EXCEL EXERCISE #11: Bivariate Regression

1. Enter the data on the attached page for STATE, ROBBERY (robberies per million residents), and POLICE (police employees per 100,000 population) into cells A1:C52.

	A	B	C
1	State Crime Data Set		
2	STATE	POLICE (X)	ROBBERY (Y)
3	AL	230	1321
4	AK	350	900
5	AZ	310	1936
6	AR	190	809
7	CA	270	3842
8	CO	270	1601
9	CT	260	2180
10	DE	280	1370
11	FL	300	3555
12	GA	240	1976
13	HI	290	1902
14	ID	240	468
15	IL	320	2170
16	IN	210	1414
17	IA	200	549
18	KS	230	1131
19	KY	200	952
20	LA	290	1970
21	ME	200	308
22	MD	310	3927
23	MA	290	2355
24	MI	250	2440
25	MN	190	991
26	MS	200	810
27	MO	280	2236
28	MT	240	340
29	NE	220	822
30	NV	360	4606
31	NH	240	420
32	NJ	350	3037
33	NM	280	1279
34	NY	370	6413
35	NC	220	823
36	ND	180	77
37	OH	210	2237
38	OK	230	1049
39	OR	240	1524
40	PA	240	1779
41	RI	280	1186
42	SC	240	1181
43	SD	200	201
44	TN	210	1806
45	TX	240	2085
46	UT	240	802
47	VT	200	389
48	VA	230	1201
49	WA	210	1351
50	WV	180	485
51	WI	240	707
52	WY	310	444

Regression Coefficients

2. Regress ROBBERY (Y) on POLICE (X). Enter the following labels.

A54: **Sum =**
A55: **N =**
A56: **Mean =**
E2: **(Y_i-Y_{bar})**
F2: **(X_i-X_{bar})**
G2: **$(Y_i-Y_{bar})(X_i-X_{bar})$**
H2: **$(X_i-X_{bar})^2$**
I2: **$(X_i-X_{bar})^2$**
F54: **Sum =**

3. Enter the following formulas.

B54: **=sum(b3:b52)**
C54: **=sum(c3:c52)**
B55: **=count(b3:b52)**
C55: **=count(c3:c52)**
B56: **=average(b3:b52)**
C56: **=average(c3:c52)**

4. Enter the following formula.

E3: **=c3-c$56**

5. Copy the formula in cell E3 to cells E4 to E52.

6. Enter the following formula.

F3: **=b3-b$56**

7. Copy the formula in cell F3 to cells F4 to F52.

8. Enter the following formula.

G3: **=f3*e3**

9. Copy the formula in cell G3 to cells G4 to G52.

10. Enter the following formula.

H3: **=f3^2**

11. Copy the formula in cell H3 to cells H4 to H52.

12. Enter the following formula.

I3: **=e3^2**

13. Copy the formula in cell I3 to cells I4 to I52.

14. Enter the following formulas.
    G54: **=sum(g3:g52)**
    H54: **=sum(h3:h52)**
    I54: **=sum(i3:i52)**

15. Enter the following labels.
    A60: **Slope =**
    A61: **Intercept =**
    B59: **Coefficient**

16. Enter the following formulas.
    B60: **=g54/h54**
    B61: **=c56-b60*b56**

Your spreadsheet should look similar to the one below.

	A	B	C	D	E	F	G	H	I
1	State Crime Data Set								
2	**STATE**	**POLICE (X)**	**ROBBERY (Y)**		(Y_i-Y_{bar})	(X_i-X_{bar})	$(X_i-X_{bar})(Yi-Y_{bar})$	$(X_i-X_{bar})^2$	$(Y_i-Y_{bar})^2$
3	AL	230	1321		-266.14	-21.2	5642.168	449.44	70830.4996
4	AK	350	900		-687.14	98.8	-67889.432	9761.44	472161.3796
5	AZ	310	1936		348.86	58.8	20512.968	3457.44	121703.2996
6	AR	190	809		-778.14	-61.2	47622.168	3745.44	605501.8596
7	CA	270	3842		2254.86	18.8	42391.368	353.44	5084393.62
8	CO	270	1601		13.86	18.8	260.568	353.44	192.0996
9	CT	260	2180		592.86	8.8	5217.168	77.44	351482.9796
10	DE	280	1370		-217.14	28.8	-6253.632	829.44	47149.7796
11	FL	300	3555		1967.86	48.8	96031.568	2381.44	3872472.98
12	GA	240	1976		388.86	-11.2	-4355.232	125.44	151212.0996
13	HI	290	1902		314.86	38.8	12216.568	1505.44	99136.8196
14	ID	240	468		-1119.14	-11.2	12534.368	125.44	1252474.34
15	IL	320	2170		582.86	68.8	40100.768	4733.44	339725.7796
16	IN	210	1414		-173.14	-41.2	7133.368	1697.44	29977.4596
17	IA	200	549		-1038.14	-51.2	53152.768	2621.44	1077734.66
18	KS	230	1131		-456.14	-21.2	9670.168	449.44	208063.6996
19	KY	200	952		-635.14	-51.2	32519.168	2621.44	403402.8196
20	LA	290	1970		382.86	38.8	14854.968	1505.44	146581.7796
21	ME	200	308		-1279.14	-51.2	65491.968	2621.44	1636199.14
22	MD	310	3927		2339.86	58.8	137583.768	3457.44	5474944.82
23	MA	290	2355		767.86	38.8	29792.968	1505.44	589608.9796
24	MI	250	2440		852.86	-1.2	-1023.432	1.44	727370.1796
25	MN	190	991		-596.14	-61.2	36483.768	3745.44	355382.8996
26	MS	200	810		-777.14	-51.2	39789.568	2621.44	603946.5796
27	MO	280	2236		648.86	28.8	18687.168	829.44	421019.2996
28	MT	240	340		-1247.14	-11.2	13967.968	125.44	1555358.18
29	NE	220	822		-765.14	-31.2	23872.368	973.44	585439.2196
30	NV	360	4606		3018.86	108.8	328451.968	11837.44	9113515.7
31	NH	240	420		-1167.14	-11.2	13071.968	125.44	1362215.78
32	NJ	350	3037		1449.86	98.8	143246.168	9761.44	2102094.02
33	NM	280	1279		-308.14	28.8	-8874.432	829.44	94950.2596
34	NY	370	6413		4825.86	118.8	573312.168	14113.44	23288924.74
35	NC	220	823		-764.14	-31.2	23841.168	973.44	583909.9396
36	ND	180	77		-1510.14	-71.2	107521.968	5069.44	2280522.82
37	OH	210	2237		649.86	-41.2	-26774.232	1697.44	422318.0196
38	OK	230	1049		-538.14	-21.2	11408.568	449.44	289594.6596
39	OR	240	1524		-63.14	-11.2	707.168	125.44	3986.6596
40	PA	240	1779		191.86	-11.2	-2148.832	125.44	36810.2596
41	RI	280	1186		-401.14	28.8	-11552.832	829.44	160913.2996
42	SC	240	1181		-406.14	-11.2	4548.768	125.44	164949.6996
43	SD	200	201		-1386.14	-51.2	70970.368	2621.44	1921384.1
44	TN	210	1806		218.86	-41.2	-9017.032	1697.44	47899.6996
45	TX	240	2085		497.86	-11.2	-5576.032	125.44	247864.5796
46	UT	240	802		-785.14	-11.2	8793.568	125.44	616444.8196
47	VT	200	389		-1198.14	-51.2	61344.768	2621.44	1435539.46
48	VA	230	1201		-386.14	-21.2	8186.168	449.44	149104.0996
49	WA	210	1351		-236.14	-41.2	9728.968	1697.44	55762.0996
50	WV	180	485		-1102.14	-71.2	78472.368	5069.44	1214712.58
51	WI	240	707		-880.14	-11.2	9857.568	125.44	774646.4196
52	WY	310	444		-1143.14	58.8	-67216.632	3457.44	1306769.06
53									
54	Sum =	12560	79357			Sum =	2008311.6	116728	73958300.02
55	N =	50	50						
56	Mean =	251.2	1587.14						
57									
58									
59		**Coefficient**							
60	Slope =	17.20505							
61	Intercept=	-2734.77							

Predicted Values and Model Goodness of Fit Measures

17. Enter the following labels.
    K2: $\mathbf{Y_{hat}}$
    L2: $\mathbf{(Y_i-Y_{bar})^2}$
    M2: $\mathbf{(Y_i-Y_{hati})^2}$
    N2: $\mathbf{(Y_{hati}-Y_{bar})^2}$

18. Enter the following formula.
    K3: **=b61+b60*b3**

19. Copy the formula in cell K3 to cells K4 to K52.

20. Enter the following formula.
    L3: **=(c3-c$56)^2**

21. Copy the formula in cell L3 to cells L4 to L52.

22. Enter the following formula.
    M3: **=(c3-k3)^2**

23. Copy the formula in cell M3 to cells M4 to M52.

24. Enter the following formula.
    N3: **=(k3-c$56)^2**

25. Copy the formula in cell N3 to cells N4 to N52.

26. Enter the following formulas.
    L54: **=sum(l3:l52)**
    M54: **=sum(m3:m52)**
    N54: **=sum(n3:n52)**

This portion of your spreadsheet should look similar to what is below.

	K	L	M	N
1				
2	Y_{hat}	$(Y_i-Y_{bar})^2$	$(Y_i-Y_{hati})^2$	$(Y_{hati}-Y_{bar})^2$
3	1222.393	70830.5	9723.371	133040.49
4	3286.999	472161.4	5697766	2889521.9
5	2598.797	121703.3	439300.1	1023450.3
6	534.1907	605501.9	75520.17	1108702.3
7	1910.595	5084394	3730325	104623.15
8	1910.595	192.0996	95849.08	104623.15
9	1738.544	351483	194883	22923.316
10	2082.646	47149.78	507863.7	245525.77
11	2426.747	3872473	1272956	704939.34
12	1394.443	151212.1	338208.1	37131.984
13	2254.696	99136.82	124394.5	445631.17
14	1394.443	1252474	858297.4	37131.984
15	2770.848	339725.8	361018	1401164
16	878.2918	29977.46	286983.3	502465.83
17	706.2412	1077735	24724.8	775982.68
18	1222.393	208063.7	8352.652	133040.49
19	706.2412	403402.8	60397.38	775982.68
20	2254.696	146581.8	81051.88	445631.17
21	706.2412	1636199	158596.1	775982.68
22	2598.797	5474945	1764123	1023450.3
23	2254.696	589609	10060.87	445631.17
24	1566.494	727370.2	763012.8	426.26002
25	534.1907	355382.9	208674.8	1108702.3
26	706.2412	603946.6	10765.89	775982.68
27	2082.646	421019.3	23517.58	245525.77
28	1394.443	1555358	1111851	37131.984
29	1050.342	585439.2	52140.21	288151.77
30	3459.05	9113516	1315494	3504046.8
31	1394.443	1362216	949539.9	37131.984
32	3286.999	2102094	62499.69	2889521.9
33	2082.646	94950.26	645846.2	245525.77
34	3631.1	23288925	7738965	4177774.4
35	1050.342	583909.9	51684.52	288151.77
36	362.1401	2280523	81304.89	1500624.7
37	878.2918	422318	1846088	502465.83
38	1222.393	289594.7	30065.08	133040.49
39	1394.443	3986.66	16784.92	37131.984
40	1394.443	36810.26	147883.8	37131.984
41	2082.646	160913.3	803973.3	245525.77
42	1394.443	164949.7	45558.08	37131.984
43	706.2412	1921384	255268.7	775982.68
44	878.2918	47899.7	860642.6	502465.83
45	1394.443	247864.6	476868.4	37131.984
46	1394.443	616444.8	350989.2	37131.984
47	706.2412	1435539	100642	775982.68
48	1222.393	149104.1	457.6538	133040.49
49	878.2918	55762.1	223453.1	502465.83
50	362.1401	1214713	15094.55	1500624.7
51	1394.443	774646.4	472578.4	37131.984
52	2598.797	1306769	4643151	1023450.3
53				
54		73958300	39405190	34553111

27. Enter the following labels.

A65: **Error =**
A66: **Regression=**
A67: **Total =**
A68: ***F* =**
A69: **Prob *F* =**
A70: **R^2 =**
B64: **df**
C64: **Sum Square**
D64: **Mean Square**
E64: **Sigma**

28. Enter the following formulas.

B65: **=b55-2**
B66: **1**
B67: **=b65+b66**
C65: **=m54**
C66: **=n54**
C67: **=l54**

29. Enter the following formulas.

D65: **=c65/(b55-2)**
D66: **=c66**
E65: **=sqrt(d65)**
B68: **=d66/d65**
B69: **=fdist(b68,b66,b65)**
B70: **=(l54-m54)/l54**

This portion of your spreadsheet should look like below.

	A	B	C	D	E
64		**df**	**Sum Square**	**Mean Square**	**Sigma**
65	Error =	48	39405190	820941.448	906.0582
66	Regression=	1	34553111	34553110.5	
67	Total =	49	73958300		
68	*F* =	42.08962			
69	Prob *F* =	4.49E-08			
70	R^2 =	0.467197			

Tests of Statistical Significance for Coefficients

30. Enter the following label.
    D2: **X_i^2**

31. Enter the following formula.
    D3: **=b3^2**

32. Copy the formula in cell D3 to cells D4 to D52.

33. Enter the following formula.
    D54: =sum(d3:d52)

34. Enter the following labels.
    C59: **Std Error**
    D59: ***t***
    E69: **df**
    F59: **Prob 1-tail**
    G59: **Prob 2-tail**

35. Enter the following formulas.
    C60: **=sqrt(d65/h54)**
    C61: **=(sqrt(d54/(b55*h54)))*e65**
    D60: **=b60/c60**
    D61: **=b61/c61**
    E60: **=b55-2**
    E61: **=b55-2**
    F60: **=tdist(abs(d60),e60,1)**
    F61: **=tdist(abs(d61),e61,1)**
    G60: **=tdist(abs(d60),e60,2)**
    G61: **=tdist(abs(d61),e61,2)**

This portion of your spreadsheet should look similar to what is below.

	A	B	C	D	E	F	G
59		**Coefficient**	**Std Error**	***t***	**df**	**Prob 1-tail**	**Prob 2-tail**
60	Slope =	17.20505	2.6519698	6.48765106	48	2.25E-08	4.4925E-08
61	Intercept=	-2734.77	678.3861	-4.0312879	48	9.87E-05	0.000197435

Your completed spreadsheet should look similar the one below.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	State Crime Data Set													
2	**STATE**	**POLICE (X)**	**ROBBERY (Y)**	X_i^2	(Y_i-Y_{bar})	(X_i-X_{bar})	$(X_i-X_{bar})(Yi-Y_{bar})$	$(X_i-X_{bar})^2$	$(Y_i-Y_{bar})^2$		Y_{hat}	$(Y_i-Y_{bar})^2$	$(Y_i-Y_{hat})^2$	$(Y_{hat}-Y_{bar})^2$
3	AL	230	1321	52900	-266.14	-21.2	5642.168	449.44	70830.4996		1222.393	70830.5	9723.371	133040.49
4	AK	350	900	122500	-687.14	98.8	-67889.432	9761.44	472161.3796		3286.999	472161.4	5697766	2889521.9
5	AZ	310	1936	96100	348.86	58.8	20512.968	3457.44	121703.2996		2598.797	121703.3	439300.1	1023450.3
6	AR	190	809	36100	-778.14	-61.2	47622.168	3745.44	605501.8596		534.1907	605501.9	75520.17	1108702.3
7	CA	270	3842	72900	2254.86	18.8	42391.368	353.44	5084393.62		1910.595	5084394	3730325	104623.15
8	CO	270	1601	72900	13.86	18.8	260.568	353.44	192.0996		1910.595	192.0996	95849.08	104623.15
9	CT	260	2180	67600	592.86	8.8	5217.168	77.44	351482.9796		1738.544	351483	194883	22923.316
10	DE	280	1370	78400	-217.14	28.8	-6253.632	829.44	47149.7796		2082.646	47149.78	507863.7	245525.77
11	FL	300	3555	90000	1967.86	48.8	96031.568	2381.44	3872472.98		2426.747	3872473	1272956	704939.34
12	GA	240	1976	57600	388.86	-11.2	-4355.232	125.44	151212.0996		1394.443	151212.1	338208.1	37131.984
13	HI	290	1902	84100	314.86	38.8	12216.568	1505.44	99136.8196		2254.696	99136.82	124394.5	445631.17
14	ID	240	468	57600	-1119.14	-11.2	12534.368	125.44	1252474.34		1394.443	1252474	858297.4	37131.984
15	IL	320	2170	102400	582.86	68.8	40100.768	4733.44	339725.7796		2770.848	339725.8	361018	1401164
16	IN	210	1414	44100	-173.14	-41.2	7133.368	1697.44	29977.4596		878.2918	29977.46	286983.3	502465.83
17	IA	200	549	40000	-1038.14	-51.2	53152.768	2621.44	1077734.66		706.2412	1077735	24724.8	775982.68
18	KS	230	1131	52900	-456.14	-21.2	9670.168	449.44	208063.6996		1222.393	208063.7	8352.652	133040.49
19	KY	200	952	40000	-635.14	-51.2	32519.168	2621.44	403402.8196		706.2412	403402.8	60397.38	775982.68
20	LA	290	1970	84100	382.86	38.8	14854.968	1505.44	146581.7796		2254.696	146581.8	81051.88	445631.17
21	ME	200	308	40000	-1279.14	-51.2	65491.968	2621.44	1636199.14		706.2412	1636199	158596.1	775982.68
22	MD	310	3927	96100	2339.86	58.8	137583.768	3457.44	5474944.82		2598.797	5474945	1764123	1023450.3
23	MA	290	2355	84100	767.86	38.8	29792.968	1505.44	589608.9796		2254.696	589609	10060.87	445631.17
24	MI	250	2440	62500	852.86	-1.2	-1023.432	1.44	727370.1796		1566.494	727370.2	763012.8	426.26002
25	MN	190	991	36100	-596.14	-61.2	36483.768	3745.44	355382.8996		534.1907	355382.9	208674.8	1108702.3
26	MS	200	810	40000	-777.14	-51.2	39789.568	2621.44	603946.5796		706.2412	603946.6	10765.89	775982.68
27	MO	280	2236	78400	648.86	28.8	18687.168	829.44	421019.2996		2082.646	421019.3	23517.58	245525.77
28	MT	240	340	57600	-1247.14	-11.2	13967.968	125.44	1555358.18		1394.443	1555358	1111851	37131.984
29	NE	220	822	48400	-765.14	-31.2	23872.368	973.44	585439.2196		1050.342	585439.2	52140.21	288151.77
30	NV	360	4606	129600	3018.86	108.8	328451.968	11837.44	9113515.7		3459.05	9113516	1315494	3504046.8
31	NH	240	420	57600	-1167.14	-11.2	13071.968	125.44	1362215.78		1394.443	1362216	949539.9	37131.984
32	NJ	350	3037	122500	1449.86	98.8	143246.168	9761.44	2102094.02		3286.999	2102094	62499.69	2889521.9
33	NM	280	1279	78400	-308.14	28.8	-8874.432	829.44	94950.2596		2082.646	94950.26	645846.2	245525.77
34	NY	370	6413	136900	4825.86	118.8	573312.168	14113.44	23288924.74		3631.1	23288925	7738965	4177774.4
35	NC	220	823	48400	-764.14	-31.2	23841.168	973.44	583909.9396		1050.342	583909.9	51684.52	288151.77
36	ND	180	77	32400	-1510.14	-71.2	107521.968	5069.44	2280522.82		362.1401	2280523	81304.89	1500624.7
37	OH	210	2237	44100	649.86	-41.2	-26774.232	1697.44	422318.0196		878.2918	422318	1846088	502465.83
38	OK	230	1049	52900	-538.14	-21.2	11408.568	449.44	289594.6596		1222.393	289594.7	30065.08	133040.49
39	OR	240	1524	57600	-63.14	-11.2	707.168	125.44	3986.6596		1394.443	3986.66	16784.92	37131.984
40	PA	240	1779	57600	191.86	-11.2	-2148.832	125.44	36810.2596		1394.443	36810.26	147883.8	37131.984
41	RI	280	1186	78400	-401.14	28.8	-11552.832	829.44	160913.2996		2082.646	160913.3	803973.3	245525.77
42	SC	240	1181	57600	-406.14	-11.2	4548.768	125.44	164949.6996		1394.443	164949.7	45558.08	37131.984
43	SD	200	201	40000	-1386.14	-51.2	70970.368	2621.44	1921384.1		706.2412	1921384	255268.7	775982.68
44	TN	210	1806	44100	218.86	-41.2	-9017.032	1697.44	47899.6996		878.2918	47899.7	860642.6	502465.83
45	TX	240	2085	57600	497.86	-11.2	-5576.032	125.44	247864.5796		1394.443	247864.6	476868.4	37131.984
46	UT	240	802	57600	-785.14	-11.2	8793.568	125.44	616444.8196		1394.443	616444.8	350989.2	37131.984
47	VT	200	389	40000	-1198.14	-51.2	61344.768	2621.44	1435539.46		706.2412	1435539	100642	775982.68
48	VA	230	1201	52900	-386.14	-21.2	8186.168	449.44	149104.0996		1222.393	149104.1	457.6538	133040.49
49	WA	210	1351	44100	-236.14	-41.2	9728.968	1697.44	55762.0996		878.2918	55762.1	223453.1	502465.83
50	WV	180	485	32400	-1102.14	-71.2	78472.368	5069.44	1214712.58		362.1401	1214713	15094.55	1500624.7
51	WI	240	707	57600	-880.14	-11.2	9857.568	125.44	774646.4196		1394.443	774646.4	472578.4	37131.984
52	WY	310	444	96100	-1143.14	58.8	-67216.632	3457.44	1306769.06		2598.797	1306769	4643151	1023450.3
53														
54	Sum =	12560	79357	3271800		Sum =	2008311.6	116728	73958300.02			73958300	39405190	34553111
55	N =	50	50											
56	Mean =	251.2	1587.14											
57														
58														
59		**Coefficient**	**Std Error**	***t***	**df**	**Prob 1-tail**	**Prob 2-tail**							
60	Slope =	17.20505	2.6519698	6.48765106	48	2.25E-08	4.4925E-08							
61	Intercept=	-2734.77	678.3861	-4.0312879	48	9.87E-05	0.000197435							
62														
63														
64		**df**	**Sum Square**	**Mean Square**	**Sigma**									
65	Error =	48	39405190	820941.448	906.0582									
66	Regression=	1	34553111	34553110.5										
67	Total =	49	73958300											
68	*F* =	42.08962												
69	Prob *F* =	4.49E-08												