

STAT 210
Applied Statistics and Data Analysis:
Homework 7

Due on November 16, 2025

Question 1

For this question use the data set `25Fhw7Q1.csv`.

- (i) Read the data and plot `vary` as a function of `varx`. Fit a simple linear regression for `vary` as a function of `varx` and add the regression line to the plot. Comment. Obtain a summary for the regression and interpret the result. Draw the diagnostic plots and use appropriate tests to check for normality and homogeneous variances. Comment on the results. What would be the predicted `vary` for a point with `varx = 2.5` using this model? Include a confidence interval at the 98% confidence level.
 - (ii) In this part you have to use the Box-Cox transformations to improve the model. To simplify this problem, you have to choose between two transformations of the output variable `vary`, a square root or a logarithm. Use the function `boxcox` on the package `MASS` with the argument `set` to the model you fitted in (i). If the confidence interval in the graph includes zero, choose a logarithmic transformation for `vary`. If the confidence interval in the graph includes 0.5 then choose a square root transformation.
 - (iii) Fit a new model with the transformation that you choose in (ii). Obtain a summary for the new regression and compare with the previous one. Draw the diagnostic plots and compare with the previous results. Use appropriate tests to check for normality and homogeneous variances.
 - (iv) Write down an equation for the final model in terms of the original variables. What would be the predicted `vary` for a point with `varx = 2.5`? Include a confidence interval at the 98% confidence level (to do this, get a confidence interval for the model with the transformed `vary` (using the logarithm or the square root) and use the inverse transformation (exponential or square) on the extremes of the confidence interval). Draw a scatterplot of `vary` against `varx` and add the regression line for the first model and the curve you obtained with the second regression.
-

Question 2

The data for this question is in the file `25Fhw7Q2.csv`

- (i) Draw a scatterplot of `w` as a function of `t`. Fit a simple linear regression model and add the line to the plot. Comment. Obtain a summary of the regression and explain the different components of the output. Write down an equation for the model. What is the estimated standard deviation for the errors?
- (ii) Draw the diagnostic plots and comment on what you observe. Use appropriate tests to check for normality and homogeneous variances. Do you identify one or more points as outliers? If you do, which points are they? Can you identify these points in the initial scatterplot?
- (iii) Fit a new regression model excluding the outlier(s) that you identified in the previous section. Draw a scatterplot with both regression lines. Compare the summary tables. Write down an equation for the new model. What is the estimated standard deviation for the errors? Do you think the outliers are influential points?

- (iv) Draw the diagnostic plots for the new model and comment. Use appropriate tests to check for normality and homogeneous variances. Compare with the results obtained for the initial model and comment.
- (v) Describe the sampling distribution for the parameters in the model that excludes the outliers. Give confidence intervals at a confidence level of 99% for the parameters of the regression. Find the predicted value for $t = 45$ with a confidence interval at the 99% level using both models and compare the results.