

# SEX DISCRIMINATION PROBLEM

## 9. The Effect of Outliers on the Inferences

In Section 4 we found that there is only one outlier in our data, the salary of a male making \$8,100. What is the effect of the observation on the t-test and confidence intervals discussed in the previous section? In order to answer the question, we will carry out the statistical analysis with and without the outlying observation.

First we will carry out the test for the whole three-year period.

<b>t-tests for Independent Samples of FSEX</b>					
<b>Variable</b>	<b>Number of Cases</b>	<b>Mean</b>	<b>SD</b>	<b>SE of Mean</b>	
-----					
BSAL					
Females	61	5138.8525	539.871	69.123	
Males	31	5887.7419	578.773	103.951	
-----					
Mean Difference = -748.8895					
Levene's Test for Equality of Variances: F= .045 P= .832					
<b>t-test for Equality of Means</b>					
<b>Variances</b>	<b>t-value</b>	<b>df</b>	<b>2-Tail Sig</b>	<b>SE of Diff</b>	<b>95%CI for Diff</b>
-----					
Equal	-6.14	90	.000	122.007	(-991.278, -506.501)
Unequal	-6.00	56.84	.000	124.835	(-998.883, -498.896)
-----					

As you see, the value of the t-statistic has changed slightly. It was -6.29, it is -6.14 without the outlying observation. The p-value of the test remains the same, zero. In fact, the p-value has changed too, but it is still so small that is rounded to zero. However, the 95% confidence interval for the mean has changed more significantly, it was (-1076.25, -559.799) before, it is now (-991.278, -506.501).

The outlying observation \$8,100 was obtained in the first year of the study data. Thus we will also carry out the test for the time period.

### t-tests for Independent Samples of FSEX

Variable	Number of Cases	Mean	SD	SE of Mean
----------	-----------------	------	----	------------

BSAL

Females	20	5350.5000	437.162	97.752
Males	5	6300.0000	424.264	189.737

Mean Difference = -949.5000

Levene's Test for Equality of Variances: F= .154 P= .698

#### t-test for Equality of Means

Variances	t-value	df	2-Tail Sig	SE of Diff	95% CI for Diff
Equal	-4.37	23	.000	217.473	(-1399.38, -499.623)
Unequal	-4.45	6.31	.004	213.437	(-1465.57, -433.425)

The p-value of the t-test remains the same (zero), but the 95% confidence interval has changed significantly. The confidence interval with the outlier was (-1766.23, -751.039), without the outlier is (-1399.38, -499.623).