SEX DISCRIMINATION PROBLEM

6. Summarizing Relationships between Variables

The Pearson correlation coefficient measures the strength and direction of a linear relationship between two quantitative variables. The scatterplots discussed in the previous section revealed a linear association between the logarithm of starting salaries and some other variables such as EDUC (education), SENIOR (seniority), and TREXP (1/(EXPER +12)). The following two tables obtained with SPSS display the values of the Pearson correlation coefficients and the p-values of the tests of significance for the correlations:

CORRELATION COEFFICIENTS

	AGE	LNBSAL	EDUC	TREXP	FSEX	SAL77	SENIO
AGE	1	.0648	-.2253	-.6522	.2618	-.5467	-.1845
LNBSAL	.0648	1	.4074	-.5077	-.5432	.4095	-.2944
EDUC	-.2253	.4074	1	-.0784	-.3273	.4210	.0598
TREXP	-.6522	-.5077	-.0784	1	.0835	.1416	.2195
FSEX	.2618	-.5432	-.3273	.0835	1	-.5242	-.0978
SAL77	-.5467	.4095	.4210	.1416	-.5242	1	.1260
SENIO	-.1845	-.2944	.0598	.2195	-.0978	.1260	1

SIGNIFICANCE OF CORRELATIONS (p-values of two-sided tests)

	AGE	LNBSAL	EDUC	TREXP	FSEX	SAL77	SENIO
AGE	NA	.537	.030	0	.011	0	.077
LNBSAL	.537	NA	0	0	0	0	.004
EDUC	.030	0	NA	.455	.001	0	.569
TREXP	0	0	.455	NA	.426	.176	.035
FSEX	.011	0	.001	.872	NA	0	.351
SAL77	0	0	0	0	0	NA	.229
SENIO	.077	.004	.569	.035	.351	.229	NA

First we analyze the correlation between the dependent variable (LNBSAL) and each the six predictor variables. The independent variable that has the highest simple correlation with the dependent variable is gender (FSEX) with the value of -.5432. As FSEX=1 for females, and 0 for males, the negative correlation between gender and starting salary shows that females tend to receive lower salaries than males. Obviously, the correlation does not enable us to claim that gender is the cause of the disparity. The p-value of the corresponding two-sided test of significance is reported as zero.

Another variable highly associated with log beginning salary is the transformed experience TREXP defined as 1/(EXPER+12). The correlation coefficient is equal to -.5077. Notice that the correlation is negative because TREXP and EXPER are negatively associated. The high correlation between the response variable LNBSAL and TREXP shows that our transformation of the EXPER variable to make its relationship with

LNBSAL approximately linear was successful. The correlation coefficient between LNBSAL and EXPER is 0.071.

Now we discuss the correlation among independent variables. There is an obvious high correlation between age and the transformed experience. The value of the correlation coefficient is -.6522. The p-value of the corresponding test of significance is reported as zero.