SEX DISCRIMINATION STUDY

## 4. Displaying and Describing the Beginning Salaries

Graphical displays and numerical summaries of the data can be very helpful for understanding the information contained in the data. We will obtain the boxplots and summary statistics for the beginning salaries.

4.1   Boxplot of Beginning Salaries on the Original Scale
4.2   Boxplot of Beginning Salaries on the Log Scale
**4.3   Summary Statistics for the Beginning Salaries**

4.1 SPSS produces the following boxplot of the beginning salaries on the original scale of measurement:

Boxplot of Beginning Salaries on the Original Scale
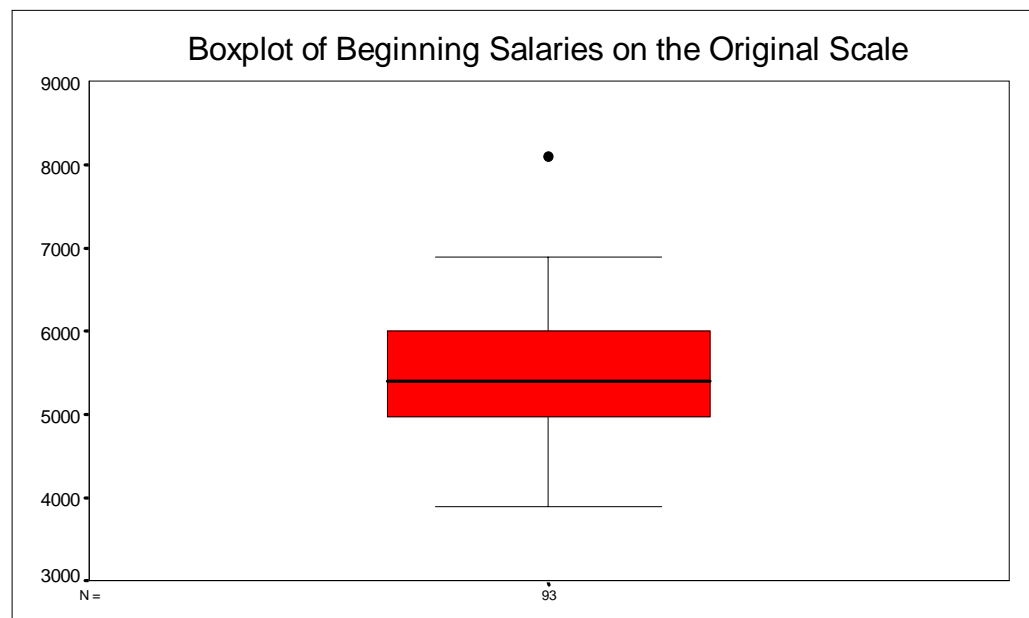
The median beginning salary indicated by the position of the horizontal line within the red box is about 5,500 dollars. The position of the median and the whiskers in the plot indicates that the distribution of the beginning salaries is approximately symmetrical with moderate length tails. There is only one outlier at the salary of over 8,000 dollars. The spread of the data, represented by the width of the box (interquartile range) is relatively small.

4.2 In one of the next sections, we will carry out a linear regression of the beginning salaries on a logarithmic scale. It will be important then to have some knowledge about the distribution of the log-transformed salaries. The following plot is a boxplot of the beginning salaries on the natural logarithm scale:

Boxplot of Beginning Salaries on the Log Scale
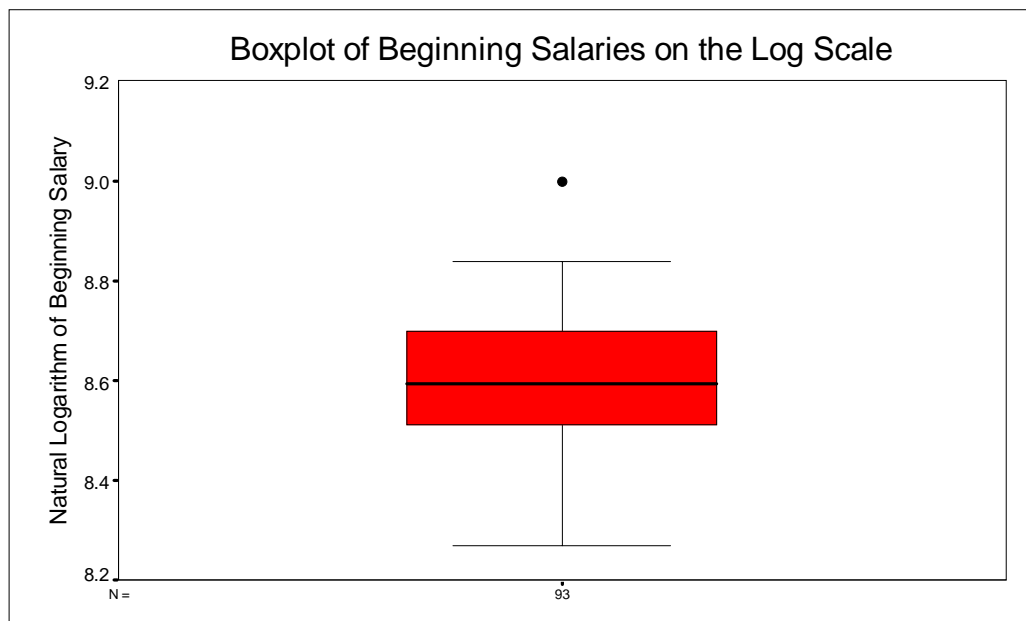
From just a quick glance, we see that the logarithm transformation has helped compress the numbers. In general, the logarithm transformation tends to pull in the long tail of the distribution on the right, but stretch it out on the left. In our case, the higher values have been pulled in, and the lower values are more spread out. The distribution is approximately symmetrical with one outlier.

4.3   Summary Statistics for the Beginning Salaries

The *Explore* procedure in SPSS can be used to obtain summary statistics for the beginning salaries for men and women:

	STATISTICS	Beginning Salary	
		MEN	**WOMEN**
CENTER	MEAN	5956.875	5138.852
	MEDIAN	6000.000	5220.000
	5% TRIM MEAN	5928.333	5136.995
	95% CI FOR MEAN	(5707.839, 6205.911)	(5000.585, 5277.120)
SPREAD	STANDARD DEV.	690.7333	539.8707
	STD ERROR	122.1056	69.1234
	VARIANCE	477112.5	291460.3
	IQR	825.000	600.000
	MINIMUM	4620.00	3900.00
	MAXIMUM	8100.00	6300.00
	RANGE	3480.00	2400.00
SHAPE	SKEWNESS	.7674	-.0780
	ST. ERROR SKEW	.4145	.3063
	KURTOSIS	1.7728	-.2863
	ST. ERROR KURT	.8094	.6038
COUNT		32	61

The numerical characteristics given in the above table confirm our earlier conclusions about the differences between the starting salaries of males and females. The measures of center (mean and median) provide strong evidence that males received larger starting salaries than the females. All measures of spread provided in the table show that the variability of male salaries was larger than the variability of female salaries.