

Accompanying this computer assignment is a new data set of 150 students who recently completed one year of college. The data set contains the following information:

1. Gender (1=female, 0=male) and Minority status (1=minority, 0=non-minority student).
2. High School GPA.
3. High School Proficiency Test scores (science, math, writing, reading) taken during junior year in high school.
4. ACT test scores (English, math, reading, science, composite) taken during senior year. Notice that the composite is the average of the four subject tests.
5. SAT test scores (math, verbal) taken during senior year.
6. College GPA in math courses, English courses, and science courses.

Based on the High School Sample data base, conduct a multivariate-multiple regression and assess the multivariate normality assumption.

Select 5 explanatory variables for X and use the three college GPAs for your outcomes Y. Compute the betas, SST, SSR, SSE, and R-Squared for each outcome.

Matrix.

```
get xpre/variables (you choose 5).
compute ones = make(nrow(xpre),1,1).
compute x = {ones,xpre}.
get y/variables gpascien gpamath gpaengl.
```

```
compute meany = csum(y)/nrow(y).
print meany.
compute onesmt=make(nrow(x),1,1).
compute meanym = onesmt*meany.
compute dy = y-meanym.
```

```
compute betas =
compute sstm =
compute ssrcm =
compute I = ident(nrow(x),nrow(x)).
compute ssem=
compute sst = diag(sstm).
compute ssrc = diag(ssrcm).
compute sse = diag(ssem).
compute dsst = inv(mdiag(sst)).
compute rsquared = dsst*ssr.
print betas.
print sst.
print ssrc.
print sse.
print rsquared.
```

End matrix.

***Answer the following Questions:***

1. What 5 explanatory variables did you select?
2. Which outcome (GPA in science, math, or English) did you explain best?
3. Do you meet the multivariate normal distribution assumption?
4. Turn in the following:
  - a. Printout from Matrix Commands
    - i. Betas
    - ii. SSs
    - iii. R-Squared
  - b. Q-Q Plot to be computed below

---

**Checking Multivariate Normal Distribution Assumption**

---

MATRIX.

```
get x /variables gspascien gpamath gpaengl (add your Xs).
```

```
Compute ones = make(nrow(x),1,1).  
Compute means = T(ones)*X/nrow(x).  
Compute d = x-(ones*means).  
Compute sigma = 1/nrow(x)*T(d)*d.  
Compute distance =
```

```
save distance  
  /outfile = *  
  /variables = distance.
```

END MATRIX.

This Matrix procedure will compute distances based on our definition of Mahalanobis multivariate distances that should be distributed Chi-Square with  $p$  degrees of freedom. These procedures will save the distances and create a new data file with one variable, distance.

Create a Q-Q plot of these distances with  $p$  degrees of freedom (if you use the three Ys and the 5 Xs, you will have 8 degrees of freedom).

In SPSS:

- a. [Graphs] → [Q-Q Plot] →
- b. Move “Distance” into the [Variables] box
- c. Select “Chi-Square” for the [Test Distribution]
- d. Enter the number of Degrees of Freedom ( $p$ , number of variables)
- e. [OK]