

STAT525 HOMEWORK#3

1. KNNL Problem 2.57.
2. KNNL Problem 2.61.
3. KNNL Problem 3.19
4. (SAS Exercise) Use the **crime rate** data described in KNNL Problem 1.28.
 - (a) Describe the distribution of the explanatory variable.
 - (b) Run the linear regression to predict the county crime rate from the percentage of individuals having at least a high school diploma.
 - (c) Plot the residuals versus the explanatory variable and briefly describe the plot noting any unusual patterns or points.
 - (d) Examine the distribution of the residuals by getting a histogram and a normal probability plot of the residuals by using the HISTOGRAM and QQPLOT statements in PROC UNIVARIATE. What do you conclude?
5. (SAS Exercise) Use the **crime rate** data described in KNNL Problem 1.28. Change the data set by changing the value of the crime rate for the last observation from 7582 to 758 (e.g., a typo). You can do this in a data step. For example,

```
DATA a2; SET a1; IF _n_ EQ 84 THEN y=758;
```

An alternative is to simply edit the data file.

- (a) Make a table comparing the results of this analysis with the results of the analysis of the original data. Include in the table the following: fitted equation, t-test for the slope, with standard error and p -value, R^2 , and the estimate of σ^2 . Briefly summarize the differences.
 - (b) Repeat parts (c) and (d) from the previous problem for this altered data set analysis and summarize how these plots help you to detect the unusual observation.
6. (SAS Exercise) Use the **sales growth** data described in KNNL Problem 3.17.
 - (a) Generate a scatterplot of the data and discuss the appropriateness of using a linear regression model.
 - (b) Using PROC TRANSREG, which power transformation of Y (i.e., value of λ) is most appropriate to use here?
 - (c) Apply this transformation of Y and generate a scatterplot. Again comment on the appropriateness of using a linear regression model.
 - (d) Run the regression model using the transformed data and generate a residual plot (using X or \hat{Y}) and a normal probability plot. What do the plots show?
 - (e) Express the estimated regression function in the original units.