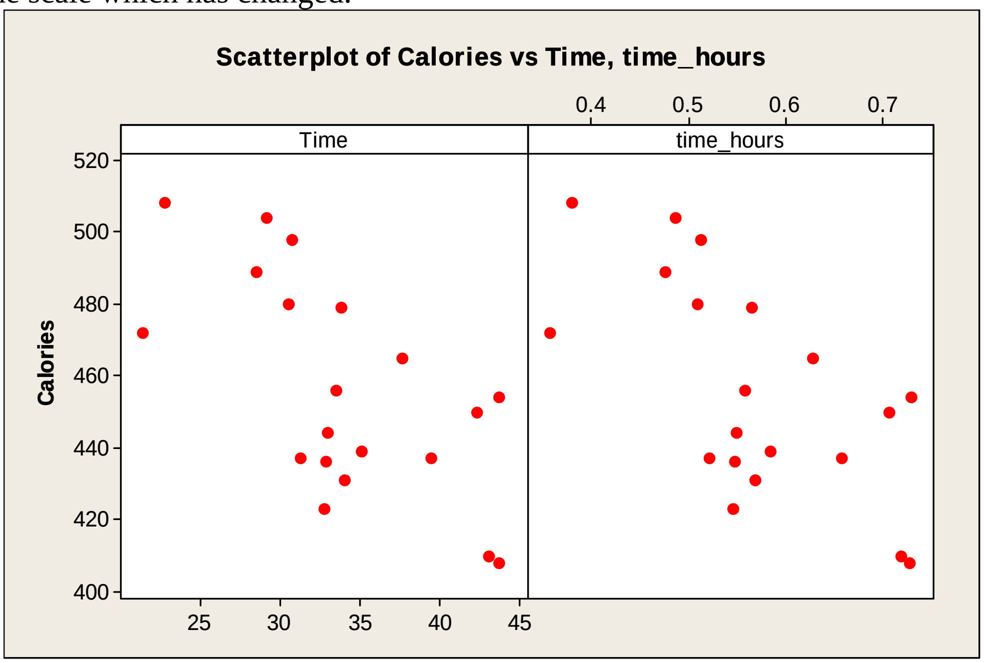
MATH 1107

Practice Test #2 – Answer key to selected problems

Problem 12 page 160. The plots are matched as follows: (a): -0.977, (b): 0.736, (c): 0.951, (d): -0.021.

Problem 15 page 161. a) The correlation between the variables time and calories is $r = -0.649$.

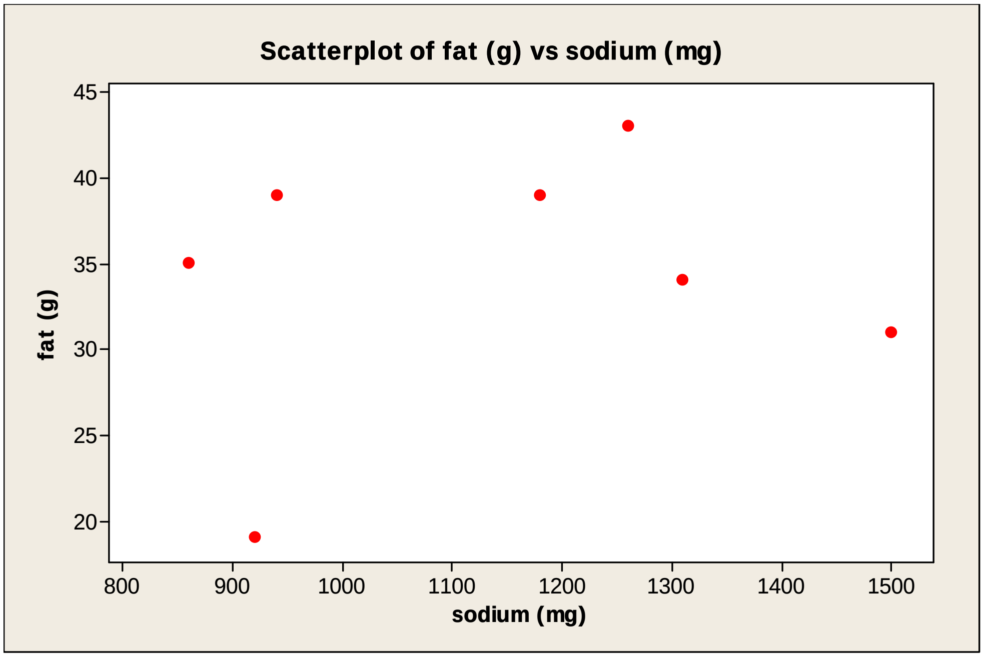
b) If time were recorded in hours rather than minutes the correlation would stay the same. See scatterplots below showing the association between variables when time is measured in minutes and hours, respectively. As it can be seen, the plots are identical, except for the time scale which has changed.



c) There is a moderately strong linear association between the time toddlers spent at lunch and the amount of calories they ingest. The more time they stay at the table the less the amount of calories (i.e. the association is negative).

d) We cannot conclude that the time spent at the table can entirely explain the average calorie intake, although an association can be observed. There may be some other factors contributing to the calorie intake.

Problem 19. Below is a scatterplot of "fat" versus "sodium". The plot shows little to no association between the two variables. Notice that the variables can be swapped when choosing the explanatory and the response variables. The corrwlation coefficient (computed with Minitab) is $r = 0.199$.



Problem 29. a) A correlation of -0.772 is actually quite strong, provided the scatterplot indicates a linear association.
b) Correlation is used in the wrong context since "continent" is not a quantitative variable.

Problem 2 page 185.

We use the formulae: $b_1 = r\frac{s_y}{s_x}$, $b_0 = \overline{y} - b_1\overline{x}$. We obtain:

a) $\hat{y} = 27 - 0.3x$; b) $\hat{y} = 10 + 0.5x$;

c) $b_1 = 15 = r\frac{s_y}{s_x} = r\frac{15}{0.8}$, so $r = 0.8$, $b_0 = -10 = \overline{y} - b_1\overline{x} = 50 - 15\overline{x}$, so $\overline{x} = 4$;

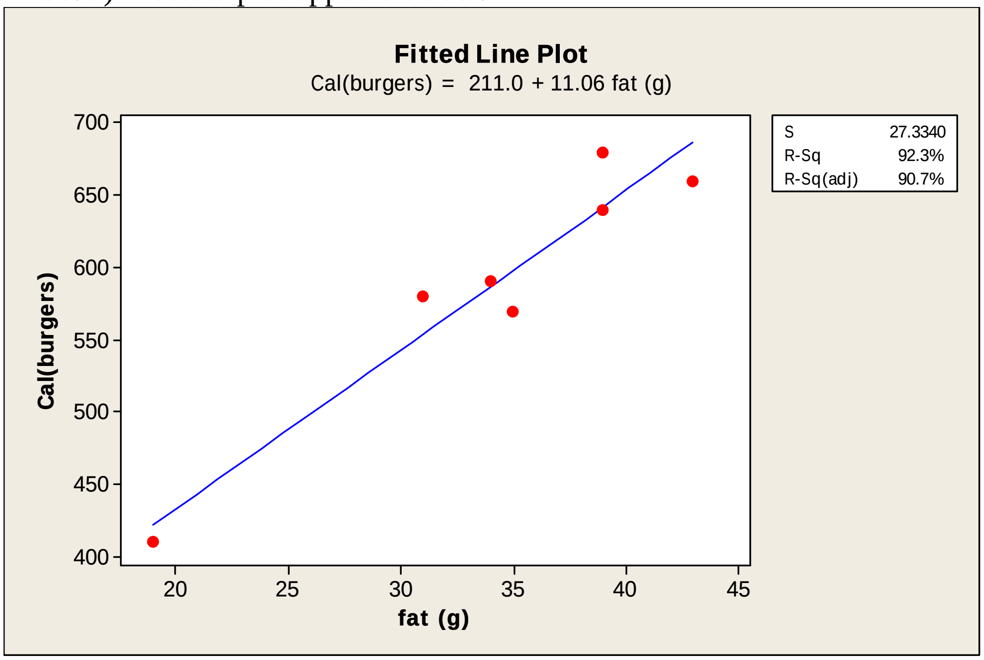
d) $b_1 = -2 = r\frac{s_y}{s_x} = -0.6\frac{4}{s_x}$, so $s_x = 1.2$, $b_0 = 30 = \overline{y} - b_1\overline{x} = 18 + 2\overline{x}$, so $\overline{x} = 6$

Problem 3 page 185. Only the residual plot (a) indicates that the linear model is appropriate. The points are scattered chaotically without any particular pattern.
(b) The residual plot has a curved shape indicating that the scatterplot is curved as well.
(c) Since the plot opens like a funnel it shows that the linear model may fit well some of the data (where the spread is small) but has greater errors for other values (where the spread is big).

Problem 15 page 186. a) The scatterplot shows a pretty strong, positive, linear association between the tar and nicotine contents. The residual plot, by the lack of pattern, also shows that the linear model is appropriate.
b) 92.4% of the variation in the nicotine content is explained by the linear dependence on the amount of tar in cigarettes.

Problem 19. a) From the output shown in problem 15 we have: $\hat{y} = 0.065x + 0.154$.
b) For a cigarette with 4 mg of tar, the predicted amount of nicotine is
$\hat{y} = 0.065 \cdot 4 + 0.154 = 0.414$.
c) For every milligram increase in tar there is an approximate increase of 0.065 mg in the average nicotine content.
d) If a cigarette has no tar content it will still have 0.154 mg of nicotine.
e) From the definition of the residual, we have $e = y - \hat{y} = y - (0.065 \cdot 7 + 0.154) = -0.5$,
and therefore $y = (0.065 \cdot 7 + 0.154) - 0.5 = 0.109mg$.

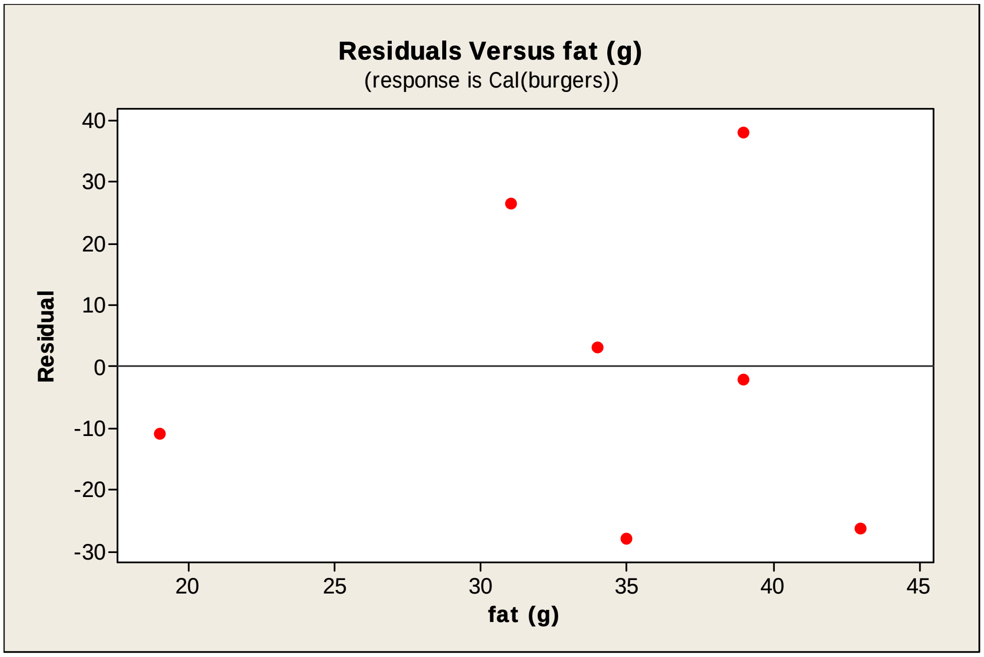
Problem 35. a) A scatterplot appears below.



We can see a very strong linear association. The association is also positive: with the increase of fat we observe an increase in the calorie content of burgers.
b) 92.3% of the variation in the calorie content of burgers is explained by the dependence on fat.
c) The equation of regression line appears in the plot above:

Cal(burgers) = 211.0 + 11.06 fat (g)

d) The plot of residuals versus the fat content appears below. Since there is no apparent pattern, the linear model seems to be appropriate.
e) Even a fat free burger contains approximately 211 calories.
f) For each additional gram of fat the average number of calories increases by approximately 11.
g) $e = y - \hat{y} = y - (211 + 11.06 \cdot 28) = 33$, so $y = 553.68$ calories.



Problem 13 page 242.

a) We look for the strongest correlation, and the table shows that weight has a -0.903 correlation with MPG.

b) The negative correlation indicates that the more the car weighs the worse (lower) its MPG will be.

c) For the relationship between weights and MPG, $-0.903^2 = 0.8154$, 81.5% of the variation in MPG is explained by the linear dependence on the weight of the car.

Chapter 14. Problem 9. We need to check that the numbers are all nonnegative and they add to 1. Therefore only a), b) and d) can be probability assignments.

Problem 11. a) Since the sum of all probabilities must be 1, it follows that $P(no\ repair) = 1-(.17+.07+.04) = 0.72$.

b) $P(no\ more\ than\ one\ repair) = P(0\ or\ 1) = 0.89$.

c) $P(some\ repairs) = 1 - P(none) = 0.28$

Problem 13. If we assume (which is perfectly conceivable) that the two cars will need repairs independent of each other, we have:

a) $P(neither\ needs\ repair) = (P(none))^2 = 0.72^2 = 0.5184$.

b) $P(both\ need\ repair) = (P(some\ repair))^2 = 0.28^2 = 0.0784$.

c) $P(at\ least\ one\ car\ needs\ repair) = 1 - P(none\ needs\ repair) = 1 - 0.72^2 = 0.4816$.

Problem 24. a) $P(AB)=1-(.45+.40+.11)=0.04$
$P(A\ or\ B)=.40+.11=.51.$
$P(not\ O)=1-P(O)=.55$.
b) $P(4\ type\ O)=.45^4=0.041$
$P(none\ is\ AB)=P(not\ AB)^4=.96^4=.8493$
$P(not\ all\ A)=1-P(all\ A)=1-.40^4=.9744$
$P(at\ least\ one\ B)=1-P(none\ B)=1-.89^4=.3726$.

Problem 31. $P(all\ R)=.29^3=.024$
$P(no\ D)=1-0.37=.63$
$P(at\ least\ one\ Indep)=1-P(no\ Indep)=1-(1-.77)^3=.988$.

Chapter 15. Problem 5.
a) $P(TV\ and\ no\ refrig)=.52-.21=.31$.
b) $P(TV\ or\ refrig\ not\ both)=.52+.38-2\cdot.21=.48$
c) $P(neither)=1-P(TV\ or\ refrig)=1-(.52+.38-.21)=.31$

Problem 6. Since half of the college graduates are married, it means that 22% of the college graduates are married. Since in all 44% are college graduates it means 22% are not married. Similarly, since in all 72% are married and 22% are college graduates, it follows that 50% of them have no college degree.
Now let's put the information in a table:

	Married	Single
College graduates	0.22	0.22
No college	0.50	0.06

The only entry that is empty is the one for workers who are not married and have no college degree. Since the percentages must add to 100% we must have 6% of workers in that category. The answer for (a) is 0.06, or 6%.
b) 50% of workers are married but have no college.
c) $P(\text{married or college graduate})=0.72+0.22=0.94$, or if you use the addition rule for probabilities of events, $P(\text{married or college graduate})=0.72+0.44-0.22=0.94$

Problem 7. a) $P(Laura\ Bush)=\frac{518}{1005}=.5154$

b) $P(under\ 50)=\frac{217+416}{1005}=.6299$

c) $P(under\ 50\ and\ Hillary)=\frac{135+158}{1005}=.2915$

d) $P(under\ 50\ or\ Hillary)=\frac{135+77+5+158+237+21+79+65}{1005}=.7731$

1. There are 3 outcomes that lead to a sum of 4: (1,3), (2,2), (3,1) and three that lead to a sum of 10: (4,6), (5,5), (6,4). Therefore $P(4\ or 10) = \frac{6}{36} = \frac{1}{6}$.
2. There are 12 face cards (at least according to some, J, Q, and K), out of which 3 are hearts, and 13 hearts, so $P(\mathit{face\ or\ heart}) = \frac{22}{52}$.
3. There are in all 135 vehicles.

A. P(no more than 6 years old)= $\frac{20+15+35+30}{135} = \frac{100}{135}$.

B. P(truck or 4-6 years old)= $\frac{15+30+25+35}{135} = \frac{105}{135}$.