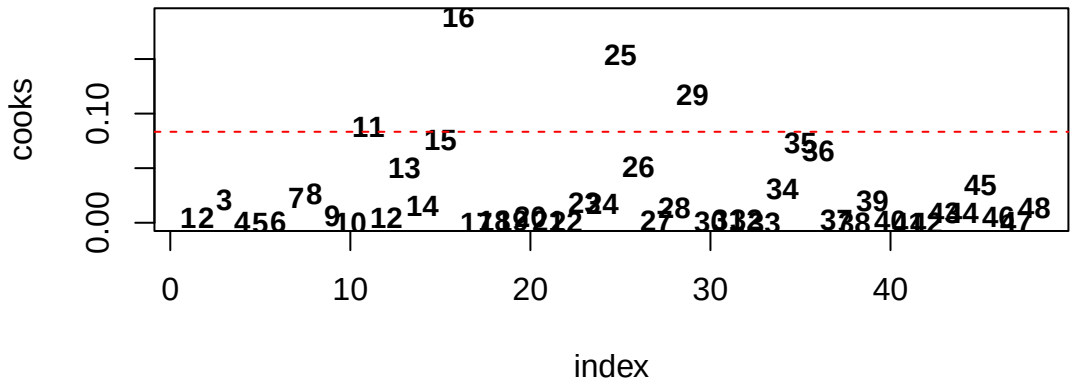
```
mod = glm(formula=Ratio ~ IT, family="gaussian", data=df)
summary(mod)
```

```
##
## Call:
## glm(formula = Ratio ~ IT, family = "gaussian", data = df)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -3.365  -0.720  -0.047   0.911   3.004
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   13.754      0.376   36.54   <2e-16 ***
## IT             0.418      0.129    3.24   0.0023 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 2.016)
##
##     Null deviance: 113.810  on 47  degrees of freedom
## Residual deviance:  92.715  on 46  degrees of freedom
## AIC: 173.8
##
## Number of Fisher Scoring iterations: 2
```

```
lev = hatvalues(mod)

n = nrow(df)
p = 3
dat = data.frame(index=seq(length(lev)), leverage=lev)
plot(leverage~index, col="white", data=dat, pch=NULL, ylim=c(0, 0.3))
text(leverage~index, labels=index, data=dat, cex=0.9, font=2)
abline(h=(p+1)/n, col="blue")
abline(h=3*(p+1)/n, col="red", lty=2)
```



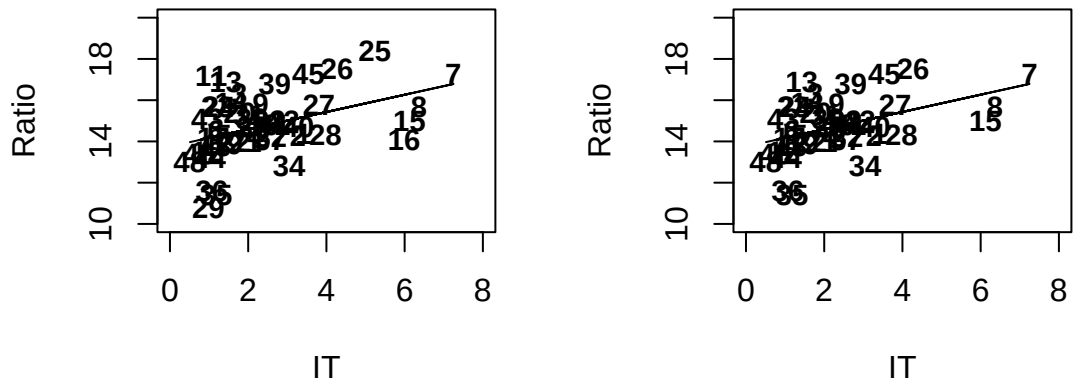

```
d = cooks.distance(mod)

dat2 = data.frame(index=seq(length(d)), cooks=d)
plot(cooks~index, col="white", data=dat2, pch=NULL)
text(cooks~index, labels=index, data=dat2, cex=0.9, font=2)
abline(h=4/n, col="red", lty=2)
```



```r
mod2 = glm(formula=Ratio ~ IT, family="gaussian", data=df[-c(11, 16, 25, 29), ])
summary(mod2)
```

```
##
## Call:
## glm(formula = Ratio ~ IT, family = "gaussian", data = df[-c(11,
##     16, 25, 29), ])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.8473  -0.6806  -0.0482   0.7498   2.5475
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   13.754      0.345   39.89   <2e-16 ***
## IT             0.419      0.124    3.38   0.0016 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.469)
##
##     Null deviance: 78.488  on 43  degrees of freedom
## Residual deviance: 61.684  on 42  degrees of freedom
## AIC: 145.7
##
## Number of Fisher Scoring iterations: 2
```

```r
par(mfrow=c(1, 2))
{
  {
    plot(Ratio~IT, data=df, col="white", pch=NULL,
         xlim=c(0, 8), ylim=c(10, 20))
    text(Ratio~IT, labels=rownames(df), data=df,
         cex=0.9, font=2)
    lines(x=df$IT, y=predict(mod, df))
  }
  {
    plot(Ratio~IT, data=df[-c(11, 16, 25, 29), ], col="white", pch=NULL,
         xlim=c(0, 8), ylim=c(10, 20))
    text(Ratio~IT, labels=rownames(df[-c(11, 16, 25, 29), ]), data=df[-c(11, 16, 25, 29), ],
         cex=0.9, font=2)
    lines(x=df[-c(11, 16, 25, 29), ]$IT, y=predict(mod2, df[-c(11, 16, 25, 29), ]))
  }
}
```




```
print(paste0("mod R2: ", round(1 - mod$deviance/mod$null.deviance, 3)))
```

```
## [1] "mod R2: 0.185"
```

```
print(paste0("mod2 R2: ", round(1 - mod2$deviance/mod2$null.deviance, 3)))
```

```
## [1] "mod2 R2: 0.214"
```

BOOTSTRAPPING

```
nboot = 10000
b0_estimates = vector(mode="numeric", length=nboot)
b1_estimates = vector(mode="numeric", length=nboot)

bootsize = 12
for(i in seq(nboot)) {
  bdat = df[sample(nrow(df), size=bootsize, replace=TRUE), ]
  bfit = update(mod, data=bdat)
  b0_estimates[i] = coef(bfit)[[1]]
  b1_estimates[i] = coef(bfit)[[2]]
}

res = data.frame(list(b0=b0_estimates, b1=b1_estimates))
```

```
print(paste0("b0 (intercept) 95% CI: [ ",
             round(mean(res$b0) - 1.96*sd(res$b0), 3), ", ",
             round(mean(res$b0) + 1.96*sd(res$b0), 3), " ]"))
```

```
## [1] "b0 (intercept) 95% CI: [ 11.92, 15.496 ]"
```

```
print(paste0("b1 (IT) 95% CI: [ ",
             round(mean(res$b1) - 1.96*sd(res$b1), 3), ", ",
             round(mean(res$b1) + 1.96*sd(res$b1), 3), " ]"))
```

```
## [1] "b1 (IT) 95% CI: [ -0.242, 1.143 ]"
```