Probability and Statistics for Psychology
and Quantitative Methods for Human Sciences

Problem Sheet 15 (HT 12): Correlation and Regression

1. The table below shows measurements of height and forced expiratory volume in one second (FEV) in a sample of male medical students.

Height (cm)	FEV (litres)
164.0	3.5
170.4	3.2
171.3	3.2
172.0	3.8
176.0	3.8
177.0	5.4
178.0	3.0
181.0	4.0
183.7	4.7

   (a) Represent the data graphically.

   (b) Calculate the sample correlation.

   (c) Perform a linear regression and include the fitted line in your graph.

   (d) Test whether the slope of the regression line differs significantly from zero.

2. A study was done in which boys' height was measured at age 6, and then again at age 18. The results may be summarized as follows:

average height at age 6 = 46 inches	SD=1.7 inches
average height at age 18 = 70 inches	SD=2.5 inches
correlation between measurements=0.8	

   Both measurements are normally distributed, and the scatterplot is oval shaped.

   (a) Write the regression equation for predicting height at age 18, given the height at age 6.

   (b) Write the regression equation for predicting height at age 6, given the height at age 18.

   (c) If we pick a random boy whose height was 48 inches at age 6, what height should you predict for him at age 18?

   (d) **[More advanced]** If we pick a random boy whose height was 48 inches at age 6, what is the probability that his height is over 72 inches at age 18?

3. A large manufacturing firm wants to determine whether a relationship exists between the number of work-hours an employee misses per year and the employee's annual wages (in thousands of dollars). A sample of 15 employees produced the following data:

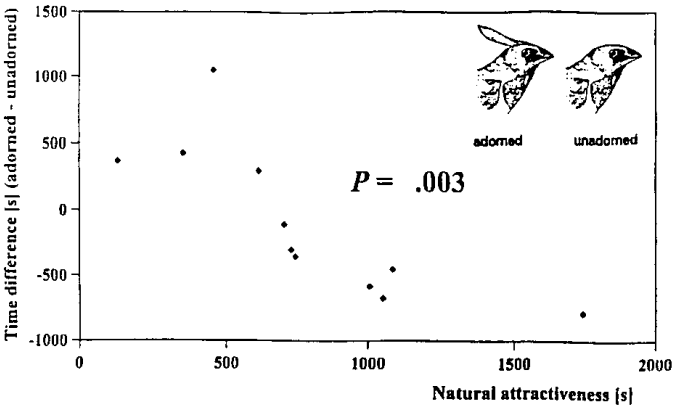
Employee	Hours	Wages
1	49	15.8
2	36	17.5
3	127	11.3
4	91	13.2
5	72	13.0
6	34	14.5
7	155	11.8
8	11	20.2
9	191	10.8
10	6	18.8
11	63	13.8
12	79	12.7
13	43	15.1
14	57	24.2
15	82	13.9

   (a) Calculate Spearman's rank correlation coefficient as a measure of the strength of the relationship between work-hours missed and annual wages.

   (b) Is there sufficient evidence to indicate that work-hours missed decreases as annual wages increases?

4. [WC99] studied the effect of ornaments on mate selection in the Javanese mannikins *Lonchura leucogastroides*. Among other measurements, they looked at a measure of attractiveness of male birds to females (X_i), and then measured the change (Y_i) in attractiveness when the male was adorned with a red feather. They found that X_i and Y_i were negatively correlated. The scatter plot is shown here:

a)



The number of individuals tested was $n = 11$, and they found the correlation coefficient between the initial attractiveness and the change resulting from the feather, was -0.73, while the p-value for a hypothesis test of no correlation between initial attractiveness and change in attractiveness was 0.003. They write that "the more attractive a male was before ornamentation, the more he lost in attractiveness after ornamentation and vice versa," and there is an "ideal" male appearance, such that birds who already approximate the ideal can only lose by changing their appearance, while those who differ from the ideal can only benefit by adding a novel feature. Do you agree that the evidence supports such an interpretation? Or is there another plausible explanation for this negative correlation?

(This example is discussed by [KP05].)

References

[KP05] Colleen Kelly and Trevor D. Price. Correcting for regression to the mean in behavior and ecology. *The American Naturalist*, 166(6):700–7, 2005.

[WC99] Klaudia Witte and Eberhard Curio. Sexes of a monomorphic species differ in preference for mates with a novel trait. *Behavioral Ecology*, 10(1):15–21, 1999.