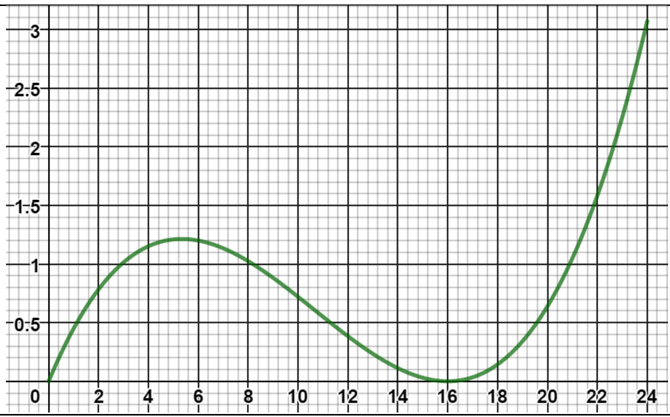
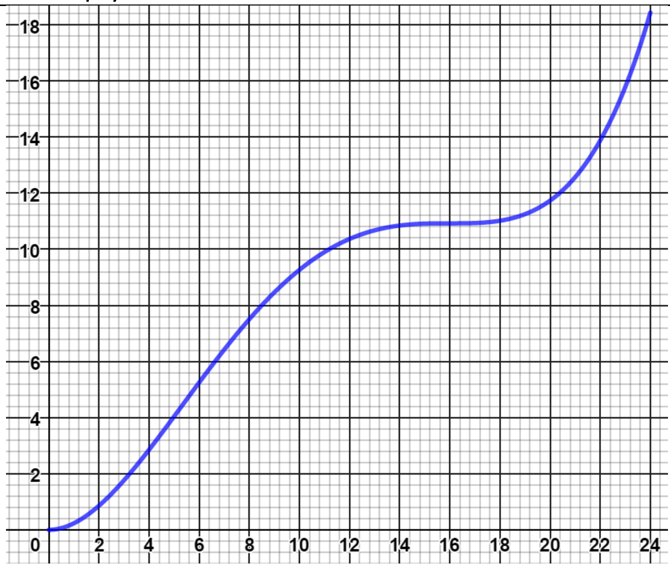
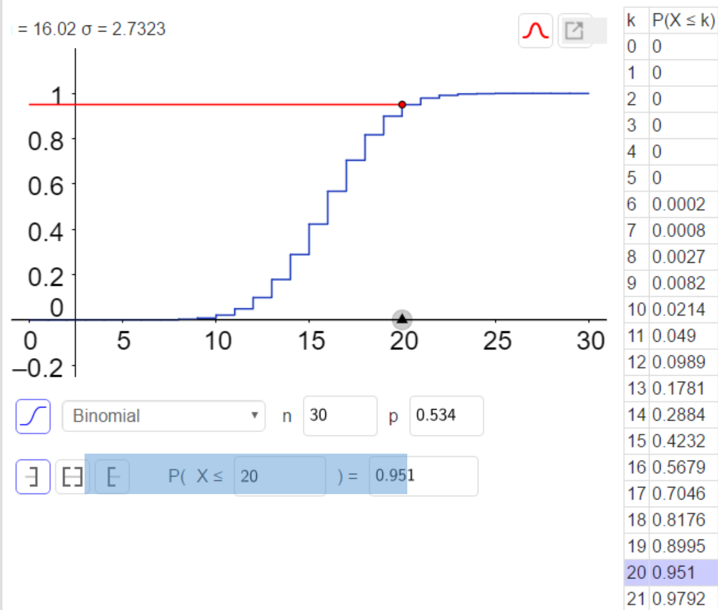
Part B Answers

	Part B1	Marks			
		KC	M	PS	I
In 2002 in Luxembourg the average temperatures per month have been recorded. It is known that January 2002 was the coldest month with 1,6°C and the highest average temperature was measured in June 2002 with 18,6°C.					
a)	**Justify**, that in Europe the monthly average temperatures for some consecutive years can be modelled with a periodic model.				
	The temperatures rise and fall within one year and this procedure repeats every year. Although the temperatures are not the same every year, on average the summer months are warm, and the winter months are cold.	2			
b)	**Give** the amplitude and the period of this model.				
	Amplitude $a = 8{,}5$ Period $T = 12$	1	1		
c)	**Determine** the parameters a, b, c and d in the model of the type $$T(x) = a \cdot sin(b \cdot (x - c)) + d$$ that describes the given data where T is the average Temperature and x is the month, starting with $x = 1$ for January 2002.				
	$a = \frac{18{,}6 - 1{,}6}{2} = 8{,}5$	1			
	$d = 8{,}5 + 1{,}6 = 10{,}1$		1		
	$b = \frac{2\pi}{12} = 0.524$		1		
	$c = 4$, because $\text{Extremum}\left(\frac{17}{2}\sin\left(\frac{131}{250}x\right) + \frac{101}{10}, -4, -2\right)$ $\rightarrow (-3, 1.6)$ there is a minimum at $x = -3$ that needs to be shifted to $x = 1$.			1 1	
On one specific day in March 2002 the rainfall was observed. The rainfall on that day can be modelled by the function $$R(t) = 0.002t^3 - 0.064t^2 + 0.512t, 0 \leq t \leq 24$$ where R(t) is the rate of rainfall in mm/h and t is the time in hours.					
d)	**Describe**, using a short text description, this day in terms of rainfall. Your answer should focus on the times with the most and the least rainfall.				
	It starts raining at midnight and the rainfall is reaching a peak between 5 and 6 o'clock. The rain gets less until 16 o'clock, when it stops for a little while. In the afternoon, the rain gets more again until it reaches a maximum at midnight.			2	1

					
An empty glass cylinder was placed outside during this day to help see how much rain had fallen.					
e)	**Sketch** the graph of a function, that shows the height of water in a glass cylinder, that was placed outside and empty before.				
	 The graph shows the derivate function of $R(t)$.	1	1	1	
f)	**Calculate** the total amount of rain on that day in mm.				
	$\int_0^{24} R(t) = 18.43$ About 18.4 mm of rain came down on that day.		1	1	
The year 2002 in Luxembourg turned out to have 195 rainy days and 170 days without rain. It can be assumed, that all days have the same chance of being a rainy day. One year later, meteorologists want to investigate, if there was more rain in 2003. Unfortunately, some data were lost, so they took only a small sample of 30 consecutive days.					
g)	**Calculate** the probability that it rains on a random day, if we assume, that the total number of rainy days in both years remains constant and the rainy days are equally distributed over the whole year.				
	$p = \frac{195}{365} = 0.534$	1			
h)	**Use** a NHST procedure to find out how many days it must rain so the meteorologists can say that there was more rain in 2003 compared to 2002 when the significance level is at 5%.				
	$H_0: p > 0.534$ $H_1: p \leq 0.534$		2		

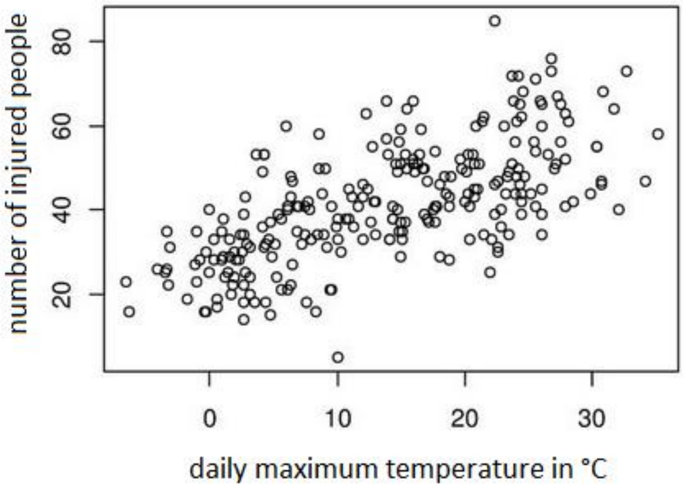


$k = 20$

So, if it was raining on more than 20 days, the scientists can be 95% sure, that in 2003 it was raining more than in 2002.

	(continued from above)		1	1	1

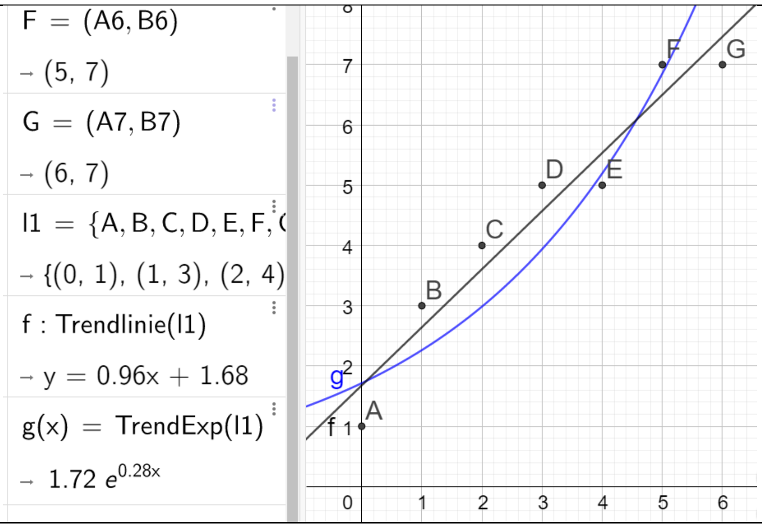
The following diagram shows the maximum temperature and the number of injured people caused by traffic accidents in Berlin on a long-term base.

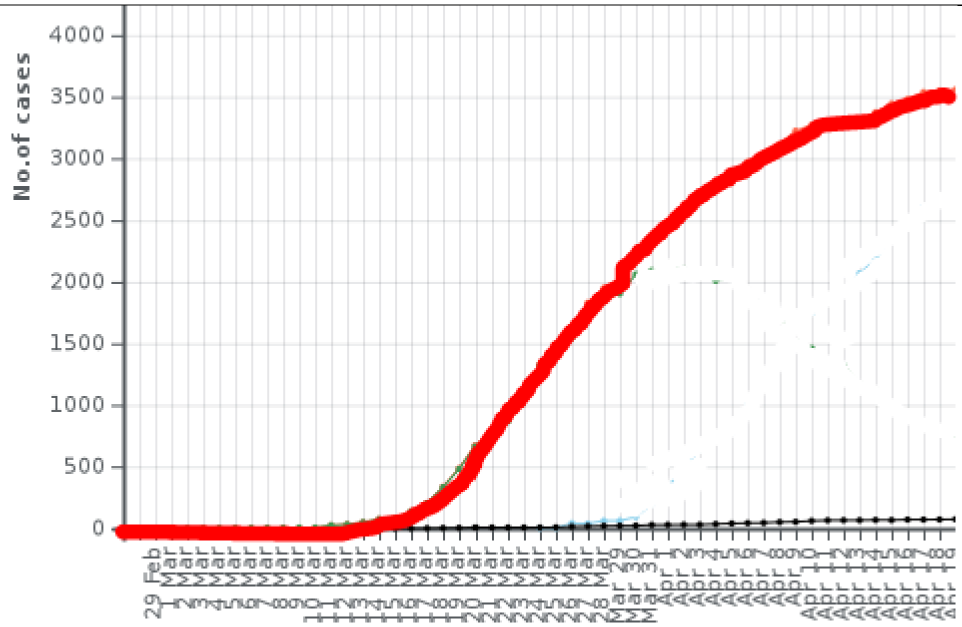


i)	**Describe** the correlation between the two values.				
	The higher the Temperature, the more injured people.				1
j)	**Explain**, why the number of injured people possibly correlates in such a way with the maximum temperature.				
	When it is warm, more people go outside, so the chance to be involved in a car accident is higher as if people stay at home.				1
	Total = 25	**6**	**8**	**7**	**4**

	Part B2	Marks			
		KC	M	PS	I
In a Covid-19 test station, 19 people with symptoms were tested on a specific day and 6 of them had a positive result. On the same day, 87 people without symptoms were tested of which 85 were tested negative.					
a)	**Show** that the probability of getting a positive result depends on whether a person has symptoms or not.				
	Consider the following events: S: "the person has symptoms" P: "the person is positive" $P(P) = \frac{8}{106}$ and $P(P\|S) = \frac{6}{19}$. $P(P) \neq P(P\|S)$ so, the probability of getting a positive result depends on whether a person has symptoms or not.		2		
To protect personal data, the test probes are labelled with a code, that contains 2 letters (out of an alphabet with 26 letters) and 4 digits (0-9). The same letters and digits may be chosen more than once.					
b)	**Calculate** the total number of different codes, that can be created by this system.				
	The total number of different codes is $26^2 \cdot 10^4 = 6\,760\,000$.	1	1		
After some months, statistics have shown, that 1.7 % of the people without symptoms are tested positive. A company with 20 employees (all without symptoms) lets everybody get tested.					
c)	**Give** two assumptions, that need to be made to model this situation with a binomial distribution.				
	• For the binomial distribution, all trials must be independent. So, we must assume, that the employees have no possibility to infect each other. • We also know that there some people (certain age group or health problems) are more likely to get infected. So, we must assume, that the chance of infection is the same for all employees. • Moreover, that we must assume, that the population of the statistic is a very large number of people. • We must assume, that the tests are reliable. • We must assume, that no employee has been vaccinated. • We must assume, that none of the employees has had Covid-19 before, so that he could be immune now.	1	1		
d)	**Calculate** the probability, that at least one of the employees is tested positive.				
	Calculate the probability that none of the employees is tested positive: $P(X=0) = B_{200;0{,}017}(0) = \binom{200}{0} \cdot 0{,}017^0 \cdot 0{,}983^{200}$ So $P(X \geq 1) = 1 - P(X=0) = 1 - 0{,}017^0 \cdot 0{,}983^{200}$, which gives $P(X \geq 1) \approx 0{,}968$		2	1	
Assuming that the situation can be modelled by a binomial distribution given by the formula $B(84; 0{,}02; k) = \binom{84}{k} \cdot 0.02^k \cdot 0.98^{84-k}$.					
e)	**Interpret** the values 84, 0.02 and 0.98 in the given context.				
	84 is the total number of employees, 0.02 is the probability that an employee is tested positive and 0.98 is the complement, i.e. the probability that an employee is not tested positive.	1	2		
On March 5 in 2020 a man who returned from Italy is the first person in Luxembourg who was tested positive on COVID-19. So, this day is marked as day 0 in the statistic. The following table shows the total number of registered infected people in Luxembourg in the days after the first case appeared.					
f)	**Draw** a scatter graph of these values together with a linear and an exponential regression model.				

Day	0	1	2	3	4	5	6
Number	1	3	4	5	5	7	7

			2	1	
g)	**Give** the equations that describe the regression lines in part f.				
	Linear: $f(x) = 0.96x + 1.68$ Exponential: $1.72 \cdot e^{0.28x}$	2			
h)	**Explain**, why it was so difficult to decide, if the spread out of the virus is best modelled with a linear or an exponential model in this early stage.				
	At an early stage, both regressions seem to be fitting well with the data. The outliers are not critically compromising any of these models. There may be an extra outlier for the exponential model, but the curve goes through two points whereas it hardly touches two points for the linear model.		2		
After seven more days other models were made to make better predictions, where t is given in days: $A(t) = 1.35567 \cdot 1.46977^t$ $B(t) = 12.4396 \cdot t - 34.8571$ On day 16, there were 670 registered cases of COVID-19 in Luxembourg.					
i)	**Calculate** the predicted number of infected people on day 16 with model A and model B and **compare** it with the true number. **Decide**, which model obviously works better for this situation and **reason** your answer.				
	$A(16) = 1.35567 \cdot 1.46977^{16}$ so $A(16) \approx 643$ $B(16) = 12.4396 \cdot t - 34{,}8571$ so $B(16) \approx 164$ The predicted value for model A is close to the observed value. The predicted value for model B is much smaller than the observed value. Given the big gap between both predicted values, model A is more suitable to the situation.		1	1	
The following diagram shows the graph of the total number of registered infections for the first 4 weeks in Luxembourg.					



j)	**Give** two possible reasons, why the curve flattens in a later stage.				
	• (Using the technological tool) we can see that $\lim_{t\to\infty} C(t) = 3404$, which shows that the curve has an horizontal tangent and therefore flattens when t increases. • $\lim_{t\to\infty} C'(t) = 0$, which shows that as t increases the tangent to the curve moves to a horizontal position and therefore the curve flattens. Also, nonmathematical reasons can be given: • In the summertime, the infection rates were going down, because many people go outside. • There was a lockdown where shops and schools closed, that caused a decrease in the infection rate. • People were wearing more masks. Other reasons that could be wrong, but that would still explain the result: • People who have had Covid-19 are immune, so the population is decreasing. • The number of tests decreased.			1	1
The curve can be modelled by the function $$C(t) = \frac{3404}{1 + 193 \cdot e^{-0.233 \cdot t}}$$					
k)	**Determine** the day with the highest infection rate by calculation.				
	We must find the maximum value for the derivative on the interval [0; 28]. Using the technological tool, we find that the maximum value for C' is achieved at $t = 22{,}59$ i.e., on the 23rd day the infection rate achieved its maximum.	1	1		
	Total = 25	**7**	**13**	**4**	**1**