

3.12 SUMMARY

We have illustrated the testing of various hypotheses in connection with the linear model. Rather than describing individual tests we have outlined a general procedure by which they can be performed. It has been shown that the various tests can also be described in terms of the appropriate sample multiple correlation coefficients. It is to be emphasized here, that before starting on any testing procedure, the adequacy of the model assumptions should always be examined. As we shall see in Chapter 4, residual plots provide a very convenient graphical way of accomplishing this task. The test procedures are not valid if the assumptions on which the tests are based do not hold. If a new model is chosen on the basis of a statistical test, residuals from the new model should be examined before terminating the analysis. It is only by careful attention to detail that a satisfactory analysis of data can be carried out.

EXERCISES

- 3.1** Using the Supervisor data, verify that the coefficient of X_1 in the fitted equation $\hat{Y} = 15.3276 + 0.7803X_1 - 0.0502X_2$ in (3.12) can be obtained from a series of simple regression equations, as outlined in Section 3.5 for the coefficient of X_2 . (equations on lecture slides)
- 3.2** Construct a small data set consisting of one response and two predictor variables so that the regression coefficient of X_1 in the following two fitted equations are equal: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1$ and $\hat{Y} = \hat{\alpha}_0 + \hat{\alpha}_1 X_1 + \hat{\alpha}_2 X_2$. Hint: The two predictor variables should be uncorrelated.
- 3.3** Table 3.10 shows the scores in the final examination F and the scores in two preliminary examinations P_1 and P_2 for 22 students in a statistics course. The data can be found at the book's Website.
- (a) Fit each of the following models to the data:

$$\text{Model 1: } F = \beta_0 + \beta_1 P_1 + \varepsilon$$

$$\text{Model 2: } F = \beta_0 + \beta_2 P_2 + \varepsilon$$

$$\text{Model 3: } F = \beta_0 + \beta_1 P_1 + \beta_2 P_2 + \varepsilon$$

- (b) Test whether $\beta_0 = 0$ in each of the three models.
- (c) Which variable individually, P_1 or P_2 , is a better predictor of F ?
- (d) Which of the three models would you use to predict the final examination scores for a student who scored 78 and 85 on the first and second preliminary examinations, respectively? What is your prediction in this case?
- 3.4** Find or construct a simple or multiple regression data set such that the resulting adjusted R_a^2 is negative.

Table 3.10 Examination Data: Scores in Final (F), First Preliminary (P_1), and Second Preliminary (P_2) Examinations

Row	F	P_1	P_2	Row	F	P_1	P_2
1	68	78	73	12	75	79	75
2	75	74	76	13	81	89	84
3	85	82	79	14	91	93	97
4	94	90	96	15	80	87	77
5	86	87	90	16	94	91	96
6	90	90	92	17	94	86	94
7	86	83	95	18	97	91	92
8	68	72	69	19	79	81	82
9	55	68	67	20	84	80	83
10	69	69	70	21	65	70	66
11	91	91	89	22	83	79	81

3.5 The relationship between the simple and the multiple regression coefficients can be seen when we compare the following regression equations:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2, \quad (3.64)$$

$$\hat{Y} = \hat{\beta}'_0 + \hat{\beta}'_1 X_1, \quad (3.65)$$

$$\hat{Y} = \hat{\beta}''_0 + \hat{\beta}'_2 X_2, \quad (3.66)$$

$$\hat{X}_1 = \hat{\alpha}_0 + \hat{\alpha}_2 X_2, \quad (3.67)$$

$$\hat{X}_2 = \hat{\alpha}'_0 + \hat{\alpha}_1 X_1. \quad (3.68)$$

Using the Examination Data in Table 3.10 with $Y = F$, $X_1 = P_1$, and $X_2 = P_2$, verify that:

- (a) $\hat{\beta}'_1 = \hat{\beta}_1 + \hat{\beta}_2 \hat{\alpha}_1$, that is, the simple regression coefficient of Y on X_1 is the multiple regression coefficient of X_1 plus the multiple regression coefficient of X_2 times the coefficient from the regression of X_2 on X_1 .
- (b) $\hat{\beta}'_2 = \hat{\beta}_2 + \hat{\beta}_1 \hat{\alpha}_2$, that is, the simple regression coefficient of Y on X_2 is the multiple regression coefficient of X_2 plus the multiple regression coefficient of X_1 times the coefficient from the regression of X_1 on X_2 .

3.6 Table 3.11 shows the regression output, with some numbers erased, when a simple regression model relating a response variable Y to a predictor variable X_1 is fitted based on 20 observations. Complete the 13 missing numbers, then compute $\text{Var}(Y)$ and $\text{Var}(X_1)$.

3.7 Table 3.12 shows the regression output, with some numbers erased, when a simple regression model relating a response variable Y to a predictor variable X_1 is fitted based on 18 observations. Complete the 13 missing numbers, then compute $\text{Var}(Y)$ and $\text{Var}(X_1)$.

Table 3.11 Regression Output When Y is Regressed on X_1 for 20 Observations

ANOVA Table				
Source	Sum of Squares	df	Mean Square	F -Test
Regression	1848.76	–	–	–
Residuals	–	–	–	–
Coefficients Table				
Variable	Coefficient	s.e.	t -Test	p -value
Constant	–23.4325	12.74	–	0.0824
X_1	–	0.1528	8.32	< 0.0001
$n = -$	$R^2 = -$	$R_a^2 = -$	$\hat{\sigma} = -$	df = –

Table 3.12 Regression Output When Y is Regressed on X_1 for 18 Observations

ANOVA Table				
Source	Sum of Squares	df	Mean Square	F -Test
Regression	–	–	–	–
Residuals	–	–	–	–
Coefficients Table				
Variable	Coefficient	s.e.	t -Test	p -value
Constant	3.43179	–	0.265	0.7941
X_1	–	0.1421	–	< 0.0001
$n = -$	$R^2 = 0.716$	$R_a^2 = -$	$\hat{\sigma} = 7.342$	df = –

3.8 Construct the 95% confidence intervals for the individual parameters β_1 and β_2 using the regression output in Table 3.5. (table on lecture slides)

3.9 Explain why the test for testing the hypothesis H_0 in (3.57) is more sensitive for detecting departures from equality of the regression coefficients than the test for testing the hypothesis H'_0 in (3.61). (equations on lecture slides)

3.10 Using the Supervisor Performance data, test the hypothesis $H_0 : \beta_1 = \beta_3 = 0.5$ in each of the following models:

(a) $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \varepsilon.$

(b) $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon.$

Table 3.13 Regression Outputs for Salary Discriminating Data

Model 1: Dependent variable is Salary				
Variable	Coefficient	s.e.	<i>t</i> -Test	<i>p</i> -value
Constant	20009.5	0.8244	24271	< 0.0001
Qualification	0.935253	0.0500	18.7	< 0.0001
Gender	0.224337	0.4681	0.479	0.6329
Model 2: Dependent variable is Qualification				
Variable	Coefficient	s.e.	<i>t</i> -Test	<i>p</i> -value
Constant	−16744.4	896.4	−18.7	< 0.0001
Gender	0.850979	0.4349	1.96	0.0532
Salary	0.836991	0.0448	18.7	< 0.0001

- 3.11** Refer to Exercise 2.10 and the data in Table 2.11, which can also be found at the book's Website. (find the table in chapter 2 lecture slides, and practice exercise)
- Using your choice of the response variable Exercise 2.10(f), test the null hypothesis that both the intercept and the slope are zero.
 - Which of the hypotheses and tests in Exercises 2.10(g), 2.10(h), and 3.11(a) would you choose to test whether people of similar heights tend to marry each other? What is your conclusion?
 - If none of the above tests is appropriate for testing the hypothesis that people of similar heights tend to marry each other, which test would you use? What is your conclusion based on this test?
- 3.12** To decide whether a company is discriminating against women, the following data were collected from the company's records: Salary is the annual salary in thousands of dollars, Qualification is an index of employee qualification, and Gender (1, if the employee is a man, and 0, if the employee is a woman). Two linear models were fit to the data and the regression outputs are shown in Table 3.13. Suppose that the usual regression assumptions hold.
- Are men paid more than equally qualified women?
 - Are men less qualified than equally paid women?
 - Do you detect any inconsistency in the above results? Explain.
 - Which model would you advocate if you were the defense lawyer? Explain.
- 3.13** Table 3.14 shows the regression output of a multiple regression model relating the beginning salaries in dollars of employees in a given company to the following predictor variables:

Table 3.14 Regression Output When Salary is Related to Four Predictor Variables

ANOVA Table				
Source	Sum of Squares	df	Mean Square	F-Test
Regression	23665352	4	5916338	22.98
Residuals	22657938	88	257477	

Coefficients Table				
Variable	Coefficient	s.e.	t-Test	p-value
Constant	3526.4	327.7	10.76	0.000
Gender	722.5	117.8	6.13	0.000
Education	90.02	24.69	3.65	0.000
Experience	1.2690	0.5877	2.16	0.034
Months	23.406	5.201	4.50	0.000
$n = 93$	$R^2 = 0.515$	$R_a^2 = 0.489$	$\hat{\sigma} = 507.4$	$df = 88$

Gender	An indicator variable (1 = man and 0 = woman)
Education	Years of schooling at the time of hire
Experience	Number of months of previous work experience
Months	Number of months with the company

In (a)–(b) below, specify the null and alternative hypotheses, the test used, and your conclusion using a 5% level of significance.

- Conduct the F -Test for the overall fit of the regression.
- Is there a *positive* linear relationship between Salary and Experience, after accounting for the effect of the variables Gender, Education, and Months?
- What salary would you forecast for a man with 12 years of education, 10 months of experience, and 15 months with the company?
- What salary would you forecast, on average, for men with 12 years of education, 10 months of experience, and 15 months with the company?
- What salary would you forecast, on average, for women with 12 years of education, 10 months of experience, and 15 months with the company?

3.14 Consider the regression model that generated the output in Table 3.14 to be a full model. Now consider the reduced model in which Salary is regressed on only Education. The ANOVA table obtained when fitting this model is shown in Table 3.15. Conduct a single test to compare the full and reduced models. What conclusion can be drawn from the result of the test? (Use $\alpha = 0.05$.)

3.15 Cigarette Consumption Data: A national insurance organization wanted to study the consumption pattern of cigarettes in all 50 states and the District of

Table 3.15 ANOVA Table When the Beginning Salary is Regressed on Education

ANOVA Table				
Source	Sum of Squares	df	Mean Square	<i>F</i> -Test
Regression	7862535	1	7862535	18.60
Residuals	38460756	91	422646	

Table 3.16 Variables in the Cigarette Consumption Data in Table 3.17

Variable	Definition
Age	Median age of a person living in a state
HS	Percentage of people over 25 years of age in a state who had completed high school
Income	Per capita personal income for a state (income in dollars)
Black	Percentage of blacks living in a state
Female	Percentage of females living in a state
Price	Weighted average price (in cents) of a pack of cigarettes in a state
Sales	Number of packs of cigarettes sold in a state on a per capita basis

Columbia. The variables chosen for the study are given in Table 3.16. The data from 1970 are given in Table 3.17. The states are given in alphabetical order. The data can be found at the book's Website.

In (a)–(b) below, specify the null and alternative hypotheses, the test used, and your conclusion using a 5% level of significance.

- Test the hypothesis that the variable Female is not needed in the regression equation relating Sales to the six predictor variables.
- Test the hypothesis that the variables Female and HS are not needed in the above regression equation.
- Compute the 95% confidence interval for the true regression coefficient of the variable Income.
- What percentage of the variation in Sales can be accounted for when Income is removed from the above regression equation? Explain.
- What percentage of the variation in Sales can be accounted for by the three variables: Price, Age, and Income? Explain.
- What percentage of the variation in Sales that can be accounted for by the variable Income, when Sales is regressed on only Income? Explain.

3.16 Consider the two models:

$$\text{RM: } H_0 : Y = \varepsilon,$$

$$\text{FM: } H_1 : Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon.$$

Table 3.17 Cigarette Consumption Data (1970)

State	Age	HS	Income	Black	Female	Price	Sales
AL	27.0	41.3	2948.0	26.2	51.7	42.7	89.8
AK	22.9	66.7	4644.0	3.0	45.7	41.8	121.3
AZ	26.3	58.1	3665.0	3.0	50.8	38.5	115.2
AR	29.1	39.9	2878.0	18.3	51.5	38.8	100.3
CA	28.1	62.6	4493.0	7.0	50.8	39.7	123.0
CO	26.2	63.9	3855.0	3.0	50.7	31.1	124.8
CT	29.1	56.0	4917.0	6.0	51.5	45.5	120.0
DE	26.8	54.6	4524.0	14.3	51.3	41.3	155.0
DC	28.4	55.2	5079.0	71.1	53.5	32.6	200.4
FL	32.3	52.6	3738.0	15.3	51.8	43.8	123.6
GA	25.9	40.6	3354.0	25.9	51.4	35.8	109.9
HI	25.0	61.9	4623.0	1.0	48.0	36.7	82.1
ID	26.4	59.5	3290.0	0.3	50.1	33.6	102.4
IL	28.6	52.6	4507.0	12.8	51.5	41.4	124.8
IN	27.2	52.9	3772.0	6.9	51.3	32.2	134.6
IA	28.8	59.0	3751.0	1.2	51.4	38.5	108.5
KS	28.7	59.9	3853.0	4.8	51.0	38.9	114.0
KY	27.5	38.5	3112.0	7.2	50.9	30.1	155.8
LA	24.8	42.2	3090.0	29.8	51.4	39.3	115.9
ME	28.0	54.7	3302.0	0.3	51.3	38.8	128.5
MD	27.1	52.3	4309.0	17.8	51.1	34.2	123.5
MA	29.0	58.5	4340.0	3.1	52.2	41.0	124.3
MI	26.3	52.8	4180.0	11.2	51.0	39.2	128.6
MN	26.8	57.6	3859.0	0.9	51.0	40.1	104.3
MS	25.1	41.0	2626.0	36.8	51.6	37.5	93.4
MO	29.4	48.8	3781.0	10.3	51.8	36.8	121.3
MT	27.1	59.2	3500.0	0.3	50.0	34.7	111.2
NB	28.6	59.3	3789.0	2.7	51.2	34.7	108.1
NV	27.8	65.2	4563.0	5.7	49.3	44.0	189.5
NH	28.0	57.6	3737.0	0.3	51.1	34.1	265.7
NJ	30.1	52.5	4701.0	10.8	51.6	41.7	120.7
NM	23.9	55.2	3077.0	1.9	50.7	41.7	90.0
NY	30.3	52.7	4712.0	11.9	52.2	41.7	119.0
NC	26.5	38.5	3252.0	22.2	51.0	29.4	172.4
ND	26.4	50.3	3086.0	0.4	49.5	38.9	93.8
OH	27.7	53.2	4020.0	9.1	51.5	38.1	121.6
OK	29.4	51.6	3387.0	6.7	51.3	39.8	108.4
OR	29.0	60.0	3719.0	1.3	51.0	29.0	157.0
PA	30.7	50.2	3971.0	8.0	52.0	44.7	107.3
RI	29.2	46.4	3959.0	2.7	50.9	40.2	123.9
SC	24.8	37.8	2990.0	30.5	50.9	34.3	103.6
SD	27.4	53.3	3123.0	0.3	50.3	38.5	92.7
TN	28.1	41.8	3119.0	15.8	51.6	41.6	99.8
TX	26.4	47.4	3606.0	12.5	51.0	42.0	106.4
UT	23.1	67.3	3227.0	0.6	50.6	36.6	65.5
VT	26.8	57.1	3468.0	0.2	51.1	39.5	122.6
VA	26.8	47.8	3712.0	18.5	50.6	30.2	124.3
WA	27.5	63.5	4053.0	2.1	50.3	40.3	96.7
WV	30.0	41.6	3061.0	3.9	51.6	41.6	114.5
WI	27.2	54.5	3812.0	2.9	50.9	40.2	106.4
WY	27.2	62.9	3815.0	0.8	50.0	34.4	132.2

- (a) Develop an F -Test for testing the above hypotheses.
- (b) Let $p = 1$ (simple regression) and construct a data set Y and X_1 such that H_0 is not rejected at the 5% significance level.
- (c) What does the null hypothesis indicate in this case?
- (d) Compute the appropriate value of R^2 that relates the above two models.

Appendix: Multiple Regression in Matrix Notation

We present the standard results of multiple regression analysis in matrix notation. Let us define the following matrices:

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} x_{10} & x_{11} & \cdots & x_{1p} \\ x_{20} & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n0} & x_{n1} & \cdots & x_{np} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

The linear model in (3.1) can be expressed in terms of the above matrices as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (\text{A.1})$$

where $x_{i0} = 1$ for all i . The assumptions made about $\boldsymbol{\varepsilon}$ for least squares estimation are

$$E(\boldsymbol{\varepsilon}) = \mathbf{0} \quad \text{and} \quad \text{Var}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \sigma^2\mathbf{I}_n,$$

where $E(\boldsymbol{\varepsilon})$ is the expected value (mean) of $\boldsymbol{\varepsilon}$, \mathbf{I}_n is the identity matrix of order n , and $\boldsymbol{\varepsilon}^T$ is the transpose of $\boldsymbol{\varepsilon}$. Accordingly, ε_i 's are independent and have zero mean and constant variance. This implies that

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}.$$

The least squares estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained by minimizing the sum of squared deviations of the observations from their expected values. Hence the least squares estimators are obtained by minimizing $S(\boldsymbol{\beta})$, where

$$S(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

Minimization of $S(\boldsymbol{\beta})$ leads to the system of equations

$$(\mathbf{X}^T \mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{Y}. \quad (\text{A.2})$$

This is the system of *normal equations* referred to in Section (3.4). Assuming that $(\mathbf{X}^T \mathbf{X})$ has an inverse, the least squares estimates $\hat{\boldsymbol{\beta}}$ can be written explicitly as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, \quad (\text{A.3})$$

from which it can be seen that $\hat{\boldsymbol{\beta}}$ is a linear function of \mathbf{Y} . The vector of fitted values $\hat{\mathbf{Y}}$ corresponding to the observed \mathbf{Y} is

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{P}\mathbf{Y}, \quad (\text{A.4})$$

where

$$\mathbf{P} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T, \quad (\text{A.5})$$

is known as the *hat* or *projection* matrix. The vector of residuals is given by

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{P}\mathbf{Y} = (\mathbf{I}_n - \mathbf{P})\mathbf{Y}. \quad (\text{A.6})$$

The properties of the least squares estimators are

1. $\hat{\beta}$ is an unbiased estimator of β (i.e., $E(\hat{\beta}) = \beta$) with variance-covariance matrix $\text{Var}(\hat{\beta})$, which is

$$\text{Var}(\hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T = \sigma^2(\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 \mathbf{C},$$

where

$$\mathbf{C} = (\mathbf{X}^T \mathbf{X})^{-1}. \quad (\text{A.7})$$

Of all unbiased estimators of β that are linear in the observations, the least squares estimator has minimum variance. For this reason, $\hat{\beta}$ is said to be the *best linear unbiased estimator* (BLUE) of β .

2. The residual sum of squares can be expressed as

$$\mathbf{e}^T \mathbf{e} = \mathbf{Y}^T (\mathbf{I}_n - \mathbf{P})^T (\mathbf{I}_n - \mathbf{P}) \mathbf{Y} = \mathbf{Y}^T (\mathbf{I}_n - \mathbf{P}) \mathbf{Y}. \quad (\text{A.8})$$

The last equality follows because $(\mathbf{I}_n - \mathbf{P})$ is a symmetric idempotent matrix.

3. An unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{\mathbf{e}^T \mathbf{e}}{n - p - 1} = \frac{\mathbf{Y}^T (\mathbf{I}_n - \mathbf{P}) \mathbf{Y}}{n - p - 1}. \quad (\text{A.9})$$

With the added assumption that the ε_i 's are normally distributed we have the following additional results:

4. The vector $\hat{\beta}$ has a $(p + 1)$ -dimensional normal distribution with mean vector β and variance-covariance matrix $\sigma^2 \mathbf{C}$. The marginal distribution of $\hat{\beta}_j$ is normal with mean β_j and variance $\sigma^2 c_{jj}$, where c_{jj} is the j th diagonal element of \mathbf{C} in (A.7). Accordingly, the standard error of $\hat{\beta}_j$ is

$$\text{s.e.}(\hat{\beta}_j) = \hat{\sigma} \sqrt{c_{jj}}, \quad (\text{A.10})$$

and the covariance of $\hat{\beta}_i$ and $\hat{\beta}_j$ is $\text{Cov}(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 c_{ij}$.

5. The quantity $W = \mathbf{e}^T \mathbf{e} / \sigma^2$ has an χ^2 distribution with $n - p - 1$ degrees of freedom.
6. $\hat{\beta}$ and $\hat{\sigma}^2$ are distributed independently of one another.
7. The vector of fitted values $\hat{\mathbf{Y}}$ has a singular n -dimensional normal distribution with mean $E(\hat{\mathbf{Y}}) = \mathbf{X}\beta$ and variance-covariance matrix $\text{Var}(\hat{\mathbf{Y}}) = \sigma^2 \mathbf{P}$.

8. The residual vector \mathbf{e} has a singular n -dimensional normal distribution with mean $E(\mathbf{e}) = \mathbf{0}$ and variance-covariance matrix $\text{Var}(\mathbf{e}) = \sigma^2(\mathbf{I}_n - \mathbf{P})$.
9. The predicted value \hat{y}_0 corresponding to an observation vector $\mathbf{x}_0 = (x_{00}, x_{01}, x_{02}, \dots, x_{0p})^T$, with $x_{00} = 1$ is

$$\hat{y}_0 = \mathbf{x}_0^T \hat{\boldsymbol{\beta}} \quad (\text{A.11})$$

and its standard error is

$$\text{s.e.}(\hat{y}_0) = \hat{\sigma} \sqrt{1 + \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0}. \quad (\text{A.12})$$

The mean response μ_0^T corresponding to \mathbf{x}_0^T is

$$\hat{\mu}_0 = \mathbf{x}_0^T \hat{\boldsymbol{\beta}} \quad (\text{A.13})$$

with a standard error

$$\text{s.e.}(\hat{\mu}_0) = \hat{\sigma} \sqrt{\mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0}. \quad (\text{A.14})$$

10. The $100(1 - \alpha)\%$ joint confidence region for the regression parameters $\boldsymbol{\beta}$ is given by

$$\left\{ \boldsymbol{\beta} : \frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T (\mathbf{X}^T \mathbf{X}) (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{\hat{\sigma}^2(p+1)} \leq F_{(p+1, n-p-1, \alpha)} \right\}, \quad (\text{A.15})$$

which is an ellipsoid centered at $\hat{\boldsymbol{\beta}}$.