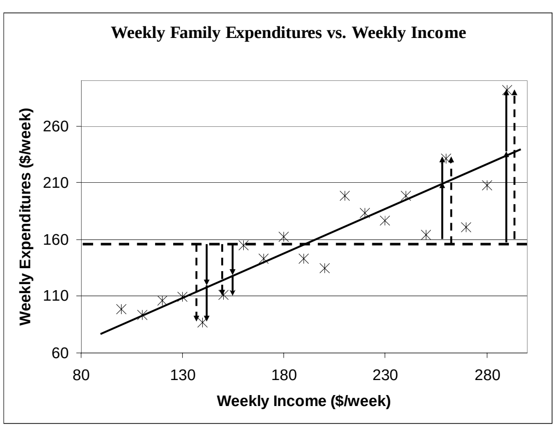
III. Simple Regression.
- A. Introduction
- B. Population Regression Equation
- C. Sample Regression Equation
- D. Ordinary Least Squares
- E. Classical Regression Model.
- F. Properties of OLS Estimators (**BLUE**)
- G. Estimator for σ^2
- H. Inference
- I. Goodness of Fit
- J. Forecasting or Prediction

I. Goodness of Fit

1. Objective – a summary measure that tells us how well our model "fits the data."
2. Measures for the Dependent Variable:

- Y_i : observed value – an actual value for the dependent variable observed in our sample.
- $\bar{Y}$: the sample mean – used when Y does not depend upon X.
- $\hat{Y}_i$: The SRF – used when Y depends on X, use SRF to estimate the population mean of Y at each level of X.

(Illustrate each of these on the graph.)



3. Sums of Squares
   - We want to know how we did on average.
   - But, summing deviations always gives zero.
   - Square them! Then sum:

$\sum (Y_i - \bar{Y})^2$ – ***Total Sum of Squares (TSS)***.

$\sum (\hat{Y}_i - \bar{Y})^2$ – ***Explained Sum of Squares (ESS)***

$\sum (Y_i - \hat{Y}_i)^2$ – ***Residual Sum of Squares (RSS)***

The Sums of Squared Deviations always add-up:

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2$$

4. R^2 a measure of Goodness of Fit.
   - What we have: ***TSS = ESS + RSS***
   - Form a ratio:
   - Interpretation:
   - Range of possible R^2 values:

Dan's Sample Regression Function (SRF) for Used 2004 Accords.

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	64053691.1	64053691	10.68	0.0029
Residual	28	167986040.4	5999501		
Total	29	232039731.5			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	23191.85	1022.62	22.68	0.00	21097.10	25286.60
Miles	-0.1050	0.0321	-3.267	0.0029	-0.1708	-0.0392

***IV. Multiple Regression*.**

A. Introduction
B. CRM
C. Estimation
D. Interpretation of Parameter Estimates
E. Properties of Estimators

IV. Multiple Regression

A. Introduction.

1. **Multiple regression** – measure the effects of several independent variables on the dependent variable.
2. Write a multiple regression population regression equation.

   Example: Demand –

   Example: Salary inequities –

B. Classical Regression Model – again the assumptions, and we add one.

1. $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots \beta_k X_{ki} + u_i$.
2. *The Xs are not random variables.*
3. $E[\, u \mid X_i \,] = 0$.
4. $Var(u \mid X_i\,] = \sigma^2$.
5. $Cov(u_i, u_j) = E[u_i u_j] = 0$.
6. *u ~ Normally.*
7. *No perfect linear associations among Xs.*

C. Estimation of population parameters.

1. How? What method??
2. Process: Describe in words.

D. Interpretation of Parameter Estimates.

1. Partial Effects – we now have several variables affecting Y. Each X has a ***partial effect on Y***.

   Example: $Y = f(X_1, X_2)$

2. Experimental vs Social Sciences
   - ***Experimental Sciences*** – in physical and biological sciences, factors are physically held constant.

   What do we do in econometrics?
   - Gather data from individuals (markets, firms, etc.) **after choices have been made**.
   - ***Holding other factors constant*** – we have to include them in the model.

3. Specification Bias

- This is what happens when ***we don't hold other factors constant.***
- Suppose ***Y*** is ***truly*** affected by ***X_1 and X_2***.

Specification Bias – what happens when we omit an important variable from our model.

Truth: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$

You specify: $Y_i = \alpha_0 + \alpha_1 X_{1i} + u_i$

Estimator: $\hat{\alpha}_1 = \dfrac{\sum x_{1i} Y_i}{\sum x_{1i}^2}$

Evaluate using the true model to see the effects of the ***specification mistake***: ***Bias***

$$E[\hat{\alpha}_1] = \beta_1 + \boxed{\beta_2 \frac{\sum x_{1i} X_{2i}}{\sum x_{1i}^2}}$$

Regression Analysis: sales versus prose, pcarn, dinc

```
The regression equation is
sales = 13355 – 3628 prose + 2634 pcarn - 19.3 dinc

Predictor      Coef  SE Coef      T      P
Constant      13355     6485   2.06  0.062
prose       -3628.2    635.6  -5.71  0.000
pcarn          2634     1013   2.60  0.023
dinc         -19.25    30.69  -0.63  0.542

S = 1076.29   R-Sq = 77.8%   R-Sq(adj) = 72.2%
```

What if we don't hold other factors constant?

```
Regression Analysis: sales versus prose

The regression equation is
sales = 16899 - 2979 prose


Predictor      Coef  SE Coef      T      P
Constant      16899     1985   8.52  0.000
prose       -2978.5    630.0  -4.73  0.000


S = 1312.19   R-Sq = 61.5%   R-Sq(adj) = 58.7%
```

E. Properties of OLS Estimators

1. ***Linear:*** The OLS estimators are still linear estimators.
2. ***Unbiased:*** If CRMAs # 1 – 3 are correct, the OLS estimators are unbiased estimators.

$$E[\hat{\beta}_1] = \beta_1$$

3. ***Minimum Variance:*** If CRMAs # 1 – 5 are correct, the OLS estimators are the ***Best Linear Unbiased Estimators.***

(Gauss-Markov Theorem)

Draw a graph illustrating properties 2 & 3.