Sec 4.1
Determine $x \rightarrow y$ Relationship

\[ \downarrow \]

Scatterplot

\[ \downarrow \]

Numerical Summaries
\( (r, r^2, LSRL) \)

\[ \downarrow \]

Residual Plot

\[ \downarrow \]

No Pattern/Random

\[ \uparrow \]

Linear $\hat{y} = a + bx$

\[ \downarrow \]

Pattern

\[ \downarrow \]

Exponential $\hat{y} = ab^x$

\[ \downarrow \]

Power $\hat{y} = ax^b$
Ex US Population (Ex 4.11, p. 213)

\[
\frac{L_1}{\text{year (1920)}} \quad \frac{L_2}{\text{pop}}
\]

\[
\frac{L_3}{\log y}
\]

\(r = 0.992\)

\(r = 0.997\)
Write exponential model:

a) Using Back Transformations

\[ \hat{y} = -8.28 + 0.005x \]

\[ \log_{10} y = -8.28 + 0.005x \]

\[ \hat{y} = 10^{-8.28 + 0.005x} \]

\[ \hat{y} = (10^{-8.28})(10^{0.005x}) \]

\[ \hat{P}_{\text{pop}} = (10^{-8.28})(10^{0.005}) \text{ Year} \]

\[ a \quad b \]
Write exponential model:

a) Using Back Transformations

\[ \hat{y} = -8.28 + 0.005x \]

\[ \log_{10} y = -8.28 + 0.005x \]

\[ \hat{y} = 10^{-8.28 + 0.005x} \]

\[ \hat{y} = (10^{-8.28})(10^{0.005x}) \]

\[ \hat{P}_{\text{pop}} = (10^{-8.28})(10^{0.005})^{\text{Year}} \]
b) Using Calculator

\[ \text{CALC} \rightarrow \text{ExpReg } L_1, L_2, Y. \]

\[ \hat{Y} = (0.0000000529)(1.01)^x \]

\[ \hat{Pop} = (0.00000000529)(1.01)^\text{year} \]

\{ May Cause "Overflow Error" \}
Exponential Regression Test

\[ L_2 = 105.7, 122.8, 131.7, 151.3, 179.3, \ldots, 248.7, 281.4 \]
Power Regression

\[ \frac{L_1}{\text{length}} \quad \frac{L_2}{\text{weight}} \]

\[ r = 0.94 \]

\[ \frac{L_3}{\log y} \quad \frac{L_4}{\log x} \]

\[ r = 0.96 \quad r = 0.99 \]

Linear

Exponential

Power

(Rockfish, p. 216)
Write Power Equation

a) Use backward transformations

\[
\hat{y} = -1.899 + 3.049 \times x \\
\log\hat{y} = -1.899 + 3.049 (\log x)
\]

\[
\hat{y} = 10^{-1.899} \times 3.049 \\
\text{Weight} = (10^{-1.899})(\text{length})^{3.049}
\]

b) Use Calculator

\[
\text{[CALL]} \rightarrow \text{PwrReg } [L_1, L_2, Y_1] \quad \text{Overflow Error?}
\]
Sec 4.2
Caution About Correlation/Regression

1) Avoid Extrapolation
2) Lurking Variables

Variable(s) that is/are not an explanatory/response variable BUT may influence either/both
a) Common Response

Lurking variable which impacts both X and Y

\[ X \rightarrow Y \]

# Ice Cream Sales  # Drownings

Z

Temp/Season

# People at Beach
b) Confounding

Lurking variable which affects the response variable (y)

\[ X \rightarrow Y \]

\[ \text{Amt of time studying} \]

\[ Z \rightarrow Y \]

\[ \text{IQ, Weighted Classes etc} \]
3) Strong Correlation ≠ Causation

- Experiments
- Correlational Studies
Sec 4.1.3
## Relations In Categorical Data (%)

<table>
<thead>
<tr>
<th></th>
<th>Student Smokes</th>
<th>Student Not Smokes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both Parents Smoke</td>
<td>400</td>
<td>1380</td>
</tr>
<tr>
<td>One Parent Smokes</td>
<td>416</td>
<td>1823</td>
</tr>
<tr>
<td>Neither Smokes</td>
<td>188</td>
<td>1168</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1004</strong></td>
<td><strong>4371</strong></td>
</tr>
</tbody>
</table>

### Marginal Distributions

- % of students smoke = \( \frac{1004}{5375} = 18.68\% \)

### Conditional Distributions

1. % of students smoke when both parents smoke = \( \frac{400}{1780} = 22.47\% \)
2. % of students smoke when neither parent smoke = \( \frac{188}{1356} = 13.86\% \)
Conclusion

- When both parents, student smoking increases from 14% to 22%.

- Percent Increase = \( \frac{8}{14} = 57\% \)

Ex Cocaine Use

1 \( \rightarrow \) 3 \( \rightarrow \) 200% increase
Simpson's Paradox

- Lurking variables change/reverse relationship

<table>
<thead>
<tr>
<th>Surgery</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Died</td>
<td>63</td>
<td>16</td>
</tr>
<tr>
<td>Survived</td>
<td>2037</td>
<td>784</td>
</tr>
</tbody>
</table>

\[
\text{Died | Hospital A} = \frac{63}{2100} = 3\% \\
\text{Died | Hospital B} = \frac{16}{800} = 2\% \leftarrow
\]
**Good Condition**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Died</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Survived</td>
<td>594</td>
<td>592</td>
</tr>
</tbody>
</table>

\[
\text{Died | Hospital A} = \frac{6}{600} = 1\% \\
\text{Died | Hospital B} = \frac{8}{600} = 1.3\% \\
\]

**Poor Condition**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Died</td>
<td>57</td>
<td>8</td>
</tr>
<tr>
<td>Survived</td>
<td>1443</td>
<td>192</td>
</tr>
</tbody>
</table>

\[
\Rightarrow 30\% \text{ increase}
\]
Independence
Marginal Distribution = Conditional Distribution

Ex

<table>
<thead>
<tr>
<th></th>
<th>Y</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>F</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

\[
P(\text{Male}) = \frac{10}{20} = 50\% \\
P(\text{Male} \mid \text{Yes}) = \frac{5}{10} = 50\%
\]

Given That