

Исследование количества автомобильных аварий.

Новиков Кирилл 10 «ИТ»

Задумывались ли вы?

Задумывались ли вы когда-нибудь как зависит количество автомобильных аварий от зарплаты и зависит ли вообще? Быть может людям с разной зарплатой одинаково все равно аккуратно ли ехать, или всё-таки нет?

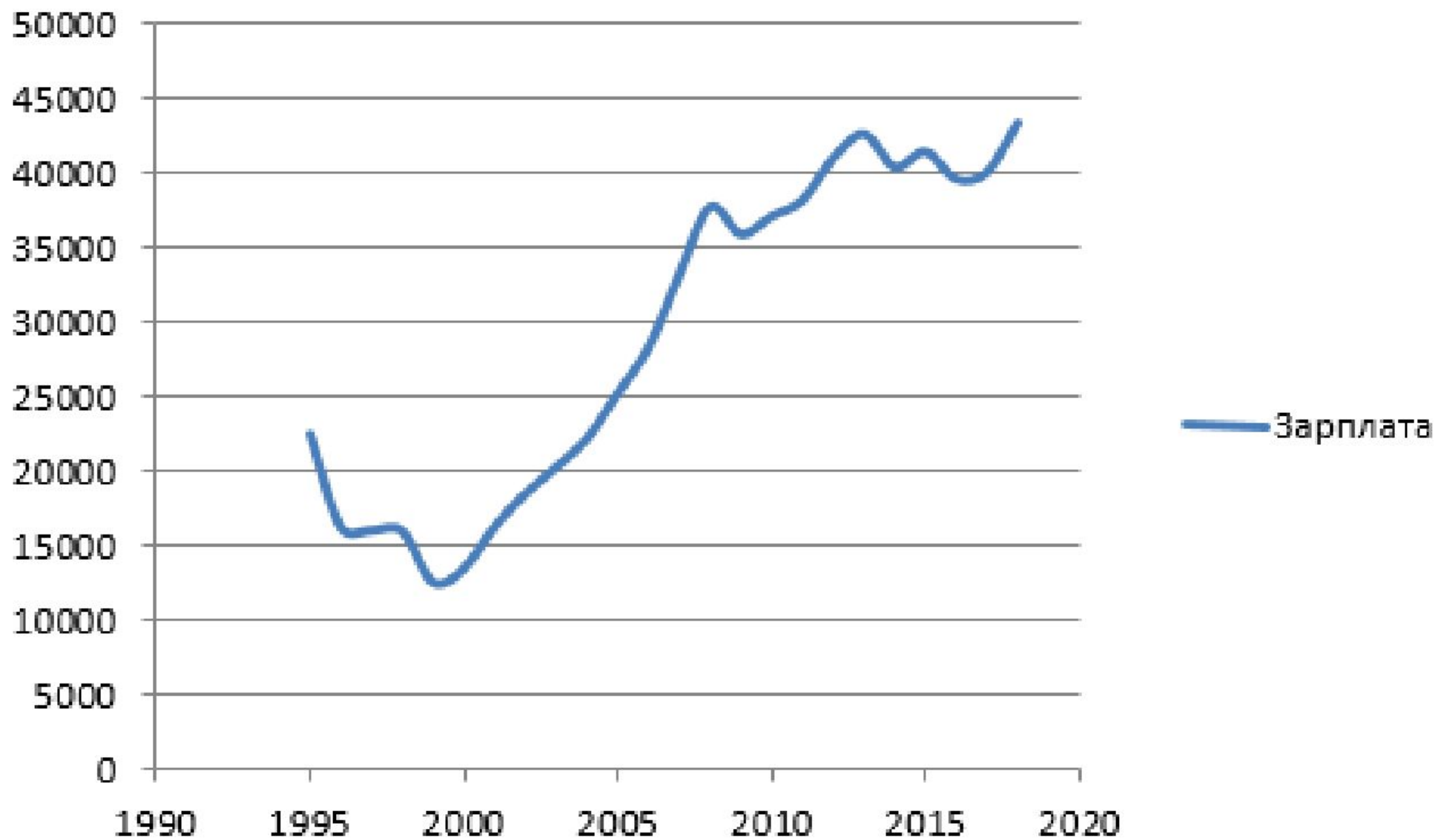


Расчёт зарплат.

Для начала найдем данные о зарплате и инфляции в России. Затем найдем коэффициент роста инфляции. Потом, благодаря нему, узнаем реальный коэффициент роста зарплаты.

	A	B	C	D	E	F	G	H	I	J	K
1	Годь	Доходь	Инфляцьья	Без замены	% инфляцы	Было бы от конца	Было бы от начала	% роста	От 1	Доход, учитывая инфляцию, беря за нуль 2019г	
2	1991	0,548	100	100		0,43	0,55	1,00000000	1,000000	93000	Не будем учитывать эти годы, потому что творилось непойми что и зарплату невозможно корректно рассчитать.
3	1992	6	6793,31	6793,31	67,9331	29,43	37,23	0,14401423	0,144014	577058	
4	1993	58,7	6850,36	6850,36	1,008397968	29,68	37,54	3,72341571	0,536225	83105	
5	1994	220,4	201192,51	201192,51	29,36962583	871,68	1102,53	0,07297933	0,039133	309433	
6	1995	472,4	465358,28	465358,28	2,313000022	2016,20	2550,16	0,72318848	0,028301	22582	Тут средне
7	1996	790,2	566806,39	566806,39	1,218000011	2455,73	3106,10	0,98725811	0,027940	16331	А вот это правильные данные
8	1997	950,2	629155,09	629155,09	1,109999995	2725,86	3447,77	0,99647111	0,027842	16123	
9	1998	1051	1160,16	1160160,00	1,843996843	5026,47	6357,68	0,78584522	0,021879	16066	
10	1999	1523	1583,62	1583620,00	1,365001379	6861,14	8678,24	1,07364621	0,023490	12626	
11	2000	2232	1903,51	1903510,00	1,201999217	8247,09	10431,23	1,20766543	0,028369	13555	
12	2001	3240	2257,56	2257560,00	1,185998498	9781,03	12371,43	1,13463804	0,032188	16370	
13	2002	4360	2598,45	2598450,00	1,150999309	11257,96	14239,51	1,09577697	0,035271	18574	
14	2003	5499	2910,27	2910270,00	1,120002309	12608,94	15948,28	1,09435256	0,038599	20353	
15	2004	6740	3250,77	3250770,00	1,116999454	14084,18	17814,22	1,13633702	0,043861	22274	
16	2005	8555	3605,1	3605100,00	1,108998791	15619,34	19755,95	1,12084503	0,049162	25311	
17	2006	10634	3929,56	3929560,00	1,090000277	17025,09	21533,99	1,17271385	0,057653	28369	
18	2007	13593	4397,18	4397180,00	1,119000601	19051,08	24096,55	1,13670915	0,065534	33269	
19	2008	17290	4982,01	4982010,00	1,133001151	21584,90	27301,41	0,95142369	0,062351	37817	
20	2009	18638	5420,43	5420430,00	1,088000626	23484,38	29703,96	1,03323006	0,064423	35980	
21	2010	20952	5875,74	5875740,00	1,083998871	25457,04	32199,06	1,02892996	0,066287	37176	
22	2011	23369	6234,16	6234160,00	1,060999976	27009,92	34163,20	1,07398782	0,071191	38251	
23	2012	26629	6645,62	6645620,00	1,066000873	28792,60	36418,00	1,03904889	0,073971	41081	
24	2013	29495	7077,58	7077580,00	1,064999202	30664,10	38785,14	0,94842278	0,070156	42686	
25	2014	29792	7884,43	7884430,00	1,114000831	34159,83	43206,68	1,02536095	0,071935	40484	
26	2015	34030	8902,31	8902310,00	1,129100011	38569,86	48784,66	0,95538450	0,068726	41511	
27	2016	36709	9382,15	9382150,00	1,053900617	40648,80	51414,18	1,01239058	0,069577	39659	
28	2017	39167	9617,64	9617640,00	1,025099791	41669,08	52704,67	1,08206500	0,075287	40150	
29	2018	43445	10027,54	10027540,00	1,042619603	43445,00	54950,92			43445	
30					D29/D28			B29/(B28*E28)	J29/H28		

Зарплата

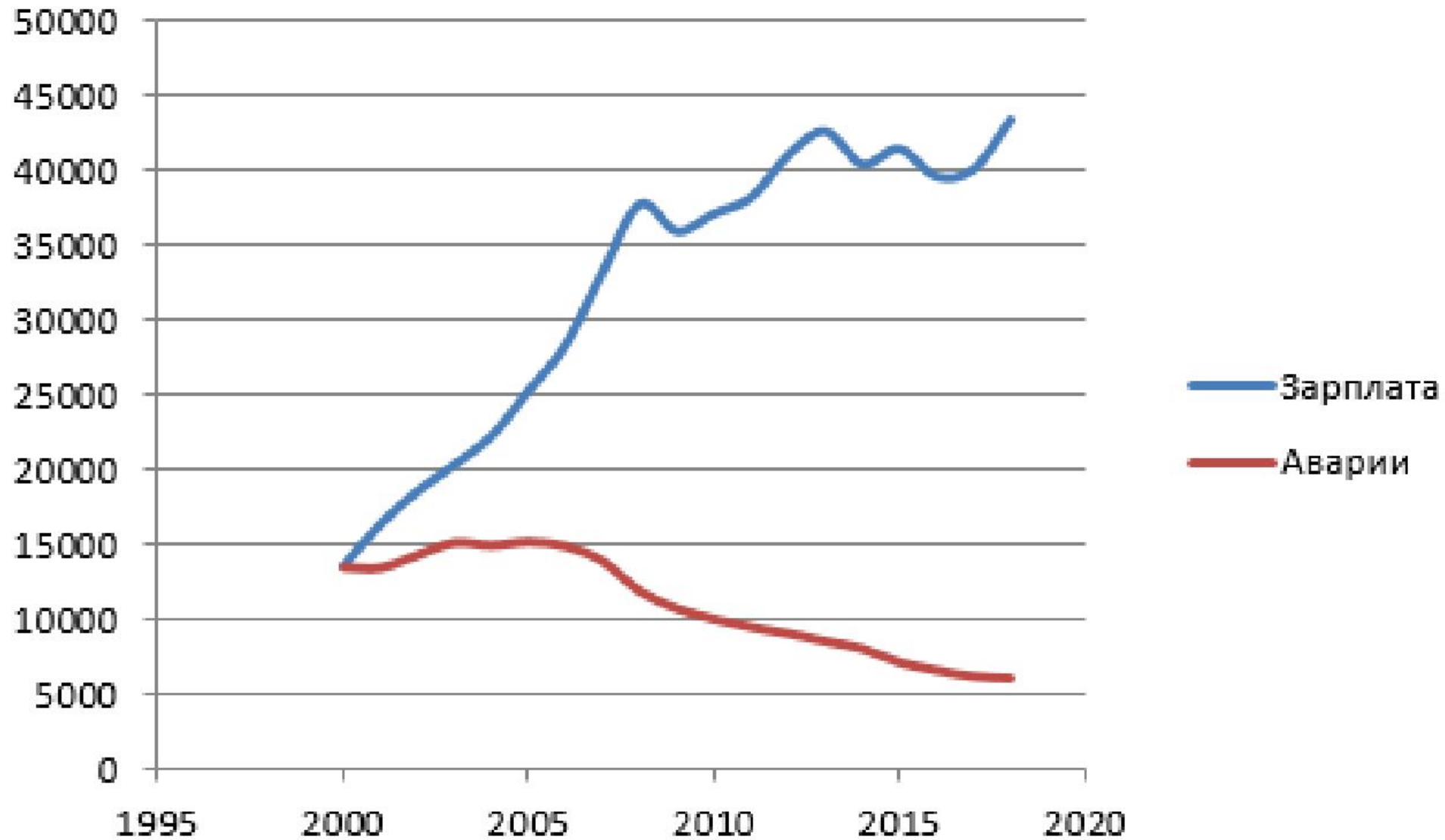


Расчёт количества аварий.

Расчёт количества аварий проще, но нужно искать больше данных. От количества ДТП, жителей в РФ и кол-ва машин на 1000 чел. мы узнаем количество ТС, а затем процент врезавшихся машин. После просто приравниваем первое значение к зарплате для простоты понимания.

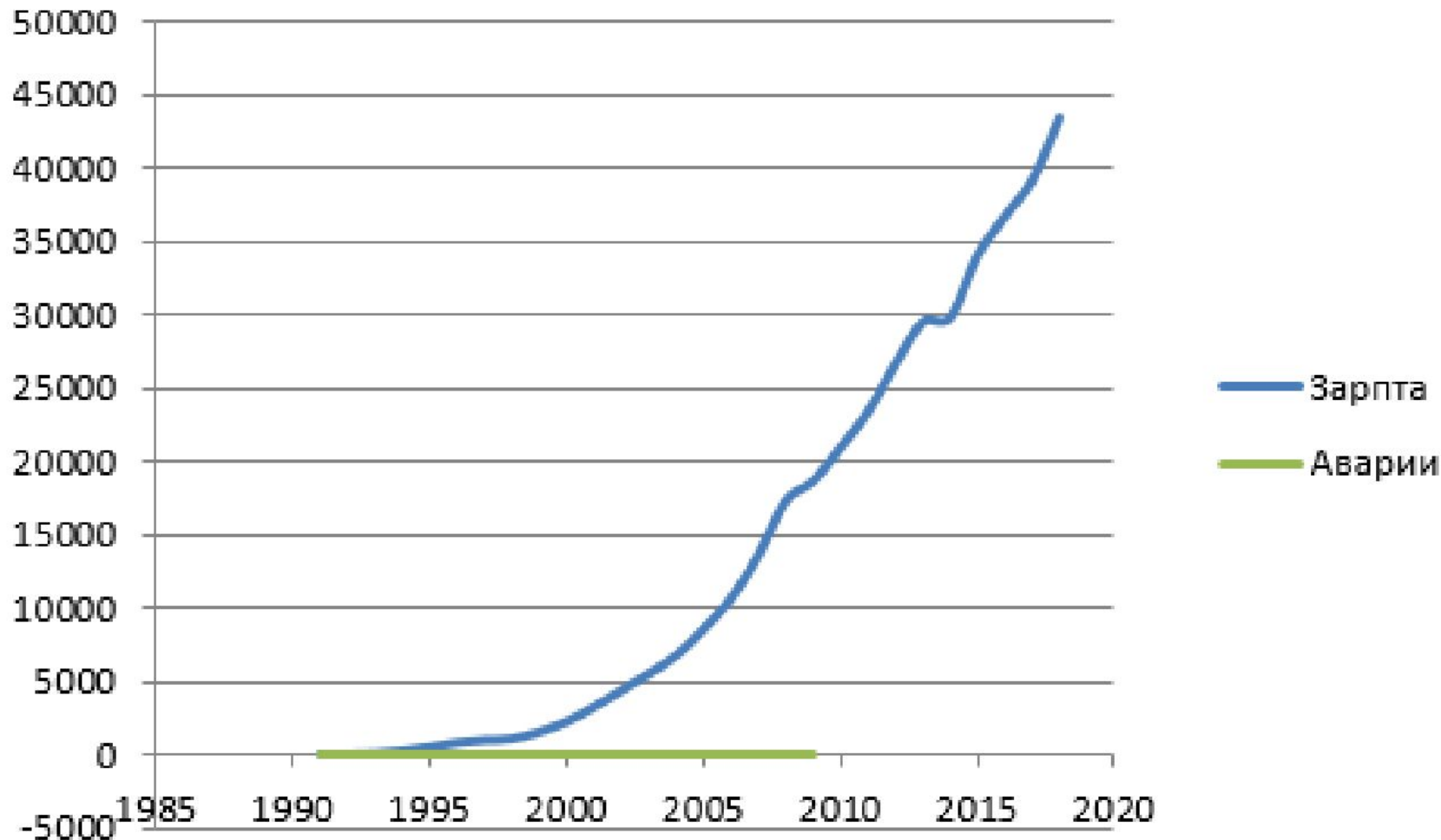
	L	M	N	O	P	Q	R	S
1	Годъ	Количество ДТП	Легковых машин у населения на 1000 ч	Население РФ	Количество ТС	"Процент" врезавшихся	"Увеличенный" процент	ДТП пропорционально зарплате
2	1991	нет данных						
3	1992	нет данных						
4	1993	нет данных						
5	1994	нет данных						
6	1995	нет данных						Коэффициент 16,49863746
7	1996	нет данных						J11/R11
8	1997	нет данных						
9	1998	нет данных						
10	1999	нет данных						
11	2000	157495	130,5	146890128	19169162	0,0082160609	821,6060902	13555
12	2001	164401	137,2	146303611	20072855	0,0081902149	819,0214919	13513
13	2002	184360	145,8	145166731	21165309	0,0087104798	871,0479809	14371
14	2003	204267	153	144963650	22179438	0,0092097462	920,9746246	15195
15	2004	208558	158,9	144168205	22908328	0,0091040255	910,402549	15020
16	2005	223342	168,4	143474219	24161058	0,0092438831	924,3883093	15251
17	2006	229140	177	142753551	25267379	0,0090686099	906,8609937	14962
18	2007	233809	194,4	142220968	27647756	0,0084567080	845,6707969	13952
19	2008	218322	212,3	142008838	30148476	0,0072415600	724,1559997	11948
20	2009	203603	219,4	141903979	31133733	0,0065396270	653,9626971	10789
21	2010	199431	228,4	142856536	32628433	0,0061121845	611,2184458	10084
22	2011	199868	242	142865433	34573435	0,0057809703	578,097031	9538
23	2012	203597	257,5	143056383	36837019	0,0055269674	552,6967372	9119
24	2013	204068	273,1	143347059	39148082	0,0052127203	521,272028	8600
25	2014	199720	283,3	143666931	40700842	0,0049070238	490,7023845	8096
26	2015	184000	288,8	146267288	42241993	0,0043558551	435,5855108	7187
27	2016	173694	294	146544710	43084145	0,0040315063	403,1506278	6651
28	2017	169432	305	146804372	44775333	0,0037840478	378,4047754	6243
29	2018	168099	309,1	146880432	45400742	0,0037025607	370,256067	6109
30					O29/1000*N29	M29/P29	Q29*100000	R29*\$S\$6

Аварии и зарплаты.



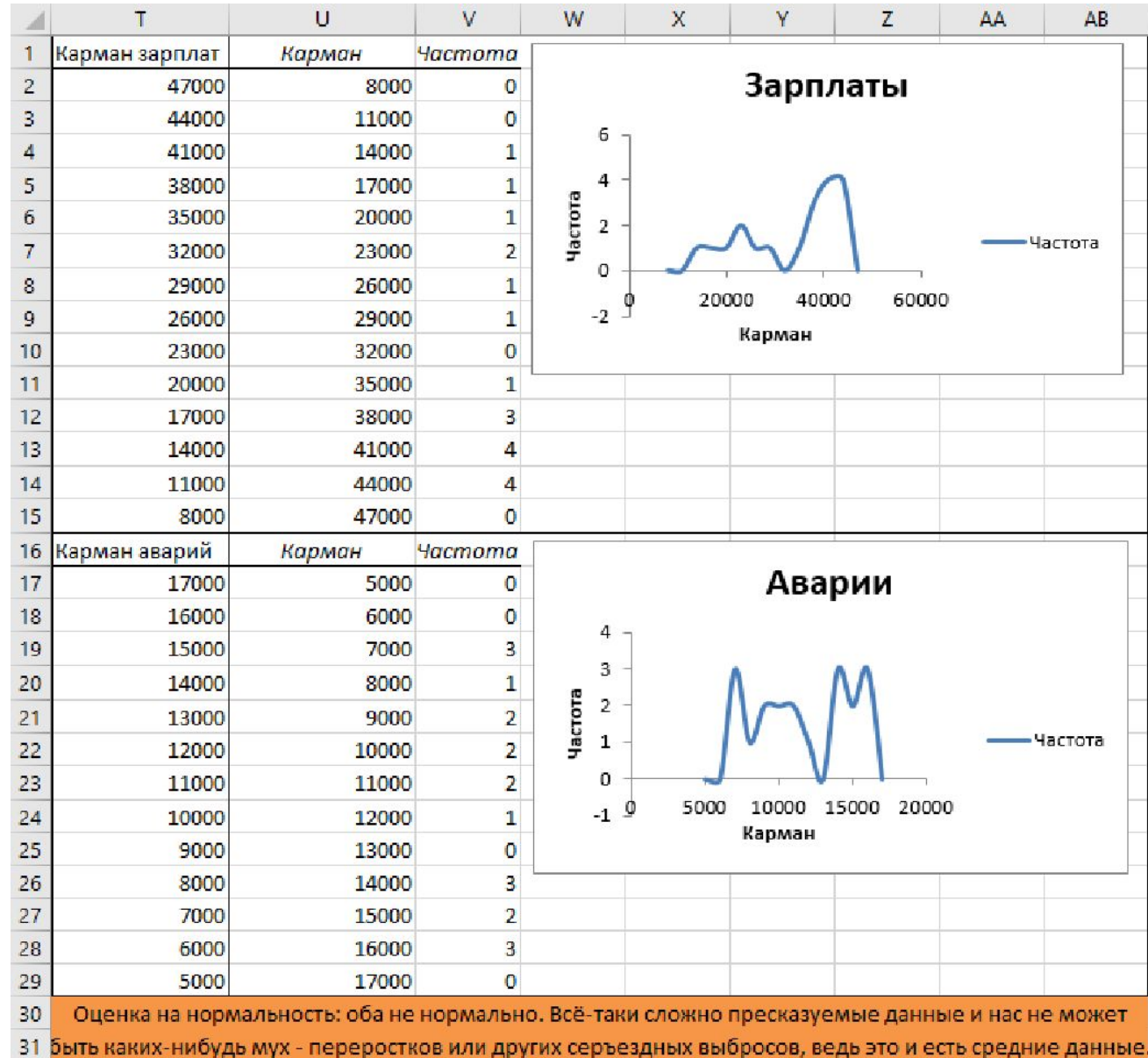
Пачаму так слажна?!

Зачем мы производили какие-либо операции над количеством аварий и зарплаты? Для того чтобы нам было лучше видно количество аварий и рост зарплаты. Если этого не сделать то вот, что будет на графике:



Постороение гистограммы.

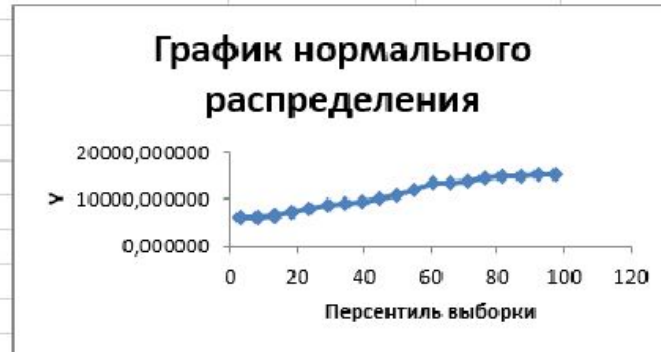
Построим то, что я называю «холмиками». Однако они не холмообразной формы, значит оценивая на нормальность надо об этом сказать. Однако это объясняется тем, что у нас не очень много данных, потому что если взять больше лет, то естественно пропуски закроются. Так же у нас ни от чего толком не зависящие данные, у нас же нет определенного среднего числа и среднего отклонения.



Регрессия.

Воспользуемся функцией регрессия, чтобы узнать коэффициенты идеальной прямой по формуле $Y=a+bX$. $a=20167$, $b=-0.28$.

	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	
1												
2												
3												
4												
5												
6	Вывод Итогов										Вывод вероятности	
7												
8	<i>Регрессионная статистика</i>										<i>Перцентиль</i>	<i>Y</i>
9	Множественный R	0,833584765								2,631578947	6108,720616	
10	R-квадрат	0,69486356								7,894736842	6243,163200	
11	Нормированный R-квадрат	0,676914358								13,15789474	6651,436048	
12	Стандартная ошибка	1904,304427								18,42105263	7186,567423	
13	Наблюдения	19								23,68421053	8095,920740	
14										28,94736842	8600,278206	
15	Дисперсионный анализ										34,21052632	9118,743089
16		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>				39,47368421	9537,813328	
17	Регрессия	1	140387079	140387079	38,71278214	9,29001E-06				44,73684211	10084,271544	
18	Остаток	17	61648380,99	3626375,353						50,00000000	10789,493449	
19	Итого	18	202035460							55,26315789	11947,587300	
20										60,52631579	13512,738663	
21	$Y=A+B*X$	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>	65,78947368	13555,381013	
22	Y-пересечение (A)	20167,96718	1527,287396	13,20508978	2,29685E-10	16945,67244	23390,26192	16945,67244	23390,26192	71,05263158	13952,415884	
23	Переменная X 1 (B)	-0,280712035	0,045116338	-6,22195967	9,29001E-06	-0,375899188	-0,185524883	-0,375899188	-0,185524883	76,31578947	14371,104843	
24										81,57894737	14961,970758	
25										86,84210526	15020,401595	
26										92,10526316	15194,826436	
27										97,36842105	15251,147584	
28												
29												



Так же можем рассчитать коэффициенты самостоятельно по формуле и проверить сойдётся ли.

$$Y_{\text{расчетн}} = a + bX$$

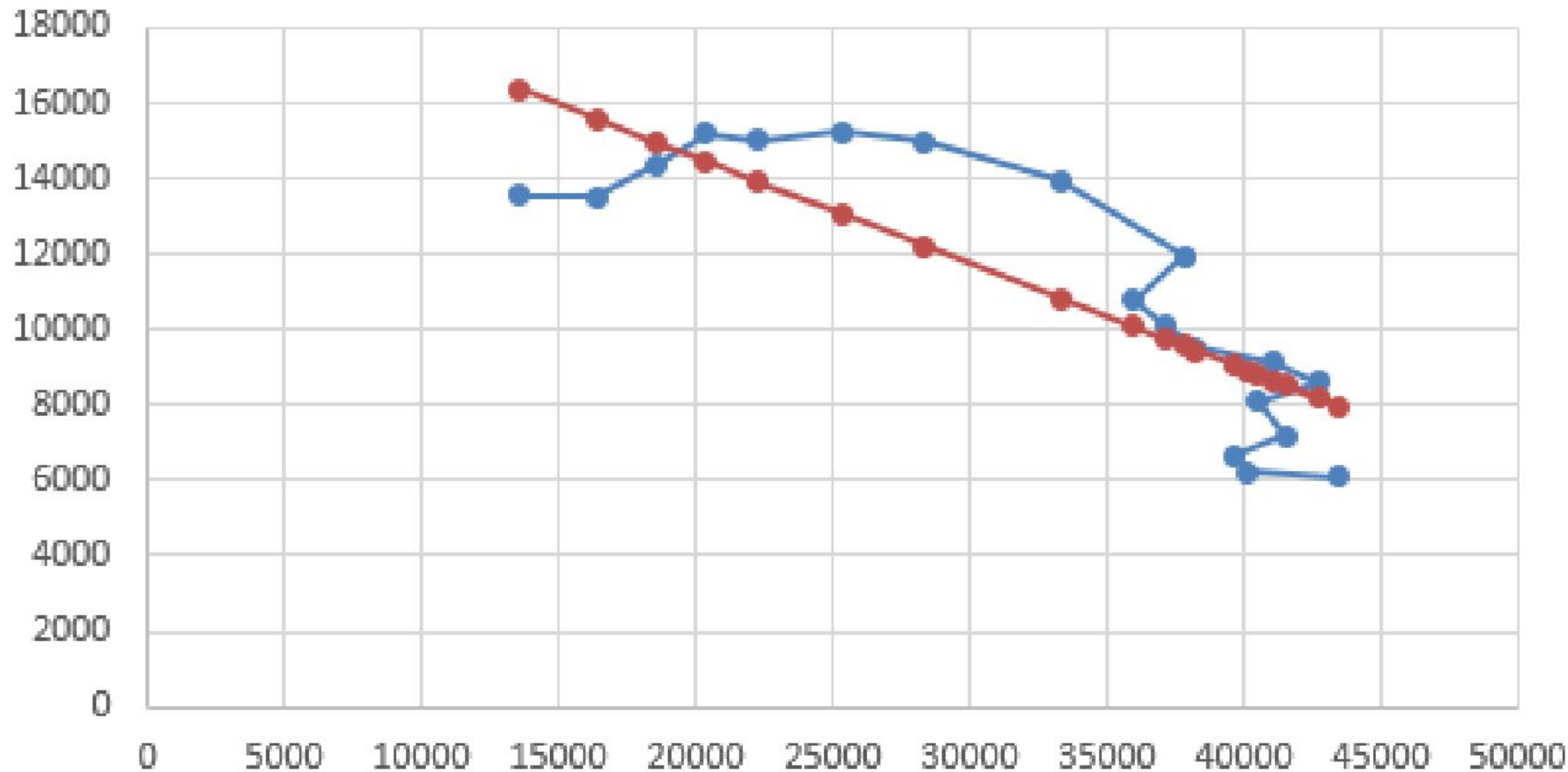
$$a = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n \sum x_i^2 - \sum x_i \sum x_i} \quad b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \sum x_i \sum x_i}$$

	AN	AO	AP	AQ	AR	AS
1		Y(ращ)=a+bX				
2	Формула	a=((сум)y(сум)x2-(сум)x(сум)xy)/(n(сум)x2-(сум)x(сум)x)				
3		b=(n(сум)xy-(сум)x(сум)y)/(n(сум)x2-(сум)x(сум)x)				
4		Наши данные		Расчётные данные		
5		X	Y	x2	xy	Y ращ
6	1	13555	13555	183748354,4	183748354,4	16362,80859
7	2	16370	13513	267988852,3	221208464,9	15572,60869
8	3	18574	14371	345009782,4	266935209,6	14953,89862
9	4	20353	15195	414262616,8	309267024,4	14454,51094
10	5	22274	15020	496124028,9	334562049	13915,43171
11	6	25311	15251	640626021,4	386015560,2	13062,97964
12	7	28369	14962	804814354,8	424459891,3	12204,37723
13	8	33269	13952	1106827189	464183097,6	10828,95498
14	9	37817	11948	1430139930	451824190	9552,226542
15	10	35980	10789	1294572743	388207824,9	10067,9
16	11	37176	10084	1382039717	374890809,4	9732,274161
17	12	38251	9538	1463161114	364833654,6	9430,369973
18	13	41081	9119	1687682941	374610900,5	8635,918594
19	14	42686	8600	1822060637	367108054,4	8185,6049
20	15	40484	8096	1638954052	327755234,8	8803,621847
21	16	41511	7187	1723139065	298319518	8515,411216
22	17	39659	6651	1572811641	263787235,9	9035,295772
23	18	40150	6243	1612029205	250663523,8	8897,35552
24	19	43445	6109	1887468025	265393367,1	7972,432802
25	Сум	616316	210184	21773460269	6317773965	
26	A ращ	20167,97				
27	B ращ	-0,28071				
28	A уст	20167,97				
29	B уст	-0,28071				

Фига, оно сошлось!

Наконец построим график с идеальной прямой по рассчитанным точи

Регрессия



Коэффициент корреляции.

Вычисляем коэффициент корреляции Пирсона и Спирмена. Я проверил оба коэффициента, потому что Пирсон очень чувствителен к выбросам, а Спирмен ещё видит явную не прямую схожесть, если она, конечно, есть.

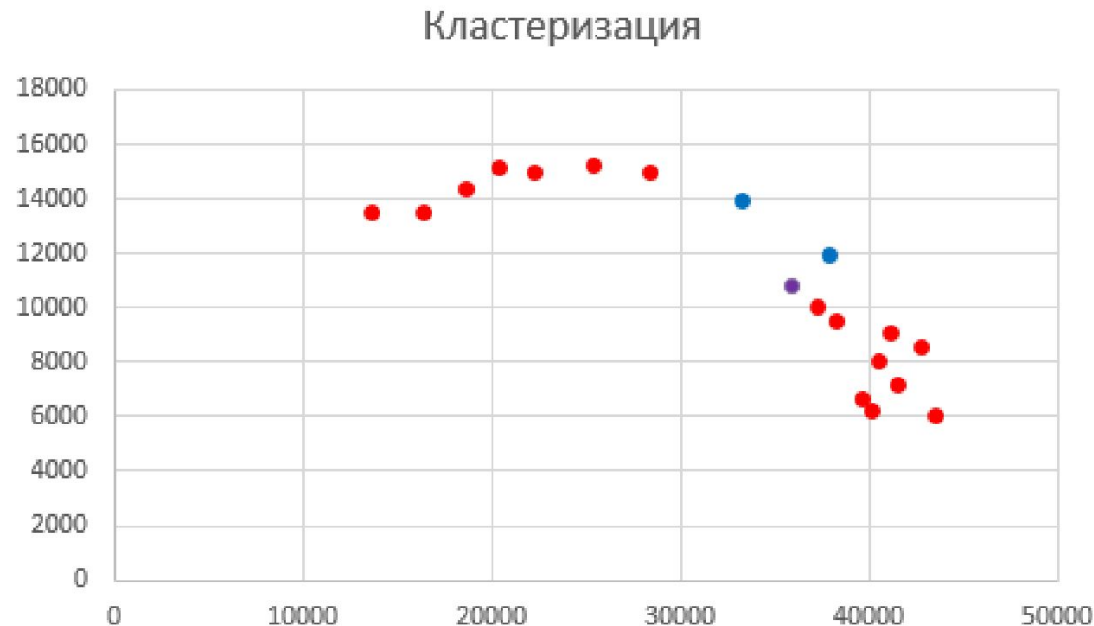
	AT	AU	AV	AW	AX	AY	AZ
1	Показатель	Значение	X	Y	Ранг(X)	Ранг(Y)	Формула
2	Коэффициент корреляции Пирсона K	-0,833584765	13555	13555	19	7	КОРРЕЛ(AO6:AO24;AP6:AP24)
3	n	19	16370	13513	18	8	СЧЁТ(AO6:AO24)
4	z	-1,199771068	18574	14371	17	5	ФИШЕР(AU2)
5	стандартная ошибка	0,25	20353	15195	16	2	1/КОРЕНЬ(AU3-3)
6	квантиль_95%	1,959963985	22274	15020	15	3	НОРМ.СТ.ОБР((1+0,95)/2)
7	Нижняя 95% граница z	-1,689762064	25311	15251	14	1	AU4-AU5*AU6
8	Верхняя 95% граница z	-0,709780072	28369	14962	13	4	AU4+AU5*AU6
9	Нижняя 95% граница K	-0,934116898	33269	13952	12	6	ФИШЕРОБР(AU7)
10	Верхняя 95% граница K	-0,610538903	37817	11948	9	9	ФИШЕРОБР(AU8)
11	Коэффициент корреляции Спирмена C	-0,836842105	35980	10789	11	10	КОРРЕЛ(AX2:AX20;AY2:AY20)
12	n	19	37176	10084	10	11	СЧЁТ(AO6:AO24)
13	z	-1,210542359	38251	9538	8	12	ФИШЕР(AU11)
14	стандартная ошибка	0,25	41081	9119	4	13	1/КОРЕНЬ(AU12-3)
15	квантиль_95%	1,959963985	42686	8600	2	14	НОРМ.СТ.ОБР((1+0,95)/2)
16	Нижняя 95% граница z	-1,700533355	40484	8096	5	15	AU13-AU14*AU15
17	Верхняя 95% граница z	-0,720551363	41511	7187	3	16	AU13+AU14*AU15
18	Нижняя 95% граница C	-0,935475712	39659	6651	7	17	ФИШЕРОБР(AU16)
19	Верхняя 95% граница C	-0,617250713	40150	6243	6	18	ФИШЕРОБР(AU17)
20	Показатель	Значение	43445	6109	1	19	Формула

Кластеризация.

Разделим на 2 кластера наши данные. Это не нужно для аналитики, а нужно по заданию. По итогам деления у меня получилось странное разделение на группы: в одной только 2-3 точки, а во второй все остальные. При чём одна точка при

разделении очень явно проявляла признаков принадлежности к одной или к другой группе.

Может быть



	BA	BB	BC
1		X	Y
2	1	13555	13555
3	2	16370	13513
4	3	18574	14371
5	4	20353	15195
6	5	22274	15020
7	6	25311	15251
8	7	28369	14962
9	8	33269	13952
10	9	37817	11948
11	10	35980	10789
12	11	37176	10084
13	12	38251	9538
14	13	41081	9119
15	14	42686	8600
16	15	40484	8096
17	16	41511	7187
18	17	39659	6651
19	18	40150	6243
20	19	43445	6109

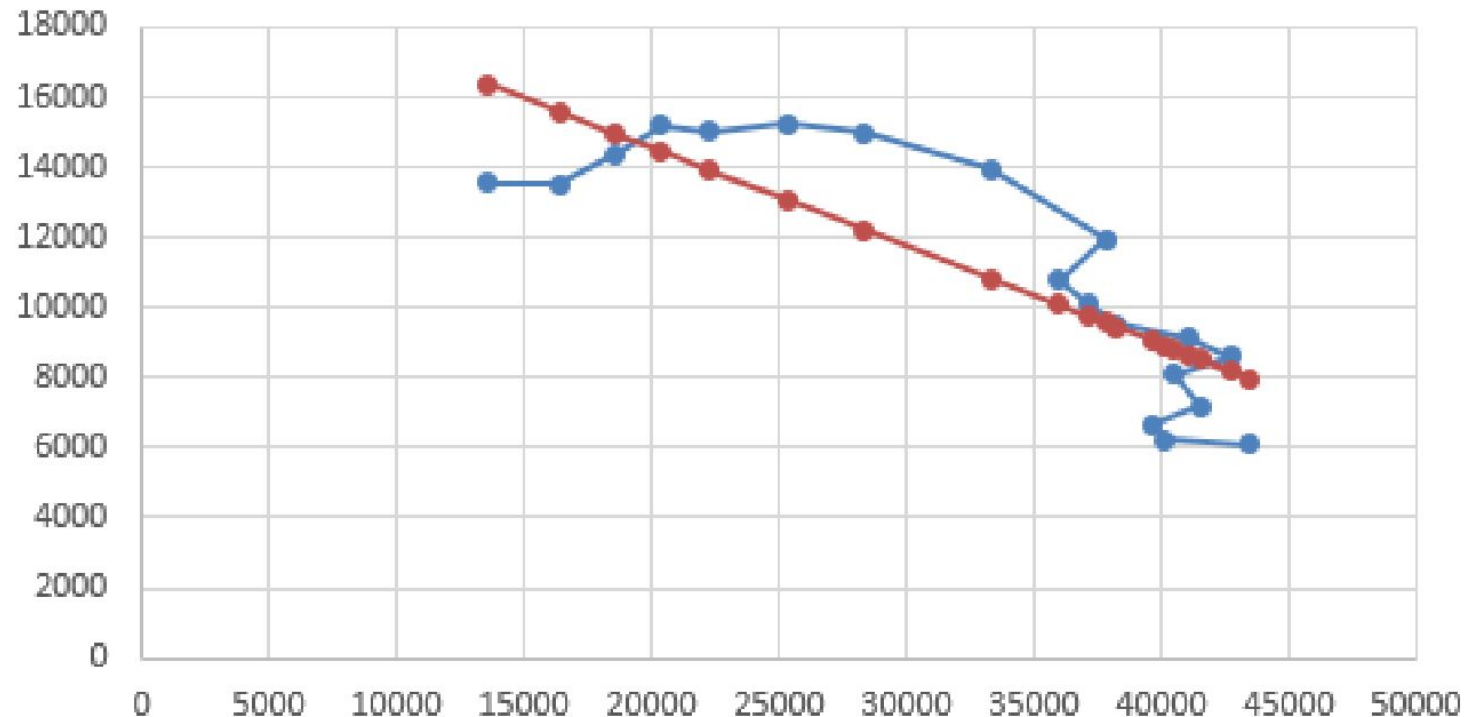
Подъитоги.

По графику не видно явной линейной зависимости, да и может при увеличении зарплаты количество аварий наоборот, уменьшаться? Я думаю нет, и вот почему: когда я искал данные по ДТП я много видел новостей о том, что благодаря внедрениям новых правил, установки камер и повышения эффективности работы ДПС аварийность снижается, и с этим мало кто поспорит. Но что если мы постараемся убрать данное влияние, просто добавив число аварий, рассчитанное алгебраической последовательностью ($a=1500$)

Коэффициент корреляции:

-0,84

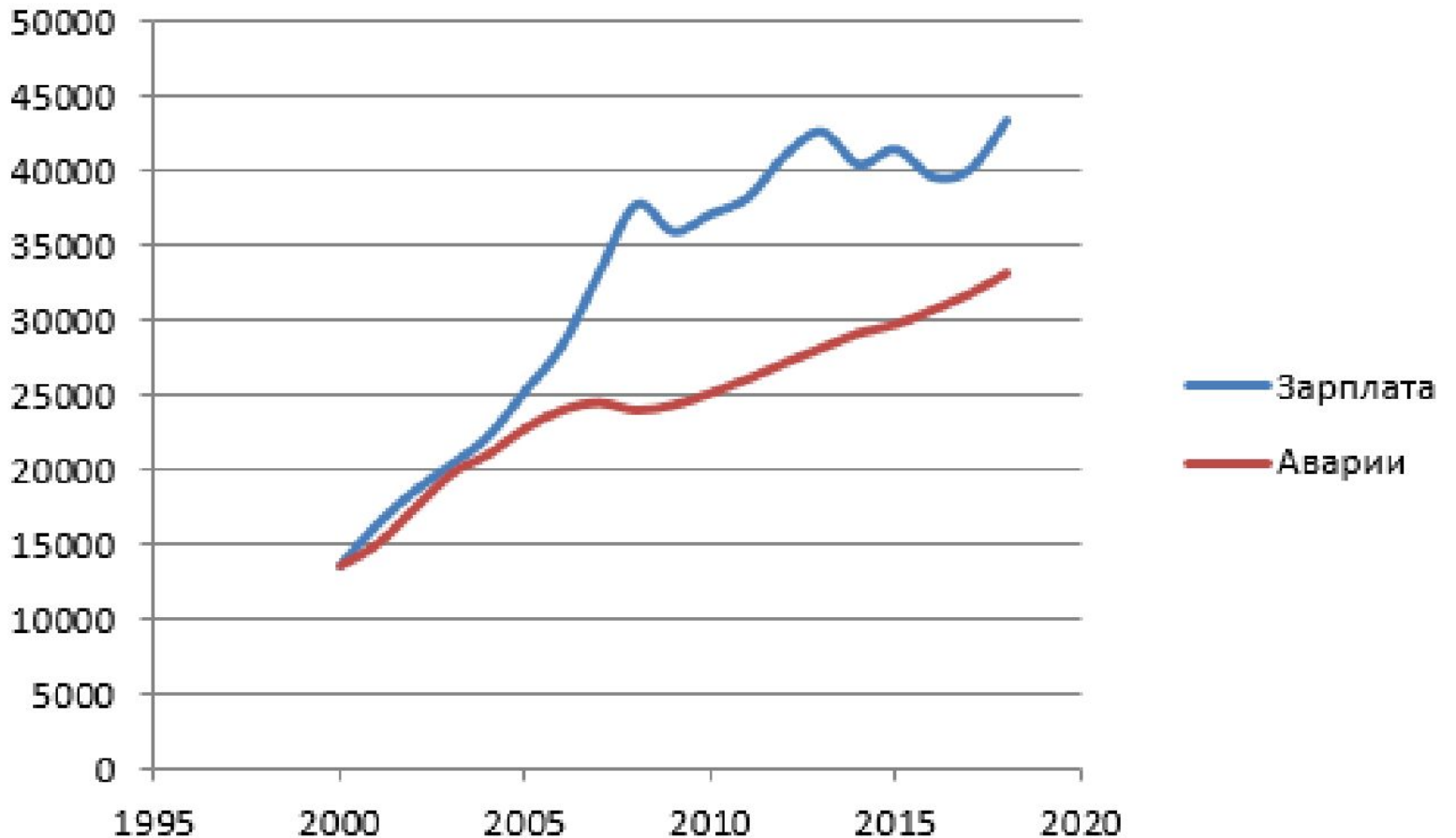
Регрессия



Итак...

Сделав всё вышесказанное мы получили такой график. По нему уже лучше видна зависимость. Сравним линии графически: до 2008 года идет подъём, а потом небольшой спад на обеих линиях. Затем ситуация снова выравнивается, но в 2014 году (обвал рубля) снижение зарплат и небольшое прекращение роста в числе аварий. Прошу заметить, что всё выше сказанное можно увидеть и по первому

Коэффициент корреляции:
0,94



ИТОГ.

В итоге можно сказать, что количество аварий **ЗАВИСИТ** от зарплат, однако это не основной фактор. Мы видели явное сходство графиков зарплат и ДТП, однако последнее все-равно неуклонно снижается, даже при росте зарплат в последнее время.

Спасибо за
внимание!