STAT 2131
Orly Olbum

HOMEWORK 7

1. $X \in \mathbb{R}^{n \times p}$, $H \in \mathbb{R}^{n \times n}$ (2c from HW5)

(i) $H^T = H$

(ii) $HH = H$

(iii) $im(H) = im(X)$

ⓐ show that H is orth. proj matrix:

for $v \in \mathbb{R}^n$, $(v - Hv)^T (Hv) = 0$

and $Hv = \underset{u \in im(x)}{\text{argmin}} \|v-u\|_2^2$

We first show that $(v-Hv)^T(Hv) = 0$:

$$(v-Hv)^T(Hv) = (v^T - v^T H^T)(Hv)$$

$$= v^T H v - v^T H^T H v = v^T H v - v^T H v = 0 \quad (i, ii)$$

Furthermore, $Hv = \underset{u \in im(x)}{\text{argmin}} \|v-u\|_2^2$:

(write $v - u = H(v-u) + Q(v-u)$)

$$\|v-u\|_2^2 = \|v - Hv + Hv - u\|_2^2$$

$$= \|(v-Hv) + (Hv-u)\|_2^2$$

$$= [(v-Hv) + (Hv-u)]^T [(v-Hv) + (Hv-u)]$$

$$= \underset{①}{(v-Hv)^T(v-Hv)} + \underset{②}{(v-Hv)^T(Hv-u)} + \underset{③}{(Hv-u)^T(v-Hv)}$$

$$+ (Hv-u)^T(Hv-u) \quad ④$$

① $(v-Hv)^T(v-Hv) = \|v-Hv\|_2^2$

② $(v-Hv)^T(Hv-u) = v^T H v - v^T u - v^T H^T H v + v^T H^T u$

$= v^T H v - v^T u - v^T H v + v^T u = 0$

③ $(Hv-u)^T(v-Hv) = v^T H^T v - v^T H H v - u^T v + u^T H v$

$= v^T H v - v^T H v - u^T v + u^T v = 0$

④ $(Hv-u)^T(Hv-u) = \|Hv-u\|_2^2$

so, $\|v-u\|_2^2 = ① + \cancel{②}^{0} + \cancel{③}^{0} + ④$

$$= \|v-Hv\|_2^2 + \|Hv-u\|_2^2 \geq \|v-Hv\|_2^2$$

making Hv the closest in im(x) to v.

next page →

1, continued

(b) Show that for tl properties, it is unique. If $P \in \mathbb{R}^{n \times n}$ is another matrix such that
(i) $P^T = P$, (ii) $PP = P$, (iii) $im(P) = im(X)$, $\underline{H = P}$

P (projection matrix) projects onto W for
$y \in W$, where $Py = y$ and
$y \in W^{\perp}$, where $Py = 0$

Py is then a linear combination of the columns of P (ie, $W = col(P)$)

Say $y = y_1 + y_2$, with $y_1 \in W$ and $y_2 \in W^{\perp}$
$Py = P(y_1 + y_2) = Py_1 + Py_2 = Py_1 + 0 = y_1$

So, the projection of P onto W is unique

we also know that, under (i), $P^T = P$ (symmetric) and from (ii), $PP = P$ (idempotent).

Also, $rank(P) = tr(P) = n$ because $P \in \mathbb{R}^{n \times n}$

For: $Y = X\beta + \epsilon$, $E(Y) = X\beta$ and $Y = y$ is an observation

P_X is the projection onto $col(X)$, and if X has full rank, $P_X = H$

Since from (iii) $im(P) = im(X)$, $P = H$
and P_X = hat matrix $\rightarrow X(X^TX)^{-1}X^T$

(c) Define $H = XX^+$, show H is hat matrix of X using properties 1-4 of Moore-Penrose pseudoinverse.

$H = XX^+$
$= (XX^+)^T$ by ①
$= (X^+)^T X^T$
$= [(X^TX)^+ X^T]^T X^T$ by ③
$= X[(X^TX)^+]^T X^T$ by ④
$= X(X^TX)^+ X^T$ by ②

and for invertible (X^TX), $(X^TX)^+ = (X^TX)^{-1}$
$\rightarrow = X(X^TX)^{-1}X^T = H$, hat matrix of X

2. $Y = X\beta + \epsilon$, $X \in \mathbb{R}^{n\times p}$ is non-random, full-rank and $\beta \in \mathbb{R}^p$ is unknown. Prove Gauss-Markov Theorem.

$E(\epsilon) = 0$, $var(\epsilon) = \sigma^2 I_n$

$var(\tilde{\beta}) = var(\hat{\beta}) + M$, M is symmetric & PSD

$\hat{\beta}$ is BLUE. Proof when X is not full-rank is almost the same as below.

@ what is OLS estimator for β? what is its variance and hat matrix H?

To find the OLS estimator, we take

$Y = X\beta + \epsilon$, solve for ϵ:

$\epsilon = Y - X\beta$

we want to minimize $\epsilon^T\epsilon$:

$$\min \epsilon^T\epsilon = (Y - X\beta)^T(Y - X\beta)$$

simplify:
$$\begin{aligned} \epsilon^T\epsilon &= (Y^T - \beta^T X^T)(Y - X\beta) \\ &= Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta \\ &= Y^TY - 2\beta^TX^TY + \beta^TX^TX\beta \end{aligned}$$

Derive wrt β:

$$\partial(\epsilon^T\epsilon)/\partial\beta = -2X^TY + 2X^TX\beta$$

Set to zero, solve for β

$$0 = -2X^TY + 2X^TX\hat{\beta}$$
$$\rightarrow 2X^TY = 2X^TX\hat{\beta}$$
$$\rightarrow X^TY = X^TX\hat{\beta}$$
$$\rightarrow \boxed{\hat{\beta} = (X^TX)^{-1}X^TY}$$

$$\begin{aligned} var(\hat{\beta}) &= var(\beta + (X^TX)^{-1}X^T\epsilon) \\ &= var[(X^TX)^{-1}X^T\epsilon] = E[(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1}] \\ &= (X^TX)^{-1}X^TE(\epsilon\epsilon^T)X(X^TX)^{-1} \\ &= (X^TX)^{-1}X^T\sigma^2 I_n X(X^TX)^{-1} \\ &= \sigma^2(X^TX)^{-1} \end{aligned}$$

$\hat{Y} = X\hat{\beta} \rightarrow \hat{Y} = X(X^TX)^{-1}X^TY = HY$

so hat matrix $H = X(X^TX)^{-1}X^T$ $\rightarrow$

2, continued

ⓑ $\tilde{\beta} = A^T Y$, linear unbiased estimator for β,
$A \in \mathbb{R}^{n \times p}$ Show that $A^T X = I_n$ & find $var(\tilde{\beta})$.
$\tilde{\beta}$ is a linear unbiased estimate for β
which would mean

$$E(\tilde{\beta}) = \beta$$

$A^T X = I$ needs to be true then, because

$$E(\tilde{\beta}) = E(A^T Y) = E(\underbrace{A^T X}_{I}\beta) = \beta$$

So, $A^T X = I$

And

$$var(\tilde{\beta}) = var(A^T Y)$$
$$= A^T var(Y)(A^T)^T = A^T var(Y) A$$
$$= A^T \sigma^2 I A = \sigma^2 \cdot A^T A$$

ⓒ $\hat{\beta}$ is OLS estimator, show that
$var(\tilde{\beta}) = var(\hat{\beta}) + M$
where M is symmetry, PSD
we know $var(\tilde{\beta}) = \sigma^2 A^T A = var(\hat{\beta}) + M$

$$var(\tilde{\beta}) = \sigma^2 A^T A = \sigma^2 (A^T H A + A^T (I - H) A)$$
$$= \sigma^2 (\underbrace{(X^T X)^{-1}}_{\text{becomes OLS } var(\hat{\beta})} + A^T (I - H) A)$$

so $M = \sigma^2 (A^T (I - H) A)$ which is PSD.
$A^T (I - H) A$ is quadratic form
and because $(I - H)$ is symmetric (previous HW)
$A^T (I - H) A$ is PSD making
M PSD and symmetric,
and

$$var(\tilde{\beta}) = var(\hat{\beta}) + M$$

→

2, continued

ⓓ For $q \in \mathbb{R}^p$, show that ⓒ implies
$\text{var}(q^T\tilde{\beta}) \geq \text{Var}(q^T\hat{\beta})$

$\text{var}(q^T\hat{\beta})$ is smaller because in the nature of **OLS**, we minimize $\hat{\beta}$

$$\begin{aligned}\text{var}(q^T\tilde{\beta}) &= q^T \text{var}(\tilde{\beta}) q \\ &= q^T(\text{var}(\hat{\beta}) + M)q \\ &= q^T \text{var}(\hat{\beta}) q + \underbrace{q^T M q}_{\text{from ⓒ, } M \text{ is PSD}}\end{aligned}$$

and since $q^T M q > 0$,
$\text{var}(q^T\tilde{\beta}) \geq \text{Var}(q^T\hat{\beta})$

ⓔ $Y = X\beta + \epsilon$, $E(\epsilon) = 0$, $\text{var}(\epsilon) = \sigma^2\Sigma$

Σ is known, invertible matrix

R is invertible $\rightarrow$ $\Sigma = RR^T$

(i) $\tilde{Y} = R^{-1}Y$, $\tilde{X} = R^{-1}X$

Find $E(\tilde{Y})$, $\text{var}(\tilde{Y})$

If $Y = X\beta + \epsilon$, $\tilde{Y} = \tilde{X}\beta + \tilde{\epsilon}$

so $R^{-1}Y = R^{-1}X\beta + \tilde{\epsilon} = R^{-1}(X\beta + \frac{\tilde{\epsilon}}{R^{-1}})$

making $Y = X\beta + \frac{1}{R^{-1}}\tilde{\epsilon}$

since $Y = X\beta + \epsilon$, $\tilde{\epsilon} = R^{-1}\epsilon$

$E(\tilde{Y}) = E(R^{-1}X\beta + R^{-1}\epsilon) = R^{-1}E(X\beta) + R^{-1}\cancelto{0}{E(\epsilon)} = R^{-1}X\beta$

$$\begin{aligned}\text{var}(\tilde{Y}) = \text{var}(R^{-1}Y) &= R^{-1}\text{var}(Y)(R^{-1})^T \\ &= R^{-1}\sigma^2\Sigma(R^{-1})^T \\ &= R^{-1}\sigma^2(RR^T)(R^T)^{-1} \\ &= R^{-1}\sigma^2 RR^T(R^T)^{-1} \\ &= \sigma^2 R^{-1}R\,R^T(R^T)^{-1} = \sigma^2\end{aligned}$$

$\rightarrow$

2 ⓔ, continued ; $Y = X\beta + \epsilon$, $E(\epsilon) = 0$, $Var(\epsilon) = \sigma^2 \Sigma$

(ii) Find BLUE for β under new Y in terms of X, Y, Σ (generalized least squares)

If $\Sigma = RR^T$, $\Sigma^{-1} = (RR^T)^{-1} = (R^T)^{-1} R^{-1}$

and $\Sigma^T = (RR^T)^T = (R^T)^T R^T = RR^T$ $\quad (\Sigma^T = \Sigma)$

and $(\Sigma^T)^{-1} = (\Sigma^{-1})^T = (R^T)^{-1} R^T = \Sigma^{-1}$

$\hat{\beta}_{OLS}$ minimizes the function $f(\hat{\beta})$

$$f(\hat{\beta}) = (\tilde{Y} - \tilde{X}\beta)^T (\tilde{Y} - \tilde{X}\beta)$$

derive, set to zero, solve for $\hat{\beta}$

we get

$$\begin{aligned}\hat{\beta}_{gls} &= (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{Y} \\ &= ((R^{-1}X)^T R^{-1}X)^{-1} (R^{-1}X)^T R^{-1} Y \\ &= (X^T (R^T)^{-1} R^{-1} X)^{-1} X^T (R^T)^{-1} R^{-1} Y \\ &= (X^T (RR^T)^{-1} X)^{-1} X^T (RR^T)^{-1} Y \\ &= (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y\end{aligned}$$

$\tilde{X}$ is full rank because X is full rank,

so $\hat{\beta}_{gls}$ is BLUE

STAT 2131 HW7 - Problems 3, 4

Problem 3: KNNL 6.10

Data is from problem 6.9:
Y = total labor hours X1 = number of cases shipped
X2 = indirect costs of total labor hours as a percentage
X3 = qualitative predictor called holiday; 1 if week has a holiday, 0 otherwise

(a) Fit regression model (6.5) to the data for three predictor variables. State the estimated regression function. How are bl, b2, and b3 interpreted here?

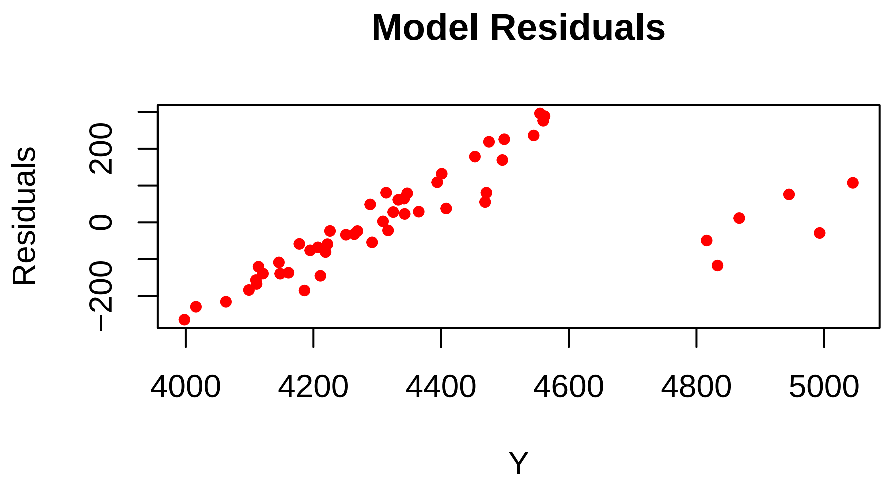
```
model1 = lm(Y ~ X1 + X2 + X3, data = grocery)
summary(model1)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X3, data = grocery)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -264.05 -110.73  -22.52   79.29  295.75
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.150e+03  1.956e+02  21.220  < 2e-16 ***
## X1           7.871e-04  3.646e-04   2.159   0.0359 *
## X2          -1.317e+01  2.309e+01  -0.570   0.5712
## X3           6.236e+02  6.264e+01   9.954 2.94e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 143.3 on 48 degrees of freedom
## Multiple R-squared:  0.6883, Adjusted R-squared:  0.6689
## F-statistic: 35.34 on 3 and 48 DF,  p-value: 3.316e-12
```

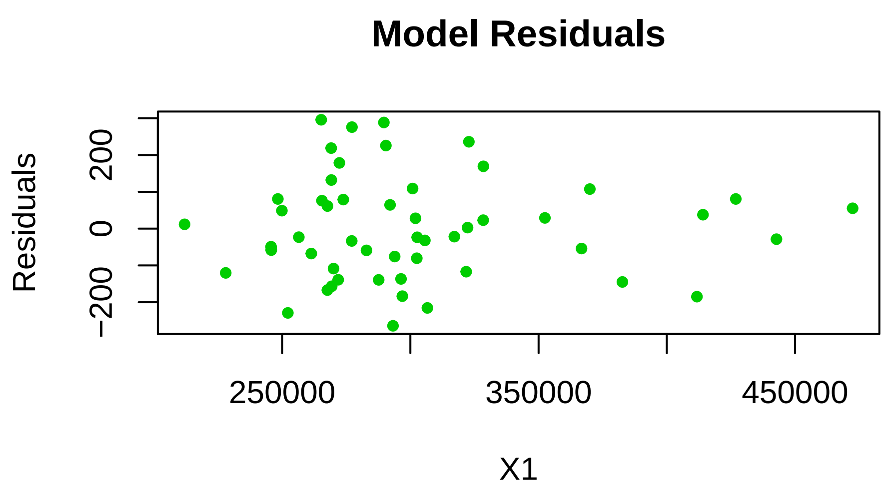
As seen in the R output, the regression function can be written as:
Y = 4,150 + .0000787X1 - 13.17X2 + 623.6X3
For every increase in cases shipped (X1), total labor hours increases by .0000787; for every increase in indirect cost of labor hours by one percentage point, total labor hours decreases by 13.17; and for a week with a holiday, total labor hours increases by 623.6.

(c) Plot the residuals against Y, X1, X2, X3, and X1X2 on separate graphs. Also prepare a normal probability plot. Interpret the plots and summarize your findings.

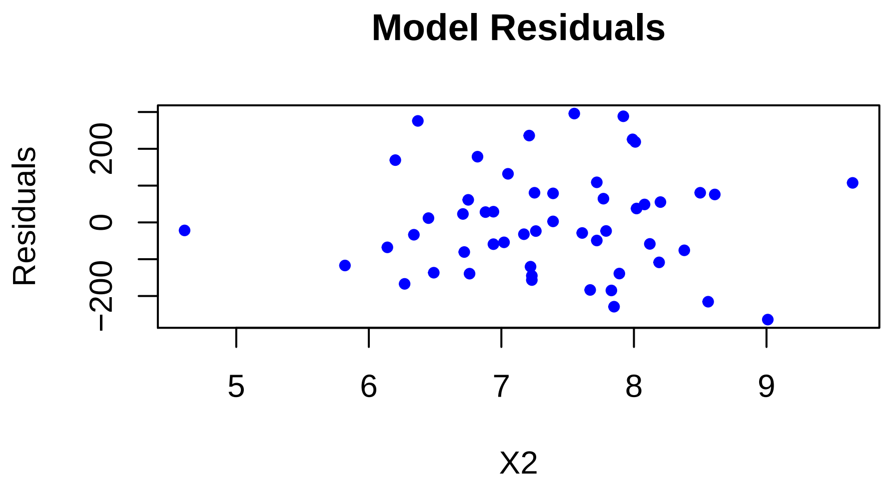
```
model1.res = resid(model1)
plot(grocery$Y, model1.res, xlab = "Y", ylab = "Residuals", main = "Model Residuals",
     pch = 20, col = 2)
```



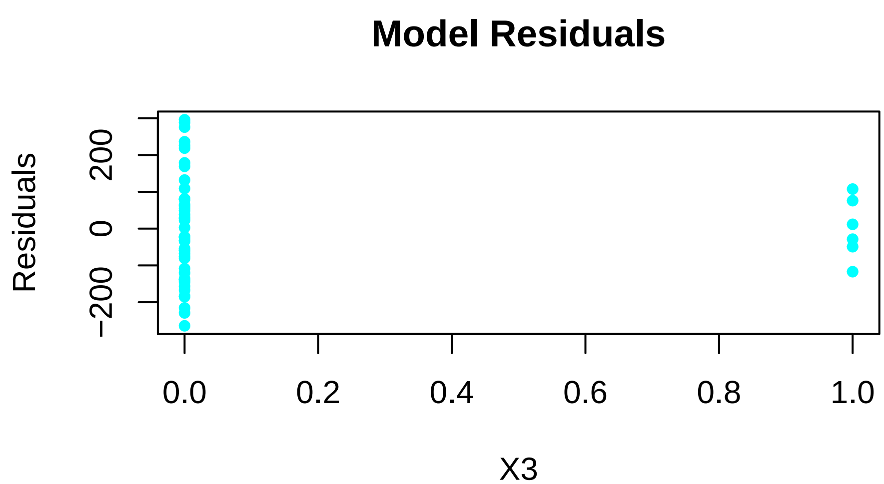

```
plot(grocery$X1, model1.res, xlab = "X1", ylab = "Residuals",
     main = "Model Residuals", pch = 20, col = 3)
```



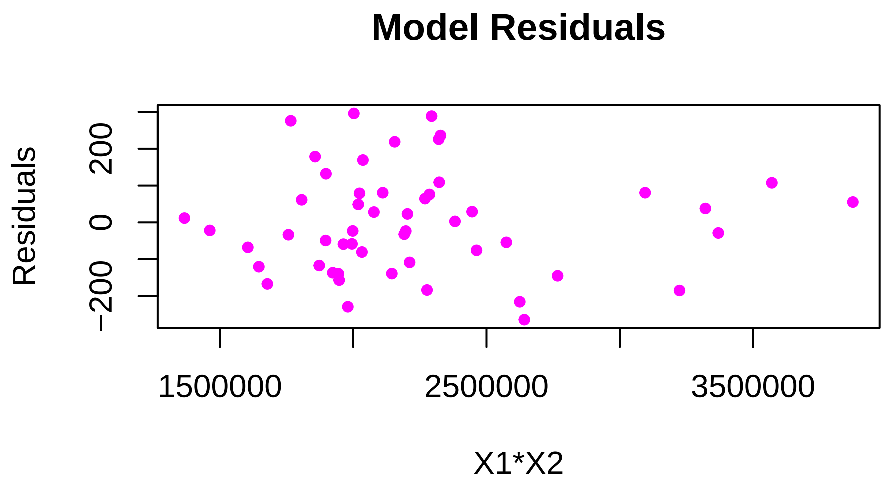

```
plot(grocery$X2, model1.res, xlab = "X2", ylab = "Residuals",
     main = "Model Residuals", pch = 20, col = 4)
```



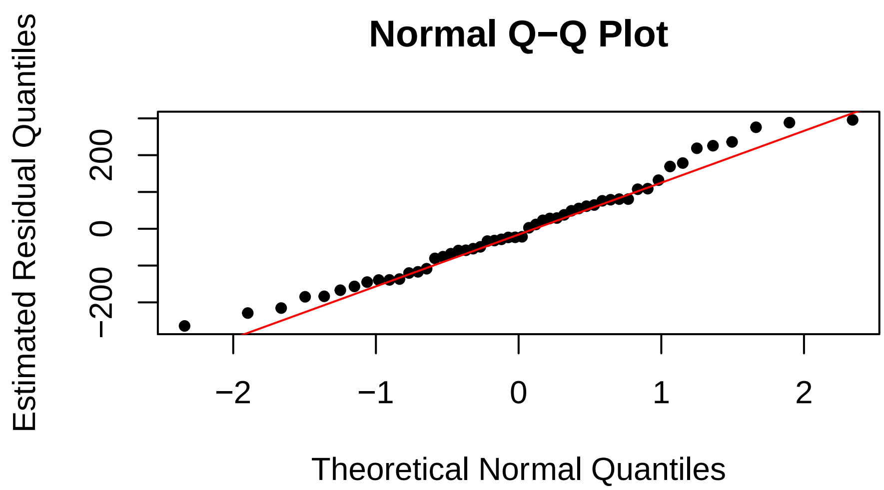

```r
plot(grocery$X3, model1.res, xlab = "X3", ylab = "Residuals",
     main = "Model Residuals", pch = 20, col = 5)
```




```r
plot(grocery$X1*grocery$X2, model1.res, xlab = "X1*X2", ylab = "Residuals",
     main = "Model Residuals", pch = 20, col = 6)
```




```
qqnorm(model1$residuals, xlab = "Theoretical Normal Quantiles",
       ylab = "Estimated Residual Quantiles", pch = 20)
qqline(model1$residuals, col = "red")
```



The first plot shows that the residuals of the model and the response variable Y are highly correlated, save for a few points outside of the linear trend. This would indicate that the model we put together is not a great fit, since the variation in Y is explained largely by the residuals rather than by the independent variables. The rest of the residual plots show no association (barring the X3 plot, which is not as easy to tell from just looking). The interaction plot between X1 and X2 additionally shows no association. The normal qq-plot shows that the residuals do follow a normal trend, which may make us feel a bit better about the model.

7.4 (b) Test whether X2 can be dropped from the regression model given that X1 and X3 are retained. Use the F test statistic and alpha = .05. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

Ho: B2 = 0 (can be dropped, no effect on response)
Ha: B2 != 0 (should be kept in the model, effect on response)

alpha = .05
Decision rule: compare model diagnostics, F statistic

```
linearHypothesis(model1, c("X2 = 0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## X2 = 0
##
## Model 1: restricted model
## Model 2: Y ~ X1 + X2 + X3
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     49 992204
## 2     48 985530  1    6674.6 0.3251 0.5712
```

Conclusion: From the hypothesis function we can see that the p-value for testing X2 = 0 is .57, the F-statistic is .3251, and the SSR for B2 is 6674.6. With alpha = 0.05 we fail to reject the Ho, and conclude that we can drop X2 from the model.

Problem 4: KNNL 6.16

Data is from problem 6.15:
Y = patient satisfaction
X1 = patient's age in years
X2 = severity of illness, index
X3 = anxiety level, index

(a) Test whether there is a regression relation; use alpha = .10. State the alternatives, decision rule, and conclusion. What does your test imply about B1, B2, and B3? What is the P-value of the test?

Ho: There is no relationship between the explanatory variables and the dependent variable.
Ha: There is some relationship between one or more of the explantory variables and the dependent variable.
alpha = .10
Decision rule: look for p-value of less than .10 to determine significance of relationship.

```
model2 = lm(Y ~ X1 + X2 + X3, data = patient)
summary(model2)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X3, data = patient)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.3524  -6.4230   0.5196   8.3715  17.1601
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 158.4913    18.1259   8.744 5.26e-11 ***
## X1           -1.1416     0.2148  -5.315 3.81e-06 ***
## X2           -0.4420     0.4920  -0.898   0.3741
```

```
## X3          -13.4702     7.0997  -1.897   0.0647 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.06 on 42 degrees of freedom
## Multiple R-squared:  0.6822, Adjusted R-squared:  0.6595
## F-statistic: 30.05 on 3 and 42 DF,  p-value: 1.542e-10
```

Conclusion: There are two significant predictors in the model, X1 and X3, that show a relationship with the dependent variable. With 90% confidence, patient's age (in years) and anxiety level (as an index) are significant predictors for patient satisfaction, severity of illness (as an index) is not, with a model p-value of about 0.000.

(b) Obtain joint interval estimates of B1, B2, and B3, using a 90 percent family confidence coefficient. Interpret your results.

```
confint(model2, level = 1 - 0.10/(2*3))
```

```
##                 0.833 %    99.167 %
## (Intercept) 113.291314 203.6911891
## X1           -1.677249  -0.6059751
## X2           -1.668803   0.7847947
## X3          -31.174356   4.2340294
```

The confidence intervals around the estimates of the betas somewhat confirm what the model diagnostics show. Since 0 is included in the interval for X2 and X3, we cannot say they are significant predictors. However, since 0 does not fall in the interval for X1, we can say that there is evidence of X1 being a significant predictor for Y.

(c) Calculate the coefficient of multiple determination. What does it indicate here?

```
summary(model2)$r.squared
```

```
## [1] 0.6821943
```

The multiple coefficient of determination, or the R^2, tells us how much of the variation in Y is explained by the independent variables in the model. For this model, about 68% of the variation in patient satisfaction is explain by patient's age (in years) and anxiety level (on an index). We saw earlier that severity of illness (on an index) does not assist in explaining this variation.