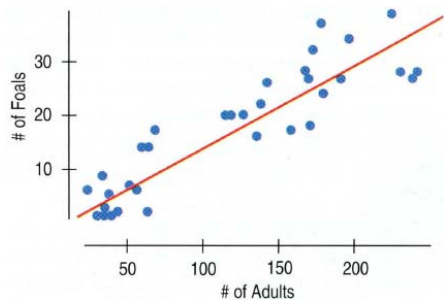
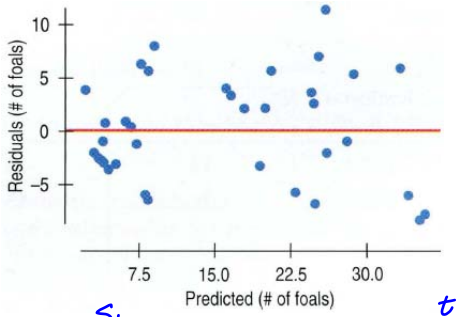
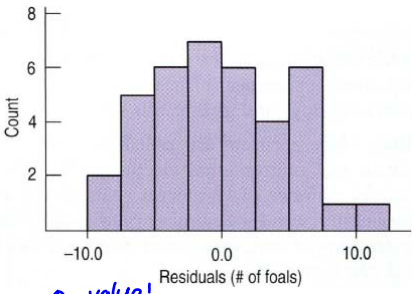
AP STATISTICS **Name:** PODUNK Mathletes **Per:** 9

Regression Slope Inference Practice

1. Large herds of wild horses can become a problem on some federal lands in the West. Researchers hoping to improve the management of these herds collected data to see if they could predict the number of foals that would be born based on the size of the current herd. Data from a random sample of 38 herds is summarized below:







b_1 S_{b_1} t p-value!

	Variable	Coefficient	S.E. of Coeff	T-ratio	Probability
x →	Adults	0.153969	0.0114	13.5	≤ 0.0001
	Constant	-1.57835	1.492	-1.06	0.2970
	S = 5.67	R-sq = 79%	R-sq adj = 80.5%		

a) Do the data present statistical evidence of a linear association between the size of herd and number of foals for these herds of wild horses?

H_0: There is NO linear association between #of adults (x) and # of foals (y) born.

H_A: There IS an linear association between # of adults and # of foals born.

t-test for regression slope

$t = 13.5$ $p \approx 0$ (Let's use $\alpha = 0.05$)

Since $p << \alpha$, we reject H_0.

there IS evidence of a linear association between #of adults per herd and # of foals born.

Conditions!

- Scatterplot of foals vs adults is fairly linear, with roughly equal variance in # of foals throughout the graph.
- The plot of residuals vs. foals has no clear pattern
- The histogram of residuals for # of foals is roughly normal.

b) Create and interpret a 95% confidence interval for the slope of the regression line relating herd size and number of foals born.

95% t-interval for slope, $df = 38 - 2 = 36$

$b_1 \pm t^* \cdot S_{b_1}$

$0.154 \pm 2.028\,(0.0114)$

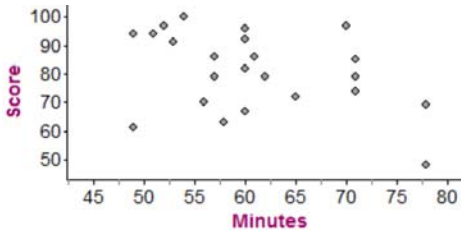
$(0.131, 0.177)$

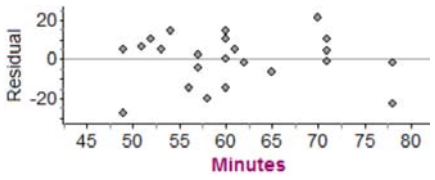
calculator: $t^*_{36} = 2.028$

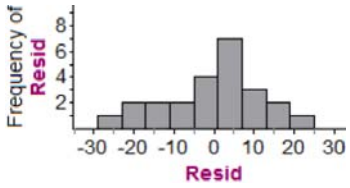
t-table: $t^*_{30} = 2.042$

We are 95% confident that for each increase of 1 adult in herd size, the model predicts a mean increase in # of foals born of between 0.131 and 0.177.

2. Can amount of time taken by a student on a test be used to predict the exam score? The test scores on a probability exam for 24 students at Podunk High School were recorded by their Statistics instructor, along with the amount of time (in minutes) that it took for each student to finish the exam. The data from the regression analysis is shown below.







Predictor	Coefficent	S.E. of Coef	T-statistic	P-value
Intercept	119.64	19.28	6.204	0.0000
Minutes	-0.6314	0.3133	-2.016	0.0562
S = 12.641	R-sq = 15.59%		R-sq adj = 11.75%	

a) Do the data present statistical evidence of a linear association between the number of minutes taken to finish the exam and exam score?

H_0: There is NO linear association between time taken to finish the exam and exam score.

H_A: There IS a linear association between...

t-test for slope, $\alpha = 0.05$

$t = -2.016$ $p = 0.0562$

Since $p > \alpha$, we FAIL to reject H_0.

We lack evidence of a linear association between # of minutes taken and exam score.

Conditions:

- The scatterplot for score vs. time is fairly linear (kinda? maybe?!)... ...with roughly equal variance in exam score throughout the scatterplot.
- The plot of residuals for score vs. minutes has no clear pattern.
- The histogram of residuals (score) has no major outliers

b) Create and interpret a 95% confidence interval for the slope of the regression line relating the amount of time taken to finish the exam and score on the exam?

$b_1 \pm t^* \cdot s_{b_1}$ $24 - 2$ $df = 22$

$-0.6314 \pm 2.074 * (0.3133)$

$(-1.2812, +0.0184)$

We are 95% confident that for each additional minute in exam time, the mean change in exam score is between -1.2812 and 0.0184 points.

(In other words, we lack evidence of any meaningful linear association...)