QMS 210 Simple Linear Regression

TRAIN TO LEARN EFFECTIVELY: TIP SHEETS

WHAT IS SIMPLE LINEAR REGRESSION?

A simple linear regression explains the relationship between an independent and dependant value, which consists of a straight line. A scatter plot is created to asses the direction of the line. The slope can be upwards (positive) or downwards (negative). We begin by finding the equation.

Prediction Line Equation: $\hat{Y}i = \beta 0 + \beta 1$

Where:

$\beta 0$ = Y Intercept for the population: represents the mean value of Y when X is zero

$\beta 1$ = Slope for the population: represents the expected change in Y per unit change in X

$X1$ = independent variable for observation i

Yi = dependant variable for observation i

COEFFICIENT OF CORRELATION AND DETERMINATION

Coefficient of correlation = r

- It is used to measure the strength of association between two variables

Coefficient of determination = r^2

- The % of variation in the dependent (y) variable is explained by its relationship to the independent (x) variable in the regression model

POSITIVE AND NEGATIVE RELATIONSHIPS

$\beta 1 > 0$ positive relationship between x and y so (x ↑, y↑)

$\beta 1 < 0$ negative relationship between x and y so (x ↑, y↓)
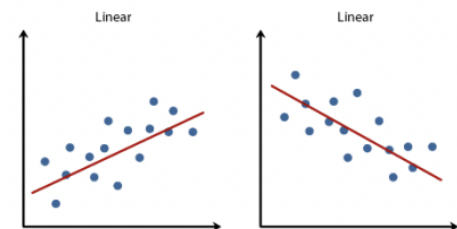
LINEAR REGRESSION

Possible equations:

$H_o: \beta_1 = 0$ $\qquad\qquad$ $H_o: \beta_1 \leq 0$ $\qquad\qquad$ $H_o: \beta_1 \geq 0$

$H_a: \beta_1 \neq 0$ $\qquad\qquad$ $H_a: \beta_1 > 0$ $\qquad\qquad$ $H_a: \beta_1 < 0$

CALCULATOR INSTRUCTIONS

Test: TEST→ T → REG → ≠ → XList: List 1 (independent variable) and YList: List 2 (dependent variable) → EXE

Critical Values: DIST→T→InvT→ "Variable"→ Area/df=n-2

Toronto Metropolitan University

TED ROGERS SCHOOL OF MANAGEMENT | Program Advising and Student Success

AREAS

H_a: $\beta_1 \neq 0$ → Area=$\frac{\alpha}{2}$ (two tail)

H_a: $\beta_1 > 0$ → Area= α (one tail)

H_a: $\beta_1 < 0$ → Area=1- α (one tail)

ANALYSIS
1) Compare t-stat with t-crit (critical value approach)
2) Compare p-value with α (p-value approach)

Then choose either:
Reject Ho and conclude that there is a linear relationship between x and y
Do not reject Ho and conclude that there is no linear relationship between x and y

HOW TO FIND THE T-STAT WITHOUT A CALCULATOR

t-stat=$\dfrac{b_1 - 0}{S_{b1}}$

b_1 = Slope coefficient

S_{b1} = Standard deviation of the slope

CONFIDENCE INTERVAL OF SLOPE
LCL= $b_1 - (t - critical)(S_{b1})$
UCL= $b_1 + (t - critical)(S_{b1})$

HOW TO FIND THE STANDARD ERROR OF THE ESTIMATE S_{YX}

$S_{YX} = \sqrt{\dfrac{SSE}{n-2}}$

SST (Sum of Squares Total) = SSR (Sum of Squares Regression) + SSE (Sum of Squares Error)
* These values can be found from the ANOVA table

Toronto Metropolitan University

TED ROGERS SCHOOL OF MANAGEMENT | Program Advising and Student Success

QMS 210 Simple Linear Regression

EXAMPLE

The sales for Sunflowers Apparel, a chain of apparel stores for women, have increased during the past 12 years as the chain has increased the number of its stores. Until now, Sunflowers senior manager selected sites based on subjective factors such as the availability of a good lease or the perception that a location seemed ideal for an apparel store. As the new director of planning, you need to develop a systematic approach to selecting new sites that will allow Sunflowers to make better-informed decisions for opening additional stores. This plan must be able to forecast annual sales for all potential stores under consideration. You believe that the size of the store significantly contributes to the success of a store and you want to use this relationship in the decision-making process. You wish to predict annual sales based on the size of the store in square feet.

Store #	Square Feet (000)	Annual Sales (in millions of dollars)	Store #	Square Feet (000)	Annual Sales (in millions of dollars)
1	1.7	3.7	8	1.1	2.7
2	1.6	3.9	9	3.2	5.5
3	2.8	6.7	10	1.5	2.9
4	5.6	9.5	11	5.2	10.7
5	1.3	3.4	12	4.6	7.6
6	2.2	5.6	13	5.8	11.8
7	1.3	3.7	14	3.0	4.1

The dependent variable, y, is the annual sales (in millions of dollars), and the independent variable, x, is the size of the store (in square feet).

1. Compute the regression coefficient – b0 (intercept) and b1 (slope)

Under Main menu, go to STAT, Input the data of store sizes (x values) in List 1 and the annual sales (y values) in List 2.

Then go to Calc (F2) and Set (F6). Go to the 3rd row and enter 2 Var XList: List 1 ; 2 Var YList: List2 ; 2 Var Freq: 1. EXE to return to the display of the data.

Now press REG (F3) and press X (F1). At this point, you are offered 2 options: ax+b (F1) or a+bx (F2). Preferably choose (F1) ax+b, then the calculator gives the following results:

Toronto Metropolitan University

TED ROGERS SCHOOL OF MANAGEMENT | Program Advising and Student Success

QMS 210 Simple Linear Regression

LinearReg (ax+b),
a = 1.66986231 =b0
b =0.96447365 =b1
r = 0.95088327
r 2 = 0.904179
MSe = 0.93388968

The values are put together to obtain the simple linear regression equation (the prediction line):
$\hat{y}$ = 0.96447365 + 1.66986231x

2. Hypothesis tests for the slope and correlation coefficient At the 0.05 level of significance, is there evidence of a linear relationship between the size of the store and annual sales?

We test for a linear relationship between the variables x and y by testing whether the population regression coefficient (slope), B, is different from zero.

H_0: β_1 = 0
H_1: $\beta_1 =\!\!/= 0$

Using the same store size and annual sales values lists from before, from MENU, STAT, TEST, t, REG, we get:

LinearReg tTest: β & ρ : =/=
XList: List1
YList: List2
Freq: 1
Save Res: None
EXE

LinearReg tTest:
$\beta =\!\!/= 0$ & $\rho =\!\!/= 0$
t = 10.6411237
p= 1.8226E-07
df = 12
a = 0.96447365
b = 1.66986232
se = 0.96637967
r = 0.95088327

r^2 = 0.904179
Conclusion: p-value = 1.8226E-07 = 0.00000018226 < 0.05 = α Hence reject H0.

Program Advising
and Student Success