Fox Module 11: Statistical inference for simple linear regression

(The attached PDF file has better formatting.)

Regression Analysis Units of Measurement Practice problems

Know how ordinary least squares estimators, their standard errors, t-values, and p-values depend on the units of measurement and displacement from the origin. The principles are

- Multiplying the explanatory variable by *k* multiplies its β by 1/*k*.
- Multiplying the response variable by *k* multiplies all the β's by *k*.
- Displacements of explanatory variables and the response variable from the origin changes α, not the ß's.

Intuition: β is in units of response variable / explanatory variable.

Illustration: Suppose claim frequency = α + β × kilometers driven.

- α is in units of claim frequency.
- β is in units of claim frequency / kilometers driven

If we write the regression equation as claim frequency = α + (β/1,000) × meters driven.

- α is in units of claim frequency.
- β is in units of claim frequency / kilometers driven

Intuition: The ß's depend on the deviations of the values from their means. A constant displacement of all the values doesn't affect the deviations. But a constant displacement of *k* raises the response variable Y by k × β. α has the same displacement as the response variable, so it also rises by k × β.

Elasticities, standardized coefficients, and *t*-values are unit-less.

- Elasticities are percentage changes: $\partial Y/Y$ / $\partial X/X$.
- The change in a value has the same units as the value itself.
  - If X is kilometers driven, then ∂X is also measured in kilometers driven.
  - If Y is claim frequency, then ∂Y is also measured in claim frequency.

Standardized coefficients are $\beta \times \sigma_x / \sigma_y$.

- β is in units of Y / X.
- σ_x is in units of X.
- σ_y. Is in units of Y.

⇛ The standardized coefficient is unit-less.

Measures of significance are not affected by units of measurement.

- The *t*-value is the ordinary least squares estimator divided by its standard deviation.
- The estimator and its standard deviation have the same units, so the *t*-value is unit-less.

The correlation between two random variables is unrelated to units of measurement, so the R^2 statistic is also unit-less.

*Question 11.1: Goodness-of-fit and Units of Measurement

We use least squares regression with N pairs of observations (X_i, Y_i) to estimate average *annual* claims cost in dollars per average *miles driven* each week, giving Y= 50 + 40X + ϵ.

If we change the parameters to annual claims costs in *Euros* and *kilometers* driven per week, which of the following is true?

A. The R^2 increases and the *t* value for kilometers driven increases
B. The R^2 increases and the *t* value for kilometers driven decreases
C. The R^2 decreases and the *t* value for kilometers driven increases
D. The R^2 decreases and the *t* value for kilometers driven decreases
E. The R^2 stays the same and the *t* value for kilometers driven stays the same

Answer 11.1: E

The R^2 and the *t* statistic are both unit-less.

- The R^2 is a proportion. If we double the units of Y, the TSS, RegSS, and RSS all increase by a factor of $2^2 = 4$. The R^2 doesn't change.
- The *t* statistic is the ordinary least squares estimator divided by its standard deviation. If we double the units of X, both the estimator and its standard deviation decrease by 50%.

*Question 11.2: Miles Driven and Annual Claim Costs

We use least squares regression with N pairs of observations (X_i, Y_i) to estimate average *annual* claims cost in dollars per average *miles driven* per day, giving $Y = 50 + 40X + \epsilon$. For instance, a policyholder who drives an average of 25 miles a day has average claim costs of $50 + 40 \times 25 = 1{,}050$ dollars a year.

If we change the parameters to annual claims costs in Euros and kilometers driven a day, what is the revised regression equation? For this problem, assume €1.00 = $1.25 and 1 kilometer = ⅝ mile (five eighths of a mile).

A. $Y = 40 + 40X + \epsilon$
B. $Y = 40 + 20X + \epsilon$
C. $Y = 40 + 64X + \epsilon$
D. $Y = 62.5 + 25X + \epsilon$
E. $Y = 62.5 + 64X + \epsilon$

Answer 11.2: B

The estimate of β is the covariance $\rho(x,y)$ divided by the variance of X.

- Using euros multiplies each Y value by 1.00 / 1.25 = 0.80.
- Using kilometer multiplies each X value by 8/5 = 1.60.

Illustration: \$10.00 = 10 × 0.80 = €8.00, and 10 miles = 10 × 1.60 = 16 kilometers.

Multiplying the Y values by 0.80 and the X values by 1.60

- Multiplies the covariance by 0.80 × 1.60 = 1.280
- Multiplies the variance of X by 1.60^2 = 2.560

This multiplies β by 1.280 / 2.560 = 0.500.

α is not affected by the units of X, since the product β × X is not affected by the units of X. But α varies directly with the units of Y: if Y is multiplied by 0.80, α is multiplied by 0.80.

Jacob: Is the product β × X unit-less?

Rachel: No; the product is in the units of Y.

We can check our result numerically:

- Before the change, if X = 0 miles, Y = \$50. Now X = 0 gives Y = €40, so α is 40.
- Before the change, if X = 5 miles, Y = \$250. Now X = 8 kilometers gives Y = \$250 × 0.8 = €200. Since α = 40, β is (200 – 40) / 8 = 20.

*Question 11.3: Displacement

We regress Y on X with a two-variable regression model $Y_i = \alpha + \beta \times X_i + \varepsilon_i$.

- X is the number of hours studied as a deviation from its mean.
- Y is the exam score as a deviation from its mean.

We change the values of X and Y to

- X is the actual number of hours studied (mean = 80 hours)
- Y is the actual exam score (mean score = 80)

Which of the following is true?

A. The R^2 increases and the adjusted R^2 increases
B. The R^2 increases and the adjusted R^2 stays the same
C. The R^2 decreases and the adjusted R^2 increases
D. The R^2 decreases and the adjusted R^2 stays the same
E. The R^2 stays the same and the adjusted R^2 stays the same

Answer 11.3: E

The displacement of X and Y does not affect the correlation between the random variables, so it does not affect the R^2 or the adjusted R^2.

*Question 11.4: Displacement

We regress Y on X with a two-variable regression model $Y_i = \alpha + \beta \times X_i + \varepsilon_i$. W
the following is true?

A. If we double each X value and decrease each Y value by 1, α increases.
B. If we double each X value but don't change the Y values, α decreases.
C. If we double each X value and increase each Y value by 1, α decreases.
D. If we double each X value and increase each Y value by 1, α increases.
E. If we double each X value and decrease each Y value by 1, α stays the same.

Answer 11.4: D

- Doubling each X value reduces β by 50% but does not change α.
- Increasing each Y value by 1 increases α by 1 but does not change β.

*Question 11.5: Standardized Coefficients and Elasticities

We regress the average auto insurance loss costs in *dollars* (the Y dependent variable) on the number of hours the auto is driven each week (the X independent variable). We estimate the ordinary least squares estimator $\hat{\beta}$, the standardized coefficie $\hat{\beta}$*, and the elasticity η.

If we use Euros for the loss costs instead of dollars, which of the following is true? Assume that one Euro is 1.25 dollars.

A. $\hat{\beta}$ increase $\hat{\beta}$* and η stay the same.

B. $\hat{\beta}$ decrease $\hat{\beta}$* and η stay the same.

C. $\hat{\beta}$ ar $\hat{\beta}$* stay the same; and η increases.

D. $\hat{\beta}$ ar $\hat{\beta}$* stay the same; and η decreases.

E. $\hat{\beta}$ and η stay the same, ar $\hat{\beta}$* increases.

Answer 11.5: B

If an hour of driving each week increases loss costs by \$10, it increases loss costs by €8, so β decreases.

The standardized coefficient and elasticity are unit-less, so they are not affected by a change in the units of measurement.