Math 10 Quiz Chapter 13 and 12: ANOVA & Linear Correlation/Regression

Note: Our quiz solutions start on the third page of this document.
The first two pages discuss the alligator data (which is real data).

The alligator data is a good example of why we can't predict for x values outside of the range of x values in our data.

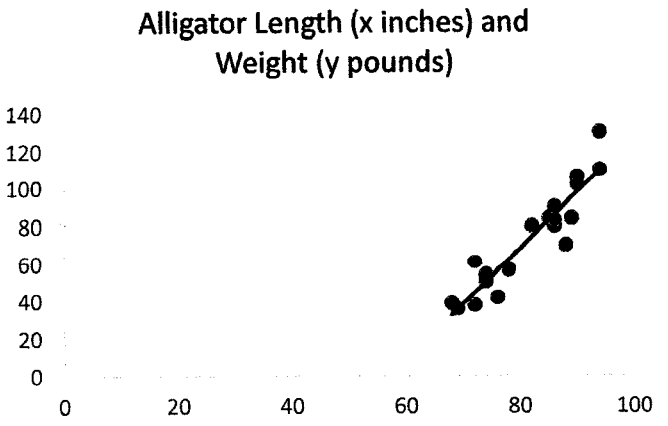
The 19 alligators in the quiz are a subsample of the sample that has 25 alligators. The 3 smallest and 3 largest alligators were removed from the original sample to create our subsample.

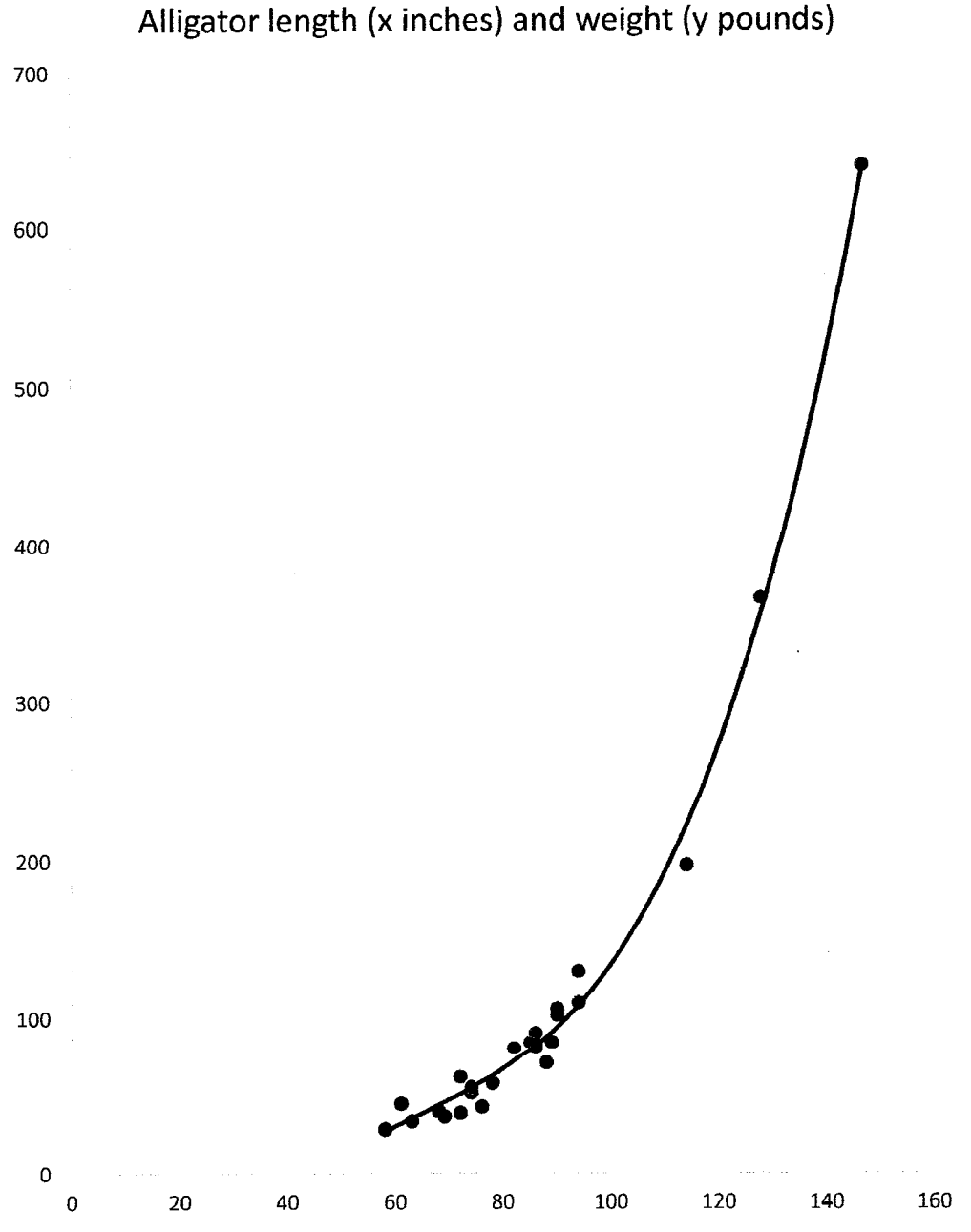
For the 19 alligators in our sample between 69 and 94 inches in length, the line DOES give a reasonable model to predict weight based on length.

However if we consider data outside that range, the linear relationship does not hold true. A curve would be a better fit.

The next page contains a graph of a line fit to our subsample of 19 alligators' data and a curve fit to the entire sample of 25 alligators' data. You can see the curve is a better fit for the entire data set. The curve is a cubic curve (third degree). This makes sense as weight is better related to volume than to length.

The 128 inch alligator referred to in one of the quiz questions actually weighs 366 pounds but our line would have predicted its weight as only 211 pounds. That's why we should not predict outside the range of the x values in our data – we have no evidence to assume the linear relationship extends beyond our data, as it in fact does not in this particular data set.


Alligator Length (x inches) and Weight (y pounds)
140
120
100
80
60
40
20
0
0
20
40
60
80
100


Alligator length (x inches) and weight (y pounds)
700
600
500
400
300
200
100
0
0
20
40
60
80
100
120
140
160

Name Last_____ First KEY Class Time: Pink Form B

Math 10 Quiz 7 Ch 12&13 Form B Winter 2019 20 points

This quiz contains two questions, one question for analysis of variance and one question for linear regression and correlation.

Question 1: 8 points; Environmental Studies

The levels of a certain type of pollutant are monitored at three different monitoring sites: Location A, Location B, Location C. Data were collected for a random sample of 6 days at each location.

Perform the appropriate hypothesis test to determine if the population average pollution levels are the same for all sites.

On the board I wrote "Use $\alpha = .05$"

Level of Pollutant		
Location A	Location B	Location C
8.9	7.4	7.2
8	8.3	6
8.3	7.1	6.8
7.9	6.5	7.7
7.4	7.6	6.7
8.1	7.5	7

For this sample data: $\bar{\bar{x}} = 7.467$ $\bar{x}_A = 8.1$ $\bar{x}_B = 7.4$ $\bar{x}_C = 6.9$

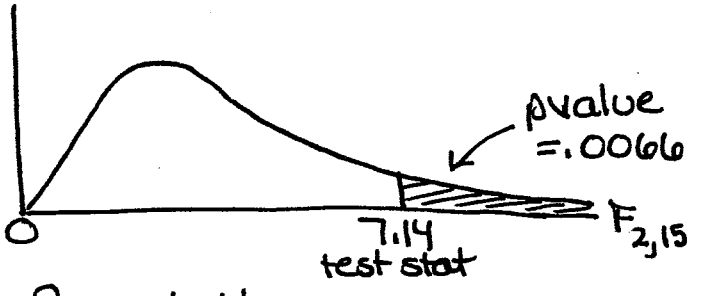
Null Hypothesis: The population mean pollution levels at all 3 locations are equal ($\mu_A = \mu_B = \mu_C$)

Alternate Hypothesis: The population mean pollution levels at some (or at least one) locations are different.

Test Statistic = 7.14 p-value = .00666

What probability distribution is used to find p-value? (Include degrees of freedom): F 2, 15

Draw the graph for this hypothesis test – label the test statistic on the graph – shade and label the p-value.



pvalue $< \alpha$
.00666 $< .05$

Decision: Reject H_0

Conclusion: At a 5% level of significance, we can conclude that the population mean pollution levels at some (or at least one) of these locations are different from each other.

Note: If we had computer output to analyze the differences between the means we'd find that only means μ_A, μ_C are different: $\mu_A \neq \mu_C$.

KEY Form B Pink

Question 2: 12 points: Environmental Studies

Biologists need to be able to estimate the weight of alligators in the wild, but weighing an alligator is a difficult and hazardous task. Length can be estimated from a photograph, which is much safer than trying to weigh an alligator.

So biologists would like a model that would allow them to estimate the weight of an alligator in the wild if they know its length. A regression line can be used to model the relationship between length and weight.

The data show lengths (in inches) and weights (in pounds) of a sample of 19 randomly selected alligators that were measured and weighed.

The calculator output is given to you. Use it! (Do not put the data into your calculator.)

X=Length	94	72	72	74	74	86	86	86	94	69	85	82	88	90	**89**	68	76	90	78
Y=Weight	130	38	61	54	51	80	90	83	110	36	84	80	70	106	**84**	39	42	102	57

Y=a+bx
$\beta \neq 0$ and $\rho \neq 0$
t=8.22
p=8.443E–9
df=17
a=–168.8
b=2.97
s=10.27
r^2=0.865
r=0.93

a. [3 *points*] **Find the correlation coefficient and write out the complete hypothesis test to determine whether it is significant.**

$H_0: \rho = 0$ $H_A: \rho \neq 0$

$r = .93$

pvalue $< \alpha$

$.000000008 < .05$ Reject H_0

Conclusion: Correlation Coefficient IS significant

b. [1 *point*] **Prediction:** *Show work; round to values to 1 decimal place*

For an alligator that is 72 inches long, what is the weight predicted by the line?

$\hat{y} = -168.8 + 2.97(72) = 45.04 \approx 45.0$ pounds

c. [2 *points*] **Interpretation of Coefficient of Determination:** Write a sentence in the context of this situation that interprets (explains) what the Coefficient of Determination tells us for this problem, and use its value in your interpretation. *Round to values to 1 decimal place as a percent (3 places in decimal form)*

86.5% of the variation in weight can be explained by the variation in length,

d. [2 *points*] **Interpretation of Slope:** Write a sentence in the context of this situation that interprets (explains) what the slope of the best fit line tells us, and use its value in your interpretation.

For every 1 inch increase in length, an alligator's weight increases by 2.97 pounds, on average.

e. [3 *points*] **"Error Analysis":** **For a length of 89 inches:** Find the residual (or "error") $y - \hat{y}$

Show work ; round to values to 1 decimal place

$\hat{y} = -168.8 + 2.97(89) = 95.5$ $y - \hat{y} = 84 - 95.5 = -11.5$

Is the observed data value above or below the best fit line? Below

Does the value predicted by the line overestimate or underestimate the observed (actual empirical) data value?

Overestimate

f. [1 *point*] **Prediction:** Suppose that you were asked to use the line to predict weight for an alligator that is 128 inches in length. What should be your response?

We should not use the line to predict the weight of a 128 inch alligator. The length of 128 inches is not within the range of x values (lengths) in the data

↑ (not between MINIMUM & MAXIMUM x values in the data)