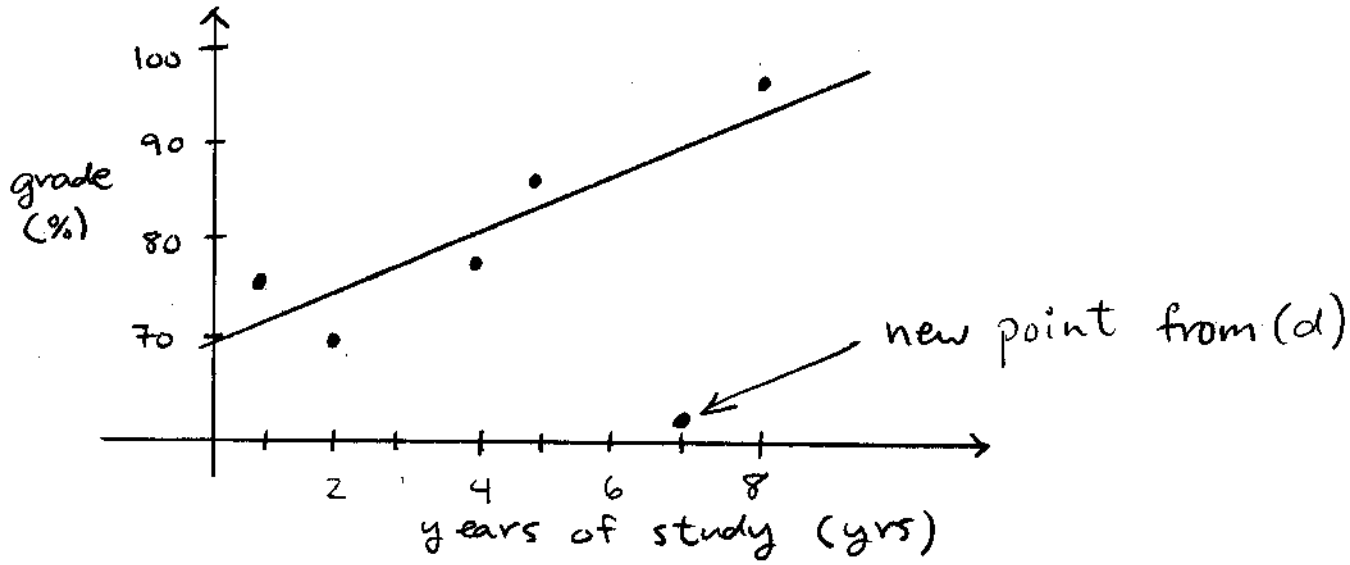
Question 1
The following is data for 5 individuals showing the number of years of piano study and the final grade in grade 12 mathematics:

grade	75	70	77	86	96
yrs	1	2	4	5	8

(a)[3 points] Sketch a scatterplot using years of study as the explanatory variable and use a straight edge to draw a regression line. Clearly label your axes and state units.



(b)[2 points] Describe the form, direction and strength of the relationship.

Relationship (or association) is strong, linear & positive.

(c)[2 points] The correlation r in this case is one of the following values:

$$-1.5, -1, -0.8, -0.1, 0.1, 0.8, 1, 1.5$$

Which is it, and why? $\boxed{r = 0.8}$

Association is positive, so $r > 0$, and $r \leq 1$ since r must be between -1 & 1 inclusive. Association is strongly, though not perfectly, linear, so $r \neq 0.1$ and $r \neq 1.0$. This leaves $r = 0.8$.

(d)[3 points] Data for a new individual is added: years of study = 7 and grade = 61. Do you expect r to increase, decrease or stay the same? Explain.

r will decrease since strength of linear association has decreased.

Question 2

Here's data for 10 Malaspina students showing, for each student, the average number of hours per week spent watching television and the grade point average (GPA) at the end of the year:

GPA	7.9	6	5	9.5	5.5	8.5	7.3	8	9.3	8.8
hours	14	10	20	7	25	9	15	13	4	21

Considering hours of television as the explanatory variable x, the least squares regression line for this data is

$$y = -0.1495x + 9.643\ ,$$

and the correlation for the data is $r = -0.6241$.

(a)[2 points] Predict the GPA if a student spends 12 hours per week watching television.

If $x = 12$ hours,

GPA $y = -0.1495(12) + 9.643$

$\doteq \boxed{7.8}$

(b)[3 points] What is the largest value of the explanatory variable beyond which the least squares model predicts invalid GPAs.

GPA cannot be negative; what is the largest value of x beyond which y is negative? Solve

$0 = -0.1495x + 9.643$

$0.1495x = 9.643$

$x = \frac{9.643}{0.1495} \doteq \boxed{64.5 \text{ hours}}$

(c)[2 points] If instead GPA is used as the explanatory variable would you expect r to increase, decrease or stay the same? Explain.

Stay the same. The calculation of r makes no distinction between explanatory & response variable.

(d)[3 points] What percentage of the variation in GPA data is not explained by the linear relationship with hours of television watching?

Percentage of variation in GPA data explained by linear relationship is r^2;

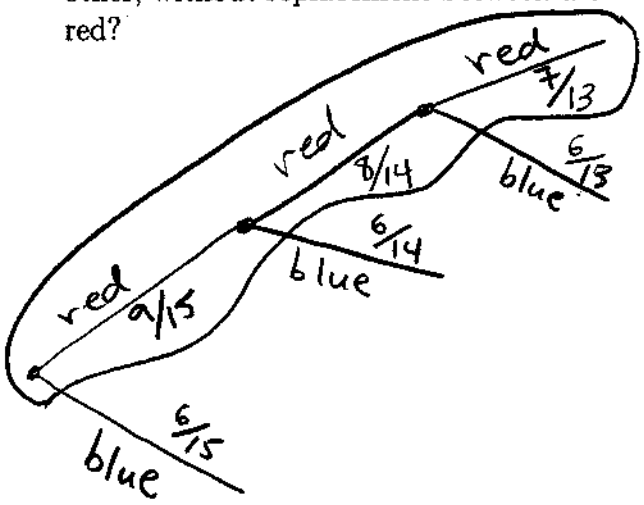
∴ Percentage not explained is $1 - r^2 = 1 - (-0.6241)^2$

$= 0.6105$

$\doteq \boxed{61.05\ \%}$

Question 3

(a)[5 points] A bag contains 9 red and 6 blue marbles. Three marbles are drawn, one after the other, without replacement between draws. What is the probability that all three marbles are red?



$$P(\text{all three red}) = \left(\frac{9}{15}\right)\left(\frac{8}{14}\right)\left(\frac{7}{13}\right) = \boxed{\frac{12}{65}}$$

(b)[5 points] Again three marbles are drawn, but this time with replacement between draws. Find the probability that exactly one of the marbles is red.

Let RBB denote a first draw of red followed by 2 blues,
BRB denote a first draw of blue followed by red followed by blue, etc.

$$P(\text{exactly one red}) = P(RBB) + P(BRB) + P(BBR)$$

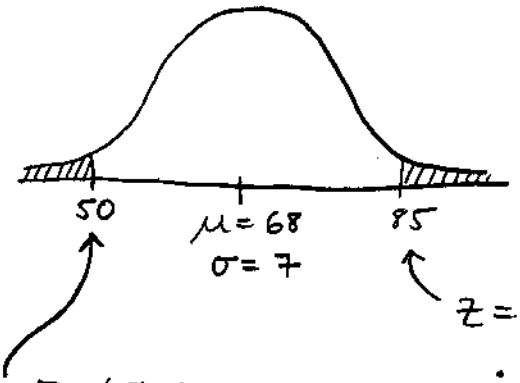
$$= \left(\frac{9}{15}\right)\left(\frac{6}{15}\right)\left(\frac{6}{15}\right) + \left(\frac{6}{15}\right)\left(\frac{9}{15}\right)\left(\frac{6}{15}\right) + \left(\frac{6}{15}\right)\left(\frac{6}{15}\right)\left(\frac{9}{15}\right)$$

$$= 3\left(\frac{9}{15}\right)\left(\frac{6}{15}\right)\left(\frac{6}{15}\right) = \boxed{\frac{36}{125}}$$

Question 4

The grades from a large math class are approximately normally distributed with mean $\mu = 68\%$ and standard deviation $\sigma = 7\%$. (Table B may be useful for this question).

(a)[5 points] A student is drawn at random from the classlist. What is the probability that the student scored below 50% or above 85%?



$z = \frac{50-68}{7} \doteq -2.6$

$z = \frac{85-68}{7} \doteq 2.4$

Let X = grade.

$P(X < 50) = P(z < -2.6)$

$= \frac{0.47}{100} = 0.0047$

$P(X > 85) = 1 - P(X \leq 85)$

$= 1 - P(z \leq 2.4)$

$= 1 - \frac{99.18}{100}$

$= 0.0082$

$\therefore P(X < 50 \text{ or } X > 85) = 0.0047 + 0.0082$

$= \boxed{0.0129}$

(b)[5 points] A student is drawn at random from the classlist. A second draw is made from the same classlist (perhaps resulting in the same student being drawn the second time). What is the probability that at least one of two corresponding scores was below 50%?

$P(\text{at least one below } 50\%)$

$= 1 - P(\text{both greater than or equal to } 50\%)$

$= 1 - P(X \geq 50)\,P(X \geq 50)$ by independence.

$= 1 - [1 - 0.0047][1 - 0.0047]$ using (a)

$= 1 - (0.9953)^2$

$\doteq \boxed{0.0094}$

Question 5

(a)[5 points] Consider the following game: you roll a six sided die. If a 1 comes up you receive \$5, if a 2 or 3 comes up you get \$4, and if a 4, 5, or 6 comes up you must pay \$2. What amount of money do you expect to win (or lose) in playing this game?

Let X = amount received in one play

$X = 5$, $P(X=5) = \frac{1}{6}$

$X = 4$, $P(X=4) = \frac{2}{6} = \frac{1}{3}$

$X = -2$, $P(X=-2) = \frac{3}{6} = \frac{1}{2}$

$$\therefore E(X) = (5)\left(\tfrac{1}{6}\right) + (4)\left(\tfrac{1}{3}\right) + (-2)\left(\tfrac{1}{2}\right)$$

$$= \frac{5}{6} + \frac{4}{3} - 1$$

$$= \frac{5+8-6}{6} = \frac{7}{6} \doteq \boxed{\$1.17}$$

(b)[5 points] The following table gives test scores from a math class. Here X represents the test score and frequency is the number of people receiving that score:

X	0	5	10	15	20
frequency	2	5	18	17	8

} total = 50.

What is the average test score?

$$E(X) = (0)\left(\tfrac{2}{50}\right) + (5)\left(\tfrac{5}{50}\right) + (10)\left(\tfrac{18}{50}\right) + (15)\left(\tfrac{17}{50}\right) + (20)\left(\tfrac{8}{50}\right)$$

$$= \frac{25}{50} + \frac{180}{50} + \frac{255}{50} + \frac{160}{50}$$

$$= \frac{620}{50}$$

$$\doteq \boxed{12.4}$$