Basic Definitions

~N(2,4) – normal, mean 2, variance 4, sigma^2

Optimal GM = 1 / | optimal price elasticity |

F-stat = R-squared / ((1- R-squared)/DF)

Three Key Characteristics of a Good Linear Model:

- Mean of Y is linear in X
- Error terms (deviations from line) are normally distributed (few deviations > 3 sd away from line)
- Error terms have constant variance

Needs to be based on many observations, not just a few
More than 1% of observations for example

e_i - least squares residual
$\hat{Y}$ - fitted value
s - sample standard dev
s^2 - sample variance
r - sample correlation coefficient
r^2 - coeff of determination, goodness of fit measure

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

SST - total sum of squares
SSR + SSE

SSR - regression sum of squares
(Predicted – Average Y)^2

SSE - error sum of squares
(Predicted – Actual Y)^2

Hypothesis Testing

t-stat = (estimate – hypo value) / standard error
standard error = (estimate – hypo value) / t-stat

$$t = \frac{b_1 - \beta_1^*}{s_{b_1}} = \frac{\text{estimate-hypo value}}{\text{std err}}$$

In Regression Output, hypo value = 0

Rejection Test

$$|t| = \left|\frac{b_1 - \beta_1^*}{s_{b_1}}\right| > t^*_{N-2,\alpha/2}$$

Confidence Interval

$$b_1 \pm t^*_{N-2,\alpha/2} s_{b_1}$$

Need to lookup t*

significance level (α) = Prob(reject when null true) by deciding what level of error of this kind is acceptable (called type I error).

The p-value is the probability of observing a value of the t statistic farther out in the tail than the observed t value.

$$p = \Pr\left[\,|t_{N-2}| \geq |t|\,\right]$$

Small p value (< α) ⟺ large | t | ⟹ reject

Large p value (≥ α) ⟺ small | t | ⟹ accept null

Slope Formulas

LS Slope

$$b_1 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \frac{s_{XY}}{s_X^2}$$

$$b_0 = \bar{Y} - b_1\bar{X}$$

$$\hat{Y}_i = b_0 + b_1 X_i$$

$$\hat{Y} - \bar{Y} = b_1(X - \bar{X})$$

$$b_1 = \frac{s_{xy}}{s_x^2} = r_{xy} \times \frac{s_y}{s_x}$$

s_{XY} = sample covariance (X, Y)

s_X^2 = sample variance of (X)

Conditional Variance of Y| X in a regression
(Conditional SD is just square root of this)

$$s^2 = \frac{1}{N-2}\sum_{i=1}^{N} e_i^2 = \frac{SSE}{N-2}$$

Residual standard error = s
s = Conditional SD of y

$$s = \sqrt{\frac{SSE}{N-2}} = \text{standard error of the regression}$$

Variance of (Y | X) = MSE = SSE/(n – 2)
Or – Residual Standard Error ^2
Residual standard error = s

Standard Error of Slope Coefficient

$$s_{b_1} = \sqrt{\frac{s^2}{(N-1)s_X^2}}$$

$$\mathrm{Var}(b_1) = \frac{\sigma^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \frac{\sigma^2}{(N-1)s_X^2}$$

three factors that affect it: N , σ^2, s_X

Output Summary

```
                                  Sb0
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  18.2259    23.8852   0.763    0.451
reboundb      0.9495     0.0438  21.675   <2e-16 ***
---
Signif.  [ s ]  0 '***' 0.001 '*  [ Sb1 ]  '*' 0.05 '.' 0.

Residual standard error: 7.67 on 33 degrees of freedom
Multiple R-squared: 0.9344, Adjusted R-squared: 0.9324
F-statistic: 469.8 on 1 and 33 DF,  p-value: < 2.2e-16

> anova(lm(rebounda~reboundb))
Analysis of Variance Table

Response: rebounda
          Df  Sum Sq Mean Sq F value    Pr(>F)
reboundb   1 27635.8 27635.8   469.8 < 2.2e-16 ***
Residuals 33  1941.2    58.8
       SSE               SSR
```

Other

Logarithmic Transformation
logarithm function has the effect of compressing large values and expanding the axis for small values.

Predictive Standard Error

$$s_{pred} = s\left(1 + \frac{1}{N} + \frac{(X_f - \bar{X})^2}{(N-1)s_X^2}\right)^{.5}$$

Definition of Prediction Error

$$e_f = Y_f - \hat{Y}_f = Y_f - b_0 - b_1 X_f$$

Sum of Errors Formulas

$$\sum (X_i - \bar{X})\left(Y_i - \left[\bar{Y} - b_1\bar{X}\right] - b_1 X_i\right) = 0$$

$$\sum e_i = \sum (Y_i - b_0 - b_1 X_i) = \sum \left(Y_i - (\bar{Y} - b_1\bar{X}) - b_1 X_i\right)$$

$$= \sum (Y_i - \bar{Y}) - b_1(X_i - \bar{X}) = \sum (Y_i - \bar{Y}) - b_1 \sum (X_i - \bar{X}) = 0 - b_1 \times 0$$

Slope Transformation

$$b_{1,yonw} = \frac{\sum (Y_i - \bar{Y})W_i}{\sum (W_i - \bar{W})^2} = \frac{\sum (Y_i - \bar{Y}).2X_i}{\sum (.2X_i - (.2\bar{X}))^2} = \frac{.2\sum (Y_i - \bar{Y})X_i}{(.2)^2 \sum (X_i - (\bar{X}))^2} = 5b_{1,yonx}$$

Calculator Functions
STAT – edit
STAT – 1-Var Stats
2nd – List – Math – stdDev, use { }

$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab*\mathrm{cov}(X,Y)$

Standard Deviation Formula

$$s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

$n =$ The number of data points
$\bar{x} =$ The mean of the x_i
$x_i =$ Each of the values of the data

Fitted Values & Residuals Formula

$$Y_i = \hat{Y}_i + (Y_i - \hat{Y}_i) = \hat{Y}_i + e_i$$

Minimize the sum of squared residuals

$$\sum_{i=1}^{N} e_i^2$$

Correlation Coefficient

$$r_{xy} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} = \frac{s_{XY}}{s_X s_Y}$$