

# Contents

# 1 SLR Example

The data give the yields of cotton and irrigation levels in the Salt River Valley for different plots of land. Each plot was on Maricopa sandy loam soil. The variables are as follows:

- **Irrigation** The amount of irrigation water applied in feet per acre. This is the predictor variable.
- **Yield** The yield of Pima cotton in pounds per acre. This is the response variable.

Observation	Irrigation	Yield	Observation	Irrigation	Yield
1	1.8	260	8	1.5	280
2	1.9	370	9	1.5	230
3	2.5	450	10	1.2	180
4	1.4	160	11	1.3	220
5	1.3	90	12	1.8	180
6	2.1	440	13	3.5	400
7	2.3	380	14	3.5	650

Table 1: Add caption

Descriptive Statistics			
	Mean	Std. Deviation	N
Yield	306.4286	149.6461	14
Irrig	1.971429	0.754911	14

Next we are given the output from the correlation analysis and the regression ANOVA. The intercept and slope estimate are determined by examining the “coefficients”.

## 2 Correlation

Pearson’s correlation coefficient ( $r$ ) is a measure of the strength of the ‘linear’ relationship between two quantitative variables. A major assumption is the normal distribution of variables. If this assumption is invalid (for example, due to outliers), the non-parametric equivalent Spearman’s rank correlation should be used.

### 2.1 Formal test of Correlation

### 2.2 Lurking variables and Spurious Correlation

Spurious Correlations. Although you cannot prove causal relations based on correlation coefficients, you can still identify so-called spurious correlations; that is, correlations that are due mostly to the influences of “other” variables. For example, there is a correlation between the total amount of losses in a fire and the number of firemen that were putting out the fire; however, what this correlation does not indicate is that if you call fewer firemen then you would lower the losses. There is a third variable (the initial size of the fire) that influences both the amount

of losses and the number of firemen. If you "control" for this variable (e.g., consider only fires of a fixed size), then the correlation will either disappear or perhaps even change its sign. The main problem with spurious correlations is that we typically do not know what the "hidden" agent is. However, in cases when we know where to look, we can use partial correlations that control for (partial out) the influence of specified variables.

## **2.3 Simpson's Paradox**

## **2.4 Rank correlation**

Spearman's Rank correlation coefficient

## **2.5 Partial Correlation**

Partial correlation analysis involves studying the linear relationship between two variables after excluding the effect of one or more independent factors.