Maths Internal Assessment

Is the rate of crime different in different regions in England and Wales?

Anon

Contents

A. Introduction

I will be looking into how the crime rate changes when you look at different regions in England and Wales. Crime has always been an interest of mine, I would like to know and understand how the crime rate is in different in the regions I have looked at. Will there be an overall increase in crime while comparing different regions, or will it be the individual crimes that alter the most? I will also be interested in whether these rates have increased or decreased in the last 9 years, between the years of 2002 - 2010.

In order to determine the relationship between the amount of crime and the regions I will be looking at the number of crimes in England and Wales. I will be looking at the rates of 12 different crimes [1], including the grand totals for crime, year and then everything put together. I also have 44 different counties, which I have separated into 7 regions [2]. All together, I have a total of 5148 pieces of data that I can use in all different ways throughout my investigation.

In order to ensure my investigation is fair and reliable, I have used all of my data from one source. All of my data comes from the police force in England and Wales; they were used for the National Statistic bulletin Crime which was published on the 15th of July 2010. I will not use any data from outside of this source because there could be many things that are different between the data, making it unreliable. For example, the data could have been collected at different times of the year, meaning the numbers are different, mistakes could have been made on any of the databases, but if I only use one then it will not affect my results, and also, different methods of collecting the data could have been used, which would mean that the data could be unreliable.

I separated the 44 counties into regions so that the data would be more compressed and therefore easier to read, understand and interpret. Doing this will allow me to look directly at the regions independently, and then make tables, graphs, box and whisper diagrams etc. In order to come to my conclusion, and therefore answer my investigation question.

My general hypothesis is that there is going to one region that has an overall crime rate higher than the other 6. However, I do not necessarily believe that this region will have all of the highest number of crimes compared to the other regions; I believe that some regions will have the highest in one crime, while others have the highest in another.

[1] Burglary in a building other than a dwelling, Burglary in a dwelling, Criminal damage, Drug offences, Fraud and forgery, Offences against vehicles, Other offences, Other theft offences, Robbery, Sexual offences, Violence against a person with injury, Violence against a person without injury.
[2] South West, East Midlands, Eastern, North East, North West, South East, Wales.

Plan

- I have done an internet search to get a database of numerical data.
- I will separate my data into 7 regions so that they are easier to read, use and analyse.
- I will plot the grand totals of all 7 regions onto 7 box plots so I can compare the distributions of crime in these 7 regions.
- I will then describe my findings and observations.
- I will plot scatter graphs of total number of crimes against year, each region.
- I will then add a regression line to each of the graphs.
- I will then use equations and correlation coefficients in order to analyse the trends of each of my scatter graphs.
- I will use my scatter graphs, equations and coefficients to make a prediction about the number of crimes in 2010.
- I will separate my data into four sections; burglary, offences, violence and other.
- I will do two chi-squared tests of independence so that I can be more confident with my results.
- I will one chi-squared for burglary and offences, and the other for violence and other.
- I will do a chi-squared test of independence on my data in order to know whether the number of crime is dependent on the region or not.

B. Data

To get this data I had to do many differently phrased searches on the internet before I found anything. I searched "Crime rates in the UK" and finally came across an interesting website [3], this website gave me a link to an interactive map of the crime rates in the UK, but also to a excel file that gave me the number of crimes for 44 different counties. I then decided that instead of working with such a large number of different places I would break it down into 7 different regions, these are the 7 regions that the database gave me;

- South West
- East Midlands
- Eastern
- North East
- North West
- South East
- Wales

And also the 12 different crimes;

- Burglary in a building other than a dwelling
- Burglary in a dwelling
- Criminal damage
- Drug offences
- Fraud and forgery
- Offences against vehicles
- Other offences
- Other theft offences
- Robbery
- Sexual offences
- Violence against a person with injury
- Violence against a person without injury.

I have used this data to plot both scatter graphs and box and whisper diagrams. Both of these graphs are very important and useful when it comes to comparing the regions and crimes. These will allow me to come to my conclusions and answer my questions.

---

3 http://rds.homeoffice.gov.uk/rds/soti.html

Regions	Crimes	2002/03	2003/04	2004/05	2005/06	2006/07	2007/08	2008/09	2009/10	Grand Total
South west	Burglary not dwelling	37365	35224	31274	29065	28450	26543	27624	24928	240473
	Burglary in a dwelling	28084	23427	20649	17315	16137	17022	17077	15577	155468
	Criminal damage	75645	84340	85762	86070	90924	81294	73035	62415	639485
	Drug offences	9506	10026	9890	11844	12820	12682	13667	13970	94405
	Fraud & forgery	26470	22457	20045	17259	15180	9636	10430	9032	130509
	Offences against vehicles	81575	69195	57452	52300	50272	44115	37972	30898	423779
	Other offences	3416	3594	3985	4989	4756	4377	4425	4893	34435
	Theft offences	94795	93249	91067	89246	90612	82764	80819	74996	697548
	Robbery	5172	4456	3504	3333	3224	2987	3072	2629	28377
	Sexual offences	4326	4609	4766	4535	4622	4293	4230	4549	35930
	Violence w. injury	27611	35540	40218	38897	36226	33733	31061	29465	272751
	Violence w/o injury	32731	34409	39878	39330	42053	36976	35550	34840	295767
	Total	426696	420526	408490	394183	395276	356422	338962	308192	3048927
East Midlands	Burglary not dwelling	46119	42569	36617	31029	30512	28974	28613	25160	269593
	Burglary in a dwelling	24162	22916	17960	15795	16191	14674	16020	14838	142556
	Criminal damage	66380	70883	68722	67737	70111	64739	59585	52551	520708
	Drug offences	5545	5785	6393	7060	7469	8166	8630	9142	58190
	Fraud & forgery	18348	18430	16740	13402	11051	9040	9074	8612	104697
	Offences against vehicles	57934	53124	43320	39544	39818	34716	33106	26696	328258
	Other offences	4298	4253	4263	4605	4369	4069	3911	4018	33786
	Theft offences	71313	68594	64644	63238	62236	63243	60798	59796	513862
	Robbery	3823	4019	3249	3314	3351	2897	3110	2863	26626
	Sexual offences	3668	3629	4029	3776	3589	3651	3563	3445	29350
	Violence w. injury	25858	31259	33986	32820	30462	28926	25284	24492	233087
	Violence w/o injury	24105	23949	26185	25358	26378	26596	25983	26377	204931
	total	351553	349410	326108	307678	305537	289691	277677	257990	2465644
Eastern	Burglary not dwelling	37885	35788	31553	30836	29560	27616	26003	24464	243705
	Burglary in a dwelling	25493	24944	20256	19429	19414	20309	21721	20158	171724
	Criminal damage	90558	99201	97426	94624	94295	82685	72900	63401	695090
	Drug offences	7295	7658	7989	9424	11343	13783	15292	14524	87308
	Fraud & forgery	27380	26945	23194	20990	18638	13670	12974	10942	154733
	Offences - vehicles	79358	72129	60171	59655	57084	51198	44890	37368	461853
	Other offences	3847	4115	4529	5124	5430	4837	5153	5068	38103
	Theft offences	101761	103146	100540	99729	95527	88616	84895	81785	755999
	Robbery	4672	4686	4287	4537	4856	4188	4194	3462	34882
	Sexual offences	4449	4794	4599	5311	4108	3806	3700	3754	34521
	Violence w. injury	28629	35594	39714	37646	34207	29903	27141	27482	260316
	violence w/o injury	32408	36828	38907	34908	36648	34411	32677	32598	279385
	total	443735	455828	433165	422213	411110	375022	351540	325006	3217619

Regions	Crimes	2002/03	2003/04	2004/05	2005/06	2006/07	2007/08	2008/09	2009/10	Grand Total
North East	Burglary not dwelling	19060	16771	13710	12566	12117	10375	9762	8693	103054
	Burglary in a dwelling	18500	16707	12469	10745	9901	8532	7831	6330	91015
	Criminal damage	56813	59481	58189	55515	54678	47303	40924	33734	406637
	Drug offences	7725	6782	6087	6981	7130	6949	7802	7686	57142
	Fraud & Forgery	8856	8099	6733	5751	4512	3708	3814	3232	44705
	Offences - vehicles	39672	35588	30684	28500	24240	20143	16844	12712	208383
	Other offences	2952	2703	2392	3268	3137	3201	3112	2753	23518
	Theft offences	49190	46423	41912	40318	39986	40220	37739	33053	328841
	Robbery	2876	2578	2074	1785	1631	1299	1136	858	14237
	Sexual offences	2215	2502	2329	2194	2013	1707	1555	1747	16262
	Violence w. injury	10612	19048	20778	20127	19473	16501	15408	13673	135620
	Violence w/o injury	20841	14948	13732	16180	18704	15959	14399	12484	127247
	Total	239312	231630	211089	203930	197522	175897	160326	136955	1556661
North West	Burglary not dwelling	70400	66037	56325	53534	51670	45447	44281	39096	426790
	Burglary in a dwelling	78155	69765	53018	49382	47596	41941	43985	39640	423482
	Criminal damage	188440	208965	208762	213439	207788	173255	154350	129467	1484466
	Drug offences	19214	19207	19692	28655	28359	33615	37437	37090	223269
	Fraud & Forgery	28838	34240	31360	28017	22489	18833	19714	18497	201988
	Offences - vehicles	155107	142827	124735	122909	119538	95004	84506	68946	913572
	Other offences	9590	10221	9821	13251	12513	11392	11348	10631	88767
	Theft offences	161358	160744	152732	152697	149310	145822	140334	129704	1192701
	Robbery	16196	13916	11629	12264	12506	10875	10417	8919	96722
	Sexual offences	7627	8640	8346	8241	7357	6983	6349	6747	60290
	Violence w. injury	61447	75516	77073	81438	73114	63171	56616	51874	540249
	Violence w/o injury	49773	59090	73727	72770	72195	64641	60364	56684	509244
	Total	846145	869168	827220	836597	804435	710979	669701	597295	6161540
South East	Burglary not dwelling	52462	51171	46487	46178	43993	40422	40467	38769	359949
	Burglary in a dwelling	40606	40879	34629	33576	32706	30098	30458	28493	271445
	Criminal damage	136714	160777	162923	162665	170023	148903	135499	115639	1193143
	Drug offences	16676	16905	17357	20328	21707	25807	24601	24798	168179
	Fraud & Forgery	39344	39349	38227	32562	35325	26319	27574	26233	264933
	Offences - vehicles	124507	116002	101006	98597	97725	83198	75656	65085	761776
	Other offences	7305	7312	7405	8495	8739	8363	8960	8879	65458
	Other theft offences	166932	173139	169622	170428	168723	158082	157957	149858	1314741
	Robbery	6538	6542	5801	5899	6388	5639	5986	5478	48271
	Sexual offences	7546	8444	9344	9167	8989	8059	7903	8048	67500
	Violence w. injury	36891	51198	62527	63131	61310	57284	54601	52957	439899
	Violence w/o injury	55602	62203	70485	77459	90382	88383	79043	75985	599542
	Total	691123	733921	725813	728485	746010	680557	648705	600222	5554836

Regions	Crimes	2002/03	2003/04	2004/05	2005/06	2006/07	2007/08	2008/09	2009/10	Grand Total
Wales	Burglary not dwelling	21885	20770	17589	16308	16196	15323	15175	14958	138204
	Burglary in a dwelling	14954	13902	11687	10010	10578	11298	10462	10092	92983
	Criminal damage	66060	66268	64345	64044	65053	59790	56693	49323	491576
	Drug offences	10265	9509	9197	9498	10528	12701	13326	13517	88541
	Fraud & forgery	12194	10654	9210	7485	6442	5672	6587	5782	64026
	Offences against vehicles	54550	52543	44687	42818	41429	34891	32298	27118	330334
	Other offences	3807	3327	3346	3665	3638	3381	3386	3380	27930
	Theft offences	54119	55515	50521	47770	47140	47358	47607	47260	397290
	Robbery	1377	1280	1084	1137	1354	1260	1215	1011	9718
	Sexual offences	2435	2481	2623	2697	2448	2574	2455	2463	20176
	Violence w. injury	27184	28451	28746	29283	26264	23539	22915	21987	208369
	Violence w/o injury	25950	24563	24589	23309	27403	25836	24495	24500	200645
	Total	294780	289263	267624	258024	258473	243623	236614	221391	2069792

This is the data that I have received from the police, and then sorted into my own chosen categories.

I have used this data to calculate the mean, Range, Upper and Lower Quartiles of all of the regions and years by creating box and whisper diagrams. Both the diagrams and the information from these will be very useful when it comes to finding out information and statistics about my investigation.

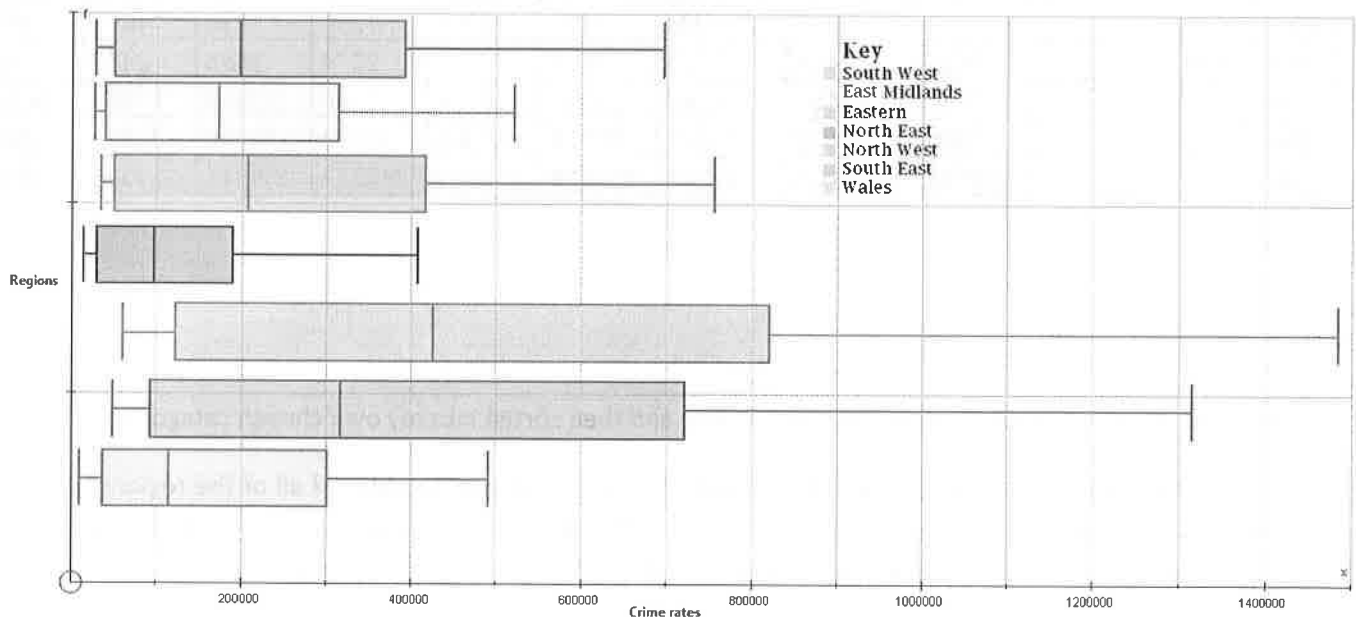# C. Analysis of the number of crimes

In order to show and compare my data I will display the results in three different ways;

Box and Whisker diagrams   - Shows numerical data through 5 different number summaries; smallest observation, lower quartile, median, upper quartile, largest observation. Box and whisker diagrams are non-parametric so they show data without any assumptions of the underlying statistics or reasons.

Scatter graphs   - A mathematical diagram using coordinates from two variables to plot them on a graph in order to be able to make conclusions and see trends. Scatter graphs are used to work out the correlations within pieces of data.

Chi-squared test   - A statistics hypothesis test, it uses all the data that has been collected and different calculations in order to reach the conclusion as to whether the null hypothesis is correct **(see page 16)**

Grand totals

This is a box plot representing the Grand totals for all 7 regions. It is easy to see just by glancing at this graph that the region with the most crimes in total is the North West, closely followed by the South East. Using these box and whisker diagrams I have been able to find out 7 different pieces of information [4] **(see appendix 1)**. The mean for North West is 513462, which is the highest.

The region with the smallest grand total of crime is the North East; it is reasonably smaller than any of the others and also has the smallest mean, with a value of 129722. In the North East they could have

---

[4] Range, Mean, Standard Deviation, Range, Lower Quartile, Median, Upper Quartile

different methods of reducing amount of crime, that are more effective than the rest of the United Kingdom. The North East might also simply be an area with less crime.
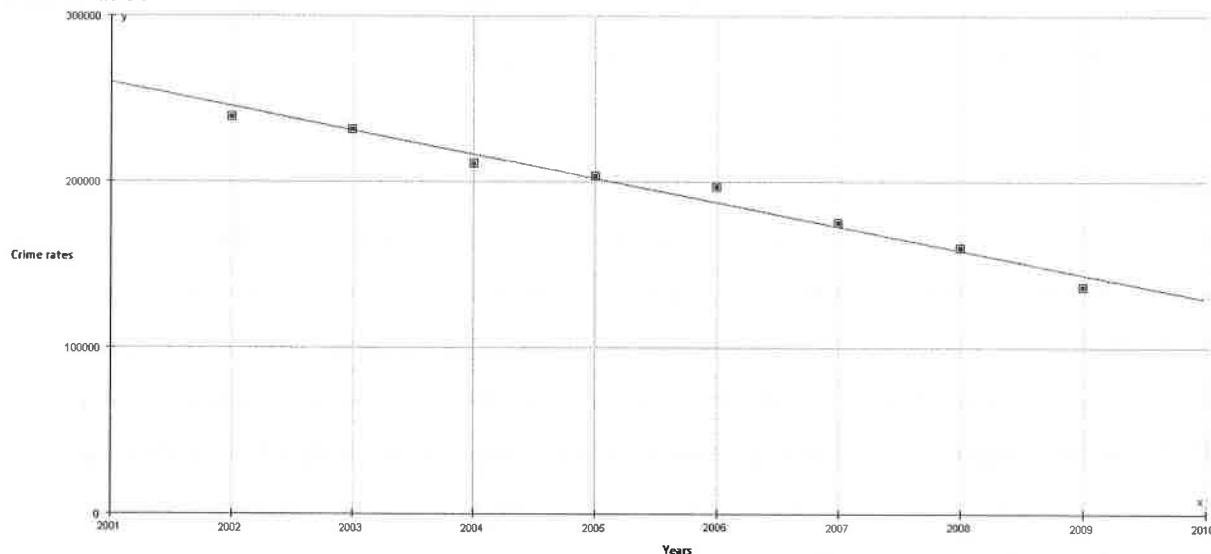
Regions

Looking at the regions separately, I am going to see whether or not the amount of crime have decreased over the years. From looking at the data I have collected **(see appendix 1)** I can come to quick judgements about my data. For example when looking at mean, the region with the highest mean is the South West, with a mean of **254077**, and the one with the lowest is the North East, with a mean of **129722**; without doing extra research into these areas and crimes I cannot state with confidence why there is such a difference between these two reasons, but I can estimate that it is down to the different levels of security which alter the rates.

In order to come to better conclusions I am going to do a scatter graph for every region and then use a regression line, so that I can clearly see the trend.

Linear regressions

A regression line is used to model a relationship between two variables, it allow forecasting/predicting of the events to come, depending on the data. It is the method of fitting a line of a set of data and then finding the equation of the line, the regression line is often known as the line of best fit. To calculate the regression line, you draw a straight line that has the smallest distance from the points on the scatter graph, to the line. I will use regression lines to outline the trends and therefore make a reliable prediction on what is to happen to the number of crimes in the years to come.
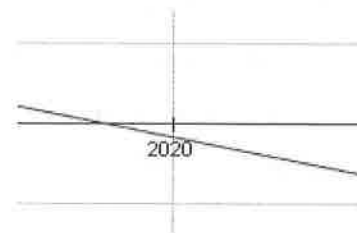
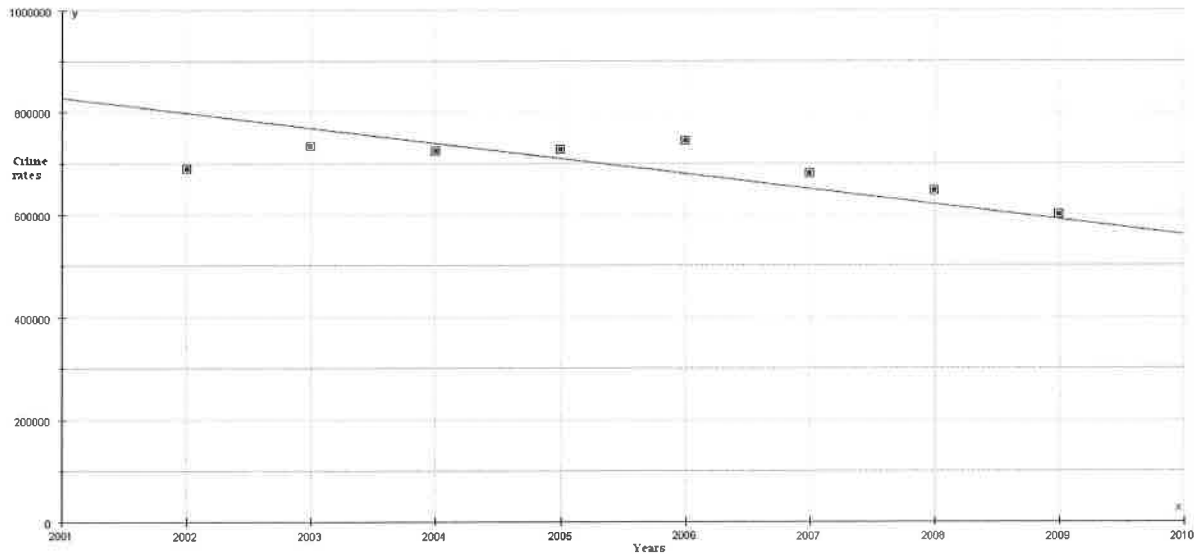
North East

Years

This graph shows the change in the amount of crime between the years of 2002 and 2010 for the North East. As you can see there is a decrease, meaning that there has been less crime as the years go on. The equation for the regression line is $y= 1.41x10^4x + 2.85x10^7$ and the correlation coefficient **-0.9867**, because the number is so close to 1, it means that it is a very strong, negative correlation. I have used a regression line to show how the decrease has happened over the years, it clearly shows a trend of decrease.

Prediction

By using the scatter graph and the regression lines I have been able to predict that the amount of crime will continue to decrease gradually until a certain point. I can use the regression line to predict number of crimes in 2020, however when I look at my graph it says that the amount of crime is going to be negative, which is impossible, so this trend isn't very reliable.
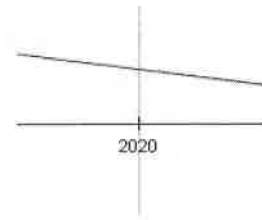
South East

This graph shows the change in the amount of time between the years of 2002 and 2010 for the South East. There is both a slight increase and decrease, but overall the crime rate does go down a small amount. The equation for the regressions line is **y=-1.41x10^{4}x + 2.89x10^7** and the correlation coefficient of this region is **-0.6879**, which means that it is a weak negative correlation. Under normal circumstances a regression line probably would not be necessary here, but I have used one so that I can see the trend of the amount of crime in this region.
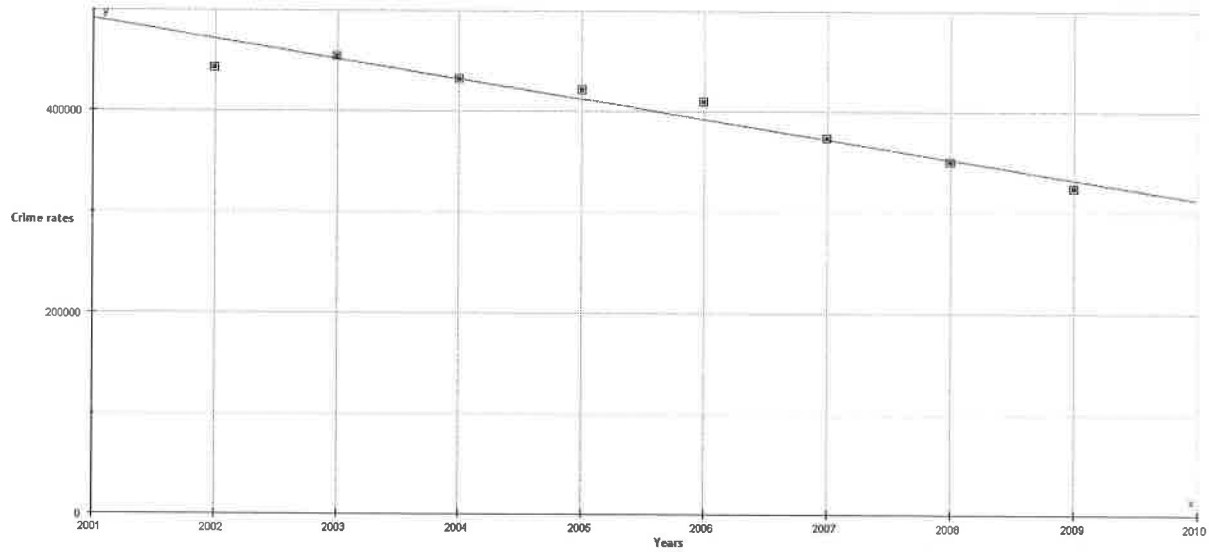
Prediction

From looking at this graph and the regression line I can see the trend, the regression line for the South East isn't as sharp as the one for the North East, implying that the amount of crime is decreasing slower as the time passes. The crime rate for 2020 is around 300000, which is half as small as the 600222 in 2009.
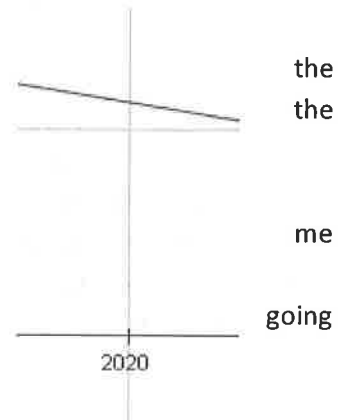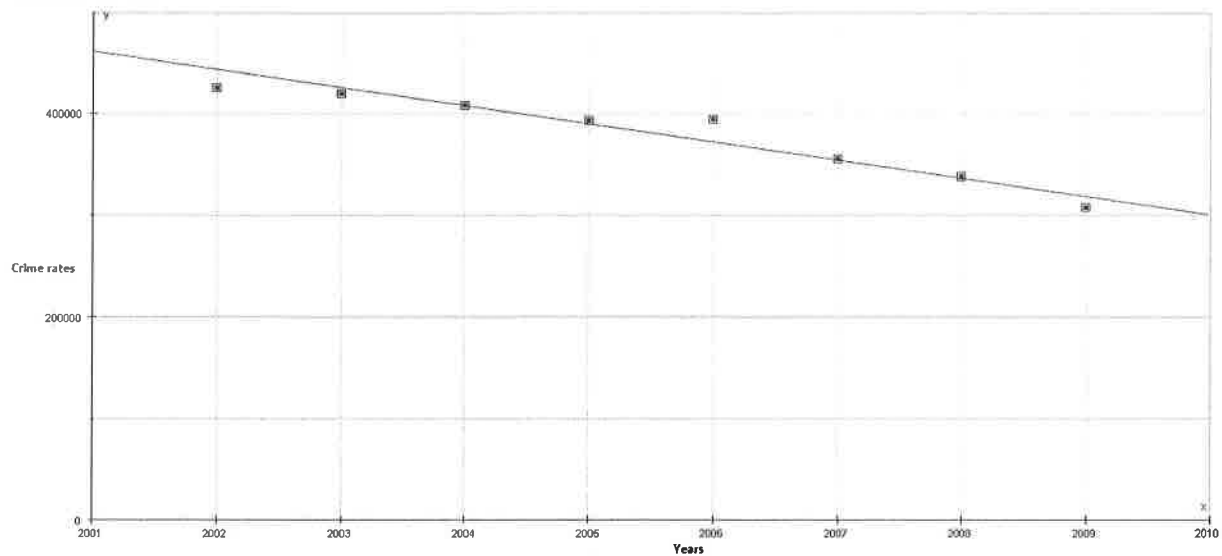
Eastern

The graph shows the change in the amount of crime between the years of 2002 and 2010 for the Eastern region of the UK. The values go down as the years go one, so it shows that the crime rate has decreased in the years. The equation for the regression line is **y = -1.83x10⁴x + 3.71x10⁷** correlation coefficient for this region **is -0.9592**, which is a very strong, negative correlation. I have used a line of best fit on this graph because it shows the decrease in the rate well and makes it clear that the pattern will continue.

Prediction

As before, I have used the regression lines to come to my conclusion about the trend of my data. The regression line is steeper than the regression line for the South East, but not as steep as the North East. From the regression line I can predict that in 2020 the amount of crime will be about 110000, which is smaller the in the South East. The trend gradually decreases, which allows me to believe that it is fairly reliable. However it is difficult to be completely curtain about the trend, because there is no way of telling exactly what is going to happen.
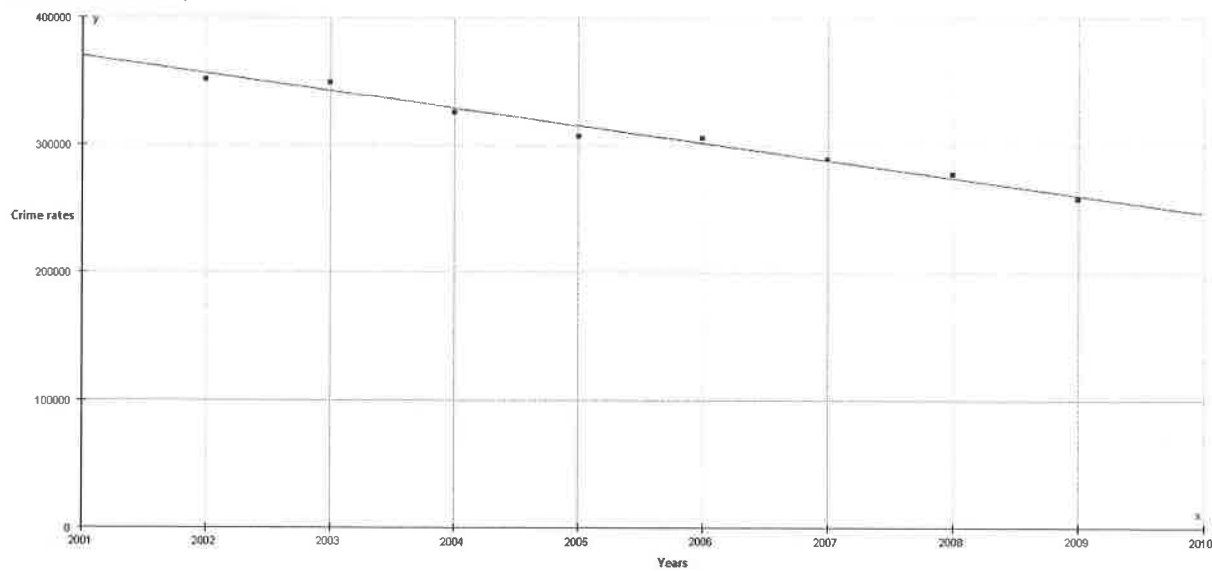
2020

South West

This graph shows the change in the amount of crime between the years of 2002 and 2010 for South West. The values decrease, showing the crime rate decreasing. The equation for this regression line is $y = -1.66 \times 10^4 x + 3.36 \times 10^7$ and the correlation coefficient for this region is **-0.9628**, which is a very strong, negative correlation. I used a line of best fit for this graph because it will show the decrease in the crime rate the best possible way.

Prediction

Again with this regression line it is obvious that there is a gradual decrease. From looking at this trend I have predicted that the crime rate in 2020 will be around 120000, as before I can be fairly confident with this trend, but I can never be 100% sure.

East Midlands

This is a graph showing the amount of crime between the years of 2002 and 2010 for the East Midlands. The amount of crime decrease as the time goes past. The equation for the regression line is
$y = -1.34 \times 10^4 x + 2.72 \times 10^7$ and the correlation coefficient for this region is **-0.9891**, which is a very strong, negative correlation. I have used a regression line to show easily and effectively how the rates decrease over the time.

Prediction

My prediction of the amount of crime in 2020 for the East Midlands is around 110000, the gradual decrease allows me to be moderately confident with the trend.

Wales

This graph shows the amount of crime between the years of 2002 and 2010 for Wales, it decreases, just like with all of the other regions, the crime rate decreases over the years. The equation for the regression line is **y= -1.01x10^{4}x + 2.05x10^7** and the correlation coefficient for this region is **-0.9843**, which is a very strong, negative correlation. The use of a line of best fit shows the decrease of the number of crimes simply and easily.
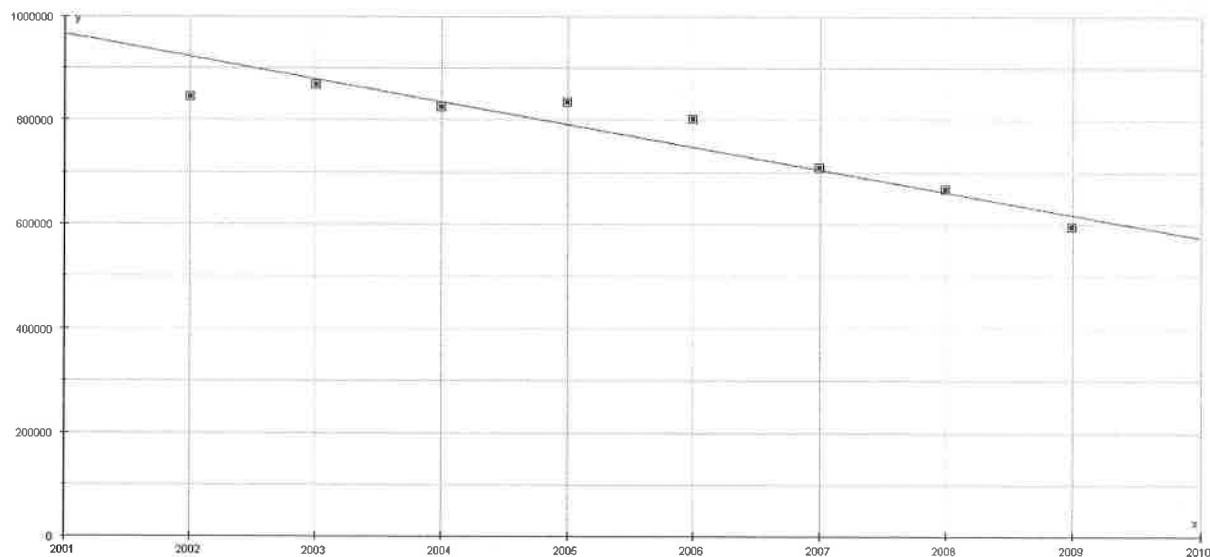
Prediction

From looking at the regression line and points on this scatter graph I have been able to conclude and predict what happens to the amount of crime over a certain period of time. If I look at the trend I'll be able to predict that in 2020 the number of crimes in Wales will be about 110000.

North West

This graph shows the amount of crime between the years of 2002 and 2010 in the North West. Just like in all of the other regions, the amount of crime decreases over time. The correlation coefficient for this region is -0.9238, which is a very strong, negative correlation. The use of a line of best fit shows the decrease over time.

Prediction
From looking at this graph I can use the regression line to look at the trend of the amount of crime, just like all the other 6 regions there is a gradual decrease. In 2020 the amount of crime in this region will be around 200000.

Overall prediction
All of these regions show a gradual decrease in the number of crimes. Following the trends that the regression lines provide shows that in the following years after 2010 the amount of crime should continue to decrease. The trend that shows in these graphs suggests that the crime rate will decrease gradually until they reach zero – and then continue into negative numbers, - but in reality this isn't possible, and it is most likely that it will decrease at a similar rate for a few years before flattening off and remaining more or less the same. Another thing that isn't possible in reality is that the line will be completely straight all the way down, there are always going to be sudden increases and decreases that makes the regression line spiky rather than one solid, straight line.

The Exponential model

I tried to model my data with an exponential function for my project, but because of the large numbers I have from my data, it was too difficult to get the exponential line to correspond with my data, so I was unable to complete the graph. However, if I was able to finish, the graph would have shown that the data would have gotten close to zero, but never reached it, which is a more reliable prediction than the ones I got from the seven graphs above.

Chi-squared Test (X²)

A chi-squared test is used to determine if there is a "significant difference between the expected frequencies and the observed frequencies." [5] It also determines the relationship between one category and the other. For example in my investigation, the chi-squared test will determine the relationship between the number of crimes and the region, the end result will tell me whether these things are dependent on one another or not.

To start a chi-squared test you need your results in a table,

	Burglary	Offences	Violence	Other	Total
South West	1093489	588549	568518	798371	3048927
East Midlands	926011	449584	438018	652031	2465644
Eastern	1171428	621785	539701	884705	3217619
North East	522910	305305	262867	465579	1556661
North West	2042973	1285898	1049493	1783176	6161540
South East	1946135	1062913	1039441	1506347	5554836
Wales	628477	466981	409014	565320	2069792
Total	8331423	4781015	4307052	6655529	24075019

This is the table of results that I will be using for my chi-squared test. I have cut the 12 different crimes into 4 categories, otherwise the test would have been so big that the conclusion from it would not have been any use to me, and I would not have been able to make accurate and reliable conclusions from it.

The next step of the chi-squared test is then to determine the degrees of freedom. You use this equation to work it out:

$$(N° of rows - 1) * (N° of columns - 1)$$

So for my investigation the degree of freedom is going to be 18, (7-1) * (4-1) = 6*3 = 18. I will use this value later on with the critical value table when I have all my results for the chi-squared test.

All of the observed frequencies, in this case are the numbers in white. You'll want to find out the observed frequencies for all of the regions, and all the crimes, the easiest way to do this is to set it out with all the possibilities:

- South west
    - + Burglary   = 1093289
    - + Offences   = 588549
    - + Violence   = 568518
    - + Other   = 798371

5 http://www.enviroliteracy.org/pdf/materials/1210.pdf

Do this for all the regions. Then you want to find the expected frequencies and to do this you need to do:

$$\frac{Total\ of\ region * total\ of\ crime}{Grand\ total}$$

I have used this equation to work out the expected frequencies of all 28 possibilities.

- South West
  - + Burglary   = (3048927*8331423) / 24075019   = 1055114.454
  - + Offences   = (3048927*478115) / 24075019   = 6055480.964
  - + Violence   = (3048927*4307053) / 24075019   = 545456.979
  - + Other   = (3048927*6655529) / 24075019   = 842874.602

Once you have the observed frequencies and the expected frequencies you have to find the difference and then square it, before dividing this value by the expected frequency.

- South West
  - + Burglary   = 1093489 – 1055114.454 = **38374.55**
    - **38374.55^2** = 1472605761,91
      - 1472605761,91/ 1055114.454 = 1395.68
  - + Offences   = 588549 – 6055480.964 = **-16931.96**
    - **-16931.96^2** = 286691416.81
      - 286691416.81/6055480.964 = 473.49
  - + Violence   = 568518 – 545456.979 = **23061.02**
    - **23061.02^2** = 531810689.38
      - 531810689.38/545456.979 = 974.98
  - + Other   = 789371 – 842874.602 = **-44503.60**
    - **-44503.60^2** = 1980570626.52
      - 1980570626.52/842874.602 = 2349.78

NOTE: if the total of all of the differences does not equal 0 then you have done something wrong.

To then find the calculated chi-squared result you have to add all of the final numbers together. Once you have this number you have to compare it to the critical value of chi-squared which is given to you in the critical value table **(see appendix 2).** Using the critical value table is pretty simple; you use the degree of freedom that you should have already calculated to know which column to look at. For the row, it is most common to look at 5% (0.95), because then it means the results will be 95% 'true'.
The point of the critical value table is to that you can determine whether your result sticks to the null hypothesis, or whether it sticks to the alternative hypothesis. If the calculated value of X^2 is bigger than the critical value then your result is according to the alternative hypothesis.

Null hypothesis – The amount of crime is not dependent on the regions.
Alternative hypothesis – The amount of crime is dependent on the regions.

My chi-squared results

	Observed frequency	Expected Frequency	Difference (fo-fe)	(fo-fe)²	(fo-fe)²/fe
South West and Burglary	1093489	1055114,454	38374,55	1472605761,91	1395,68
South West and Offences	588549	605480,9644	-16931,96	286691416,81	473,49
South West and Violence	568518	545456,979	23061,02	531810689,38	974,98
South West and Other	798371	842874,6024	-44503,60	1980570626,52	2349,78
East Midlands and Burglary	926011	853263,0081	72747,99	5292270329,06	6202,39
East Midlands and Offences	449584	489647,8357	-40063,84	1605110933,90	3278,09
East Midlands and Violence	438018	441106,8968	-3088,90	9541283,44	21,63
East Midlands and Other	652031	681626,2594	-29595,26	875879378,27	1284,98
Eastern and Burglary	1171428	1113492,161	57935,84	3356561395,34	3014,45
Eastern and Offences	621785	638981,2072	-17196,21	295709541,07	462,78
Eastern and Violence	539701	575636,1957	-35935,20	1291338291,86	2243,32
Eastern and Other	884705	889509,4357	-4804,44	23082602,51	25,95
North East and Burglary	522910	538699,5233	-15789,52	249309045,86	462,80
North East and Offences	305305	309134,5262	-3829,53	14665270,63	47,44
North East and Violence	262867	278488,6638	-15621,66	244036381,31	876,29
North East and Other	465579	430338,2867	35240,71	1241907874,07	2885,89
North West and Burglary	2042973	2132268,144	-89295,14	7973622679,11	3739,50
North West and Offences	1285898	1223609,218	62288,78	3879892408,71	3170,86
North West and Violence	1049493	1102307,466	-52814,47	2789367785,03	2530,48
North West and Other	1783176	1703355,173	79820,83	6371364416,75	3740,48
South East and Burglary	1946135	1922311,605	23823,40	567554154,61	295,25
South East and Offences	1062913	1103124,954	-40211,95	1617001283,30	1465,84
South East and Violence	1039441	993767,3363	45673,66	2086083554,00	2099,17
South East and Other	1506347	1535632,104	-29285,10	857617334,38	558,48
Wales and Burglary	628477	716274,1045	-87797,10	7708331551,08	10761,71
Wales and Offences	466981	411036,2945	55944,71	3129810077,65	7614,44
Wales and Violence	409014	370288,4626	38725,54	1499667244,96	4050,00
Wales and Other	565320	572193,1385	-6873,14	47240032,22	82,56
			0,00 $X^2 =$		66108,70

Because my degree of freedom is 18, and I want to look at 5% on the critical value table I can see that the critical value for X^2 is 28.869. Comparing this to the calculated value of 66108.70 it is very obvious that there is strong relationship between the amount of crime and the region. Seeing as the two values are so far apart it gives me a high confidence in the result, and I feel that it is correct that these two variables are related.

Smaller chi-squared result

Seeing as I calculated such a large (yet correct) number for the chi-squared test, I have decided that it would be more appropriate and accurate if I broke my chi-squared down into two groups of two, rather than one group of one. I have decided to do this so that I can see easier and make more reliable conclusions on my data. In order to do this, I will keep all the regions together and do the first test with Burglary and Offences, and the second with Violence and Other.

Burglary and Offences

	Burglary	Offences	Total
South West	1093489	588549	1682038
East Midlands	926011	449584	1375595
Eastern	1171428	621785	1793213
North East	522910	305305	828215
North West	2042973	1285898	3328871
South East	1946135	1062913	3009048
Wales	628477	466981	1095458
Total	8331423	4781015	13112438

	Observed frequency	Expected Frequency	Difference (fo-fe)	(fo-fe)²	(fo-fe)²/fe
South west and Burglary	1093489	1068738,711	24750,28947	612576828,7	573,1774
East Midlands and Burglary	926011	874029,9723	51981,02772	2702027243	3091,4583
Eastern and Burglary	1171428	1139377,439	32050,56065	1027238438	901,5787
North East and Burglary	522910	526233,9086	-3323,908595	11048368,35	20,9952
North West and Burglary	2042973	2115108,755	-72135,75502	5203567152	2460,1889
South East and Burglary	1946135	1911898,589	34236,41064	1172131813	613,0722
Wales and Burglary	628477	696035,6249	-67558,62486	4564167793	6557,3767
South west and Offences	588549	613299,2895	-24750,28947	612576828,7	998,8220
East Midlands and Offences	449584	501565,0277	-51981,02772	2702027243	5387,1923
Eastern and Offences	621785	653835,5606	-32050,56065	1027238438	1571,0960
North East and Offences	305305	301981,0914	3323,908595	11048368,35	36,5863
Norht West and Offences	1285898	1213762,245	72135,75502	5203567152	4287,1387
South East and Offences	1062913	1097149,411	-34236,41064	1172131813	1068,3429
Wales and Offences	466981	399422,3751	67558,62486	4564167793	11426,9207
				X² calc value =	38993,9463
				X² crit value =	12.592

These are the results that I received for the test with Burglary and Offences. Because my table is 2 x 7 my degree of freedom will be 6, this means that my critical value is 12.592. There is still a large difference between the calculated value and the critical, this means that we ignore the null hypothesis and take the alternative hypothesis. The difference being so large also gives me confidence in my answer and allows

me to believe that there is no way the burglary and Offence crime cannot be dependent on the region. The amount of burglary and offences are dependent on the region.

Violence and Other.

	Violence	Other	Total
South West	568518	798371	1366889
East Midlands	438018	652031	1090049
Eastern	539701	884705	1424406
North East	262867	465579	728446
North West	1049493	1783176	2832669
South East	1039441	1506347	2545788
Wales	409014	565320	974334
Total	4307052	6655529	10962581

	Observed frequency	Expected Frequency	Difference (fo-fe)	(fo-fe)²	(fo-fe)²/fe
South West and Violence	568518	537032,4745	31485,52551	991338316,6	1845,956
East Midlands and Violence	438018	428265,7273	9752,272655	95106821,93	222,074324
Eastern and Violence	539701	559630,1374	-19929,13738	397170516,9	709,701802
North East and Violence	262867	286196,727	-23329,72696	544276159,8	1901,75536
North West and Violence	1049493	1112917,905	-63424,90517	4022718596	3614,56903
South East and Violence	1039441	1000206,183	39234,81708	1539370871	1539,05354
Wales and Violence	409014	382802,8457	26211,15427	687024608,1	1794,72179
South West and Other	798371	829856,5255	-31485,52551	991338316,6	1194,59001
East Midlands and Other	652031	661783,2727	-9752,272655	95106821,93	143,712943
Eastern and Other	884705	864775,8626	19929,13738	397170516,9	459,275674
North East and Other	465579	442249,273	23329,72696	544276159,8	1230,69995
North West and Other	1783176	1719751,095	63424,90517	4022718596	2339,12838
South East and Other	1506347	1545581,817	-39234,81708	1539370871	995,981484
Wales and Other	565320	591531,1543	-26211,15427	687024608,1	1161,43436
				X² calc value =	19152,6547
				X² crit value =	12.592

These are the results that I calculated for Violence and Other crimes. Just like with the test before the degree of freedom is 6 because it is a 2 x 7 table. This means that the critical value for this is the same as the one before; 12.592. Again, just like above, it is obvious because of the huge difference in numbers that the null hypothesis is wrong and the alternative hypothesis is right. The amount of violence and other crimes are dependent on region.

Validity
The chi-squared test is a very valid test; it uses calculations to provide an answer on whether or not it is dependent. This test is valid because it is a well respected form of determining the dependence, and it is

simple to understand and use with nearly all types of data. However I question the validity of this test, with my investigation, the numbers are so large that seems very unlikely that my variables will be independent, small dif add up. I could have divided all of the data that I received by 1000 so that the data I have with would be smaller. This would have meant that the calculated value would have been a lot smaller, and therefore the results might have been more easily interoperated.

Conclusion

I have had to use all of the data that I have collected, and all the graphs and tests that I have done on it to come to the conclusion about whether or not the amount of crime is dependent on the region. After doing this investigation I have come to the conclusion that the amount of crime is dependent on the region.

The chi-squared test of independence was one of the major tests that helped me reach this conclusion, seeing as the calculated value and the critical value are so far apart it means that there is no way that these two variables cannot be linked, so I am extremely confident with my conclusion.

Looking at the box and whisker diagram, and then all of the scatter graphs and regression lines it is obvious that there are different ranges for different regions, showing that the amount of crime changes according to the region. Using regression lines is a valid way of discovering the trend and therefore the correct conclusion, regression lines use all of the data they are given and generate a line showing the general trend of this information. However, there are limitations, in my investigation it is impossible for the number of crimes to get to zero, or to go below, but with the regression lines it continues decreasing, meaning eventually it does get the number that are impossible for my investigation. What would really happen is it would continue decreasing at a gradual rate until it reaches a certain point, and then it will stop decreasing. Also, it might not actually be a completely straight line down to this point, there will always be sudden increases and falls that means the line will not be straight.

I believe that all of the process I have done throughout my investigation has allowed me to reach a good, reliable conclusion about the amount of crime in Britain and Wales. If I was to do this investigation again there isn't anything I would change, because seeing as I didn't collect the data myself there would not be a problem with that. It would have been practically impossible for me to gather information about the amount of crime; I believe that getting my data off the internet was the simplest, but also most reliable way of doing it.

The validity of my conclusion is affected by the validity of the chi-squared test, if I could be certain about its reliability and whether or not it is valid would have made my conclusion more valid. However, I think that using the data source that I did to get all the crime numbers improves the validity. I trust the website that I got my data from because it has all of the information of crimes and amount of crime all over the UK for many different things. All of the data that is given I calculated by local police, so they are going to know about the crime that is happening in their region. Another reason why I trust this source is because it is collected and calculated by the Home Office, which is the leading government for many times such as; immigration, drug policies, counter-terrorism and crime.

Bibliography

http://rds.homeoffice.gov.uk/rds/soti.html (Accessed 5th October 2010)

http://www.enviroliteracy.org/pdf/materials/1210.pdf (Accessed 6th October 2010)

Bedding, Stephen; Coad, Mal; Forrest, Jane; Fussey, Beryl and Waldman de Tokman, Paula, *Mathematical Studies Course Companion*, Oxford University Press, 2007.

Appendices

Appendix 1

Overall	Topic	Mean	Standard Deviation	Range	Lower Quartile	Median	Upper Quartile
Grand total	South West	254077	218620	669171	50548.8	197971	391776
	East Midlands	205470	168996	494082	39887	173744	313592
	Eastern	268135	237235	721478	50404.3	207715	416236
	North East	129722	120676	392400	28814.8	97034,5	190192
	North West	513462	439818	1,42E+06	123039	425136	820241
	South East	462903	410989	1,27E+06	92669.8	315697	721218
	Wales	172483	151725	481858	36954	115594	299843

Appendix 2

Critical values of the χ^2 distribution (topic 6.9)

$$p = P(X \leq c)$$

p	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
$v=1$	0.00004	0.0002	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

v = number of degrees of freedom

Possible Scores for the Crime Rate Project

Your criteria ↓

Name	Project 2	Project	Old Criteria		New criteria	
Strand	Name	Mark	Justification	Mark	Justification	
1	Introduction	2	Clear title, description of task and plan	3/3	Plan is highly detailed	
2	Information / measurement	3	Candidate had found extensive and relevant data sufficient in both quality and quantity. Data is organised and ready for use. Candidate has grouped data in to regions to help.	2/3	No description of the sampling process is given	
3	Mathematical analysis	4	Simple processes averages, box and whisker diagrams. Chi2 and regression lines done and correct and relevant. The student has discussed the validity of the chi2 test when big numbers are used. Also, the candidate might consider that these are crime frequencies and not rates. This does not make the test irrelevant though and the candidate is penalised in D for not considering this in the interpretation. Processes are relevant and correct.	5/5	The new criteria make this possible. At least one further process and all processes are correct.	
4	Interpretation of Results	2	Candidate correctly interprets most results in the project. As mentioned above, the candidate does not achieve 'comprehensive discussion' I would have expected comment on the fact that data was crime frequency and not rate for example.	2/3	As before	
5	Validity	1	Candidate considers validity of linear model leading to zero or negative predictions. Candidate also discusses the impact of big numbers on chi2 independence tests. On reflection this may even have been enough to award 2 here.	1/1	Considering the previous justification	
6	Structure and communication	3	Project follows logical structure, uses correct notation and reads well. It is a good piece of work. Note, the candidate does use * for multiply on a couple of occasions. This does not detract from the feel and I site the first example project in the TSM from IBO where candidate does the same and is still awarded 3. I am aware it is not always appropriate to include whole database in the body of the project, but in this case it is part of a section on data and is not so long as to detract from the flow of the project.	3/3	Considering the previous justification	
7	Commitment	2	Candidate embraced the whole idea of project work, met obligations and produced a good bit of work.	2/2	As above	
Total	Score	17		17		

read all comments