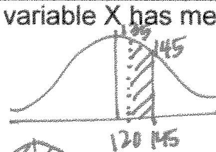
1. A normally distributed random variable X has mean 120 and standard deviation 25.

(a) Find P(135 ≤ X ≤ 145).



normalcdf (135, 145, 120, 25) = 0.115597805...

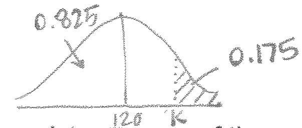
$\boxed{P(135 \le X \le 145) = 0.116}$

(b) Find P(X ≥ 90).

normalcdf (90, 1 x 10^99, 120, 25) = 0.8849302684...

$\boxed{P(X \ge 90) = 0.885}$

(c) Find k given that P(X ≥ k) = 0.175



invNorm (0.825, 120, 25) = 143.364732

$\boxed{K = 143}$ P(X < K) = 0.825

2. Members of a certain club are required to register for one of three games, billiards, snooker or darts.
The number of club members of each gender choosing each game in a particular year is shown in the table:

OBSERVED VALUES:

	Billiards	Snooker	Darts	
Male	39	16	8	63
Female	21	14	17	52
	60	30	25	115

(a) Use a χ^2 (Chi-squared) test at the 5% significance level to test whether choice of games is independent of gender. State clearly the null and alternative hypotheses tested, the expected values, and the number of degrees of freedom used.

H_0 = gender and choice of games are independent factors.
H_1 = gender and choice of games are not independent factors.

EXPECTED VALUES:

	Billiards	Snooker	Darts	
Male	32.9	16.4	13.7	63
Female	27.1	13.6	11.3	52
	60	30	25	115

d.f. = (2−1)(3−1) = 2 ⟹ χ^2_{crit} = 5.991 = 5.99 @ 5% significance level

χ^2_{calc} = 7.79245523... = 7.79 (used GDC)

p-value = 0.0203184159... = 0.0203

Because $\chi^2_{calc} = 7.79 > \chi^2_{crit} = 5.99$, reject H_0 and conclude that gender and choice of games are not independent factors. This conclusion is also supported by the fact that the p-value = 0.0203 is less than the significance level = 0.05. Therefore, based upon this data, there appears to be a statistically significant relationship between gender and choice of games.

The following year the choice of games was widened and the figures for that year are as follows:

	Billiards	Snooker	Darts	Fencing	
Male	4	15	8	10	37
Female	10	21	17	37	85
	14	36	25	47	122

(b) If the χ^2 test were applied to this new set of data,

(i) why would it be necessary to combine billiards with another game? The expected value for male/billiards is 4.24. Because there is an expected value less than 5 the χ^2 test will be considered unreliable as the data is currently organized.

(ii) which other game would you combine with billiards and why?
Billiards and snooker are similar games (Americans may need to look this up. :)).

~~(iii)~~ If a club member is selected at random, what is the probability that the club member selected is female who chose billiards or snooker?

OMIT for this unit... oops.

3. The heights and weights of 10 students selected at random are shown in the table below:

Student	1	2	3	4	5	6	7	8	9	10
Height x cm	155	161	173	150	182	165	170	185	175	145
Weight y kg	50	75	80	46	81	79	64	92	74	108

(a) **Using separate graph paper**, plot this information on a scatter graph. Use a scale of 1 cm to represent 20 cm on the x-axis and 1 cm to represent 10 kg on the y-axis.

(b) Calculate the mean height. $\bar{x} = 166.1 \rightarrow \boxed{166 \text{ cm}}$ (3 sig. fig) $\boxed{\bar{x} = 166\text{cm}}$

(c) Calculate the mean weight. $\boxed{\bar{y} = 74.9 \text{ Kg}}$

(d) It is given that $S_{xy} = 44.31$.

(i) By first calculating the standard deviation~~s~~ of the heights, correct to two decimal places, show that the gradient of the line of regression of y on x is 0.276. **SHOW WORK!**

$S_x (= \sigma_x) = 12.68$ gradient = slope = $\frac{S_{xy}}{S_x^2} = \frac{44.31}{(12.68)^2} = 0.276$

(use σ_x^2)

(ii) Calculate the equation of the line of best fit using the gradient above and the means of the height and weight.

$y - \bar{y} = \frac{S_{xy}}{S_x^2}(x - \bar{x})$

$y - 74.9 = 0.276(x - 166)$

$y - 74.9 = 0.276x - 45.816 \rightarrow \boxed{y = 0.276x + 29.1}$

also with calc: $y = 0.276x + 29.1$
$r = 0.200$

(iii) Draw the line of best fit on your graph.

(e) Use your line to estimate

(i) The weight of a student of height 190 cm.
By graph: approximately 82 Kg.

$y = 0.276(190) + 29.1$
$y = 81.54 \text{ kg}$

(ii) The height of a student of weight 72 kg.
By graph: approximately 154 cm

$72 = 0.276x + 29.1$
$42.9 = 0.276x \rightarrow x = 155.43\ldots$
$x = 155 \text{cm}$

(f) It is decided to remove the data for student number 10 from the calculations. Explain **briefly** what effect this will have on the line of best fit.

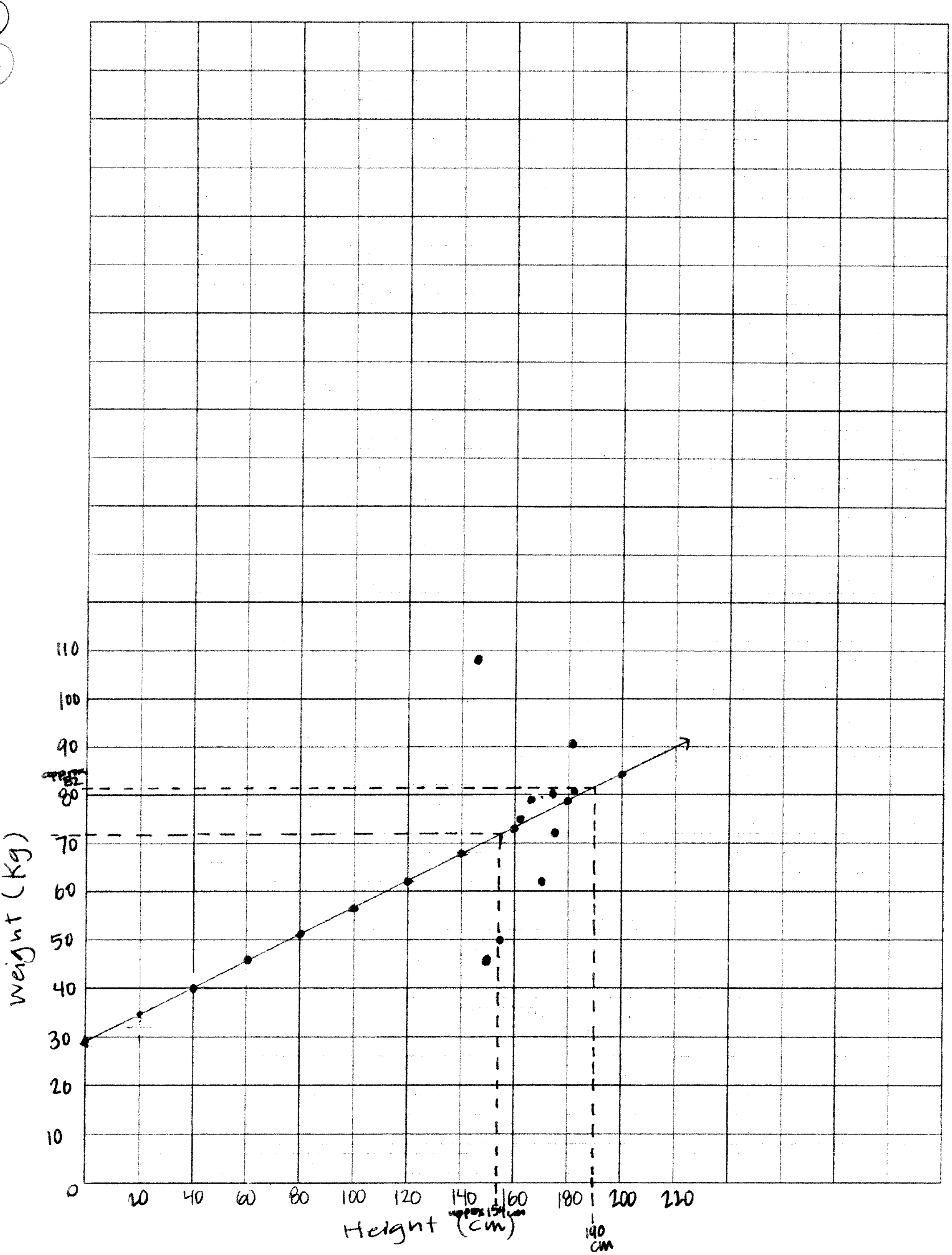
Because the data for student #10 is so different (an outlier) from the other values (the others are more closely clustered together), the line of best fit will have a greater gradient/slope and the line of fit will model the data more accurately.

new line of fit → $y = 1.10x - 113$
$r = 0.856$

*****USE SEPARATE GRAPH PAPER FOR THESE PROBLEMS!!**

8

3

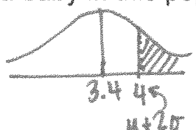


Weight (Kg)
110
100
90
approx 82
80
70
60
50
40
30
20
10
0
20
40
60
80
100
120
140
160
180
200
220
approx 154 cm
Height (cm)
140 cm

4. The mean birth weight of babies in a population is normally distributed with mean 3.4 kg and standard deviation $\sigma =$ 300 grams. ← grams! 1000 grams = 1 Kilogram so... 300 grams = 0.3 Kg

(a) What is the probability that a baby in this population has birth weight: $X \sim N(3.4, 0.3^2)$

(i) in excess of 4 kg

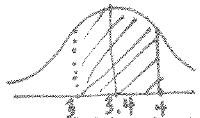


normalcdf $(4, 1\times10^{99}, 3.4, 0.3) = 0.022750062\ldots = 0.0228$

$\boxed{P(X > 4) = 0.0228}$

*about 2.28% of babies weigh more than 4g.

(ii) between 3 and 4 kilograms?

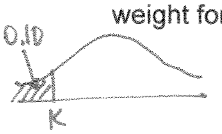


normalcdf $(3, 4, 3.4, 0.3) = 0.8860386561\ldots = 0.886$

$\boxed{P(3 \le X \le 4) = 0.886}$

*about 88.6% of babies weigh between 3g and 4g.

(b) A *low birth weight* corresponds to any newborn weighing in the lowest 10% of birth weights. Find the weight for this population below which a baby is classified as having a *low birth weight*.



$P(X < K) = 0.10$

invNorm $(0.10, 3.4, 0.3) = 3.01553453\ldots = 3.02$

$K = 3.02$ Kg

Babies in this population classified as having low birth weight weigh 3.02 Kg or less.

5. A nursery has developed a new hybrid plant. They claim that this hybrid will grow equally well in any light conditions. They have provided the following data to support their claim.

	Height < 60 cm	Height ≥ 60 cm
Sunlight	37	43
Shade	22	18
Dark	25	19

(a) Write suitable null and alternate hypotheses for a χ^2 test.

H_0 = Plant height is independent of light conditions.

H_1 = plant height is not independent of light conditions.

(b) Find the χ^2 statistic (χ^2_{calc}) for the plant data.

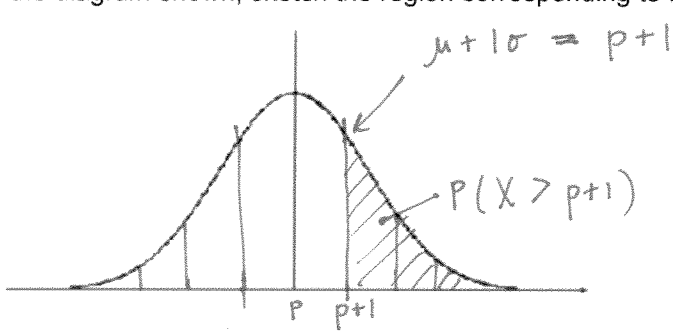
Using GDC: $\chi^2_{calc} = 1.571555736\ldots$ $\boxed{\chi^2_{calc} = 1.57}$

(c) Given that $\chi^2_{crit} = 5.991$, is there evidence at a 5% significance level to support the nursery's claim? Explain.

Because $\chi^2_{calc} = 1.57 < \chi^2_{crit} = 5.99$, accept H_0 and conclude that plant height is independent of light conditions. This conclusion is supported by the p-value since the p-value $= 0.456 < 0.05$ significance level. Therefore, the nursery's claim is justified according to this data.

6. Consider $X \sim N(p, 1)$.

(a) On the diagram shown, sketch the region corresponding to $P(X > p + 1)$.



(b) Given that $P(X < k) = P(X > p + 1)$, write down the value of k.

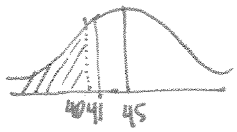
Because the normal distribution is symmetric about the mean $P(X > p+1) = P(X < p-1)$

Therefore, $\boxed{k = p - 1}$

Name: Key
Date: ________ Block: ____ page 4

7. The time taken for a skier to complete a particular downhill run is normally distributed with mean 45 seconds and standard deviation 4 seconds. $X \sim N(45, 4^2)$

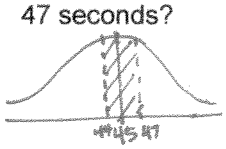
(a) Find the probability that the skier completes one downhill run in under 40 seconds.



normalcdf $(-1\times10^{99}, 40, 45, 4) = 0.105649839\ldots \approx 0.106$

$\boxed{P(X<40) = 0.106}$

(b) If the skier completes a total of 60 runs, how many times would you expect the run to take between 44 and 47 seconds?



normalcdf $(44, 47, 45, 4) = 0.2901687421\ldots \cdot (60\text{ runs}) = 17.41012453\ldots$

$P(44 \le X \le 47) = 0.290$

$\boxed{\text{we would expect that approximately 17 runs would take between 44 and 47 seconds.}}$

8. The average height *h* (in mm) of grass *t* days after being mowed is shown in the table below.

Days since mowed, *t*	0	1	2	3	4	5	6	7	8	9
Height, *h* (in mm)	5	5.7	5.7	6.2	6.8	7.1	8	8.3	9	9.3

(a) Calculate the value of Pearson's product-moment correlation coefficient *r*.

GDC: $r = 0.9926500227 \Rightarrow \boxed{r = 0.993}$

(b) Explain the significance of the size and sign of *r*.

Because $0.75 \le r < 1$, there is a strong, positive correlation between height and days since the grass was mowed.

(c) The regression line for *h* on *t* is $h = 0.4879t + 4.9145$. Using this equation, estimate the

(i) height of the grass after 14 days

$h = 0.4879(14) + 4.9145$ $\quad h = 11.7451$ $\quad \boxed{11.7\text{ mm}}$

(ii) time required for the height of the grass to reach 20 mm.

$20 = 0.4879t + 4.9145$

$15.0855 = 0.4879t$

$t = 30.91924575$

$t = 30.9$

$\boxed{\text{approximately 31 days}}$