

Test 1
STAT 320
Dr. Bensmail

February 20,2006

Name:.....

ID:.....

Data on a sample of **57** pitchers have been collected, which include the **Salary (Y)** and **Years of experience (X)**. The goal is to fit a regression line to explain the variability in these salaries using the predictor **X**. (**Download Data "Pitcher" from my home page <http://web.utk.edu/~hbensmai>**)

Each question is counted 5 points, Total = 100 points

(1) Provide the followings summary statistics for the variable "Years":

(a) Mean, (b) Std deviation, First Quantile and Third Quantile

(2) What percent of players have less than 2 years of experience?

(3) Generate the following plots for variable "Years":

(a) Histogram of "Years"

(b) Normal quantile plot of "Years"

(4) Do you think that the distribution of "Years" is normal? Symmetric? Explain clearly?

(5) Provide a scatterplot of the "Salary" against "Years".

(6) Do you think that there is a linear relationship between "Salary" and "Years"? why or why not?

(7) Using the summary statistics from JMP (R, S_X, S_Y, \bar{X} and \bar{Y}) calculate the estimated slope and estimated intercept of the regression line.

(8) Using JMP, what is the estimated slope and intercept of the regression line.

- (9) Give the the three variations SST, SSE and SSR.
- (10) Calculate the standard error of the regression line \hat{Y} (**predicted Salary**).
- (11) Is "**Years**" important to predict "**Salary**". Answer this question by providing a **95%** confidence interval of the slope and discuss.
- (12) If a player has **17** years of experience, predict his salary.
- (13) **What is the meaning of the estimated slope related to the Years?**
- (14) **What is the meaning of the coefficient of determination?**
- (15) Provide the residual plot (**Residuals** versus "**Years**" or **Residuals** versus **predicted Salary**). Do you detect any problem? If yes, which assumption on the residuals is clearly violated?
- (16) We decided to transform "Salary" using a logarithm transformation. We note this new variable: LogSalary. Fit a regression model on LogSalary using Years as explanatory variable.
- (17) Comment on the residual plot from the output of the regression of "**LogSalary**" versus "**Years**".
- (18) Using the regression model in part (16) what is the predicted salary of a player with **17** years of experience? Compare your answer with the point estimate obtained in part (12).