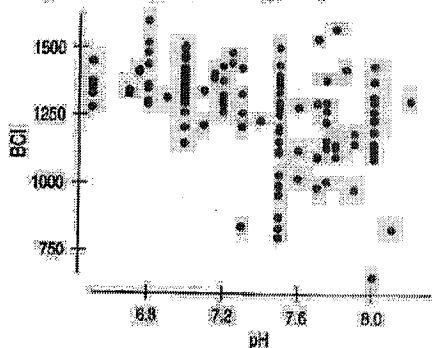**19. Acid rain.** Biologists studying the effects of acid rain on wildlife collected data from 163 streams in the Adirondack Mountains. They recorded the *pH* (acidity) of the water and the *BCI*, a measure of biological diversity. Following is a scatterplot of *BCI* against *pH*:

And here is part of the regression analysis:

Dependent variable is: BCI
R-squared = 27.1%
s = 140.4 with 163 − 2 = 161 degrees of freedom

Variable	Coefficient	SE(Coeff)
Intercept	2733.37	187.9
pH	−197.694	25.57

a) State the null and alternative hypotheses under investigation.
b) Assuming that the assumptions for regression inference are reasonable, find the *t* and P-value for the test.
c) State your conclusion.

a) $H_0: \beta = 0$ There is no linear relationship between PH and BCI. ← a little more precision than just saying "association"
   $H_A: \beta \neq 0$ There is a linear relationship between PH and BCI.

b) $t = \dfrac{b}{s_b} = \dfrac{-197.694}{25.57} = -7.73$ $p\text{-value} = (2)tcdf(-999, -7.73, 161) \approx 0$

c) with $\alpha = .05$, p-value ≈ 0 is low so we reject H_0.
   we do have sufficient statistical evidence to conclude that there is a linear relationship between PH and BCI.

22. Sales and profits. A business analyst was interested in the relationship between a company's sales and its profits. She collected data (in millions of dollars) from a random sample of Fortune 500 companies, and created the regression analysis and summary statistics shown. The assumptions for regression inference appeared to be satisfied.

	Profits	Sales
Count	79	79
Mean	209.839	4178.29
Variance	635,172	49,163,000
Std Dev	796.977	7011.63

Dependent variable is: Profits
R-squared = 66.2%   s = 466.2

Variable	Coefficient	SE(Coeff)
Intercept	−176.644	61.16
Sales	0.092498	0.0075

a) Is there a significant association between sales and profits? Test an appropriate hypothesis and state your conclusion in context.
b) Do you think that a company's sales serve as a useful predictor of their profits? Use the values of both R^2 and *s* in your explanation.

a) $H_0: \beta = 0$ No assoc. b/w sales & profits
   $H_A: \beta \neq 0$ Is an assoc. b/w sales & profits
   $t = \dfrac{b}{s_b} = \dfrac{.092498}{.0075} = 12.33$
   p-value $= tcdf(12.33, 999, 77) \approx 0$
   with $\alpha = .05$, p-value is low so reject H_0.
   We do have SSE to conclude that there is an assoc between sales & profits

b) $r^2 = .662$ so about 66% of the variability in profits is explained by the LSRL which relates it to sales.
   $s = 466.2$ is standard deviation of the residuals, so error in model (difference between actual & predicted profits for a given sales is ≈466 million which seems pretty high).
   I would say this is medium—useful—need to keep in mind the high error for any given sales level.

3. No opinion. Here's a regression of the percentage of respondents whose response to the question about voting for a woman president was "no opinion." We wonder if the percentage of the public who have no opinion on this issue has changed over the years. Assume that the conditions for inference are satisfied.

Dependent variable is: No Opinion
R-squared = 9.5%
s = 2.280 with 16 − 2 = 14 degrees of freedom

Variable	Coefficient	SE(Coeff)	t-ratio	P-value
Intercept	7.69262	2.445	3.15	0.0071
Year	−0.042708	0.0353	−1.21	0.2458
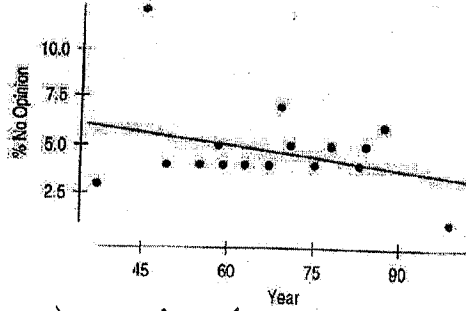
a) State the appropriate hypothesis for the slope.
b) Test your hypothesis and state your conclusion in the proper context.
c) On the next page is the scatterplot corresponding to the regression for No Opinion. How does the scatterplot change your opinion of the trend in "no opinion" responses? Do you think the true slope is negative? Does this change the conclusion of your hypothesis test of part b? Explain.

a) $H_0: \beta = 0$ (no change in % saying no opinion)
$H_A: \beta \neq 0$ (linear-rel (change) in % saying no opinion).

b) $t = -1.21$ p-value = .2458
High p-value, cannot reject H_0. There is no evidence of a change in % of people saying "no opinion" to a woman president.

c) LSRL seems influenced by outliers on end. w/o those, obviously no trend. (no change to what we stated).

15. Cereal. A healthy cereal should be low in both calories and sodium. Data for 77 cereals were examined and judged acceptable for inference. The 77 cereals had between 50 and 160 calories per serving and between 0 and 320 mg of sodium per serving. The regression analysis is shown.

Dependent variable is: Sodium
R-squared = 9.0%
s = 80.49 with 77 − 2 = 75 degrees of freedom

| Variable	Coefficient	SE(Coeff)	t-ratio	P-value
Intercept	21.4143	51.47	0.416	0.6786
Calories	1.29357	0.4738	2.73	0.0079

a) Is there an association between the number of calories and the sodium content of cereals? Explain.
b) Do you think this association is strong enough to be useful? Explain.

a) w/ $t = 2.73$, p-value = .0079
there is strong evidence that slope is not zero (there is a relationship)
$b = 1.29$ suggest higher calorie cereals also have higher sodium.

b) $r^2 = .09$ so only 9% of variation in sodium explained by calories (LSRL)
The regression LSRL not very useful.

27. Printers. In March 2002, *Consumer Reports* reviewed several models of inkjet printers. Shown are the speed of the printer (in pages per minute) and the cost per page printed. Is there evidence of an association between *speed* and *cost?* Test an appropriate hypothesis and state your conclusion.

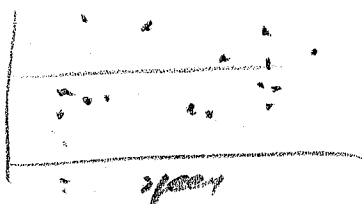| Speed (ppm)	Cost (cents/page)
4.6	12.0
5.5	8.5
4.5	6.2
3.8	3.4
4.6	2.6
3.7	4.0
4.7	5.8
4.7	8.1
4.0	9.4
3.1	14.9
1.9	2.6
2.2	4.3
1.8	4.6
2.0	14.8
2.0	4.4

Scatterplot

cost

Speed
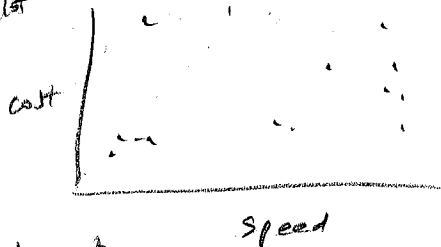
Straight enough? no pattern here, very scattered ✓
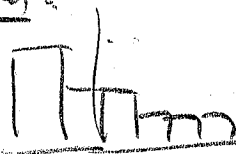
indep. no pattern in residuals;
residuals

fanning? no ✓

Nearly Normal?

skewed X

Should not proceed.