Answers

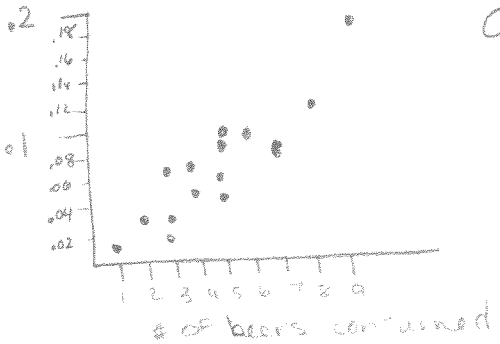
Linear Regression Data Study- Beer and BAC

How much alcohol can one consume before one's blood alcohol content (BAC) is above the legal limit? An undergraduate statistics project conducted at The Ohio State University in Columbus, Ohio explored the relationship between BAC and other variables such as amount of alcohol consumed, weight, gender and age.

Protocol The experiment took place in February of 1986 at a student dormitory. Sixteen students volunteered to be the subjects in the experiment. Each student blew into a breathalyzer to indicate that his or her initial BAC was zero. The number (between 1 and 9) of 12 ounce beers to be drunk was assigned to each of the subjects by drawing tickets from a bowl. Thirty minutes after consuming their final beer, students had their BAC measured by a police officer of the OSU police department. The officer also administered a road sobriety test before and after the alcohol consumption. This involved performing four simple tasks, graded on a scale of 1 to 10 (ten being a perfect rating), demonstrating coordination: balancing on one foot, touching the tip of one's nose with a forefinger, placing one's head back with one's eyes closed, and walking heel to toe. The police officer was not aware of how much alcohol each subject had consumed.

ID_OSU	Gend_OSU	Wght_OSU	Beers	BAC	1st-Sobr	2nd-Sobr
1	female	132	5	0.1	10	6
2	female	128	2	0.03	9.5	9.25
3	female	110	9	0.19	9.75	4.75
4	male	192	8	0.12	10	7.5
5	male	172	3	0.04	10	9.75
6	female	250	7	0.095	9.5	6.5
7	female	125	3	0.07	9.5	7
8	male	175	5	0.06	9.75	8.75
9	female	175	3	0.02	9.5	6
10	male	275	5	0.05	9.75	8.5
11	female	130	4	0.07	9.5	8.5
12	male	168	6	0.1	9.5	7.75
13	female	128	5	0.085	9.75	8.25
14	male	246	7	0.09	10	7.75
15	male	164	1	0.01	9.5	9.5
16	male	175	4	0.05	10	9

(a) Make a scatterplot of BAC versus number of beers consumed. Describe what you see.



a strong, positive, linear association between beer consumed & BAC.

(b) Use you calculator to compute two-variable statistics. Record the summary statistics below.

Beer Cons. $n=16$
$\bar{x} = 4.8$
$S_x = 2.2$

BAC $n=16$
$\bar{y} = .074$
$S_y = .044$

(d) How strong is the *linear* association? Find the correlation $r = \frac{1}{n-1}\sum\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right)$.

r = .894

Relatively strong since close to 1.

As beer consumptions increases, BAC increases

(e) Use your calculator to find a linear regression model for predicting BAC from number of beers consumed. Interpret the slope and y-intercept of the model in context.

$\widehat{BAC} = -.0127 + .01796(\text{beer consumed})$

Regression Analysis: BAC versus Number of Beers

Predictor	Coef	SE Coef	T	P
Constant	-0.01270	0.01264	-1.00	0.332
Number of Beers	0.017964	0.002402	7.48	0.000

S = 0.0204410 R-Sq = 80.0% R-Sq(adj) = 78.6%

Slope = .01796

For each additional beer consumed, the predicted BAC increases by .01796.

Yint = -.0127

For 0 beers consumed, predicted BAC = -.0127. This is not reasonable.

(f) Use your model from (e) to predict the BAC for a student who consumes 6 beers.

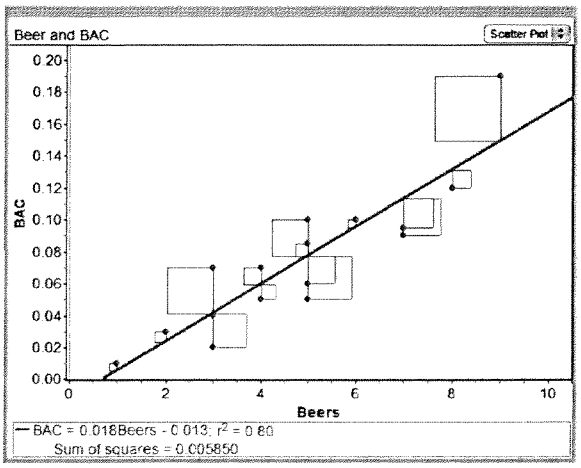
$\widehat{BAC} = -.0127 + .01796(6) = .095$

(g) Find and interpret the residual for the student in the study who consumed 6 beers.

residual = .1 − .095 = .005

The actual BAC was .005 more than predicted by LSRL for the person who consumed 6 beers.

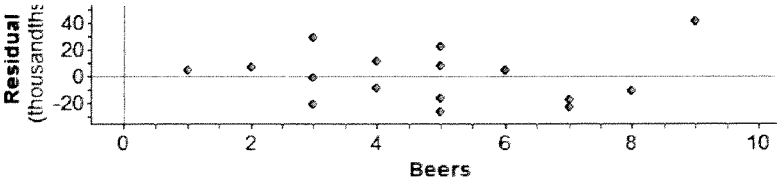
(h) The graph below displays the same linear model that you obtained from your calculator. Explain why this model is known as the *least-squares regression line*.



minimizes the sum of the squared residuals (b/c Σ residuals in 0).

(i) The graph below plots the residual for each of the 16 original data points versus the number of beers consumed. It is known as a *residual plot*. Describe what you see.

a random scatter



(j) Is a linear model appropriate for describing the relationship between number of beers consumed and BAC? Justify your answer.

Yes, scatterplot is linear & residual plot is random

How well does the linear model do at predicting BAC from number of beers?

(k) How large is a typical prediction error? Compute $s = \sqrt{\frac{\sum residuals^2}{n-2}}$ and interpret the result.

$s = \sqrt{\frac{.005850}{14}} = .02$ The predicted BAC from the LSRL is typically off by .02

(l) What does r^2 tell us?

$r^2 = .8$

80% of the variation in BAC is explained by the LSRL of BAC on beer consumed