

AP Statistics 4.1

1. The table below shows stopping distances in feet for a car tested 3 times at each of 5 speeds. We hope to create a model that predicts stopping distance from the speed of the car.

Speed (mph)	Stopping Distances (ft)
20	64, 62, 59
30	114, 118, 105
40	153, 171, 165
50	231, 203, 238
60	317, 321, 276

a. What should you do first? Do it!

b. Explain why a linear model is not appropriate.

c. Re-express the data to straighten the scatterplot.

d. Create an appropriate model.

2. Imagine that you have been put on charge of organizing a fishing tournament in which prizes will be awarded for the heaviest fish caught. You know that many of the fish caught during the tournament will be measured and released. You are also aware that using delicate scales to try to weigh a fish that is flopping around in a moving boat will probably not yield very accurate results. It would be much easier to measure the length of the fish while on board. You want a model that will predict the weight of the fish from its length. Here is data supplied by a nearby marine research laboratory for the Atlantic Ocean Rockfish:

Age (yr)	Length (cm)	Weight (g)
1	5.2	2
2	8.5	8
3	11.5	21
4	14.3	38
5	16.8	69
6	19.2	117
7	21.3	148
8	23.3	190
9	25.0	264
10	26.7	293
11	28.2	318
12	29.6	371
13	30.8	455
14	32.0	504
15	33.0	518
16	34.0	537
17	34.9	651
18	36.4	719
19	37.1	726
20	37.7	810

AP Statistics 4.2

Linear growth

Exponential growth

1. Consider these data for the growth of a bean plant.

Number of weeks	1	2	3	4	5	6	7	8	9	10
Height of plant in cm	0.5	0.7	1.1	1.5	2.4	3.5	5.0	7.4	10.0	17.0

a. Make a scatterplot; observe the pattern.

b. Transform the data by taking \ln of y . Complete the 3rd row of the table above and calculate the least squares regression line.

c. Use this model to predict the height in 12 weeks.

d. What kind of transformation linearizes exponential data?

2. Remember properties of logs?

$$\log(MN) =$$

$$\log\left(\frac{M}{N}\right) =$$

$$\log_b x^p =$$

3. Using the line of best fit that you calculated in part 1, write an **exponential model** for the plant growth.

4. Florida is one of the fastest-growing states in the United States. The population figures for each census year from 1830 through 2000 are given.

Year	Population
1830	34,730
1840	54,477
1850	87,445
1860	140,424
1870	187,748
1880	269,493
1890	391,422
1900	528,542
1910	752,619
1920	968,470
1930	1,468,211
1940	1,897,414
1950	2,771,305
1960	4,951,560
1970	6,789,443
1980	9,746,961
1990	12,938,071
2000	15,982,824

a.) Construct a scatterplot of the relationship between # of years since 1830 and the population. Comment on the pattern.

b.) Find a transformation that straightens the relationship. Sketch the scatterplot and residual plot of the transformed data.

c.) Find the LSRL, the coefficient of determination, and the correlation for the transformed data.

d.) Write a model to predict population based on the number of years since 1830. What is the rate of population growth?

e.) Predict the population of Florida in 2015.

AP Statistics 4.3

1. For each of these relationships first write the equation that relates x and y . Then use this equation to find a power of y that you could plot against x in order to get a linear plot.
 - a) y is the area of a circle, and x is the radius of the circle.
 - b) y is the volume of a block whose sides all have equal lengths, and x is the side length.
 - c) y is the volume of an 8-ft section of log with a circular cross section, and x is the diameter of the log's cross section.

Power Law models are of the form:

When is a power law model appropriate?

2. The table on the following page shows the brain weights and body weights for a collection of mammals. The goal is to establish the relationship of brain weight to body weight.
 - a) Construct a scatter plot of brain weight to body weight. Assuming that there is a power relationship here; can you guess what it is from the scatter plot? If y is written as a function of x to some power, should the power be greater than 1 or less than 1?
 - b) Plot $\log(\text{brain})$ versus $\log(\text{body})$. Describe the pattern of the plot.
 - c) Fit a line to the plot in part b. Write an equation relating y to x . Does your equation support your answer part a?

Species	Brain Weight (g)	Body Weight (kg)	Species	Brain Weight (g)	Body Weight (kg)
African elephant	5712	6654	Horse	655	521
African giant pouched rat	6.6	1	Human	1320	62
Arctic fox	44.5	3.385	Jaguar	157	100
Arctic ground squirrel	5.7	0.92	Kangaroo	56	35
Asian elephant	4603	2547	Lesser short-tailed shrew	0.14	0.005
Baboon	179.5	10.55	Little brown bat	0.25	0.01
Big brown bat	0.3	0.023	Mole rat	3	0.122
Cat	25.6	3.3	Mountain beaver	8.1	1.35
Chimpanzee	440	52.16	Mouse	0.4	0.023
Chinchilla	6.4	0.425	Musk shrew	0.33	0.048
Cow	423	465	Nine-banded armadillo	10.8	3.5
Desert hedgehog	2.4	0.55	North American opossum	6.3	1.7
Donkey	419	187.1	Owl monkey	15.5	0.48
Eastern American mole	1.2	0.075	Pig	180	192
European hedgehog	3.5	0.785	Rabbit	12.1	2.5
Giant armadillo	81	60	Raccoon	39.2	4.288
Giraffe	680	529	Rat	1.9	0.28
Goat	115	27.66	Red fox	50.4	4.235
Golden hamster	1	0.12	Rhesus monkey	179	6.8
Gorilla	406	207	Roe deer	98.2	14.83
Gray seal	325	85	Sheep	175	55.5
Gray wolf	119.5	36.33	Tree shrew	2.5	0.104
Ground squirrel	4	0.101	Water opossum	3.9	3.5
Guinea pig	5.5	1.04	Yellow-bellied marmot	17	4.05

3. *Chimp hunting parties.* After Jane Goodall discovered that chimpanzees are not solely vegetarian, much research began into the behavior of chimpanzees as hunters. Some animals hunt alone or in small groups, while others hunt in large groups. Where does the chimp fit in, and what is the success rate of chimps' hunting parties? Not surprisingly, the success of the hunt depends in part on the size of the hunting party. The table below gives some data on the number of chimps in a hunting party and the success rate of parties of that size.

- a) Plot the data to examine the relationship between size of hunting party and success. Describe the pattern.

Number of Chimps	Percentage Successful
1	20
2	30
3	28
4	42
5	40
6	58
7	45
8	62
9	65
10	63
12	75
13	75
14	78
15	75
16	82

- b) Transform the data to produce a model with better predicting ability than the simple linear one. Write the model.

- c) Investigate the residual plot from the model in part b. Are you happy with the fit of model?

4. How is the birthrate of countries related to their economic output? Do richer countries have higher birthrates, perhaps because families can afford more children? Or do poorer countries have higher birthrates, perhaps due to the need for family workers and a lack of education? Display 3.118 shows the birthrates (number of births per thousand population) and the GNP (in thousands of dollars per capita) for a selection of countries from around the world.

a) Construct a scatter plot and residual plot of these data and comment on the pattern you observe.

b) Transform the data and write a model to predict birthrates.

c) Interpret the slope and intercept of the model in the context of the data.

Country	Birthrate (per 1000)	GNP
Algeria	18.9	1.7
Argentina	17.7	4.2
Australia	12.7	19.5
Brazil	18.1	2.8
Canada	11.1	22.4
China	12.8	1.0
Colombia	22	1.8
Denmark	12	30.3
Egypt	24.9	1.5
France	12.7	22.2
Germany	8.8	22.7
India	23.8	0.5
Indonesia	21.9	0.7
Israel	18.9	16.0
Japan	9.6	34.0
Malaysia	24.2	3.5
Mexico	22.3	5.9
Nigeria	39.2	0.3
Pakistan	32.8	0.4
Philippines	26.8	1.0
Russia	9.2	2.1
South Africa	19.4	2.5
Spain	10	14.6
United Kingdom	11.1	25.5
United States	14.2	35.4

5. Exponential or Power? Determine whether an exponential or power model is appropriate.

a) Create an appropriate model for the relationship between the body weight of a mammal and its life span.

TABLE 4.4 Body weight and lifetime for several species of mammals

Species	Weight (kg)	Life span (years)	Species	Weight (kg)	Life span (years)
Baboon	32	20	Guinea pig	1	4
Beaver	25	5	Hippopotamus	1400	41
Cat, domestic	2.5	12	Horse	480	20
Chimpanzee	45	20	Lion	180	15
Dog	8.5	12	Mouse, house	0.024	3
Elephant	2800	35	Pig, domestic	190	10
Goat, domestic	30	8	Red fox	6	7
Gorilla	140	20	Sheep, domestic	30	12
Grizzly bear	250	25			

Source: G. A. Sacher and E. F. Staffelt, "Relation of gestation time to brain weight for placental mammals: implications for the theory of vertebrate growth," *American Naturalist*, 108 (1974), pp. 593–613. We found these data in F. L. Ramsey and D. W. Schafer, *The Statistical Sleuth: A Course in Methods of Data Analysis*, Duxbury, 1997.

b) How does mold grow? Below is the data from an investigation on the growth of the mold *Aspergillus nidulans*. Create an appropriate model for the relationship between the number of hours elapsed and the mean size of the mold colony.

Hours	Colony sizes			Mean
0	1.25	1.60	0.85	1.23
3	1.18	1.05	1.32	1.18
6	0.80	1.01	1.02	0.94
9	1.28	1.46	2.37	1.70
12	2.12	2.09	2.17	2.13
15	4.18	3.94	3.85	3.99
18	9.95	7.42	9.68	9.02
21	16.36	13.66	12.78	14.27
24	25.01	36.82	39.83	33.89
36	138.34	116.84	111.60	122.26

AP Statistics 4.4

1. In 2006, an electronic replay system debuted in both men's and women's professional tennis. Each player is allowed two unsuccessful challenges per set. Here are some data on the results of challenges made during the first few months of the new system.

	Successful	Unsuccessful
Men	201	288
Women	126	224

- a. Calculate the marginal distribution (in percents) for each of the two variables.
- b. Write a sentence describing what each marginal distribution tells you.
- c. Calculate the two conditional distributions of challenge results, given gender. Display your results in a table.

- d. Write a few sentences describing the relationship between gender and challenge results.

2. A researcher suspected a relationship between people's preference in movies and preference in pizza. A random sample of 100 people produced the following two-way table:

Favorite Movie	Ground		
	Pepperoni	beef	Mushrooms
<i>Jurassic Park</i>	20	5	10
<i>Lethal Weapon (I)</i>	8	15	12
<i>Gone with the Wind</i>	15	2	13

- a. Calculate the marginal distributions in percents for each variable.
- b. Compute (in percents) the conditional distribution of favorite movie among those who prefer ground beef topping. Show the distribution in a table.

c. Briefly describe your findings in words.

d. Compute (in percents) the conditional distribution of favorite pizza topping among those who chose *Gone with the Wind* as their favorite movie. Show the distribution in a table.

e. Briefly describe your findings in words.

3. The local newspaper reports the following data on patient mortality at two local hospitals.

	Lived	Died	Total		percent who lived
MERCY HOSPITAL	790	210			
COUNTY HOSPITAL	900	100			

a. Complete the table and give you impression of how the two hospitals compare.

b. Later that week, the paper releases some more details. This time, the **same information** is **disaggregated** (broken down into parts) with the following result:

For patients who were admitted in fair condition or better:

	Lived	Died	Total		percent who lived
MERCY HOSPITAL	580	10			
COUNTY HOSPITAL	860	30			

For patients who were admitted in poor condition or worse:

	Lived	Died	Total		percent who lived
MERCY HOSPITAL	210	200			
COUNTY HOSPITAL	40	70			

Complete these tables. Compare the percentages in the first table with those in the next two tables. Do you observe anything strange?

What is this result called?



What is a **lurking variable**?

When there is a strong **association** between two variables, list and define the three types of explanations:

- 1.
- 2.
- 3.

For each association described below state which explanation from above applies, then draw a diagram that shows a plausible relationship. Don't forget to **define** each variable used.

Ex. There is a positive correlation between the heights of mothers and daughters.	Ex. A soccer coach wanted to improve the team's playing ability, so he had them run two miles a day. At the same time the players decided to take vitamins. In two weeks the team was playing noticeably better.
1. Sharks ♥ Ice Cream! At a local beach, there is a strong positive correlation for ice cream sales and the number of shark attacks.	2. In one study, students at a large high school were offered an SAT preparation course, and SAT scores of students who completed the course were higher than scores of students who chose not to take the course.
3. People who smoke are 10 to 20 times more likely to develop lung cancer.	4. READING FIGHTS CAVITIES. The number of cavities in elementary school children is negatively related to the size of their vocabulary.
5. Allergic to Hot Chocolate? A strong correlation has been found in a certain city in the northeastern United States between weekly sales of hot chocolate and weekly sales of facial tissues. 78	6. STOP GLOBAL WARMING: BECOME A PIRATE! For average global temperature and number of pirates $r = -.93$

When CAN we imply causation?

Chapter 4 Review

Multiple Choice

1. A study covering many countries found a strong positive correlation between the life expectancy in a country and the percentage of households in the country with telephones. The best explanation of this observed correlation is that
- ☐ A. both life expectancy and telephone ownership are exhibiting a common response to the lurking variable of the country's socioeconomic condition.
 - ☐ B. telephone ownership and use is a major contributing cause of longer life.
 - ☐ C. in countries where life expectancy is high, the rate of telephone ownership tends to be low, and in countries where life expectancy is low, the rate of telephone ownership tends to be high.

2. Suppose we classify 315 randomly selected college students according to their general major field and their self-described political viewpoint. The table presents the results.

	Sciences	Business	Humanities	Social Sciences
Liberal	17	12	32	30
Moderate	33	40	23	20
Conservative	35	38	17	18

What percentage of liberals surveyed were humanities majors?

- ☐ A. 44.4%
 - ☐ B. 35.2%
 - ☐ C. 28.9%
3. Suppose that the variables x and y are related by the power function model $\hat{y} = 16x^2$. If you apply base 10 logarithms to both sides to transform the model into a linear model, then the resulting linear model will have which of the following forms?
- ☐ A. $\log \hat{y} = \log 16 + 2x$
 - ☐ B. $\log \hat{y} = 2 \log 16 + 2 \log x$.
 - ☐ C. $\log \hat{y} = \log 16 + 2 \log x$.
4. Suppose that a scatterplot of $\log \hat{y}$ versus x shows a strongly linear relationship with a value of r close to 1. Which of the following will then occur?
- ☐ A. A residual plot for the original data (x, y) will display a random scattering of points.
 - ☐ B. A scatterplot of y vs. x will resemble an exponential curve.
 - ☐ C. The correlation between y and x will be close to 0.

Free Response

The following table shows the federal debt for the years 1980 through 1991.

Year	Federal Debt (in trillions)
1980	0.909
1981	0.994
1982	1.1
1983	1.4
1984	1.6
1985	1.8
1986	2.1
1987	2.3
1988	2.6
1989	2.9
1990	3.2
1991	3.6

1. Determine if an exponential or power model is more appropriate for predicting Federal Debt based on the number of years since 1980. How did you decide?

2. Transform the data appropriately using log. Then perform a least square regression on the transformed data. Write the LSRL equation for the transformed data:

3. Write the model for the federal debt based on years since 1980.

4. Use your model to predict the national debt in the year 2000.

Below is data on the distance between the observer and a 100-W light bulb and the intensity of that light bulb.

Distance (m)	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
Intensity	0.297	0.252	0.205	0.175	0.153	0.135	0.114	0.102	0.093	0.083	0.073

5. Determine if an exponential or power model is more appropriate. How did you decide?

6. Transform the data appropriately using \ln . Then perform a least square regression on the transformed data. Write the LSRL equation for the transformed data:

7. Write the model.

8. Use your model to predict the intensity of a 100-W light bulb from 4 meters away.

A random selection of alumni from University and Woodbridge High Schools who graduated high school in 2005 were asked if they have student loans. Below is the results:

	Has Student Loans	Does Not Have Student Loans
Uni	52	41
Woodbridge	62	31

9.) Calculate the marginal distributions for each variable.

10.) Find the conditional distribution of having student loans or not having student loans given the high school.

11.) Write a sentence about your findings.

The same students from the table above were then broken up based on if they went to a Cal State or UC school. Below are the results:

Cal State

	Has Student Loans	Does Not Have Student Loans
Uni	11	26
Woodbridge	31	15

UC

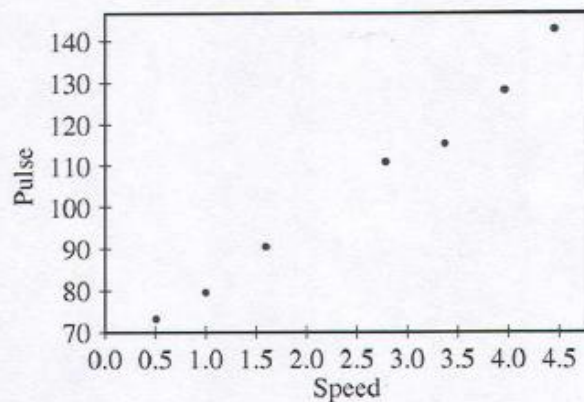
	Has Student Loans	Does Not Have Student Loans
Uni	41	15
Woodbridge	31	16

12.) Recalculate the conditional distributions you found in #10.

13.) Write a sentence about your findings.

AP STATISTICS 2005B

5. John believes that as he increases his walking speed, his pulse rate will increase. He wants to model this relationship. John records his pulse rate, in beats per minute (bpm), while walking at each of seven different speeds, in miles per hour (mph). A scatterplot and regression output are shown below.



Regression Analysis: Pulse Versus Speed

Predictor	Coef	SE Coef	T	P
Constant	63.457	2.387	26.58	0.000
Speed	16.2809	0.8192	19.88	0.000

S = 3.087

R-Sq = 98.7%

R-Sq (adj) = 98.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3763.2	3763.2	396.13	0.000
Residual	5	47.6	9.5		
Total	6	3810.9			

- Using the regression output, write the equation of the fitted regression line.
- Do your estimates of the slope and intercept parameters have meaningful interpretations in the context of this question? If so, provide interpretations in this context. If not, explain why not.
- John wants to provide a 98 percent confidence interval for the slope parameter in his final report. Compute the margin of error that John should use. Assume that conditions for inference are satisfied.