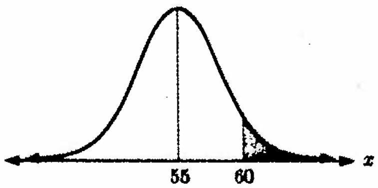
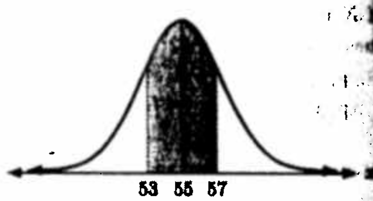
c



$P(X \geqslant 60) \approx 0.238$

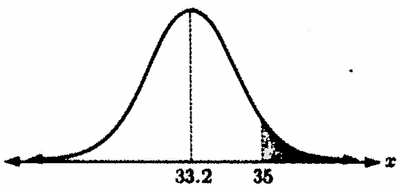
d



$P(53 \leqslant X \leqslant 57) \approx 0.225$

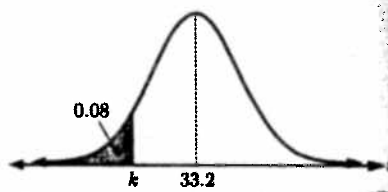
5 If X is the life of a battery in weeks, then $X \sim N(33.2, 2.8^2)$.

a



$P(X \geqslant 35) \approx 0.260$

b



We need to find k such that

$P(X \leqslant k) = 0.08$

$\therefore \quad k \approx 29.3$

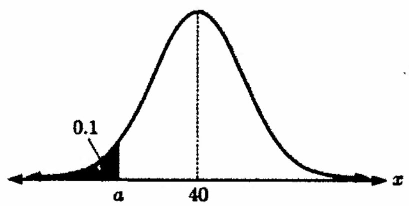
So, the manufacturer can expect the batter
to last 29.3 weeks before 8% of them fail

6 If X is the speed of a typist in words per minute, then $X \sim N(40, 16.7^2)$.

$P(X \leqslant a) = 0.10$

$\therefore \quad a \approx 18.6$

So, typists with speeds between 0 and 18.6 words per minute are enrolled.



7 If T is the temperature in degrees Celsius, then $T \sim N(25.4, 4.8^2)$.

$P(X \leqslant b) = 0.05$

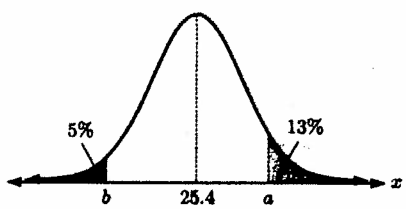
$b \approx 17.5$

$P(X \geqslant a) = 0.13$

$\therefore \quad P(X \leqslant a) = 1 - 0.13$

$= 0.87$

$\therefore \quad a \approx 30.8$

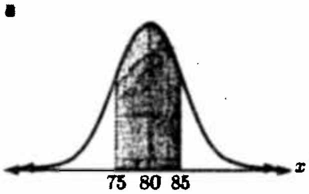


So, the range of Alison's suitable walking temperatures are from 17.5°C to 30.8°C.

REVIEW SET 10C

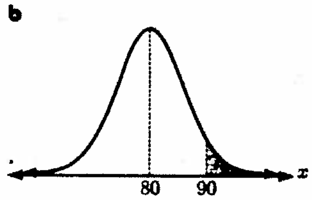
1 X is normally distributed with mean 80 and standard deviation 14.

a



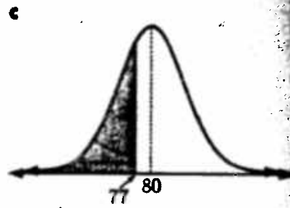
$P(75 \leqslant X \leqslant 85) \approx 0.279$

b



$P(X > 90) \approx 0.238$

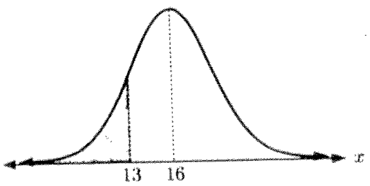
c



$P(X < 77) \approx 0.415$

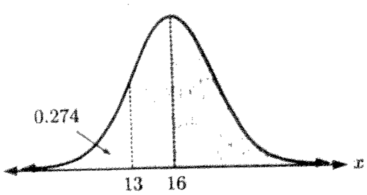
1 X is normally distributed with mean 16 and standard deviation 5.

a



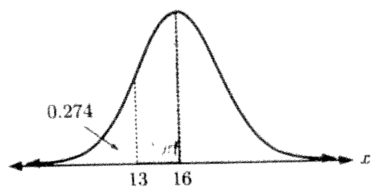
$P(X < 13) \approx 0.274$

b **i**



$P(X \geqslant 13)$
$= 1 - P(X < 13)$
$\approx 1 - 0.274$
≈ 0.726

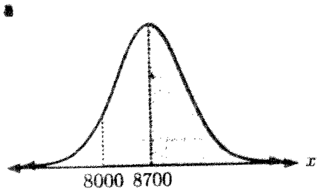
ii



$P(13 \leqslant X \leqslant 16)$
$= P(X \leqslant 16) - P(X < 13)$
$\approx 0.5 - 0.274$
≈ 0.226

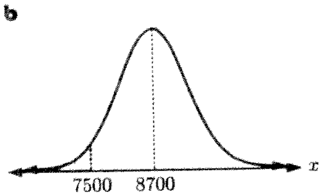
3 If X is the daily energy intake of a Canadian adult in kilojoules, then $X \sim N(8700, 1000^2)$.

a



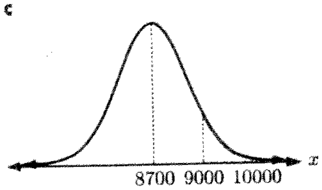
$P(X > 8000) \approx 0.758$

b



$P(X < 7500) \approx 0.115$

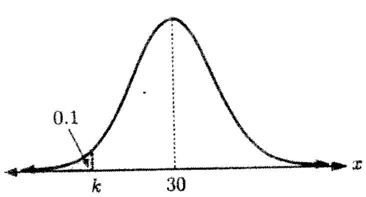
c



$P(9000 < X < 10\,000)$
≈ 0.285

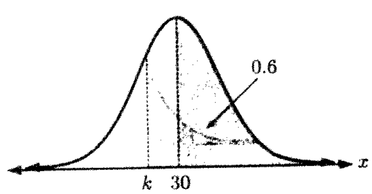
4 X is normally distributed with mean 30 and standard deviation 8.

a



$P(X \leqslant k) = 0.1$
$\therefore \; k \approx 19.7$

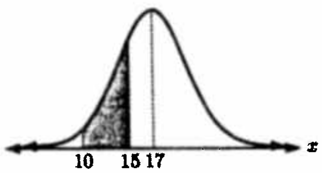
b



$P(X \geqslant k) = 0.6$
$\therefore \; P(X \leqslant k) = 0.4$
$\therefore \; k \approx 28.0$

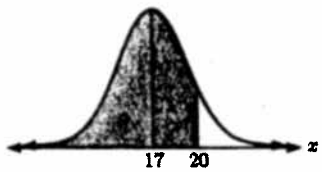
5 If X is the weight of a suitcase in kg, then $X \sim N(17, 3.4^2)$.

a



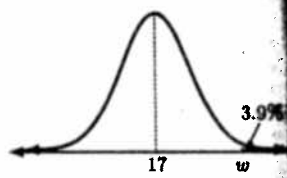
$P(10 < X < 15) \approx 0.258$

b



$P(X < 20) \approx 0.811$
and $300 \times 0.811 \approx 243$
So, you would expect about 243 suitcases out of 300 to be lighter than 20 kg.

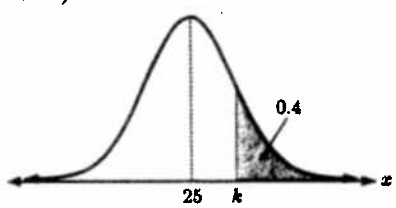
c



$P(X > w) = 0.039$
$\therefore\ P(X \leqslant w) = 0.961$
$\therefore\ w \approx 23.0$
So, the maximum weight limit is about 23.0 kg.

6 X is normally distributed with mean 25 and standard deviation 7.

a $P(X > k) = 0.4$

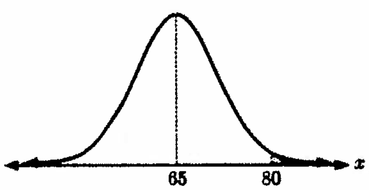


$\therefore\ k > 25$

b $P(X > k) = 0.4$
$\therefore\ P(X \leqslant k) = 0.6$
$\therefore\ k \approx 26.8$

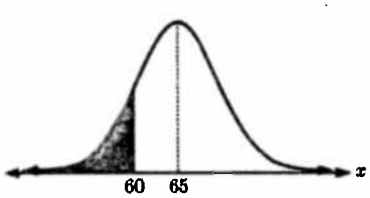
7 If X is the time of a participant in minutes, then $X \sim N(65, 9^2)$.

a



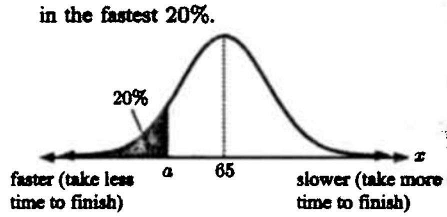
$P(X > 80) \approx 0.0478$

b **i**



$P(X < 60) \approx 0.289$
and $5000 \times 0.289 \approx 1446$
So, 1446 people are expected to complete the fun-run in less than one hour.

ii Simon is the 1000th person out of 5000 people to complete the fun-run.
Now, $\frac{1000}{5000} = \frac{1}{5} = 0.20$, so Simon is in the fastest 20%.



$P(X \leqslant a) = 0.20$
$\therefore\ a \approx 57.4$
So, Simon takes about 57.4 minutes to complete the fun-run.

c H_0 is that the *result* and *guess* are independent.
H_1 is that the *result* and *guess* are not independent.
At a 5% significance level and with 1 degree of freedom, the critical value is 3.84.
Since $\chi^2_{calc} \approx 1.35 < 3.84$, we do not reject H_0.
So, at a 5% significance level, Horace's *guess* and *result* are independent.

d As our conclusion was that Horace's *guess* and the *result* are independent, we can say that Horace's claim is not valid.

2 **a**

Result

Country

	Pass	Fail	Sum
France	$\frac{232\times 85}{309} \approx 63.8$	$\frac{77\times 85}{309} \approx 21.2$	85
Germany	$\frac{232\times 224}{309} \approx 168.2$	$\frac{224\times 77}{309} \approx 55.8$	224
Sum	232	77	309

b $\text{df} = (2-1)(2-1) = 1$
At a 10% significance level, and with 1 degree of freedom, the corresponding critical value from the provided table is 2.71.

c Use $\chi^2_{calc} = \sum \frac{(|f_o - f_e| - 0.5)^2}{f_e}$.

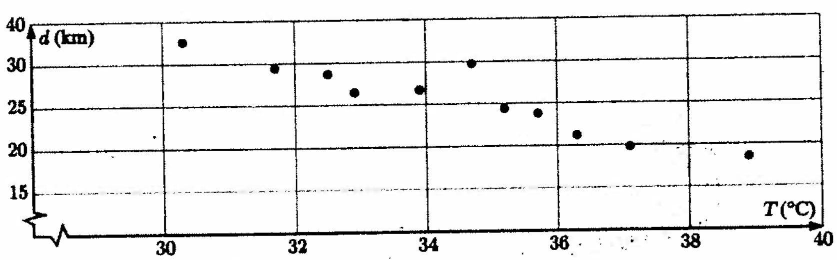
f_o	f_e	$f_o - f_e$	$\|f_o - f_e\|$	$\|f_o - f_e\| - 0.5$	$(\|f_o - f_e\| - 0.5)^2$	$\frac{(\|f_o - f_e\| - 0.5)^2}{f_e}$
56	63.8	−7.8	7.8	7.3	53.29	0.835 266 457
29	21.2	7.8	7.8	7.3	53.29	2.513 679 245
176	168.2	7.8	7.8	7.3	53.29	0.316 825 208
48	55.8	−7.8	7.8	7.3	53.29	0.955 017 921
					Total	4.620 788 831

$\chi^2_{calc} \approx 4.62$

d H_0 is that the *result* and *country* are independent.
H_1 is that the *result* and *country* are not independent.
Since $\chi^2_{calc} \approx 4.62 > 2.71$, we reject H_0.
So, at a 10% significance level, the motorbike *test result* and *country* are not independent.

REVIEW SET 11A

1 **a**



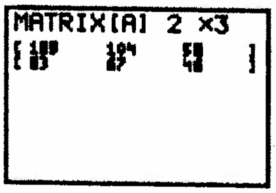
b $r \approx -0.928$ {using technology}
Since $-0.95 < r \leqslant -0.87$, there is a strong negative linear correlation between the temperature of that day and the number of kilometres ridden.
In general, the higher the temperature is, the less number of kilometres Thomas tends to ride.

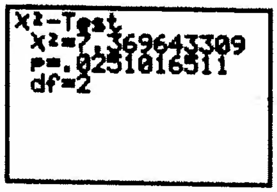
c $d \approx -1.64T + 82.3$ {using technology}

d If Thomas does not ride at all, $d = 0$
$\therefore\ 0 \approx -1.644T + 82.25$
$\therefore\ 1.644T \approx 82.25$
$\therefore\ T \approx \frac{82.25}{1.644} \approx 50.0°C$

Note that this is not a reliable extrapolation. Fifty degrees lies well outside the poles of the data set (the closest data value we have is for 38.9 degrees) and it may be that Thomas stops riding at some other temperature.

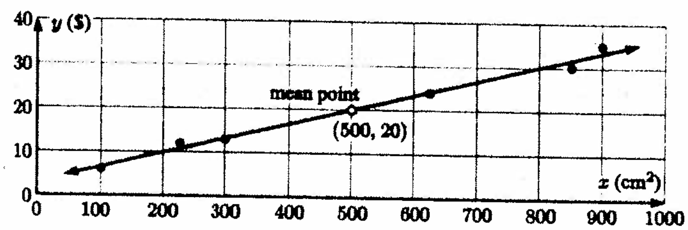
2 H_0 is that *wearing a seatbelt* and *severity of injury* are independent.
H_1 is that *wearing a seatbelt* and *severity of injury* are not independent.
$\text{df} = (2-1)(3-1) = 2$ The significance level is 5% or 0.05.
We reject H_0 if $\chi^2_{calc} > 5.99$.





$\chi^2_{calc} \approx 7.3696$ So, $\chi^2_{calc} \approx 7.37$ which is > 5.99, so we reject H_0.
$p \approx 0.0251$ which is < 0.05, again evidence for rejecting H_0.
So we conclude at a 5% significance level that *wearing a seatbelt* and *severity of injury* are not independent.

3 **a, e**



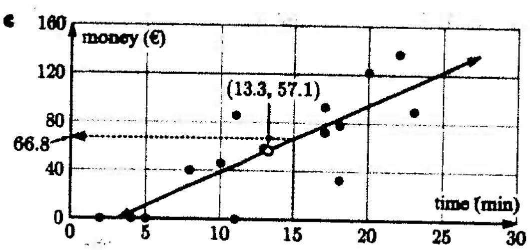
b Using technology, $r \approx 0.994$.

c Since $0.95 \leqslant r < 1$, there is a very strong positive linear correlation between *area* and *price* of each canvas type. In general, the larger the area of the canvas, the more expensive the price.

d Using technology, $y \approx 0.0335x + 3.27$

f When $x = 1200$, $y \approx 0.033\,46(1200) + 3.27 \approx 43.42$
So, a canvas with area 1200 cm² will cost around $43.42. However, this estimate is an extrapolation (1200 lies outside the known area values), so it may be unreliable.

4 **a, c**



b $\overline{x} = \dfrac{\sum x}{n} = \dfrac{8+18+5+\dots+17}{15}$
≈ 13.3

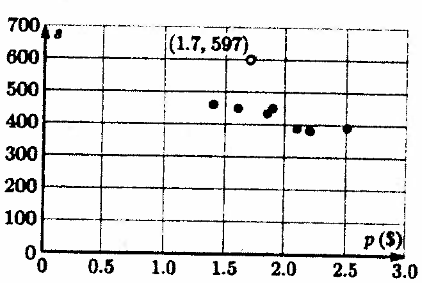
$\overline{y} = \dfrac{\sum y}{n} = \dfrac{40+78+0+\dots+93}{15}$
≈ 57.1

So, the mean point is approximately (13.3, 57.1).

d There is a moderate positive linear correlation between *time in the store* and *money spent.*

e When a person is in the store for 15 minutes, then using the graph we can estimate they will spend €66.80. This estimate is reliable as it is an interpolation, not an extrapolation.

5 **a**



b There is one outlier, (1.7, 597).
It should not be deleted as there is no evidence of an error.

c $s \approx -116p + 665$ {using technology}

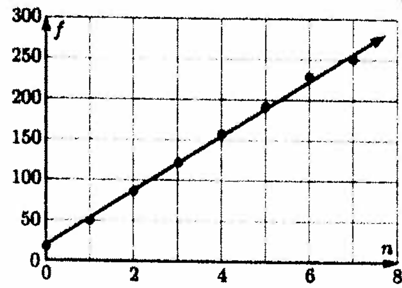
d For this data, $r^2 \approx 0.346$, so there is only a weak correlation between the variables. In addition, 50 cents is a long way outside the poles, so this extrapolation is unreliable.
Therefore this would not be a reliable prediction.

6 **a** The independent variable is the number of waterings, n.

b $f \approx 34.0n + 19.3$ {using technology}

c Yes, plants need water to grow, so it is expected that an increase in watering will result in an increase in flowers.

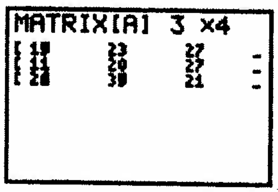
d

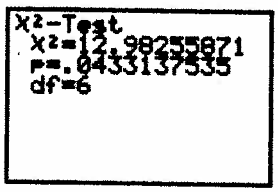


e **i** If $n = 2.5$, $f \approx 34.0 \times 2.5 + 19.3 \approx 104$
If $n = 10$, $f \approx 34.0 \times 10 + 19.3 \approx 359$

ii The case $n = 10$ is unreliable as it lies outside the poles and over watering could be a problem. The case $n = 2.5$ is a much more reliable estimate.

7 H_0 is that factors P and Q are independent.
H_1 is that factors P and Q are not independent.
$\text{df} = (3-1)(4-1) = 6$
For **a**, the significance level is 5% or 0.05, and we reject H_0 if $\chi^2_{calc} > 12.59$.
For **b**, the significance level is 1% or 0.01, and we reject H_0 if $\chi^2_{calc} > 16.81$.





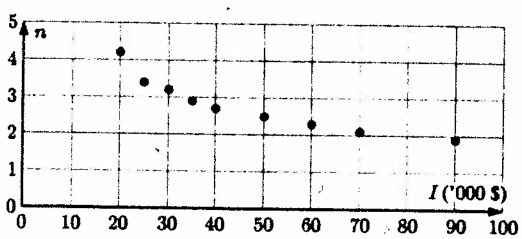
$\chi^2_{calc} \approx 12.9826$

a Since $\chi^2_{calc} \approx 13.0$ which is > 12.59, we reject H_0.
$p \approx 0.0433$ which is < 0.05, again saying we should reject H_0.
$\therefore$ we conclude that at a 5% level of significance, P and Q are not independent.

b Since $\chi^2_{calc} \approx 13.0$ which is < 16.81, we do not reject H_0.
$p \approx 0.0433$ which is > 0.01, again saying we should not reject H_0.
$\therefore$ we conclude that at a 1% level of significance, P and Q are independent.

REVIEW SET 11B

1 **a**



b $r \approx -0.908$ {using technology}

c $n \approx -0.0284I + 4.12$ {using technology}

d **i** When $I = 45$,
$n \approx -0.0284I + 4.12$
$\approx (-0.0284 \times 45) + 4.12$
≈ 2.842
So, for a family with an income of \$45 000, the average number of children is ≈ 2.84.

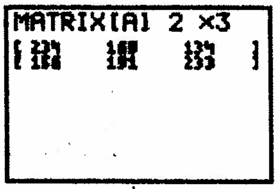
ii When $I = 140$,
$n \approx -0.0284I + 4.12$
$\approx (-0.0284 \times 140) + 4.12$
≈ 0.144
So, for a family with an income of \$140 000, the average number of children is ≈ 0.144.

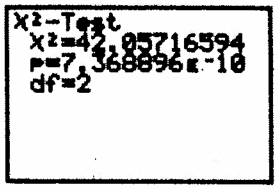
- The first estimate is interpolated, and therefore reliable. The second estimate is extrapolated, and therefore may not be reliable.

2 **a** **i** Negative correlation. As prices increase, the number of tickets sold is likely to decrease.
ii Causal. Less people will be able to afford tickets as the prices increase.

b **i** Positive correlation. As icecream sales increase, number of drownings is likely to increase.
ii Not causal. Both these variables are dependent on the number of people at the beach.

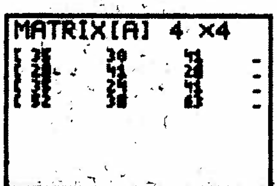
3 H_0 is that the *age of a driver* and *increasing the speed limit* are independent.
H_1 is that the *age of a driver* and *increasing the speed limit* are not independent.
df $= (2-1)(3-1) = 2$ Level of significance is 10% or 0.10.
We reject H_0 if $\chi^2_{calc} > 4.61$

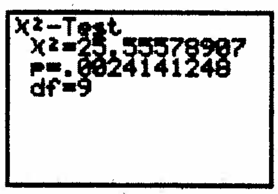




$\chi^2_{calc} \approx 42.057$
Since $\chi^2_{calc} \approx 42.1$ which is greater than 4.61, we reject H_0.
$p \approx 7.37 \times 10^{-10}$ which is < 0.10, again saying we should reject H_0.
So, we can conclude that at a 10% level of significance, *age of a driver* and *increasing the speed limit* are not independent.

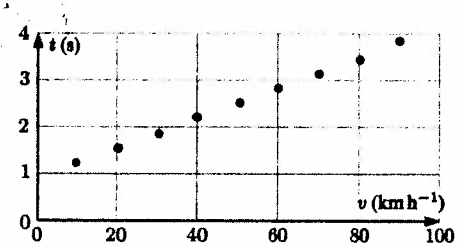
4 H_0 is that *intelligence level* and *business success* are independent.
H_1 is that *intelligence level* and *business success* are not independent.
df $= (4-1)(4-1) = 9$ The significance level is 1% or 0.01.
We reject H_0 if $\chi^2_{calc} > 21.67$





$\chi^2_{calc} \approx 25.556$
Since $\chi^2_{calc} \approx 25.6$ which is > 21.67, we reject H_0.
$p \approx 0.002\,41$ which is < 0.01, again saying we should reject H_0.
So, we conclude that at a 1% level of significance, *intelligence level* and *business success* are not independent.

5 **a**



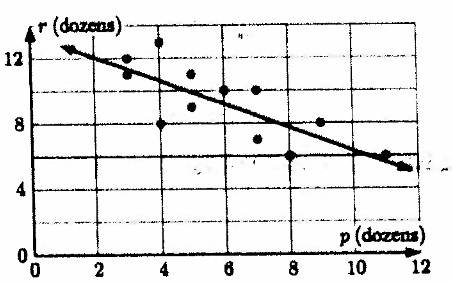
A linear model seems appropriate.

b $t \approx 0.0322v + 0.906$ {using technology}

c **i** If $v = 55$ km h^{-1}
$t \approx 0.0322(55) + 0.906$
≈ 2.68 seconds

ii If $v = 110$ km h^{-1}
$t \approx 0.0322(110) + 0.906$
≈ 4.44 seconds

d The vertical intercept (0.906) is the driver's reaction time.

a 

b $r \approx -0.706p + 13.5$ dozen maidens {using technology}

c $r \approx -0.763$ {using technology}
There is a moderate negative association between the variables. So, the more maidens Furry Reaper abducts the fewer Silent Predator takes.
Thus there is some evidence of collaboration, which supports Superman's suspicions, but it is only moderately strong.

d $r \approx -0.7059(6) + 13.49$
≈ 9.25 dozen maidens
≈ 111 maidens

e If $r = 20$, $20 \approx -0.7059p + 13.49$
$\therefore\ 0.7059p \approx 13.49 - 20$
$\therefore\ p \approx \dfrac{-6.51}{0.7059}$
$\therefore\ p \approx -9.23$

So if Furry Reaper abducts 20 or more dozen maidens, p will be negative, which is impossible since Silent Predator cannot abduct a negative number of maidens. Therefore the model is inappropriate for this data range.

f When $r = 0$, $0 \approx -0.7059p + 13.49$
$\therefore\ 0.7059p \approx 13.49$
$\therefore\ p \approx \dfrac{13.49}{0.7059}$
$\therefore\ p \approx 19.1$
$\therefore$ p-intercept is 19.1

and when $p = 0$
$r \approx 0 + 13.5$
$\therefore\ r \approx 13.5$
$\therefore$ r-intercept is 13.5
These represent how many dozen maidens we would expect one villain to abduct if the other villain did not abduct any.

g Superman should capture Silent Predator, as this would make $p = 0$ and thus enforce the r-intercept of 13.5 dozen maidens.
(Capturing the Furry Reaper enforces the p-intercept, 19.1 dozen maidens.)