medical school

Chris Parrish

January 23, 2016

medical school

reference:
- Cannon, et al., Stat2, chapter 09, example 9.4-9.5, 9.11-9.12
- Cannon, et al., Student R Manual, chapter 9

Import the data.

```
data <- read.csv("MedGPA.csv", header=TRUE)
head(data, 4)
```

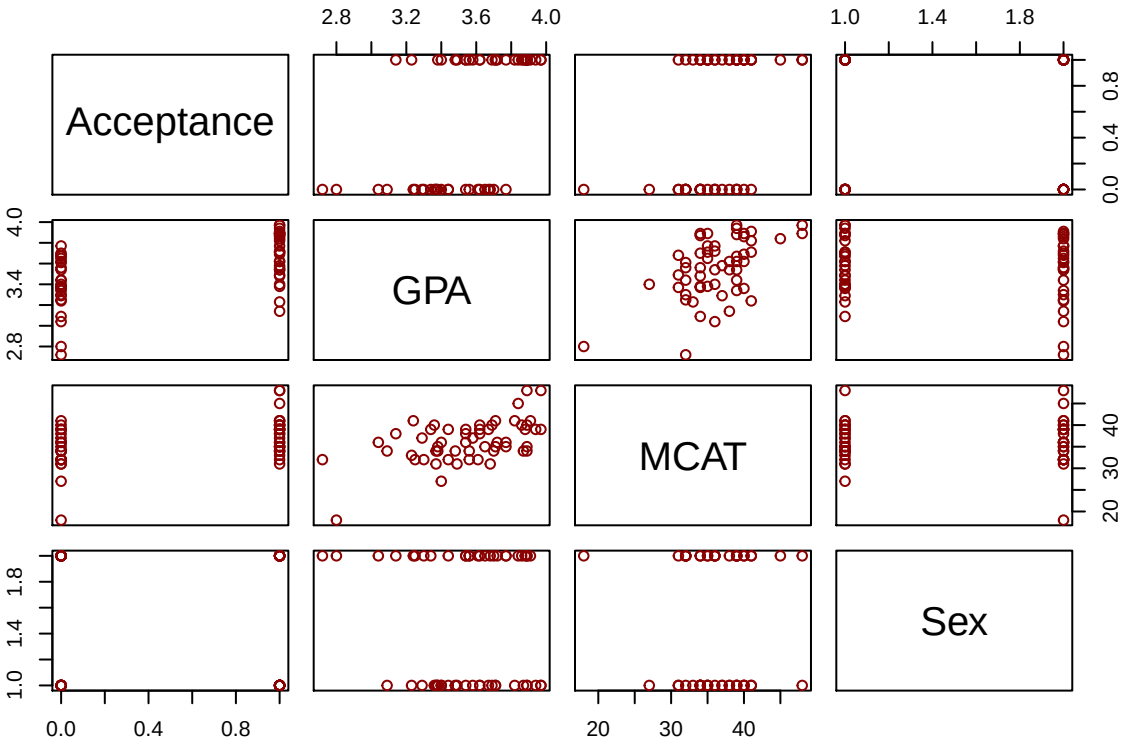
```
##   Accept Acceptance Sex BCPM  GPA VR PS WS BS MCAT Apps
## 1      D          0   F 3.59 3.62 11  9  9  9   38    5
## 2      A          1   M 3.75 3.84 12 13  8 12   45    3
## 3      A          1   F 3.24 3.23  9 10  5  9   33   19
## 4      A          1   F 3.74 3.69 12 11  7 10   40    5
```

```
dim(data)
```

```
## [1] 55 11
```

Scatterplot matrix.

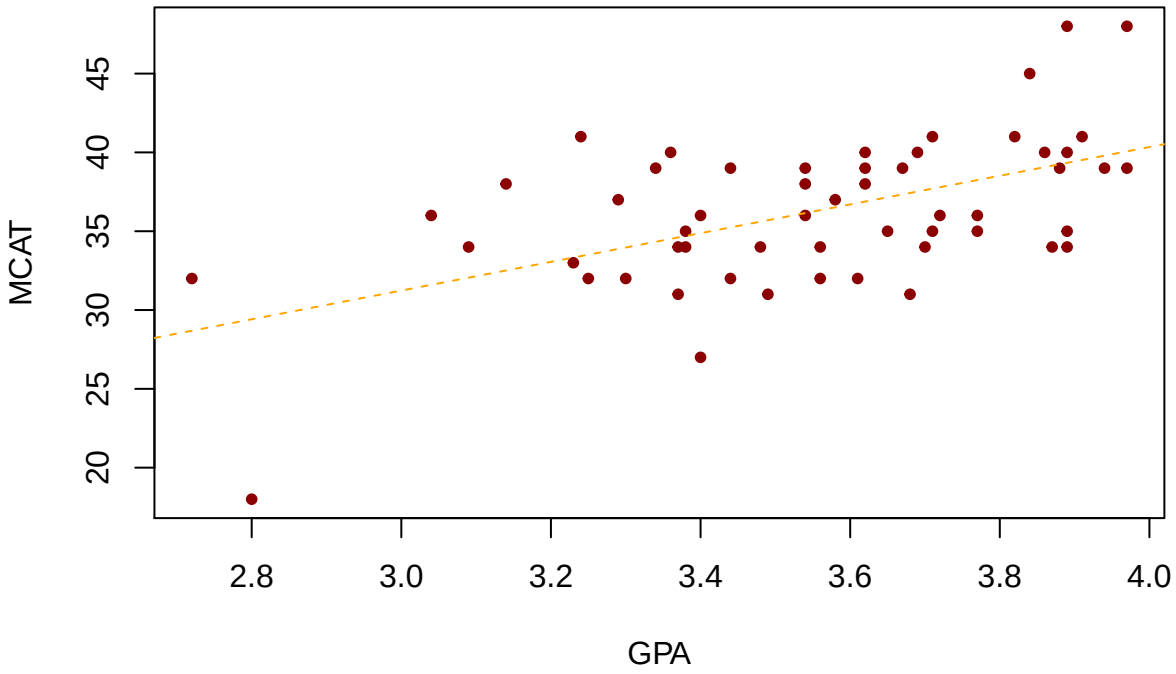
```
pairs(~ Acceptance + GPA + MCAT + Sex, data=data, col="darkred")
```



Simple regression.

Note: Coefficients of the regression line do not agree with the values shown on p.458.

```
plot(MCAT ~ GPA, data=data,
     pch=20, col="darkred")
MCAT.lm <- lm(MCAT ~ GPA, data=data)
abline(MCAT.lm, col="orange", lty=2)
```



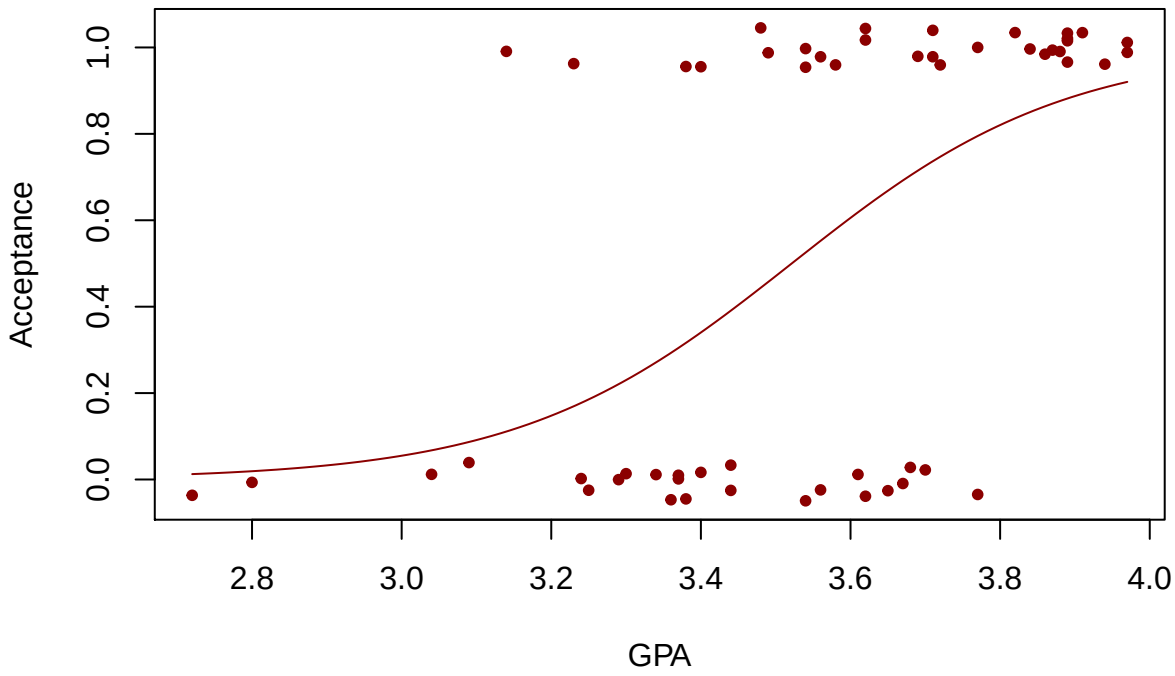

```
options(show.signif.stars=FALSE)
summary(MCAT.lm)
```

```
##
## Call:
## lm(formula = MCAT ~ GPA, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.4148  -2.5168  -0.1519   2.6653   8.6616
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.923      6.922   0.567    0.573
## GPA            9.104      1.942   4.688 1.97e-05
##
## Residual standard error: 4.088 on 53 degrees of freedom
## Multiple R-squared:  0.2931, Adjusted R-squared:  0.2798
## F-statistic: 21.98 on 1 and 53 DF,  p-value: 1.969e-05
```

Logistic regression with `glm`

reference: Cannon, et al., Student R Manual, chapter 9

```r
with(data,
     plot(GPA, jitter(Acceptance, amount=.05),
          pch=20, col="darkred", ylab="Acceptance"))
MCAT.glm <- glm(Acceptance ~ GPA, data=data, family=binomial)
b0 <- coef(MCAT.glm)[1]
b1 <- coef(MCAT.glm)[2]
library(boot)
curve(inv.logit(b0 + b1 * x), col="darkred", add=TRUE)
```




```r
summary(MCAT.glm)
```

```
##
## Call:
## glm(formula = Acceptance ~ GPA, family = binomial, data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7805  -0.8522   0.4407   0.7819   2.0967
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -19.207      5.629  -3.412 0.000644
## GPA            5.454      1.579   3.454 0.000553
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 75.791  on 54  degrees of freedom
## Residual deviance: 56.839  on 53  degrees of freedom
## AIC: 60.839
##
## Number of Fisher Scoring iterations: 4
```

Prediction.

```
new.data <- data.frame(GPA=3.6)
predict(MCAT.glm, new.data, type="response")
```

```
##         1
## 0.6055141
```

Odds ratio.

$\beta_1 = log(OR)$

$exp(\beta_1) = OR$

A one unit increase in GPA (for instance, from 3.0 to 4.0) increases the odds ratio of acceptance by a factor of 233! A one-tenth unit increase in GPA (for instance, from 3.0 to 3.1) increases the odds ratio of acceptance by a factor of 1.73.

```
coef(MCAT.glm)
```

```
## (Intercept)         GPA
##  -19.206503    5.454166
```

```
beta1 <- coef(MCAT.glm)[2]
OR <- exp(beta1)
OR
```

```
##      GPA
## 233.7298
```

```
OR <- exp(0.1 * beta1)
OR
```

```
##      GPA
## 1.725327
```

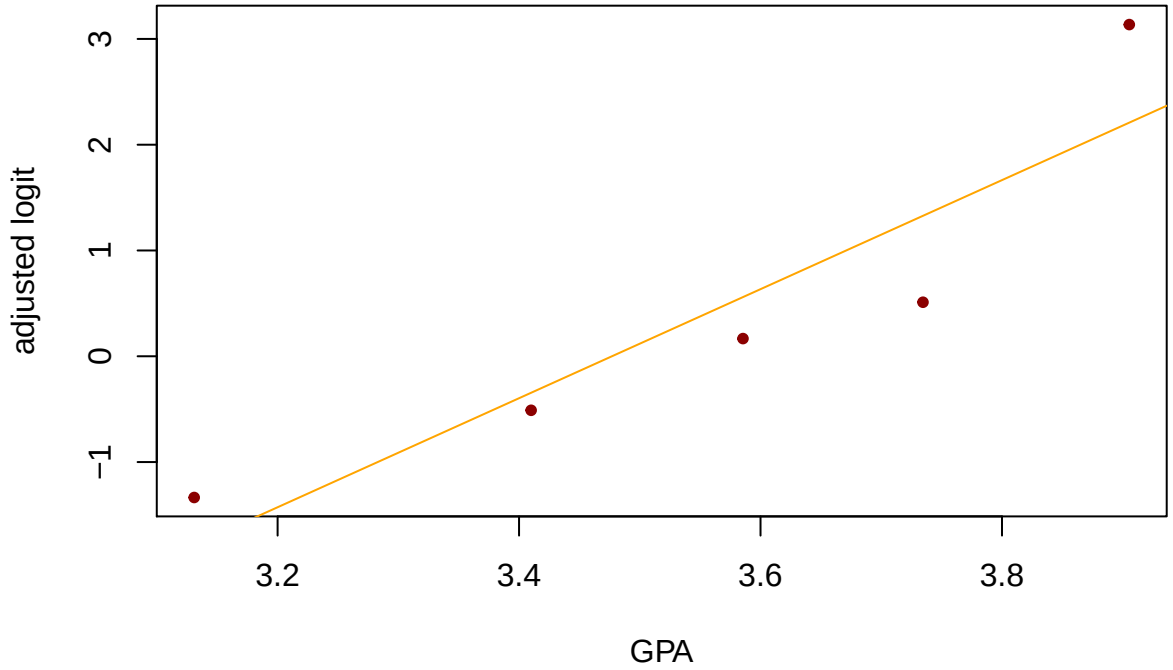
Slicing.

reference: Cannon, et al., Student R Manual, chapter 9

Compare with Figure 9.18, p.473.

```
sorted.MedGPA.df <- data[order(data$GPA), ]
x <- sorted.MedGPA.df$GPA
y <- sorted.MedGPA.df$Acceptance
x.mat <- matrix(x, ncol=11, nrow=5, byrow=TRUE)
x.means <- apply(x.mat, 1, mean)
y.mat <- matrix(y, ncol=11, nrow=5, byrow=TRUE)
y.yes <- apply(y.mat, 1, sum)
y.no <- 11 - y.yes
y.prop <- y.yes / (y.yes + y.no)
y.prop.adj <- (.5 + y.yes) / (1 + y.yes + y.no)
y.logit.adj <- log(y.prop.adj / (1 - y.prop.adj))
```

```r
plot(x.means, y.logit.adj,
     pch=20, col="darkred",
     xlab="GPA", ylab = "adjusted logit")
abline(lm(y.logit.adj ~ x.means), col="orange")
```



Formal inference: tests and intervals.

```r
data <- read.csv("MedGPA.csv", header=TRUE)
data$GPA10 <- data$GPA * 10
med.school.glm <- glm(Acceptance ~ GPA10, data=data, family=binomial)
summary(med.school.glm)
```

```
##
## Call:
## glm(formula = Acceptance ~ GPA10, family = binomial, data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7805  -0.8522   0.4407   0.7819   2.0967
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -19.2065     5.6287  -3.412 0.000644
## GPA10         0.5454     0.1579   3.454 0.000553
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 75.791  on 54  degrees of freedom
## Residual deviance: 56.839  on 53  degrees of freedom
## AIC: 60.839
##
## Number of Fisher Scoring iterations: 4
```

CI for odds ratio.

$\beta_1 = log(OR)$

$exp(\beta_1) = OR$

```
confint(med.school.glm)        # CI for beta1
```

```
## Waiting for profiling to be done...
```

```
##                    2.5 %     97.5 %
## (Intercept) -31.7132662 -9.3761026
## GPA10         0.2696316  0.8965621
```

```
exp(confint(med.school.glm))   # CI for exp(beta1) = OR
```

```
## Waiting for profiling to be done...
```

```
##                     2.5 %       97.5 %
## (Intercept) 1.686955e-14 8.472476e-05
## GPA10       1.309482e+00 2.451162e+00
```