

# QUANTITATIVE METHODS IN PUBLIC POLICY

## Spring 2005

PROBLEM SET #4 – February 24, 2005

---

**STATISTICAL EVIDENCE OF DISCRIMINATION.** In 1973 approximately 44% of all men who applied to the University of California (Berkeley) were admitted while about 30% of female applicants were admitted. This raised concern that there was discrimination in the admissions process, and an investigation was launched. Admissions decisions at Berkeley were made by department, so data was collected on the number of male and female applicants to each department and the number of each gender that were admitted by each department. The data for the University's six largest departments is contained in the following table.

Major Dept.	# Males Applied	# Males Admitted	% Males Admitted	# Females Applied	# Females Admitted	% Females Admitted
A	825	512	0.62	108	89	0.82
B	560	353	0.63	25	17	0.68
C	325	120	0.37	593	202	0.34
D	417	138	0.33	375	202	0.54
E	191	53	0.28	393	94	0.24
F	373	22	0.06	341	24	0.07

Is the fact that males were admitted at a significantly higher percentage than females evidence of gender-based discrimination at Berkeley?

Look at the admissions decisions of the individual departments. Does this table suggest that Berkeley admitted a higher percentage of males than females? Does this data provide evidence that any of the six departments discriminated against women?

How is it possible that most departments admitted a greater percentage of females than males, yet the University as a whole admitted a much lower percentage of females?

Do you believe that Berkeley discriminated against female applicants?

**USING STATISTICS TO INFER CORRELATION AND CAUSATION.** Statisticians often use “ $r^2$ ” as one measure of the strength of the linear relationship between two variables. A scatter plot of two variables that are perfectly correlated ( $r^2=1$ ) would form a straight line. In nearly all cases, however, plotting two variables against each other will result in “clouds” of data points that may or may not resemble a line. The larger the value of  $r^2$ , the less the data points will stray from the line.

For the following scatter plots of hypothetical data, first state whether you think the  $r^2$  for the data is high (close to 1), low (close to 0) or moderate (in between). Second, state whether you believe there is actually a correlation between the variables. Finally, determine whether this data suggests a causal link between the two variables.

**Aggregate data from a sample of U.S. cities**

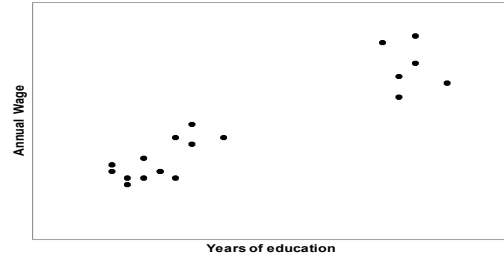


$r^2$ ?

Correlation?

Causation?

**Data from a sample of recent immigrants**

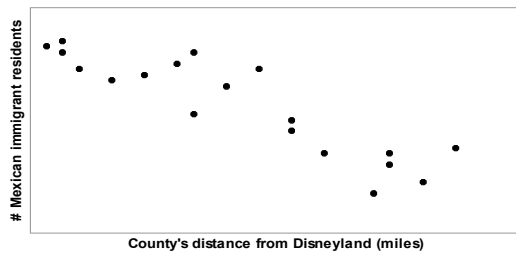


$r^2$ ?

Correlation?

Causation?

**Data from a sample of counties in California**

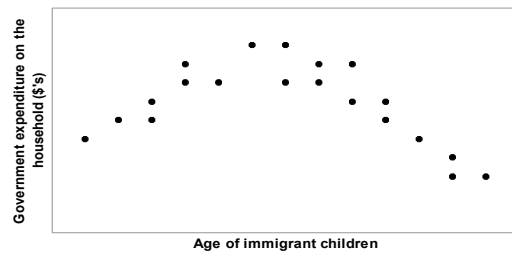


$r^2$ ?

Correlation?

Causation?

**Data from a sample of immigrant households**



$r^2$ ?

Correlation?

Causation?