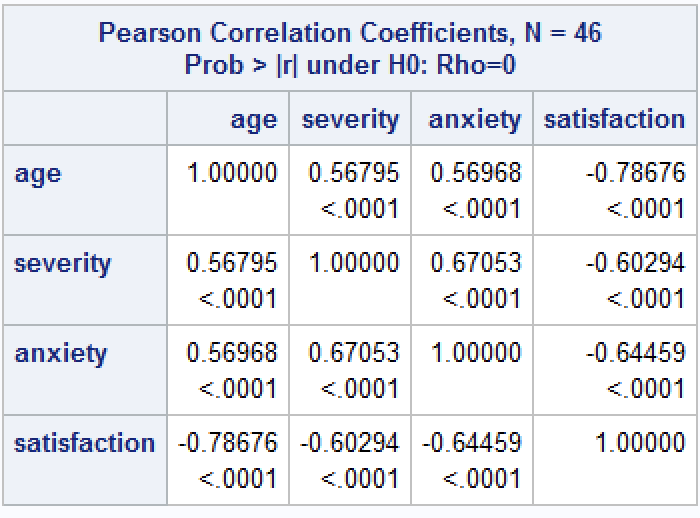
**1. Use the patient satisfaction data described in KNNL Problem 6.15.**

1. Compute the pairwise correlations between the ’s and between each and . Which variable appears to be the best individual predictor?



Age appears to be the best individual predictor since satisfaction varies the strongest relative to age () than severity and anxiety ( and , respectively).

1. Run the linear regression with age, severity of illness and anxiety level as the explanatory variables and satisfaction as the response variable. Summarize the regression results.

A screenshot of a computer

Description automatically generated

The regression model is , but the only significant predictor seems to be age ().

1. Plot the residuals versus the predicted satisfaction and each of the explanatory variables. Are there any unusual patterns?

A graph of different age groups

Description automatically generated with medium confidenceA graph with blue dots

Description automatically generated

Residuals appear to have constant variance and more or less evenly distributed about , so we conclude that so far, none of our assumptions about regression have been violated.

1. Examine the assumption of normality for the residuals using a qqplot or histogram. State your conclusions.

A graph of a normal line

Description automatically generated

We see that the residuals do not significantly deviate from the 45-degree line, so we conclude our assumption of normality is appropriate.

1. Predict the satisfaction for a year old patient with illness severity and anxiety level . Provide a 95% prediction interval with your prediction.

. , , and is singular… how do I proceed?

**2. Refer to Patient satisfaction Problem 6.15.**

1. Obtain the analysis of variance table that decomposes the regression sum of squares into extra sums of squares associated with ; with, given ; and with , given and .

A table with numbers and symbols

Description automatically generated

1. Test whether can be dropped from the regression model given that and are retained. Use the test statistic and level of significance . State the alternatives, decision rule, and conclusion. What is the -value of the test?

Hypotheses: , . We want to conclude , and to conclude . Then . This is less than , so we conclude ; anxiety not useful in predicting satisfaction and can be dropped.

**3. Refer to Patient satisfaction Problem 6.15. Test whether both and can be dropped from the regression model given that is retained. Use . State the alternatives, decision rule, and conclusion. What is the -value of the test?**

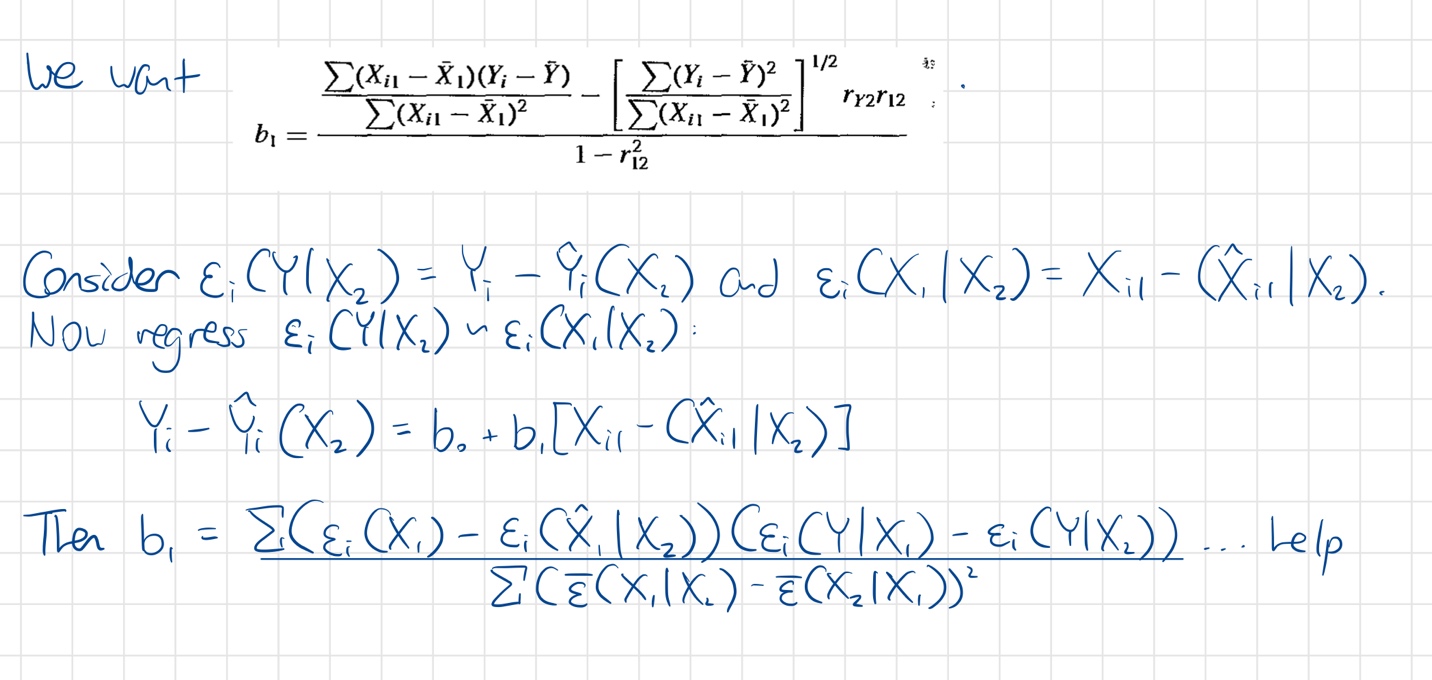
**A screenshot of a computer

Description automatically generatedA table with numbers and symbols

Description automatically generated**

Hypotheses: , At least one for . The test statistic is This is greater than , so we conclude among the predictor variables severity and anxiety, at least one of them is statistically useful in predicting satisfaction. We can further justify our conclusion using “thimid test severity, anxiety” in SAS. After doing so, we get a -value of , so we also conclude .

**4. Derive the equation (7.56) on page 281 (HINT: Recall that can be obtained by regressing the residuals of vs the residuals of ).**



**5. Steroid level. An endocrinologist was interested in exploring the relationship between the level of a steroid () and age () in healthy female subjects whose ages ranged from to years. She collected a sample of healthy females in this age range.**

1. Fit regression model (8.2). Plot the fitted regression function and the data. Does the quadratic regression function appear to be a good fit here? Find .

A screenshot of a computer

Description automatically generated

and the predictors “age” and “age2” are significant, so the quadratic regression function appears to be a good fit.

1. Test whether or not there is a regression relation; use . State the alternatives, decision rule, and conclusion. What is the -value of the test?

Hypotheses: , At least one for . The decision rule is as follows: , conclude , and , conclude . Then and , so we conclude that there exists a regression relationship between age and steroid level.

1. Obtain joint interval estimates for the mean steroid level of females aged , , and respectively. Use the most efficient simultaneous estimation procedure and a percent

family confidence coefficient. Interpret your intervals.

We know , so , , . Consider the Bonferroni simultaneous prediction limits for : Then … once again, is singular and I can’t proceed. If I know the value of , then I know the simultaneous mean response CIs are for and . The way we would interpret these intervals is that we are 99% confident the true mean steroid level is within those intervals for ages and .

1. Predict the steroid levels of females aged using a percent prediction interval.

Interpret your interval.

We know , and the 99% PI is given by . We are 99% confident that a new observation of a steroid level for a 15-year-old woman will lie within this interval.

1. Test whether the quadratic term can be dropped from the model; use . State the alternatives, decision rule, and conclusion.

A screenshot of a graph

Description automatically generated

Hypotheses: , . The decision rule is as follows: , conclude , and , conclude . , which is greater than , so we conclude that the quadratic term has statistically significant predictive power in the model.

1. Express the fitted regression function obtained in part (a) in terms of the original variable .

The fitted regression function is . Let . Then .

**6. Refer to Copier maintenance Problem 1.20. The users of the copiers are either training institutions that use a small model, or business firms that use a large, commercial model An analyst at Tri-City wishes to fit a regression model including both number of copiers serviced () and type of copier () as predictor variables and estimate the effect of copier model (S-small, L-large) on number of minutes spent on the service call. Assume that regression model (8.33) is appropriate, and let if small model and if large, commercial model.**

1. Explain the meaning of all regression coefficients in the model.

(8.33), the regression model is . is the expected number of service minutes when there are no copiers to be serviced and the user is a business firm. is the expected increase in service minutes for each additional copier that requires work. is the expected increase in service minutes for a training institution. Finally, is the error term.

1. Fit the regression model and state the estimated regression function.

A screenshot of a computer

Description automatically generated

The regression function is .

1. Estimate the effect of copier model on mean service time with a 95 percent confidence interval. Interpret your interval estimate.

We want . We are 95% confident that the mean service time for a small copier lies within the interval .

1. Why would the analyst wish to include , number of copiers, in the regression model when interest is in estimating the effect of type of copier model on service time?

Because the number of copiers is also statistically significant in predicting service time, so accounting for this variable lets us make better estimates based on type of copier.

1. Obtain the residuals and plot them against . Is there any indication that an interaction term in the regression model would be helpful?

A graph with numbers and dots

Description automatically generated

The “type\_serviced” variable is the interaction term . Indeed, an interaction term in the regression model would appear helpful as evidenced by the distribution of residuals.