

Final Exam

AMS 256

Wyara Vanesa Moura e Silva

Spring 2018

June 10

Question 1

Consider the linear regression model

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i, \quad \epsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$$

$$\mathbf{y} = [82, 79, 74, 83, 80, 81, 84, 81]^T$$

$$\mathbf{x}_1 = [10, 9, 9, 11, 11, 10, 10, 12]^T$$

$$\mathbf{x}_2 = [15, 14, 13, 15, 14, 14, 16, 13]^T$$

See the R code used to do the simple algebra in Appendix 1.

(a)

Provide the least square estimates of β_1 , β_2 and σ^2 .

Solution:

Then, by definition we have that the least square estimate are given by: $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$.
Thus, we can get:

$$\mathbf{X}^T = \begin{pmatrix} 10 & 9 & 9 & 11 & 11 & 10 & 10 & 12 \\ 15 & 14 & 13 & 15 & 14 & 14 & 16 & 13 \end{pmatrix}^T$$

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} 848 & 1168 \\ 1168 & 1632 \end{pmatrix}$$

$$(\mathbf{X}^T \mathbf{X})^{-1} = \frac{1}{19721} \begin{pmatrix} 1632 & -1168 \\ -1168 & 848 \end{pmatrix} = \begin{pmatrix} 0.08279 & -0.05925 \\ -0.05925 & 0.043019 \end{pmatrix}$$

$$\mathbf{X}^T \mathbf{y} = \begin{pmatrix} 6612 \\ 9194 \end{pmatrix}$$

Therefore,

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \begin{pmatrix} 2.647727 \\ 3.738636 \end{pmatrix}$$

Now, for the mean squared error (MSE), we know that:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \mathbf{X}_i \hat{\beta})^2}{n - \text{rank}(\mathbf{X})}$$

In which $\text{rank}(\mathbf{X}) = 2$. Then,

$$(y - \mathbf{X} \hat{\beta})^2 = [0.31004, 8.00632, 2.45919, 4.86002, 2.14888, 4.76033, 5.26911, 0.39062]^T$$

$$\hat{\sigma}^2 = \frac{28.20455}{8 - 2} = 4.70075$$

(b)

Provide 95% confidence intervals for β_1 and $2 * \beta_1 + \beta_2$.

Solution:

In order to get the confidence intervals, we can use the t-statistic given by:

$$t = \frac{\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} - \boldsymbol{\lambda}^T \boldsymbol{\beta}}{\sqrt{\hat{\sigma}^2 \boldsymbol{\lambda}^T (\mathbf{X}^T \mathbf{X})^{-1} \boldsymbol{\lambda}}} \sim t_{n-\text{rank}(\mathbf{X})}$$

- First, β_1 .

In this case, $\boldsymbol{\lambda}^T = [1, 0]^T$, and all the other estimates and statistics we already have from (a). Thus,

$$P \left(\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} - t_{\frac{\alpha}{2}, n-2} \sqrt{\hat{\sigma}^2 \boldsymbol{\lambda}^T (\mathbf{X}^T \mathbf{X})^{-1} \boldsymbol{\lambda}} \leq \boldsymbol{\lambda}^T \boldsymbol{\beta} \leq \boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} + t_{\frac{\alpha}{2}, n-2} \sqrt{\hat{\sigma}^2 \boldsymbol{\lambda}^T (\mathbf{X}^T \mathbf{X})^{-1} \boldsymbol{\lambda}} \right) = 1 - \alpha$$

So, as we want a 95% confidence interval, the $t_{\frac{\alpha}{2}, n-2} = 2.44691$.

$$\sqrt{\hat{\sigma}^2 \boldsymbol{\lambda}^T (\mathbf{X}^T \mathbf{X})^{-1} \boldsymbol{\lambda}} = 0.62384$$

Then, we can find the interval, which is given by:

$$P \left(2.64772 - t_{\frac{\alpha}{2}, n-2} * 0.62384 \leq \boldsymbol{\lambda}^T \boldsymbol{\beta} \leq 2.64772 + t_{\frac{\alpha}{2}, n-2} * 0.62384 \right) = 0.95$$

Therefore, $[1.12122, 4.17422]$ is the 95% confidence interval for β_1 .

- Second, for $2 * \beta_1 + \beta_2$.

In this case, $\boldsymbol{\lambda}^T = [2, 1]^T$, and all the other estimates and statistics we already have from (a). Thus,

$$P \left(\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} - t_{\frac{\alpha}{2}, n-2} \sqrt{\hat{\sigma}^2 \boldsymbol{\lambda}^T (\mathbf{X}^T \mathbf{X})^{-1} \boldsymbol{\lambda}} \leq \boldsymbol{\lambda}^T \boldsymbol{\beta} \leq \boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} + t_{\frac{\alpha}{2}, n-2} \sqrt{\hat{\sigma}^2 \boldsymbol{\lambda}^T (\mathbf{X}^T \mathbf{X})^{-1} \boldsymbol{\lambda}} \right) = 1 - \alpha$$

So, as we want a 95% confidence interval, the $t_{\frac{\alpha}{2}, n-2} = 2.44691$.

$$\sqrt{\hat{\sigma}^2 \boldsymbol{\lambda}^T (\mathbf{X}^T \mathbf{X})^{-1} \boldsymbol{\lambda}} = 0.80301$$

Then, we can find the interval, which is given by:

$$P \left(9.03409 - t_{\frac{\alpha}{2}, n-2} * 0.80301 \leq \boldsymbol{\lambda}^T \boldsymbol{\beta} \leq 9.03409 + t_{\frac{\alpha}{2}, n-2} * 0.80301 \right) = 0.95$$

Therefore, $[7.06919, 10.99899]$ is the 95% confidence interval for $2 * \beta_1 + \beta_2$.

(c)

Perform a $\alpha = 0.01$ level test for $H_0: \beta_2 = 3$.

Solution:

In order, to perform the test, first we know that $\lambda^T = [0, 1]$, and then we can do the test using the t-statistic given by:

$$t = \frac{\lambda^T \hat{\beta} - \lambda^T \beta}{\sqrt{\hat{\sigma}^2 \lambda^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda}} \sim t_{n-\text{rank}(\mathbf{X})}$$

In this case, $\lambda^T \beta = 3$, then using the estimate and statistics from (a),

$$\sqrt{\hat{\sigma}^2 \lambda^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda} = 0.44969$$

Then, the t-statistic and p-value are given by:

$$t_{obs} = \frac{3.73863 - 3}{0.44969} = 1.64253$$

$$p\text{-value} = P_{H_0}(|t| > t_{obs}) = P_{H_0}(|t| > 1.64253) = 0.15158$$

Therefore, we fail to reject the null hypothesis, as $p\text{-value} > 0.1$.

(d)

Find $p\text{-value}$ for the test $H_0: \beta_1 = \beta_2$.

Solution:

In order, to perform the test, first we know that $\lambda^T = [1, -1]$, and then we can do the test using the t-statistic given by:

$$t = \frac{\lambda^T \hat{\beta} - \lambda^T \beta}{\sqrt{\hat{\sigma}^2 \lambda^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda}} \sim t_{n-\text{rank}(\mathbf{X})}$$

In this case, $\lambda^T \beta = 0$, then using the estimate and statistics from (a),

$$\sqrt{\hat{\sigma}^2 \lambda^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda} = 1.07167$$

Then, the t-statistic and p-value are given by:

$$t_{obs} = \frac{-1.09091}{1.07167} = -1.01795$$

$$p\text{-value} = P_{H_0}(|t| > t_{obs}) = P_{H_0}(|t| > 1.01795) = 0.34797$$

Therefore, we fail to reject the null hypothesis, as $p\text{-value} > 0.1$.

Question 2

Consider the setting and the dataset in the previous question. Use the R package to run linear regression. Provide

(a)

p -value for testing $\beta_2 = 0$.

Solution:

Then, for the `lm` function in R, we get as result from the output:

```
> fit.2 = lm(y ~ -1 + x1 + x2)
> summary(fit.2)

Call:
lm(formula = y ~ -1 + x1 + x2)

Residuals:
    Min       1Q   Median       3Q      Max
-2.29545 -1.65057  0.03409  1.72159  2.82955

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
x1      2.6477      0.6238   4.244 0.005415 **
x2      3.7386      0.4497   8.314 0.000164 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.168 on 6 degrees of freedom
Multiple R-squared:  0.9995, Adjusted R-squared:  0.9993
F-statistic: 5518 on 2 and 6 DF,  p-value: 1.604e-10
```

In which we can check that the p -value for testing $\beta_2 = 0$ is equal to 0.000164, in which we have strong evidence to reject the hypothesis that $\beta_2 = 0$, using a $\alpha = 0.001$.

(b)

Draw the joint confidence set for (β_1, β_2) .

Solution:

In order to get the joint confidence interval, we are going to use the `ellipse` function in R, This function happens to have a specialized method for `lm` models. then, we get as output:

```
> library(ellipse)

> plot(ellipse(fit.2, which = c(1,2), level = 0.95), type = 'l',
+ cex.lab=1.5,cex.main = 2,cex.axis=1.8)

> points(fit.2$coefficients[1], fit.2$coefficients[2], col=2)
> segments(fit.2$coefficients[1],0,fit.2$coefficients[1],fit.2$coefficients[2])
> segments(0,fit.2$coefficients[2],fit.2$coefficients[1],fit.2$coefficients[2])
> legend(3.5,4.5, pch=1, box.lty = 0,cex=1.6,
+ legend=c(expression(paste(hat(beta)[1]:hat(beta)[2]))),bty='n',col=c(2))
```

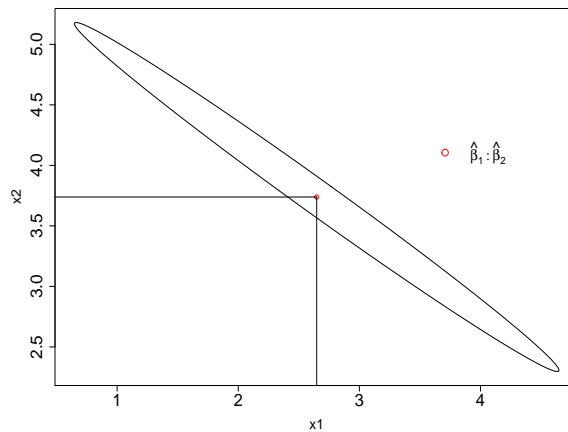


Figure 1: Joint confidence set of (β_1, β_2) .

Thus, Figure 1 display the 95% contour of the joint distribution of $\hat{\beta}_1$ and $\hat{\beta}_2$. So, we have 95% of the probability mass of the joint distribution falls within this ellipse.

(c)

Add an intercept to the model and check if predictor coefficients are significant.

Solution:

Then, as output from the `lm` function in R, we get as result:

```
> summary(fit.3)
```

Call:

```
lm(formula = y ~ x1 + x2)
```

Residuals:

1	2	3	4	5	6	7	8
0.125	1.125	-1.500	-0.500	-1.125	1.500	-0.250	0.625

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	30.0000	8.2583	3.633	0.01502 *
x1	1.6250	0.4556	3.567	0.01610 *
x2	2.3750	0.4556	5.213	0.00343 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.245 on 5 degrees of freedom

Multiple R-squared: 0.8826, Adjusted R-squared: 0.8356

F-statistic: 18.79 on 2 and 5 DF, p-value: 0.004725

Therefore, the predictors coefficients are significant, as the p -value are less than 0.01 for all three coefficients (including the intercept), it can be seen in column named $\text{Pr}(>|t|)$ in the output given by the `lm` function above.

Question 3

Considering a model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, where $\mathbf{X} = [\mathbf{1} : \mathbf{x}_1 : \mathbf{x}_2 : \dots : \mathbf{x}_p]$. Show that the model fitting statistic R^2 for this model is simply the square of the correlations between observed and predicted values of y .

Solution:

The observed values are: $\mathbf{y} = [y_i, \dots, y_n]^T$ and the predicted values are: $\hat{\mathbf{y}} = [\hat{y}_i, \dots, \hat{y}_n]^T$. Then, by definition we know that:

$$\text{corr}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{\text{cov}(\mathbf{y}, \hat{\mathbf{y}})}{\sqrt{\text{Var}(\mathbf{y})\text{Var}(\hat{\mathbf{y}})}} = \frac{S_{\mathbf{y}, \hat{\mathbf{y}}}}{\sqrt{S_{\mathbf{y}}S_{\hat{\mathbf{y}}}}}$$

Also, by definition, we have that:

$$\begin{aligned} S_{\mathbf{y}} &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 &= \frac{1}{n} \mathbf{y}^T (\mathbf{I} - \mathbf{P}_1) \mathbf{y} \\ S_{\hat{\mathbf{y}}} &= \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2 &= \frac{1}{n} \hat{\mathbf{y}}^T (\mathbf{I} - \mathbf{P}_1) \hat{\mathbf{y}} \\ S_{\mathbf{y}, \hat{\mathbf{y}}} &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}) &= \frac{1}{n} \mathbf{y}^T (\mathbf{I} - \mathbf{P}_1) \hat{\mathbf{y}} \end{aligned}$$

In which \mathbf{P}_1 is the perpendicular projection matrix onto $\mathcal{C}(\mathbf{W})$, $\mathbf{W} = [1, \dots, 1]^T$. In addition, we know that \mathbf{P}_x is the perpendicular projection matrix onto $\mathcal{C}(\mathbf{X})$. Another information would be, $\hat{\mathbf{y}} = \mathbf{P}_x \mathbf{y}$. Therefore,

Proof.

$$\begin{aligned} \text{corr}(\mathbf{y}, \hat{\mathbf{y}})^2 &= \frac{S_{\mathbf{y}, \hat{\mathbf{y}}}^2}{S_{\mathbf{y}}S_{\hat{\mathbf{y}}}} &= \frac{[\mathbf{y}^T (\mathbf{I} - \mathbf{P}_1) \hat{\mathbf{y}}]^2}{[\mathbf{y}^T (\mathbf{I} - \mathbf{P}_1) \mathbf{y}][\hat{\mathbf{y}}^T (\mathbf{I} - \mathbf{P}_1) \hat{\mathbf{y}}]} \\ &= \frac{[\mathbf{y}^T (\mathbf{I} - \mathbf{P}_1) \mathbf{P}_x \mathbf{y}]^2}{[\mathbf{y}^T (\mathbf{I} - \mathbf{P}_1) \mathbf{y}][\mathbf{y}^T \mathbf{P}_x (\mathbf{I} - \mathbf{P}_1) \mathbf{P}_x \mathbf{y}]} \\ &= \frac{[\mathbf{y}^T \mathbf{P}_x \mathbf{y} - \mathbf{y}^T \mathbf{P}_1 \mathbf{y}]^2}{[\mathbf{y}^T (\mathbf{I} - \mathbf{P}_1) \mathbf{y}][\mathbf{y}^T \mathbf{P}_x \mathbf{y} - \mathbf{y}^T \mathbf{P}_1 \mathbf{y}]} \\ &= \frac{[\mathbf{y}^T \mathbf{P}_x \mathbf{y} - \mathbf{y}^T \mathbf{P}_1 \mathbf{y}]}{[\mathbf{y}^T (\mathbf{I} - \mathbf{P}_1) \mathbf{y}]} \\ &= \frac{\mathbf{y}^T (\mathbf{P}_x - \mathbf{P}_1) \mathbf{y}}{\mathbf{y}^T (\mathbf{I} - \mathbf{P}_1) \mathbf{y}} = R^2 \end{aligned}$$

□

This ratio correspond to the regression sum of squared by total sum of squared, in which correspond to the definition of the coefficient of determination R^2 , ending the proof.

Question 3

Christensen presents mathematics ineptitude scores (Score y_{ijk}) for a group of $N = 35$ students categorized by:

- Major i (1 = Economics, 2 = Anthropology, and 3 = Sociology);
- High school background ("BG") j (1 = Rural and 2 = Urban).

The model is a two-way ANOVA, being:

$$y_{ijk} = \mu + \alpha_i + \nu_j + \gamma_{ij} + \epsilon_{ijk}$$

The constraints used were: $\alpha_1 = \nu_1 = 0$. Also $\gamma_{ij} = 0$ if $i = 1$ and $j = 1$.

See the R code used to do the analysis in Appendix 1, as well as the data used.

(a)

Which group of students has the lowest average score? (What is it?) Which group of students has the highest average score? (What is it?)

Solution:

We can see the graph of means of scores by the groups, in Figure 2 that the group lowest average score is given by students from Economics major with BG Rural, and the highest average score is students from Anthropology major with BG Rural.

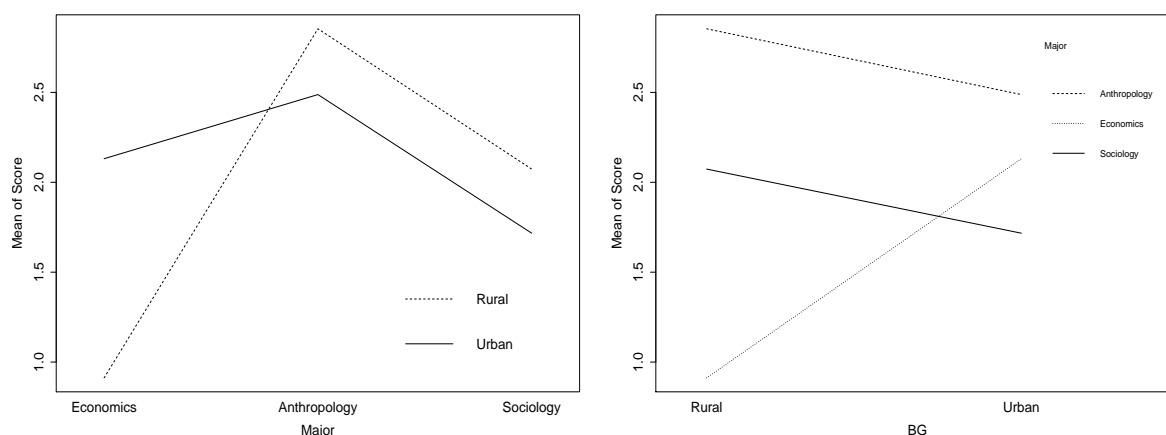


Figure 2: Mean Scores of the students by interaction between the groups.

In which, we can get from the output of the `lm` function in R, that: The highest average score gets value: $\mu + \text{Major } 2 = 0.9122 + 1.9418 = 2.8540$. The lowest average score gets value: $\mu = 0.9122 = 0.9122$. We have this value because the constraints in R is Major 1 = BG 1 = 0.

This means, that we expect to have the higher scores from students that have major in Anthropology and their high school background was Rural. And a lower scores from students hat have major in Economics and their high school background was Rural.

(b)

In the `summary(.)` output there is an F-statistic, $F = 2.553$ with 5 and 29 degrees of freedom.

Solution:

1. (i) What are the null and alternative hypotheses being tested?

From default the F-test tests from the `summary(.)` function in R, test the reduced model with $y_{ijk} = \mu$ versus the full model. We can check that we get the same result if we fit two models, one with $y_{ijk} = \mu$ and the other with all the groups, and then use the `anova` function to compare the two models.

2. (ii) What conclusion would you make? (Please state in general terms that relate to the groups rather than parameters).

We can check the output from the `summary(.)` function in R, that the p -value from the F-test is given by 0.0533, thus we reject the null hypothesis of the reduced model using $\alpha = 0.1$. Then, we have low evidence, but there is still evidence that there are differences between groups. And in Figure 2 we can observe a little of this behavior.

And there is low evidence of interaction between Economics and Sociology, whether the student with BG Rural or Urban. In Figure 2, we can note that the effect in the mean scores of BG being Rural, are always higher than Urban, if evaluate the majors Anthropology and Sociology. As well as the mean of scores of majors Anthropology are always higher than Sociology if we evaluated them separated. Moreover, we have an interaction between Economics and Sociology. As we point out in the letter (a) we can expect a real change in the mean scores, of a student with BG rural or urban, depending on whether the major is Economics or Sociology.

1 Appendix

See below all the R Code used to perform the analysis.

```
rm(list=ls(all=TRUE))
set.seed(27)

# take home - final - AMS 206

y = c(82, 79, 74, 83, 80, 81, 84, 81)
x1 = c(10, 9, 9, 11, 11, 10, 10, 12)
x2 = c(15, 14, 13, 15, 14, 14, 16, 13)

X = cbind(x1, x2)

fit.1 = lm(y ~ -1 + x1 + x2)
summary(fit.1)
p = 2
n = length(y)
XtX = t(X)%*%X
invXtX = solve(XtX)
Xty = t(X)%*%y

# Question 1
# 1.a
beta.hat = solve(t(X)%*%X)%*%t(X)%*%y
sigma.hat = sum((y-X%*%beta.hat)^2)/(n-p)

# 1.b
lambda.1 = c(1,0)
var.est.1 = sqrt((sigma.hat)%*%(t(lambda.1)%*%invXtX)%*%lambda.1))

t(lambda.1)%*%beta.hat + qt(0.975,6)*var.est.1
t(lambda.1)%*%beta.hat - qt(0.975,6)*var.est.1

lambda.2 = c(2,1)
var.est.2 = sqrt((sigma.hat)%*%(t(lambda.2)%*%invXtX)%*%lambda.2))

t(lambda.2)%*%beta.hat + qt(0.975,6)*var.est.2
t(lambda.2)%*%beta.hat - qt(0.975,6)*var.est.2

# 1.c
lambda.3 = c(0,1)
var.est.3 = sqrt((sigma.hat)%*%(t(lambda.3)%*%invXtX)%*%lambda.3))

t(lambda.3)%*%beta.hat + qt(0.975,6)*var.est.3
t(lambda.3)%*%beta.hat - qt(0.975,6)*var.est.3

t.val.3 = (t(lambda.3)%*%beta.hat - 3) / var.est.3
```

```

pval.3 = (1-pt(abs(t.val.3),n-p))*2

# 1.d
lambda.4 = c(1,-1)
var.est.4 = sqrt((sigma.hat)%*(t(lambda.4)%*invXtX%*lambda.4))

t(lambda.4)%*beta.hat + qt(0.975,6)*var.est.4
t(lambda.4)%*beta.hat - qt(0.975,6)*var.est.4

t.val.4 = (t(lambda.4)%*beta.hat - 0) / var.est.4
pval.4 = (1-pt(abs(t.val.4),n-p))*2

# Question 2
# 2.a
fit.2 = lm(y ~ -1 + x1 + x2)
summary(fit.2)

lambda.5 = c(0,1)
var.est.5 = sqrt((sigma.hat^2)%*(t(lambda.5)%*invXtX%*lambda.5))

t.val.5 = (t(lambda.5)%*beta.hat - 0) / var.est.5
pval.5 = (1-pt(abs(t.val.5),n-p))*2

# 2.b
library(ellipse)
plot(ellipse(fit.2, which = c(1,2), level = 0.95), type = 'l',
cex.lab=1.5,cex.main = 2,cex.axis=1.8)
points(fit.2$coefficients[1], fit.2$coefficients[2], col=2)
segments(fit.2$coefficients[1],0,fit.2$coefficients[1],fit.2$coefficients[2])
segments(0,fit.2$coefficients[2],fit.2$coefficients[1],fit.2$coefficients[2])
legend(3.5,4.5, pch=1, box.lty = 0,cex=1.6,
legend=c(expression(paste(hat(beta)[1]:hat(beta)[2]))),bty='n',col=c(2))

# 2.c
fit.3 = lm(y ~ x1 + x2)
summary(fit.3)

# Question 4
data = matrix(c(0.620, 1, 1,
1.342, 1, 1,
0.669, 1, 1,
0.687, 1, 1,
0.155, 1, 1,
2.000, 1, 1,
1.228, 1, 2,
3.762, 1, 2,
2.219, 1, 2,
4.207, 1, 2,
0.615, 1, 3,

```

```

2.245, 1, 3,
2.077, 1, 3,
3.357, 1, 3,
1.182, 2, 1,
1.068, 2, 1,
2.545, 2, 1,
2.233, 2, 1,
2.664, 2, 1,
1.002, 2, 1,
2.506, 2, 1,
4.285, 2, 1,
1.696, 2, 1,
3.080, 2, 2,
2.741, 2, 2,
2.522, 2, 2,
1.647, 2, 2,
1.999, 2, 2,
2.939, 2, 2,
2.240, 2, 3,
0.330, 2, 3,
3.453, 2, 3,
1.527, 2, 3,
0.809, 2, 3,
1.942, 2, 3), nrow = 35, ncol = 3, byrow = TRUE,
  dimnames=list(c(seq(1,35,1)),c("Scores","BG","Major")))
data = as.data.frame(data)

fit.4 = lm(Scores ~ as.factor(Major)*as.factor(BG), data=data)
summary(fit.4)
anova(fit.4)

names <- c('Major','BG')
data[, names] = lapply(data[, names], factor)
levels(data$Major)[c(1,2,3)] = c('Economics', 'Anthropology', "Sociology")
levels(data$BG)[c(1,2)] = c("Rural", "Urban")

interaction.plot(data$Major, data$BG, data$Scores, trace.lab="BG",
  legend=F,cex.axis=1.5, cex.lab=1.5, xlab="Major", ylab="Mean of Score", cex=1.5)
legend("bottomright", lty = c(2,1), box.lty = 0,cex=1.5, legend=c("Rural","Urban"))

interaction.plot(data$BG, data$Major, data$Scores, trace.lab="Major",cex.axis=1.5,
  cex.lab=1.5, xlab="BG", ylab="Mean of Score", cex=1.5)

fit.5 = lm(Scores ~ 1, data=data)

anova(fit.4, fit.5)

```