HW 1

Jared Yu

January 15, 2019

```
jp = read.delim(file = 'JobProficiency.txt', header = TRUE, sep = '') # read data

### Histograms
hist(jp$X1, main = 'Histogram of Test 1 Scores', xlab = 'Test Scores')
hist(jp$X2, main = 'Histogram of Test 2 Scores', xlab = 'Test Scores', breaks = 10)
hist(jp$X3, main = 'Histogram of Test 3 Scores', xlab = 'Test Scores', breaks = 10)
hist(jp$X4, main = 'Histogram of Test 4 Scores', xlab = 'Test Scores', breaks = 10)

### pairs matrix, correlation matrix
pairs(jp, labels = c('JP Score', 'Test 1', 'Test 2', 'Test 3', 'Test 4'))
cor(jp)

# fit MLR model
mlr <- lm(Y~., data = jp)

# t-statistics and ANOVA SSE
summary(mlr)
anova(mlr) # SSE = residual SS

# Calculate SSTO, SSR, etc.
nt=25
p=5
sse = 336
ssto = sum((jp$Y-mean(jp$Y))^2)
ssr = ssto - sse
df.sse = nt-p
df.ssr = p-1
df.ssto = nt-1
mse = sse/df.sse
msr = ssr/df.ssr
msto = ssto/df.ssto
f.star = msr/mse
1-pf(f.star, 4, 20) # p-value

# 2a
plot(x = mlr$fitted.values, y = jp$Y, main = 'Observed vs Fitted',
     xlab = 'Fitted', ylab = 'Observed')
plot(x = mlr$fitted.values, y = mlr$residuals, main = 'Residuals vs Fitted',
     xlab = 'Fitted', ylab = 'Residuals')
abline(h=0)

# 2b
hist(mlr$residuals, main = 'Histogram of Residuals', xlab = 'Residual Values')
qqnorm(mlr$residuals); qqline(mlr$residuals)
qq <- qqnorm(mlr$residuals)
cor(qq$x, qq$y)
```

```r
# 3a
summary(mlr); AIC(mlr) # AIC
summary(lm(Y~X1+X3+X4, data = jp)); AIC(lm(Y~X1+X3+X4, data = jp)) # best
summary(lm(Y~X1+X3, data = jp)); AIC(lm(Y~X1+X3, data = jp))
summary(lm(Y~X3, data = jp)); AIC(lm(Y~X3, data = jp))
summary(lm(Y~1, data = jp)); AIC(lm(Y~1, data = jp))

sink('bic.txt') # BIC
BIC(mlr)
BIC(lm(Y~X1+X3+X4, data = jp)) # best
BIC(lm(Y~X1+X3, data = jp))
BIC(lm(Y~X3, data = jp))
BIC(lm(Y~1, data = jp))
sink()
best_fit <- lm(Y~X1+X3+X4, data = jp)
sink('best.txt')
summary(best_fit)
sink()

sink('aic_bic.txt') # 3b
AIC(lm(Y~X1+X2+X3+X4, data = jp))

AIC(lm(Y~X1+X2+X3, data = jp))
AIC(lm(Y~X1+X2+X4, data = jp))
AIC(lm(Y~X1+X3+X4, data = jp)) # BEST
AIC(lm(Y~X2+X3+X4, data = jp))

AIC(lm(Y~X1+X2, data = jp))
AIC(lm(Y~X1+X3, data = jp))
AIC(lm(Y~X1+X4, data = jp))
AIC(lm(Y~X2+X3, data = jp))
AIC(lm(Y~X2+X4, data = jp))
AIC(lm(Y~X3+X4, data = jp))

AIC(lm(Y~X1, data = jp))
AIC(lm(Y~X2, data = jp))
AIC(lm(Y~X3, data = jp))
AIC(lm(Y~X4, data = jp))

AIC(lm(Y~1, data = jp))

BIC(lm(Y~X1+X2+X3+X4, data = jp))

BIC(lm(Y~X1+X2+X3, data = jp))
BIC(lm(Y~X1+X2+X4, data = jp))
BIC(lm(Y~X1+X3+X4, data = jp)) # BEST
BIC(lm(Y~X2+X3+X4, data = jp))

BIC(lm(Y~X1+X2, data = jp))
BIC(lm(Y~X1+X3, data = jp))
BIC(lm(Y~X1+X4, data = jp))
BIC(lm(Y~X2+X3, data = jp))
BIC(lm(Y~X2+X4, data = jp))
```

```
BIC(lm(Y~X3+X4, data = jp))

BIC(lm(Y~X1, data = jp))
BIC(lm(Y~X2, data = jp))
BIC(lm(Y~X3, data = jp))
BIC(lm(Y~X4, data = jp))

BIC(lm(Y~1, data = jp))
sink()
```