Printout

Sunday, November 7, 2021 2:01 PM

Section 1

MANE 6313

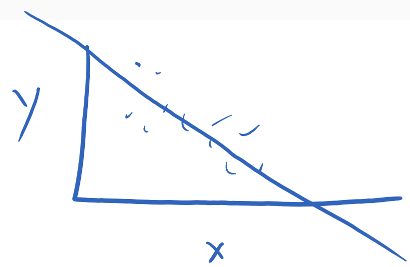
Subsection 1

Week 12, Module A

Student Learning Outcome

- Select an appropriate experimental design with one or more factors,
- Select an appropriate model with one or more factors,
- Evaluate statistical analyses of experimental designs,
- Assess the model adequacy of any experimental design, and
- Interpret model results.

Module Learning Outcome



Describe linear regression.

Resources for the Week 12, Module A micro-lecture are:

- Week 12, Module A Micro-lecture
- Week 12, Module A Marked Notes

Introduction to Linear Regression

- We are interested in a relationship between a single *dependent variable* or *response* y that depends on k *independent* or *regressor variables*.
- We assume that there is some mathematical function $y = \phi(x_1, x_2, \ldots, x_k)$. In general, we don't know this function
  $= f(x_1, x_2, \ldots, x_k)$
- We'll use low order polynomial equations as an approximating function. This is called *empirical modeling*.
- What are methods that we can determine if there is a relationship between two (or more) variables?

Relationship between two or more variables

Example 12.8 Suppose that a scientist takes experimental data on the radius of a propellant grain Y as a function of powder temperature x_1, extrusion rate x_2, and die temperature x_3. Fit a linear regression model for predicting grain radius, and determine the effectiveness of each variable in the model. The data are as follows:

$2^3 = 8$

y Grain radius	x_1 Powder temperature	x_2 Extrusion rate	x_3 Die temperature
82	150 (−1)	12 (−1)	220 (−1)
93	190 (1)	12 (−1)	220 (−1)
114	150 (−1)	24 (1)	220 (−1)
124	150 (−1)	12 (−1)	250 (1)
111	190 (1)	24 (1)	220 (−1)
129	190 (1)	12 (−1)	250 (1)
157	150 (−1)	24 (1)	250 (1)
164	190 (1)	24 (1)	250 (1)

Figure 1: Example 12.8

Linear regression models

- In general they look like

$$y = \underbrace{\beta_0 + \beta_1 x_1 + \beta_2 x_2}_{\text{linear}} + \underbrace{\beta_{11} x_1^2 + \beta_{22} x_2^2}_{\text{full quadratic}} + \beta_{12} x_1 x_2 + \varepsilon$$

interaction

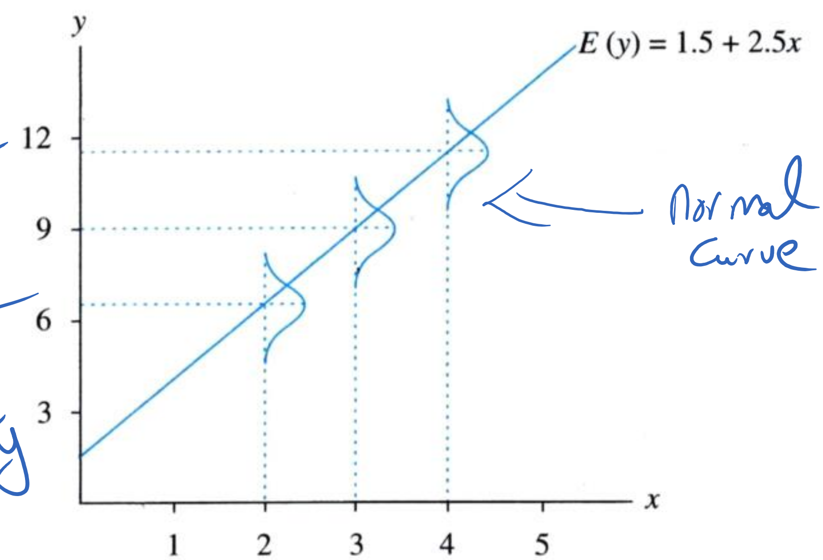
- This model is *linear* in the parameters β

$z_2 = x_2^2$

$\beta_{22} z_2$

- See graphical explanation from Ott.

FIGURE 11.2
Theoretical distribution of y in regression



Normality is not required to fit model

all inference on model parameters require normality

Figure 2: Figure 11.2

Estimation of Parameters

- Parameter estimates are derived using least squares. Goal is to minimize the squared error.
- In general, the matrix formulation is used. Model is defined to be

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

Y, β & ε are vectors
X matrix

- The least squares estimates are found by minimizing

$$L = \sum_i \varepsilon_i^2 = \varepsilon'\varepsilon = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

- The least squares estimates must satisfy

$$\frac{\partial L}{\partial \beta}\Big|_{\widehat{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\widehat{\beta} = 0$$

- Which leads to the solution

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

- We can define the predicted response to be $\hat{y}$ - predicted values

$$\widehat{\mathbf{y}} = \mathbf{X}\widehat{\beta}$$

- The residuals are defined to be

$$\mathbf{e} = \mathbf{y} - \widehat{\mathbf{y}}$$

- Thus the sum of squares errors can be shown to be

$$\begin{aligned} SS_E = & \quad (\mathbf{y} - \mathbf{X}\widehat{\beta})'(\mathbf{y} - \mathbf{X}\widehat{\beta}) \\ = & \quad \mathbf{y}'\mathbf{y} - \widehat{\beta}'\mathbf{X}'\mathbf{y} \end{aligned}$$

Matrix Formulation Example

Matrix Notation

4 vectors

$$y = \begin{bmatrix} 82 \\ 93 \\ 114 \\ 124 \\ 111 \\ 129 \\ 157 \\ 164 \end{bmatrix} \qquad X = [1 \vdots x_1 \vdots x_2 \vdots x_3] = \begin{array}{c} \begin{array}{cccc} B_0 & X_1 & X_2 & X_3 \end{array} \\ \begin{bmatrix} 1 & 150 & 12 & 220 \\ 1 & 190 & 12 & 2\ \ 0 \\ 1 & 150 & 24 & 220 \\ 1 & 150 & 12 & 250 \\ 1 & 190 & 24 & 220 \\ 1 & 190 & 12 & 250 \\ 1 & 150 & 24 & 250 \\ 1 & 190 & 24 & 250 \end{bmatrix} \end{array}$$

Column vector of size 8

another Representation

Let $c_1' = \frac{\text{Powder} - 170}{20}$, $c_2' = \frac{\text{rate} - 18}{6}$, $c_3' = \frac{\text{die} - 235}{15}$

$$C' = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Coding Variables

- From the example, there were two ways to represent the same problem, coded and uncoded
- Why use coded variables[a]
  - Computational ease and increased accuracy in estimating the model coefficients
  - Enhanced interpretability of the coefficient estimates in the model.
- Internally most statistical software codes for estimating parameters

[a] Khuri, and Cornell (1987). Response Surfaces: Designs and Analyses. Dekker

correlation

Sunday, November 7, 2021 2:23 PM

Correlations

	radius (y)	x1	x2
x1	0.094		
x2	0.554	0.000	
x3	0.817	0.000	0.000

Screen clipping taken: 11/7/2021 2:24 PM

none, designed experiment

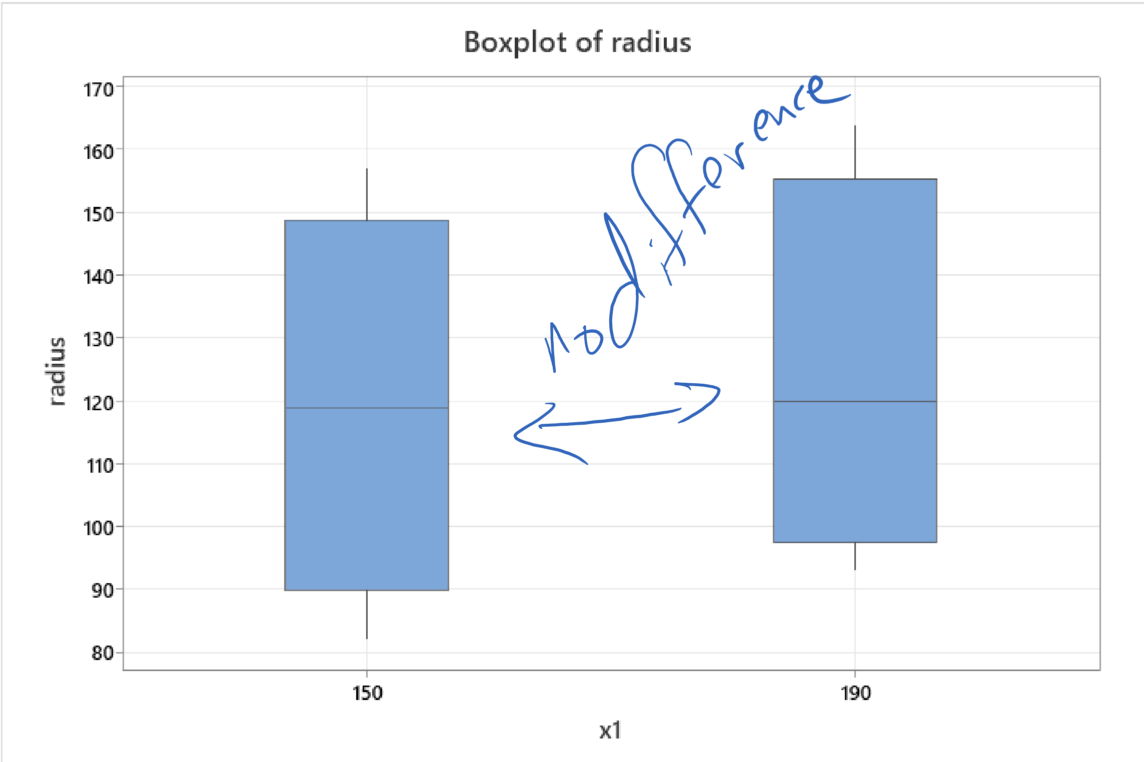
y & x_1 → weak, none

y & x_2 → moderative +

y & x_3 → strong +

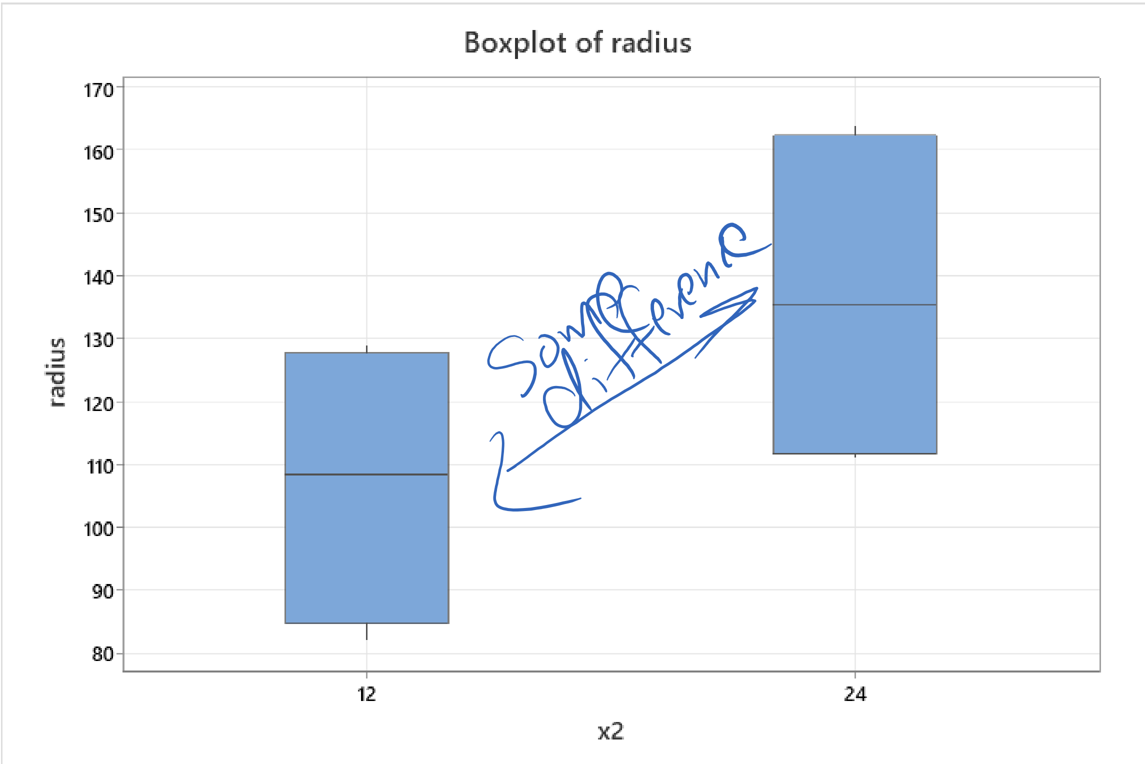
Box plot

Sunday, November 7, 2021 2:25 PM



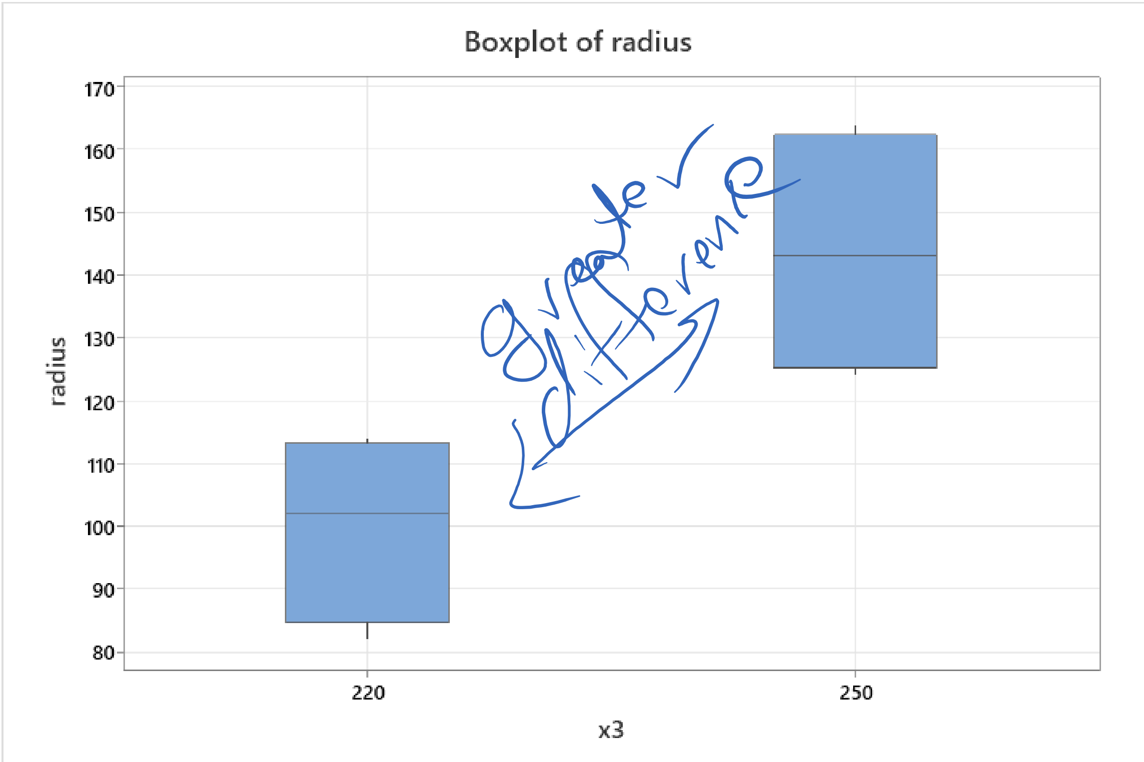
Screen clipping taken: 11/7/2021 2:26 PM

Sunday, November 7, 2021 2:27 PM



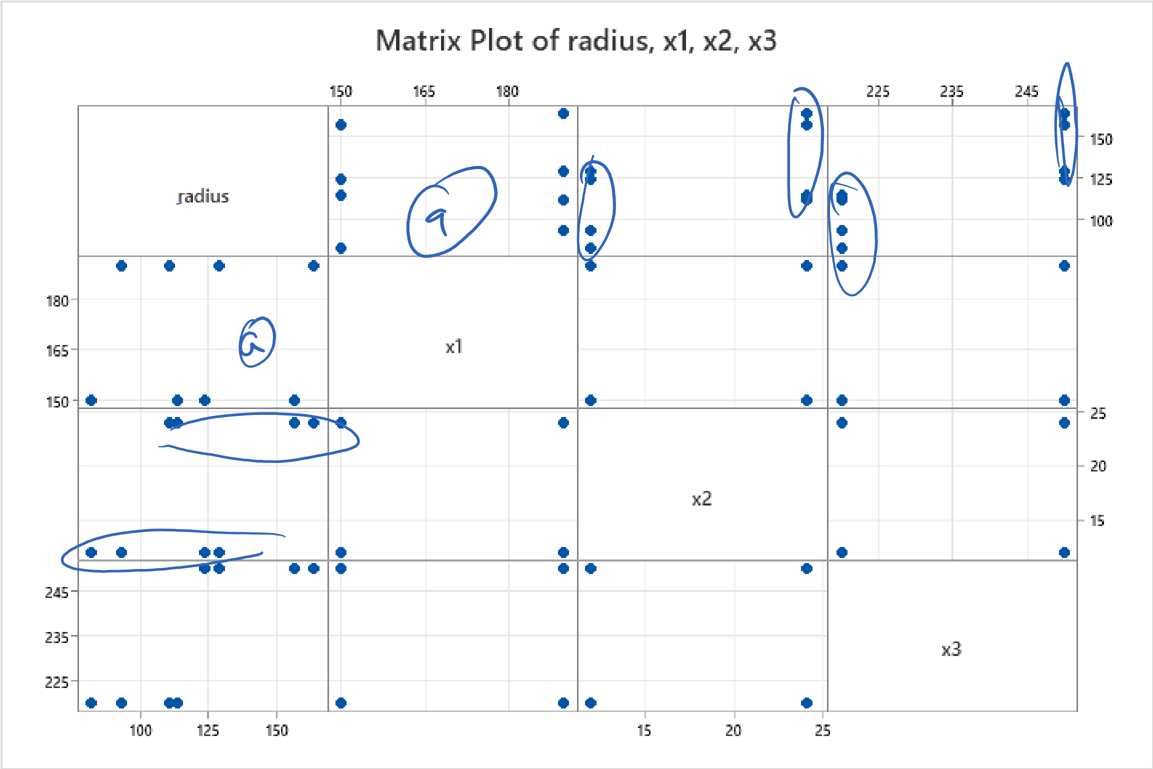
Screen clipping taken: 11/7/2021 2:27 PM

Sunday, November 7, 2021 2:27 PM



Screen clipping taken: 11/7/2021 2:28 PM

Sunday, November 7, 2021 2:28 PM



Screen clipping taken: 11/7/2021 2:29 PM