# Other Regression Models






1. **Multiple Linear Regression**: More than one predictor variables
2. **Categorical Predictors**: Predictor variables are categories such as CPU type, disk type, and so on.
3. **Curvilinear Regression**: Relationship is nonlinear





## Multiple Linear Regression Models

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k + e$$

- Given a sample of *n* observations with *k* predictors

$$\{(x_{11}, x_{21}, \ldots, x_{k1}, y_1), \ldots, (x_{1n}, x_{2n}, \ldots, x_{kn}, y_n)\}$$

$$\begin{aligned}
y_1 &= b_0 - b_1x_{11} - b_2x_{21} - \cdots - b_kx_{k1} + e_1 \\
y_2 &= b_0 - b_1x_{12} - b_2x_{22} - \cdots - b_kx_{k2} + e_2 \\
&. \\
&. \\
&. \\
y_n &= -b_0 - b_1x_{1n} - b_2x_{2n} - \cdots - b_kx_{kn} + e_n
\end{aligned}$$





## Vector Notation

In vector notation, we have:

- $$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ . \\ . \\ . \\ b_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{bmatrix}$$
- or $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$
- All elements in the first column of **X** are 1.

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$





## Example 15.1

- Seven programs were monitored to observe their resource demands. In particular, the number of disk I/O's, memory size (in kBytes), and CPU time (in milliseconds) were observed.

| CPU Time $y_i$ | Disk I/O's $x_{1i}$ | Memory Size $x_{2i}$ |
|---|---|---|
| 2 | 14 | 70 |
| 5 | 16 | 75 |
| 7 | 27 | 144 |
| 9 | 42 | 190 |
| 10 | 39 | 210 |
| 13 | 50 | 235 |
| 20 | 83 | 400 |





## Example 15.1 (Cont)

$$\text{CPU time} = b_0 + b_1(\text{number of disk I/O's}) + b_2(\text{memory size})$$

- In this case:

$$\mathbf{X} = \begin{bmatrix} 1 & 14 & 70 \\ 1 & 16 & 75 \\ 1 & 27 & 144 \\ 1 & 42 & 190 \\ 1 & 39 & 210 \\ 1 & 50 & 235 \\ 1 & 83 & 400 \end{bmatrix}$$

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 7 & 271 & 1324 \\ 271 & 13,855 & 67,188 \\ 1324 & 67,188 & 326,686 \end{bmatrix}$$

## Example 15.1 (Cont)

$$\mathbf{C} = (\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 0.6297 & 0.0223 & -0.0071 \\ 0.0223 & 0.0280 & -0.0058 \\ -0.0071 & -0.0058 & 0.0012 \end{bmatrix}$$

$$\mathbf{X}^T\mathbf{y} = \begin{bmatrix} 66 \\ 3375 \\ 16,388 \end{bmatrix}$$

- The regression parameters are:

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = (-0.1614, 0.1182, 0.0265)^T$$

- The regression equation is:

$$\text{CPU time} = -0.1614 + 0.1182(\text{number of disk I/O's}) + 0.0265(\text{memory size})$$





## Regression with Categorical Predictors

- Note: If all predictor variables are categorical, use one of the experimental design and analysis techniques for statistically more precise (less variant) results Use regression if most predictors are quantitative and only a few predictors are categorical.
- Two Categories: $x_j = \begin{cases} 0 & \Rightarrow \text{First value} \\ 1 & \Rightarrow \text{Second value} \end{cases}$
- $b_j$ = difference in the effect of the two alternatives
  $b_j$ = Insignificant ⇒ Two alternatives have similar performance
- Alternatively: $x_j = \begin{cases} -1 & \Rightarrow \text{First value} \\ +1 & \Rightarrow \text{Second value} \end{cases}$

  $b_j$ = Difference from the average response Difference of the effects of the two levels is $2b_j$





## Categorical Predictors (Cont)

- Three Categories: Incorrect:

$$x_1 = \begin{cases} 1 & \Rightarrow \text{Type A} \\ 2 & \Rightarrow \text{Type B} \\ 3 & \Rightarrow \text{Type C} \end{cases}$$

  This coding implies an order ⇒ B is half way between A and C. This may not be true.
- Recommended: Use two predictor variables

$$x_1 = \begin{cases} 1, & \text{If type A} \\ 0, & \text{Otherwise} \end{cases}$$

$$x_2 = \begin{cases} 1, & \text{If type B} \\ 0, & \text{Otherwise} \end{cases}$$





## Categorical Predictors (Cont)

Thus, $(x_1, x_2) = (1, 0) \Rightarrow \text{Type A}$

$(x_1, x_2) = (0, 1) \Rightarrow \text{Type B}$

$(x_1, x_2) = (0, 0) \Rightarrow \text{Type C}$

- This coding does not imply any ordering among the types. Provides an easy way to interpret the regression parameters.

$$y = b_0 + b_1x_1 + b_2x_2 + e$$





## Categorical Predictors (Cont)

- The average responses for the three types are:

$$\bar{y}_A = b_0 + b_1$$

$$\bar{y}_B = b_0 + b_2$$

$$\bar{y}_C = b_0$$

- Thus, $b_1$ represents the difference between type A and C.
  $b_2$ represents the difference between type B and C.
  $b_0$ represents type C.





## Categorical Predictors (Cont)

- Level = Number of values that a categorical variable can take
- To represent a categorical variable with k levels, define k-1 binary variables:

$$x_j = \begin{cases} 1, & \text{If } j\text{th value} \\ 0, & \text{otherwise} \end{cases}$$

- *k*th (last) value is defined by $x_1 = x_2 = \cdots = x_{k-1} = 0$.
- $b_0$ = Average response with the *k*th alternative.
- $b_j$ = Difference between alternatives *j* and *k*.
- If one of the alternatives represents the status quo or a standard against which other alternatives have to be measured, that alternative should be coded as the *k*th alternative.

## Case Study 15.1: RPC performance

- RPC performance on Unix and Argus

$y = b_0 + b_1 x_1 + b_2 x_2$

where, y is the elapsed time, $x_1$ is the data size and

$$x_2 = \begin{cases} 1 & \Rightarrow \text{UNIX} \\ 0 & \Rightarrow \text{ARGUS} \end{cases}$$

| UNIX | | ARGUS | |
|---|---|---|---|
| Data Bytes | Time | Data Bytes | Time |
| 64 | 26.4 | 92 | 32.8 |
| 64 | 26.4 | 92 | 34.2 |
| 64 | 26.4 | 92 | 32.4 |
| 64 | 26.2 | 92 | 34.4 |
| 234 | 33.8 | 348 | 41.4 |
| 590 | 41.6 | 604 | 51.2 |
| 846 | 50.0 | 860 | 76.0 |
| 1060 | 48.4 | 1074 | 80.8 |
| 1082 | 49.0 | 1074 | 79.8 |
| 1088 | 42.0 | 1088 | 58.6 |
| 1088 | 41.8 | 1088 | 57.6 |
| 1088 | 41.8 | 1088 | 59.8 |
| 1088 | 42.0 | 1088 | 57.4 |





## Case Study 15.1 (Cont)

| Parameter | Mean | Std. Dev. | Confidence Interval |
|---|---|---|---|
| $b_0$ | 36.739 | 3.251 | ( 31.1676, 42.3104) |
| $b_1$ | 0.025 | 0.004 | ( 0.0192, 0.0313) |
| $b_2$ | -14.927 | 3.165 | ( -20.3509, -9.5024) |

- All three parameters are significant. The regression explains 76.5% of the variation.
- Per byte processing cost (time) for both operating systems is 0.025 millisecond.
- Set up cost is 36.73 milliseconds on ARGUS which is 14.927 milliseconds more than that with UNIX.





## Curvilinear Regression

- If the relationship between response and predictors is nonlinear but it can be converted into a linear form
  ⇒ curvilinear regression.

  Example:

$$y = bx^a$$

Taking a logarithm of both sides we get:

$$\ln y = \ln b + a \ln x$$

Thus, ln *x* and ln *y* are linearly related. The values of ln b and a can be found by a linear regression of ln *y* on ln *x*.





## Curvilinear Regression: Other Examples

| Nonlinear | Linear |
|---|---|
| y = a + b/x | y = a + b (1/x) |
| y=x/(a+bx) | (1/y)=a+bx |
| y = x/(a+bx) | (x/y) = a + bx |
| y = abx | $\ln(y) = \ln(a) + (\ln(b))$x |
| y=a+$bx_n$ | y=a+b($x^n$) |

- If a predictor variable appears in more than one transformed predictor variables, the transformed variables are likely to be correlated ⇒ multicollinearity.
- Try various possible subsets of the predictor variables to find a subset that gives significant parameters and explains a high percentage of the observed variation.





## Example 15.4

- Amdahl's law: I/O rate is proportional to the processor speed. For each instruction executed there is one bit of I/O on the average.

| System No. | MIPS Used | I/O Rate |
|---|---|---|
| 1 | 19.63 | 288.60 |
| 2 | 5.45 | 117.30 |
| 3 | 2.63 | 64.60 |
| 4 | 8.24 | 356.40 |
| 5 | 14.00 | 373.20 |
| 6 | 9.87 | 281.10 |
| 7 | 11.27 | 149.60 |
| 8 | 10.13 | 120.60 |
| 9 | 1.01 | 31.10 |
| 10 | 1.26 | 23.70 |





## Example 15.4 (Cont)

- Let us fit the following curvilinear model to this data:

$$\text{I/O Rate} = \alpha(\text{MIPS Rate})^{b_1}$$

- Taking a log of both sides we get:

$$\log(\text{I/O Rate}) = \log(\alpha) + b_1 \log(\text{MIPS Rate})$$

$$b_0 = \log(\alpha)$$

## Example 15.4 (Cont)

| Obs. No. | $x_1$ | y |
|---|---|---|
| 1 | 1.293 | 2.460 |
| 2 | 0.736 | 2.069 |
| 3 | 0.420 | 1.810 |
| 4 | 0.916 | 2.552 |
| 5 | 1.146 | 2.572 |
| 6 | 0.994 | 2.449 |
| 7 | 1.052 | 2.175 |
| 8 | 1.006 | 2.081 |
| 9 | 0.004 | 1.493 |
| 10 | 0.100 | 1.375 |

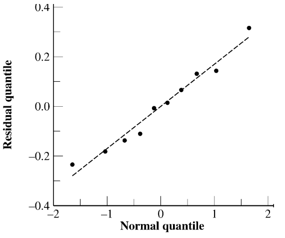
| Parameter | Mean | Std. Dev. | Confidence Interval |
|---|---|---|---|
| $b_0$ | 1.423 | 0.119 | ( 1.20, 1.64) |
| $b_1$ | 0.888 | 0.135 | ( 0.64, 1.14) |

- Both coefficients are significant at 90% confidence level.
- The regression explains 84% of the variation.
- At this confidence level, we can accept the hypothesis that the relationship is linear since the confidence interval for $b_1$ includes 1.





## Example 15.4 (Cont)




- Errors in log I/O rate do seem to be normally distributed.





## Summary

- Too many predictors may make the model weak.
- Categorical predictors are modeled using binary predictors
- Curvilinear regression can be used if a transformation gives linear relationship.
- Transformation: s = g(y) $\Rightarrow \; w = h(y) = \int \frac{1}{g(y)} dy$
- Outliers: Use your system knowledge. Check measurements.
- Common mistakes:No visual verification, control vs correlation