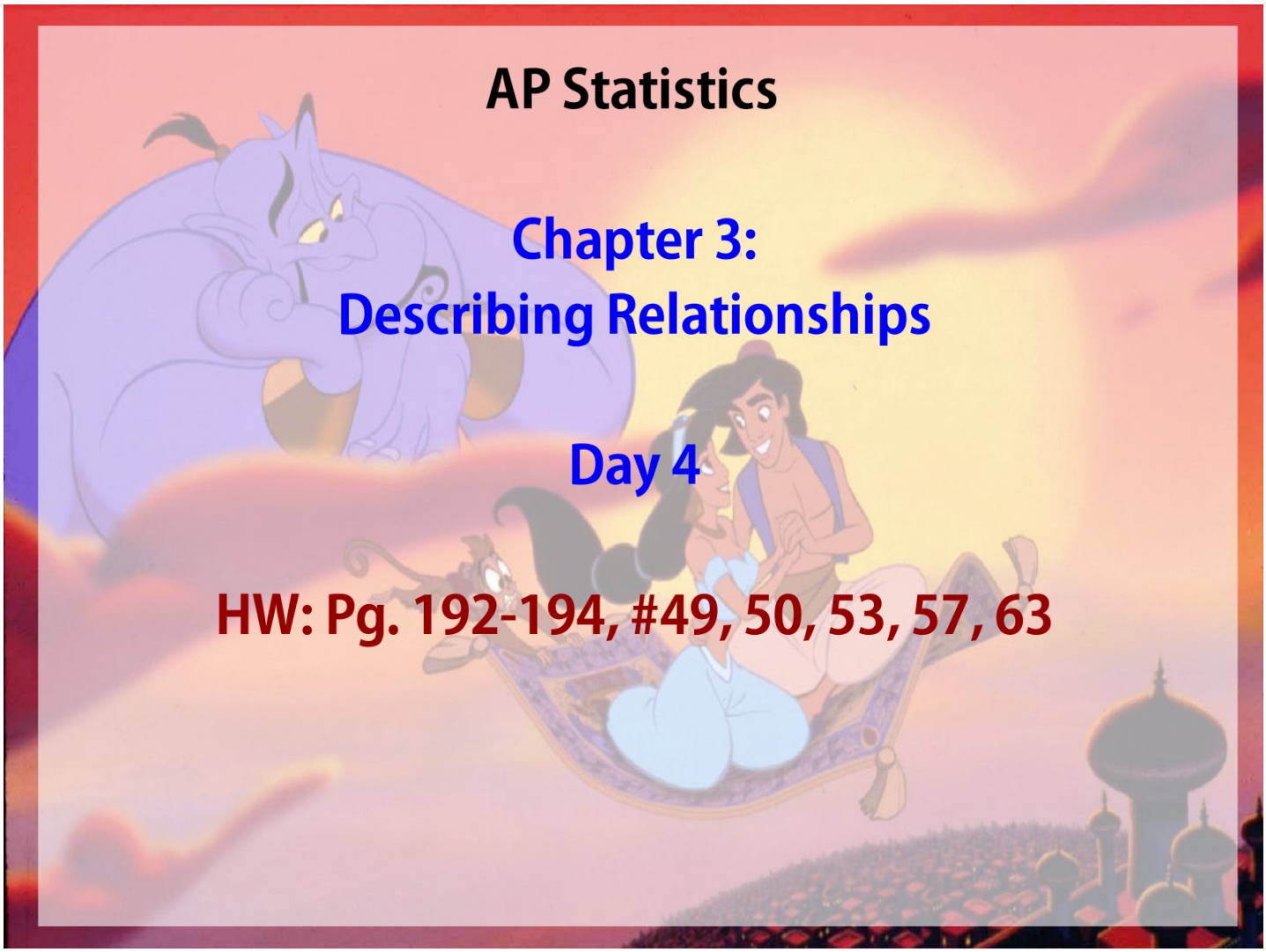


AP Statistics

Chapter 3:
Describing Relationships

Day 4

HW: Pg. 192-194, #49, 50, 53, 57, 63

Coefficient of Determination

- Is the percent of variation in the response variable that can be explained by its linear relationship with the explanatory variable.
- We can find what percentage of the predictions should be pretty accurate and what percentage of predictions might give us errors.
- It is denoted by r^2. It is just the correlation coefficient squared.
- Always between 0 and 1.

Example

A random sample of 10 office assistants hired within the last six months was selected from a large company. Each assistant's experience (in months) at the time of hire and annual starting salary (in thousands of dollars) were recorded in the table below.

Experience (in months)	Starting Salary (in $1,000 dollars)
5	28
12	34
2	24
0	19
2	24
10	32
5	25
1	20
10	29
5	23

Example

Compute the coefficient of determination. Interpret it.

$r^2 = .8747$

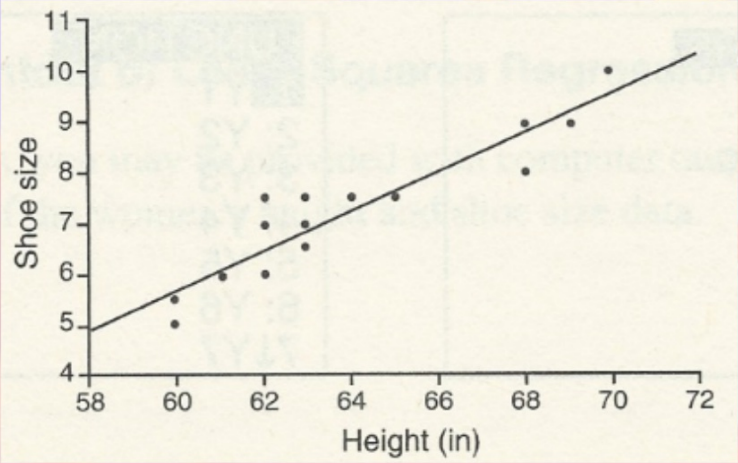
87.47% of the variability in starting salaries can be explained by its linear relation with the experience.

12.53% of the data can be attributed to others factors than experience.

Experience (in months)	Starting Salary (in $1,000 dollars)
5	28
12	34
2	24
0	19
2	24
10	32
5	25
1	20
10	29
5	23

Example

A scatterplot of the heights (in inches) and shoe sizes for 15 women is shown below, along with the least-squares regression line. The value of the correlation coefficient is .9225. Find and interpret the coefficient of determination.



$r^2 = (.9225)^2 = .8510$

85.10% of the variability in shoe sizes can be explained by its linear relationship with the height of woman.

Computer Printout of Least-Squares Regression

On the exam, you may be provided with computer output to interpret.

The output below is of the women's height and shoe size data.

Shoe = -17.7 + 0.390(63)
Shoe = 6.87

6.87
-.5274
6.3424

6.87
+.5276
7.3976

The regression equation is

$\hat{y}$ =

Shoesizes = – 17.7 + 0.390 Heights — least squares equation

Predictor	Coef	StDev	T	P
Constant	–17.693	2.760	–6.41	0.000
Heights	0.39000	0.04308	9.05	0.000

intercept — slope

r^2 =

S = 0.5276 — standard error about the regression line

R – Sq = 86.3% — coefficient of determination

R – Sq (adj) = 85.3%

Computer Printout of Least-Squares Regression

- The correlation coefficient r does not show up on the printout. It can be found by taking the square root of R-Sq.
- Do not forget that the sign of the correlation is the same as the slope.
- Ignore R-Sq(adj), as it is not used in this part of the AP Statistics curriculum.
- The value of s is the standard error about the regression line (average distance that the observed values fall from the regression line).

Example

What is the coefficient of correlation? What does it tell us about the data?

$r = .9290$ The data is strong, linear, and positive

Interpret the coefficient of determination.

$r^2 = 86.3\%$ 86.3% of the variability in shoe sizes can be explained by its linear relationship with the heights.

Explain what S = 0.5276 means.

The average error from the prediction line will be about .5276 of a shoe size.

The regression equation is

Shoesizes = – 17.7 + 0.390 Heights — least squares equation

Predictor	Coef	StDev	T	P
Constant	–17.693	2.760	–6.41	0.000
Heights	0.39000	0.04308	9.05	0.000

intercept — slope

S = 0.5276 — standard error about the regression line

R – Sq = 86.3% — coefficient of determination

R – Sq (adj) = 85.3%