# Linear Regression Examples

**Table 1**. GPA versus GMAT

| Students | GPA | GMAT |
|---|---|---|
| 1 | 3.60 | 708 |
| 2 | 3.57 | 713 |
| 3 | 3.48 | 713 |
| 4 | 3.50 | 710 |
| 5 | 3.45 | 703 |
| 6 | 3.50 | 701 |
| 7 | 3.35 | 672 |
| 8 | 3.40 | 700 |
| 9 | 3.35 | 680 |
| 10 | 3.39 | 665 |
| 11 | 3.28 | 654 |
| 12 | 3.20 | 668 |
| 13 | 3.31 | 650 |
| 14 | 3.30 | 637 |

**Question 1**. Generate a linear regression equation for GMAT as a function of GPA:
$Y = B_1X + B_0$

**Table 2**. Regression Equation Table

| Students | GPA (X) | GMAT (Y) | $X - M_X$ (3) | $Y - M_Y$ (4) | $(X - M_X)^2$ (5) | $(Y - M_Y)^2$ (6) | DE (5)(6) |
|---|---|---|---|---|---|---|---|
| 1 | 3.60 | 708 | 0.194286 | 24.14286 | 0.037747 | 582.87755 | 4.6906122 |
| 2 | 3.57 | 713 | 0.164286 | 29.14286 | 0.02699 | 849.30612 | 4.7877551 |
| 3 | 3.48 | 713 | 0.074286 | 29.14286 | 0.005518 | 849.30612 | 2.16489896 |
| 4 | 3.50 | 710 | 0.094286 | 26.14286 | 0.00889 | 683.44898 | 2.46489796 |
| 5 | 3.45 | 703 | 0.044286 | 19.14286 | 0.001961 | 366.44898 | 0.8477551 |
| 6 | 3.50 | 701 | 0.094286 | 17.14286 | 0.00889 | 293.87755 | 1.61632653 |
| 7 | 3.35 | 672 | -0.05571 | -11.8571 | 0.003104 | 140.59184 | 0.66061224 |
| 8 | 3.40 | 700 | -0.00571 | 16.14286 | 3.27E-05 | 260.59184 | -0.0922449 |
| 9 | 3.35 | 680 | -0.05571 | -3.85714 | 0.003104 | 14.877551 | 0.21489796 |
| 10 | 3.39 | 665 | -0.01571 | -18.8571 | 0.000247 | 355.59184 | 0.29632653 |
| 11 | 3.28 | 654 | -0.12571 | -29.8571 | 0.015804 | 891.44898 | 3.75346939 |
| 12 | 3.20 | 668 | -0.20571 | -15.8571 | 0.042318 | 251.44898 | 3.26204082 |
| 13 | 3.31 | 650 | -0.09571 | -33.8571 | 0.009161 | 1146.3061 | 3.24061224 |
| 14 | 3.30 | 637 | -0.10571 | -46.8571 | 0.011176 | 2195.5918 | 4.95346939 |
| | $M_X$ | $M_Y$ | | | $SS_X$ | $SS_Y$ | SP |
| Mean (M) | 3.4057 | 683.8571 | | | 0.1749 | 8881.7143 | 32.8614 |

$\mathbf{B_1} = SP/SS_X = 32.8614/0.1749 =$ **187.84*** and
$\mathbf{B_0} = M_Y - B_1(M_X) = 683.8571 - 187.89(3.4057) =$ **44.13**
* Note: calculation based on more significant digits
So regression equation is ***GMAT = 1.87.84(GPA)+ 44.13***

**Question 2**. Predict a student's GMAT score if that student's GPA is 3.43.

***GMAT = 187.84(3.43) + 44.13 = 688***

**Question 3**. What is the Residual sum of square ($SS_{Res}$) for the regression model or equation?

**Table 3**. Residual Sum of Square Table

| Students | GPA (X) | GMAT (Y) | Predicted Y *187.84X + 44.13* (Y) | Residuals Y - Y | Residual$^2$ $(Y - Y)^2$ |
|---|---|---|---|---|---|
| **1** | 3.60 | 708 | 720.35195 | -12.352 | 152.5707 |
| **2** | 3.57 | 713 | 714.71672 | -1.71672 | 2.947141 |
| **3** | 3.48 | 713 | 697.81104 | 15.18896 | 230.7045 |
| **4** | 3.50 | 710 | 701.56786 | 8.432141 | 71.101 |
| **5** | 3.45 | 703 | 692.17581 | 10.82419 | 117.163 |
| **6** | 3.50 | 701 | 701.56786 | -0.56786 | 0.322464 |
| **7** | 3.35 | 672 | 673.39172 | -1.39172 | 1.936884 |
| **8** | 3.40 | 700 | 682.78377 | 17.21623 | 296.3987 |
| **9** | 3.35 | 680 | 673.39172 | 6.60828 | 43.66937 |
| **10** | 3.39 | 665 | 680.90536 | -15.9054 | 252.9804 |
| **11** | 3.28 | 654 | 660.24285 | -6.24285 | 38.97324 |
| **12** | 3.20 | 668 | 645.21558 | 22.78442 | 519.1298 |
| **13** | 3.31 | 650 | 665.87808 | -15.8781 | 252.1135 |
| **14** | 3.30 | 637 | 663.99967 | -26.9997 | 728.9824 |
| | $M_X$ | $M_Y$ | | | $SS_{Res}$ |
| Mean (M) | 3.4057 | 683.8571 | | | **2708.993** |

$SS_{Residual} = SS_{Res} = \sum(Y - Y)^2 =$ **2708.993**

**Question 4**. Calculate Pearson r and the regression sum of square for the linear regression model.

**Pearson r:**

$$r = \frac{SP}{\sqrt{SS_X SS_Y}} = \frac{32.8614}{\sqrt{0.1749(8881.7143)}} = 0.8338, \text{ and } r^2 = 0.6952$$

$$SS_{Regression} = SS_{Reg} = r^2 SS_Y = 0.6952(8881.7143) = 6174.58$$

$$SS_{Residual} = SS_{Res} = (1 - r^2)SS_Y = 0.3048(8881.7143) \simeq 2707.15$$

(Compare to 2708 from Table 3)

So $r =$ **0.83** and $SS_{Reg} =$ **6174**

**Question 5.** Produced the regression analysis of variance table and test the significance of the regression equation.

Analysis of Regression is similar to Analysis of variance:

***F*-ratio** = *Mean Square Regression divided by Mean Square of Residual*

| **Source** | ***SS*** | ***df*** | **MS** | ***F*** |
|---|---|---|---|---|
| **Regression** | $SS_{Reg}$ | 1 | $MSR = SS_{Reg}/1$ | MSR/MSE |
| **Residual** | $SS_{Res}$ | n – 2 | $MSE = SS_{Res}/(n-2)$ | |
| **Total** | $SS_{Reg} + SS_{Res}$ | n - 1 | | |

**Table 4**. Analysis of Regression Table

| **Source** | ***SS*** | ***df*** | **MS** | ***F*** |
|---|---|---|---|---|
| **Regression** | 6174.58 | 1 | 6174.58 | **27.35** |
| **Residual** | 2708.99 | 12 | 225.75 | |
| **Total** | 8893.57 | 13 | | |

$H_0$: The regression equation does not account for a significant portion of the variability for the Y scores

$H_a$: The regression equation account for a significant portion of the variability for the Y scores (i.e. the regression equation is a good model for predicting Y from X).

$F_{Statistics}$ = **27.52** and $F_{CV}$ = **4.75** (*df* = 1, 12 and a = 0.05)

Since $F_{Stat} > F_{CV}$, Reject the Null Hypothesis and conclude that the regression equation is a good one.

Question 6. Compute the standard error of the estimate.

$$\text{Std error est} = \sqrt{\frac{SS_{Res}}{n-2}} = \sqrt{\frac{2708.993}{12}} = 15.02$$

SPSS Outputs from Linear Regression Analysis of GPA and GMAT:

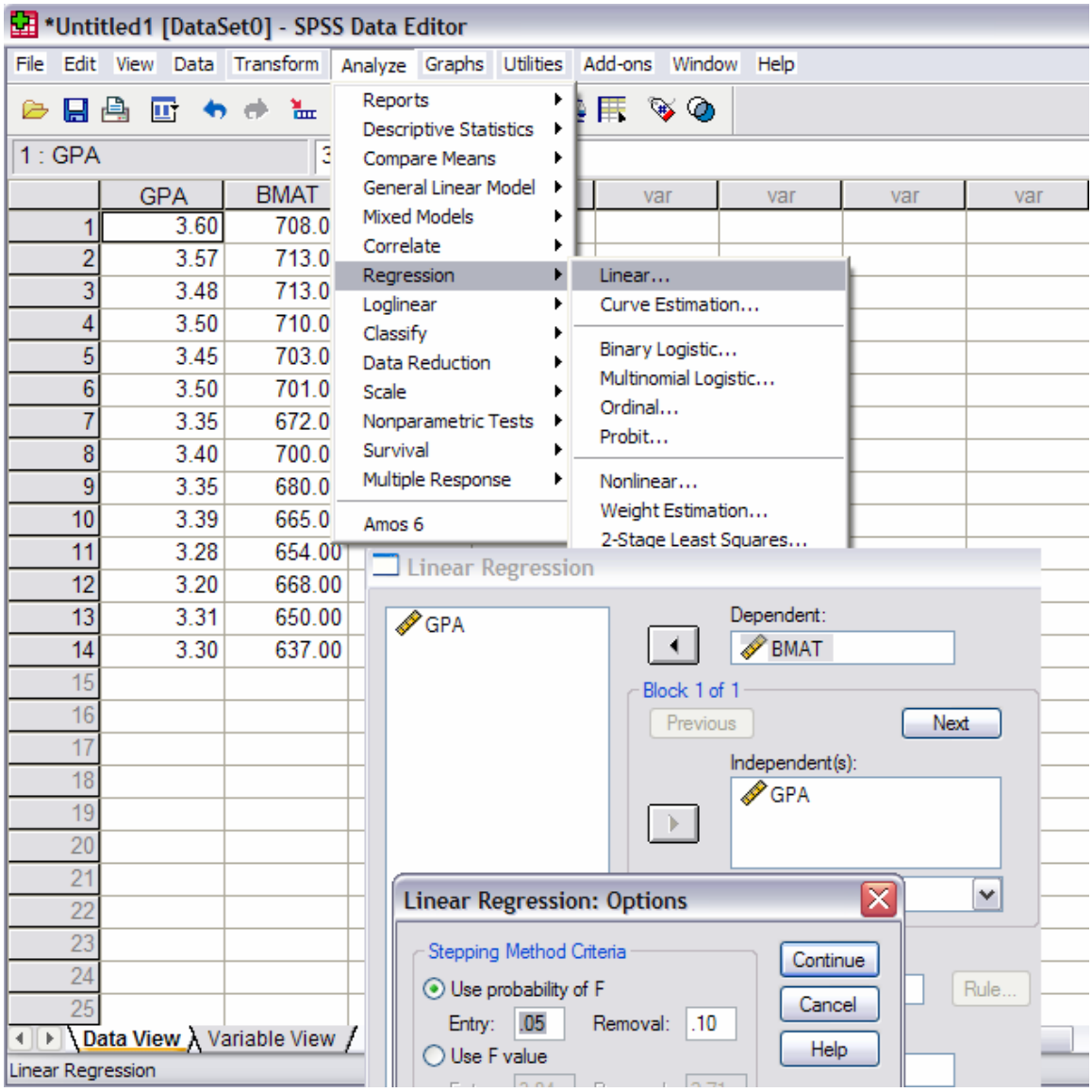


**Figure 1**. SPSS: Linear Regression Procedure.

## SPSS Outputs: Linear Regression

**Variables Entered/Removed(b)**

| Model | Variables Entered | Variables Removed | Method |
|---|---|---|---|
| 1 | GPA(a) | . | Enter |

a All requested variables entered.
b Dependent Variable: BMAT

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | **.834**(a) | **.695** | .670 | **15.02496** |

a Predictors: (Constant), GPA

**ANOVA(b)**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | **6172.721** | **1** | **6172.721** | **27.343** | .000(a) |
| | Residual | **2708.993** | **12** | **225.749** | | |
| | Total | **8881.714** | **13** | | | |

a Predictors: (Constant), GPA
b Dependent Variable: BMAT

**Coefficients(a)**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | | |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | t | Sig. |
| 1 | (Constant) | **44.125** | 122.407 | | .360 | .725 |
| | GPA | **187.841** | 35.922 | .834 | 5.229 | .000 |

a Dependent Variable: BMAT

SPSS Output: Linear regression Plot

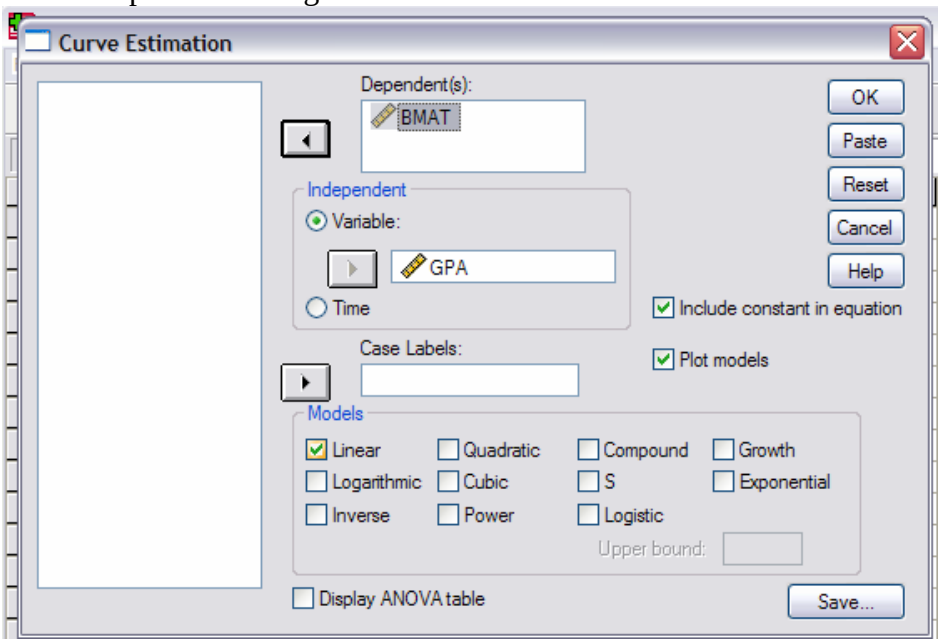


Figure 2. SPSS Linear Plot: Analyze -> Regression -> Curve Estimate (Linear)

**Model Summary and Parameter Estimates**

Dependent Variable: BMAT

| Equation | Model Summary | | | | | Parameter Estimates | |
|---|---|---|---|---|---|---|---|
| | R Square | F | df1 | df2 | Sig. | Constant | B1 |
| Linear | **.695** | **27.343** | 1 | 12 | .000 | 44.125 | 187.841 |

The independent variable is GPA.

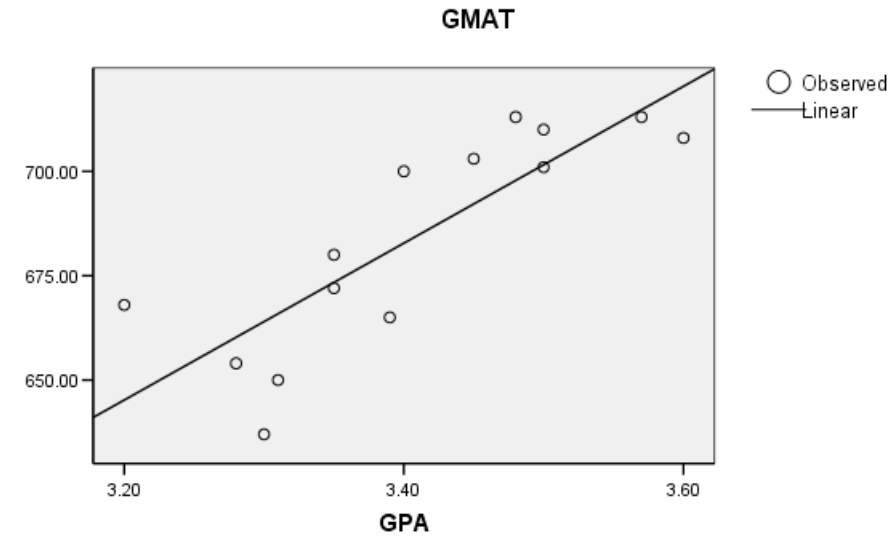

**Table 5**. *F*-Distribution Table a = 0.05

| a= 0.05 | *df1* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ***df2*** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **1** | 161.45 | 199.50 | 215.71 | 224.58 | 230.16 | 233.99 | 236.77 | 238.88 |
| **2** | 18.5128 | 19.0000 | 19.1642 | 19.2467 | 19.2963 | 19.3295 | 19.3531 | 19.3709 |
| **3** | 10.1280 | 9.5521 | 9.2766 | 9.1172 | 9.0134 | 8.9407 | 8.8867 | 8.8452 |
| **4** | 7.7086 | 6.9443 | 6.5914 | 6.3882 | 6.2561 | 6.1631 | 6.0942 | 6.0410 |
| **5** | 6.6079 | 5.7861 | 5.4094 | 5.1922 | 5.0503 | 4.9503 | 4.8759 | 4.8183 |
| **6** | 5.9874 | 5.1432 | 4.7571 | 4.5337 | 4.3874 | 4.2839 | 4.2067 | 4.1468 |
| **7** | 5.5915 | 4.7374 | 4.3468 | 4.1203 | 3.9715 | 3.8660 | 3.7871 | 3.7257 |
| **8** | 5.3176 | 4.4590 | 4.0662 | 3.8379 | 3.6875 | 3.5806 | 3.5005 | 3.4381 |
| **9** | 5.1174 | 4.2565 | 3.8625 | 3.6331 | 3.4817 | 3.3738 | 3.2927 | 3.2296 |
| **10** | 4.9646 | 4.1028 | 3.7083 | 3.4780 | 3.3258 | 3.2172 | 3.1355 | 3.0717 |
| **11** | 4.8443 | 3.9823 | 3.5874 | 3.3567 | 3.2039 | 3.0946 | 3.0123 | 2.9480 |
| **12** | **4.7472** | 3.8853 | 3.4903 | 3.2592 | 3.1059 | 2.9961 | 2.9134 | 2.8486 |
| **13** | 4.6672 | 3.8056 | 3.4105 | 3.1791 | 3.0254 | 2.9153 | 2.8321 | 2.7669 |
| **14** | 4.6001 | 3.7389 | 3.3439 | 3.1122 | 2.9582 | 2.8477 | 2.7642 | 2.6987 |
| **15** | 4.5431 | 3.6823 | 3.2874 | 3.0556 | 2.9013 | 2.7905 | 2.7066 | 2.6408 |
| **16** | 4.4940 | 3.6337 | 3.2389 | 3.0069 | 2.8524 | 2.7413 | 2.6572 | 2.5911 |
| **17** | 4.4513 | 3.5915 | 3.1968 | 2.9647 | 2.8100 | 2.6987 | 2.6143 | 2.5480 |
| **18** | 4.4139 | 3.5546 | 3.1599 | 2.9277 | 2.7729 | 2.6613 | 2.5767 | 2.5102 |
| **19** | 4.3808 | 3.5219 | 3.1274 | 2.8951 | 2.7401 | 2.6283 | 2.5435 | 2.4768 |
| **20** | 4.3513 | 3.4928 | 3.0984 | 2.8661 | 2.7109 | 2.5990 | 2.5140 | 2.4471 |
| **21** | 4.3248 | 3.4668 | 3.0725 | 2.8401 | 2.6848 | 2.5727 | 2.4876 | 2.4205 |
| **22** | 4.3009 | 3.4434 | 3.0491 | 2.8167 | 2.6613 | 2.5491 | 2.4638 | 2.3965 |
| **23** | 4.2793 | 3.4221 | 3.0280 | 2.7955 | 2.6400 | 2.5277 | 2.4422 | 2.3748 |
| **24** | 4.2597 | 3.4028 | 3.0088 | 2.7763 | 2.6207 | 2.5082 | 2.4226 | 2.3551 |
| **25** | 4.2417 | 3.3852 | 2.9912 | 2.7587 | 2.6030 | 2.4904 | 2.4047 | 2.3371 |
| **26** | 4.2252 | 3.3690 | 2.9752 | 2.7426 | 2.5868 | 2.4741 | 2.3883 | 2.3205 |
| **27** | 4.2100 | 3.3541 | 2.9603 | 2.7278 | 2.5719 | 2.4591 | 2.3732 | 2.3053 |
| **28** | 4.1960 | 3.3404 | 2.9467 | 2.7141 | 2.5581 | 2.4453 | 2.3593 | 2.2913 |
| **30** | 4.1709 | 3.3158 | 2.9223 | 2.6896 | 2.5336 | 2.4205 | 2.3343 | 2.2662 |
| **40** | 4.0847 | 3.2317 | 2.8387 | 2.6060 | 2.4495 | 2.3359 | 2.2490 | 2.1802 |
| **50** | 4.0343 | 3.1826 | 2.7900 | 2.5572 | 2.4004 | 2.2864 | 2.1992 | 2.1299 |
| **60** | 4.0012 | 3.1504 | 2.7581 | 2.5252 | 2.3683 | 2.2541 | 2.1665 | 2.0970 |
| **120** | 3.9201 | 3.0718 | 2.6802 | 2.4472 | 2.2899 | 2.1750 | 2.0868 | 2.0164 |
| **100000** | 3.8415 | 2.9958 | 2.6050 | 2.3720 | 2.2142 | 2.0987 | 2.0097 | 1.9385 |