

# Brief Introduction to the Interpretation of Regression Analysis

<i>CTRY</i>	<i>GovPart</i>	<i>GSpend</i>
<i>US</i>	6.16	22.83
<i>JA</i>	8.86	16.39
<i>GE</i>	5.71	30.44
<i>FR</i>	5.82	42.95
<i>IT</i>	5.09	38.70
<i>UK</i>	6.06	34.56
<i>CA</i>	5.09	21.99
<i>AU</i>	4.75	40.16
<i>BE</i>	5.88	50.29
<i>DE</i>	5.21	39.49
<i>FI</i>	5.13	30.40
<i>GR</i>	7.52	40.11
<i>IR</i>	6.00	38.74
<i>NE</i>	5.68	54.86
<i>NO</i>	3.99	39.42
<i>PO</i>	8.45	41.53
<i>SP</i>	8.38	23.70
<i>SW</i>	3.64	40.02
<i>SZ</i>	5.78	8.98
<i>AL</i>	6.38	25.79
<i>NZ</i>	5.64	42.38

$$GSpend = 48.30 - 2.32 GovPart$$

(s.e.) (1.82)

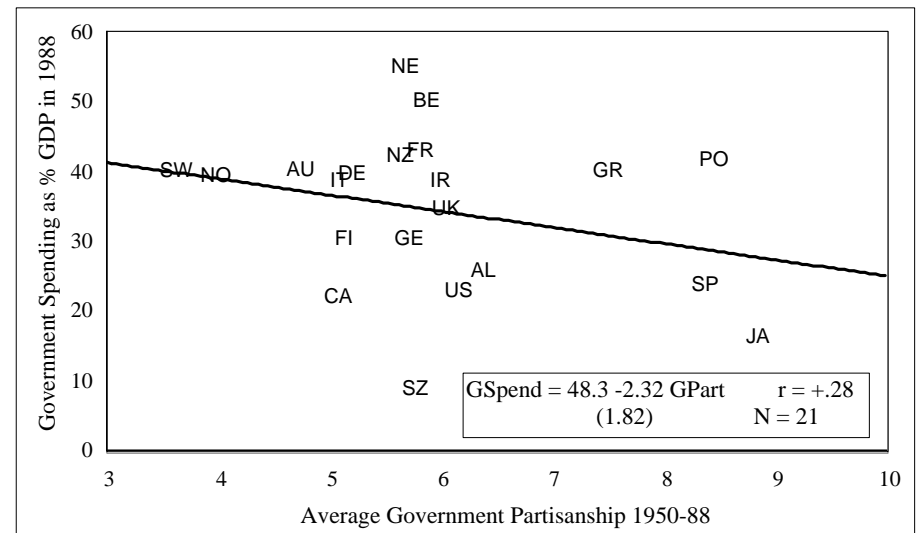
t-stat -1.27

p-level 0.22

$$R^2 = .0787 ; r = .28$$

*GovPart* = 1950-87 average of an index of left-right (0-10) partisanship of government

*GSpend* = 1988 total gov't expenditure as a percent of GDP



**Model to estimate:**  $GSpend = a + b(GovPart) + error$

<i>CTRY</i>	<i>GovPart</i>	<i>NPGov</i>	<i>Spend</i>
<i>US</i>	6.16	1.00	22.83
<i>JA</i>	8.86	1	16.39
<i>GE</i>	5.71	2.2	30.44
<i>FR</i>	5.82	4.30	42.95
<i>IT</i>	5.09	2.70	38.70
<i>UK</i>	6.06	1.00	34.56
<i>CA</i>	5.09	1.00	21.99
<i>AU</i>	4.75	1.80	40.16
<i>BE</i>	5.88	3.00	50.29
<i>DE</i>	5.21	1.80	39.49
<i>FI</i>	5.13	2.90	30.40
<i>GR</i>	7.52	1.10	40.11
<i>IR</i>	6.00	1.60	38.74
<i>NE</i>	5.68	3.30	54.86
<i>NO</i>	3.99	1.80	39.42
<i>PO</i>	8.45	1.60	41.53
<i>SP</i>	8.38	1.00	23.70
<i>SW</i>	3.64	1.40	40.02
<i>SZ</i>	5.78	3.80	8.98
<i>AL</i>	6.38	1.70	25.79
<i>NZ</i>	5.64	1.00	42.38

Model to estimate:  $GSpend = b_0 + b_1GPart + b_2NPart \dots$

...Estimated model:

$$GSpend = 43.5 - 2.0GPart + 1.6NPart$$

(s.e.) (1.9) (2.6)

t-stat -1.0 0.59

p-level .31 .56

$R^2 = .0965$  ;  $r = .31$

Will also often see results represented in a table, thus:

	<i>GSpend, Model 1</i>	<i>GSpend, Model 2</i>
Constant (Intercept)	<b>48.3</b>	<b>43.5</b>
	<b>(s.e.)</b>	<b>(s.e.)</b>
GPart	<b>-2.32</b>	<b>-2.0</b>
	<b>(1.82)</b>	<b>(1.9)</b>
NPart		<b>+2.6</b>
		<b>(.59)</b>
Observations	<b>21</b>	<b>21</b>
$R^2$	<b>.079</b>	<b>.097</b>

1. Notice that the output refers to an equation where the name to the left, here “GSpending” or in the top row of the table or indicated in a note to the table, is the *dependent variable* [thing being explained], Y, and the rest of the names to the right are *independent variables* [things doing the explaining], X: “GPart”, “NPart”.

2. The *(estimated) coefficients* on those independent variables tell you how much the dependent variable tends to move for a 1-unit increase in the independent variable. Thus, the **sign tells the direction of the relationship** & the **size tells you the magnitude** (recognizing the substantive scales of independent & dependent variables).

3. The next row contains the *(estimated) standard errors of the coefficient estimates*. They convey the **precision, certainty, sureness, or tightness of the estimated relationship**: with how great a certainty do increases in the independent variable lead to movements in the dependent variable? You can read these as loosely something like the +/- number that comes with every survey, as a kind of “give-or-take” number. *E.g.*, “when the average partisanship of government increases by 1, government spending tends to decrease by about 2 percent of GDP [coefficient on *GPart*] plus or minus (give or take) 1.9 [standard error on that coefficient].”

4. The next row gives the standard *t-statistics*. Standard errors should at the very least be smaller than the coefficient for us to lend any credence to the sign of the coefficient estimate. We’d prefer them to be no larger than half the coefficient. This t-statistic is simply the coefficient divided by the standard error. So, larger t-statistics imply more certainly signed relationships (more surely positive or negative relationships, not necessarily larger ones). By convention, we look for t-stats approaching or exceeding about 1.65 or about 2.

5. The next row is the probability, under certain assumptions, of having estimated a coefficient this far or farther from 0 if the truly no relationship existed. We can also look to these to see if we

can say with any statistical certainty that a relationship (positive or negative) exists. We like these “*p-levels*” to be approaching or smaller than 0.10 or .05 (corresponding to t-stats of 1.65 or 2).

6. The  $R^2$  often reported with regression also relates to the tightness of the relationship, but it does so for the relationship between the entire set of right-hand-side variables, whereas the std errs, t-stats, and p-levels do so for each independent variable’s relationship to the dependent variable separately.  $R^2$  is literally the square of the correlation between the predicted values (what you get if you use the estimated coefficients and the X values) and the dependent-variable Y’s actual values.  $R^2$  is also the share of the sample variation in the dependent variable, Y, that is “explained” (linearly accounted) by the explanatory variables, X. It thus summarizes how well the *set* of explanators predict the dependent variable.

7. With more than one regressor (right-hand-side variable), this is *multiple regression*. We use this to explore relationships between some x’s and Y *controlling for* other x’s. Graphically, it’s just like the scatter-plots for 2 variables, except in more than 2 dimensions because we are considering the relationship between a *set* of independent variables & the dependent variable. In the simple linear-additive case, each coefficient is “the effect of that x on Y, controlling for the other x’s, i.e., holding all else constant, i.e., after having ‘netted out’ all relationships between the other x’s & Y”.

8. More generally, the “effect of X on Y” is  $\frac{\Delta Y}{\Delta X}$ , which in this case is just the coefficient on X.

(Once we move beyond this simplest kind of equation, purely linear-additive, effects, which are always and everywhere  $\frac{\Delta Y}{\Delta X}$ , can differ from coefficients.