

EXCEL EXERCISE #10: Statistical Analysis

1. Enter the data for URBAN (percent living in urban areas), MURDER (murders per million residents), ROBBERY (robberies per million residents), POLICE (police employees per 100,000 population), and REGION for the 50 states that are presented on the attached page. Label the sheet DATA.
2. Copy the data in A1:B52 (STATE and MURDER) onto the same cells on Sheet2. Perform the steps 3 through 22 below on Sheet2. Label the sheet UNIVARIATE.

Variance and Standard Deviation

3. Calculate the variance for MURDER by breaking down the formula $\Sigma(X_i - X_{\text{bar}})^2 / (N - 1)$, where X_{bar} is the mean, X_i is each case of the variables, and N is the number of cases. Enter the following labels in the appropriate cells.

A54: **Mean =**
A55: **N =**
A57: **Variance=**
A58: **Std Dev=**
C2: **$(X_i - X_{\text{bar}})$**
D2: **$(X_i - X_{\text{bar}})^2$**

4. Enter the following formulas.
B54: **=average(b3:b52)**
B55: **=count(b3:b52)**
5. Enter the following formula.
C3: **=b3-\$b\$54**
6. Copy the formula in cell C3 into cells C4 to C52.
7. Enter the following formula.
D3: **=c3^2**
8. Copy the formula in cell D3 into cells D4 to D52.
9. Enter the following formula.
D54: **=sum(d3:d52)**
This is the sum of the squared deviations, or the numerator for the variance equation in step 3 above. (Note: Steps 5 and 6 could be accomplished with the function: =sumproduct(c3:c52,c3:c52).)
10. Enter the following formula to calculate the variance for the variable MURDER.
B57: **=d54/(b55-1)**

11. The standard deviation is simply the square root of the variance which can be calculated with the following formula.

B58: **=sqrt(b57)**

Skewness and Kurtosis

12. Enter the following labels into the cells indicated below

A59: **Skewness=**

A60: **Kurtosis=**

E2: **$(X_i - X_{\text{bar}})^3$**

F2: **$(X_i - X_{\text{bar}})^4$**

13. Enter the following formula.

E3: **=c3^3**

14. Copy the formula in cell E3 into cells E4 to E52.

15. Enter the following formula.

F3: **=c3^4**

16. Copy the formula in cell F3 into cells F4 to F52.

17. Enter the following formula.

E54: **=sum(e3:e52)**

F54: **=sum(f3:f52)**

18. Enter the following formulas.

B59: **=(b55/((b55-1)*(b55-2)))*(e54/(b58^3))**

B60: **=(((b55*(b55+1))/((b55-1)*(b55-2)*(b55-3)))*(f54/(b58^4)))-((3*(b55-1)*(b55-1))/((b55-2)*(b55-3)))**

The completed worksheet should look similar to the one below.

	A	B	C	D	E	F
1	State Crime Data Set					
2	STATE	MURDER	$(X_i - \bar{X}_{\text{bar}})$	$(X_i - \bar{X}_{\text{bar}})^2$	$(X_i - \bar{X}_{\text{bar}})^3$	$(X_i - \bar{X}_{\text{bar}})^4$
3	AL	132	54.76	2998.658	164206.5	8991947
4	AK	97	19.76	390.4576	7715.442	152457.1
5	AZ	103	25.76	663.5776	17093.76	440335.2
6	AR	92	14.76	217.8576	3215.578	47461.93
7	CA	143	65.76	4324.378	284371.1	18700242
8	CO	69	-8.24	67.8976	-559.476	4610.084
9	CT	47	-30.24	914.4576	-27653.2	836232.7
10	DE	69	-8.24	67.8976	-559.476	4610.084
11	FL	145	67.76	4591.418	311114.5	21081116
12	GA	138	60.76	3691.778	224312.4	13629222
13	HI	87	9.76	95.2576	929.7142	9074.01
14	ID	31	-46.24	2138.138	-98867.5	4571632
15	IL	106	28.76	827.1376	23788.48	684156.6
16	IN	89	11.76	138.2976	1626.38	19126.23
17	IA	22	-55.24	3051.458	-168563	9311393
18	KS	69	-8.24	67.8976	-559.476	4610.084
19	KY	88	10.76	115.7776	1245.767	13404.45
20	LA	157	79.76	6361.658	507405.8	40470687
21	ME	28	-49.24	2424.578	-119386	5878577
22	MD	95	17.76	315.4176	5601.817	99488.26
23	MA	41	-36.24	1313.338	-47595.4	1724856
24	MI	102	24.76	613.0576	15179.31	375839.6
25	MN	26	-51.24	2625.538	-134533	6893448
26	MS	145	67.76	4591.418	311114.5	21081116
27	MO	111	33.76	1139.738	38477.54	1299002
28	MT	40	-37.24	1386.818	-51645.1	1923263
29	NE	44	-33.24	1104.898	-36726.8	1220799
30	NV	20	-57.24	3276.418	-187542	10734912
31	NH	25	-52.24	2729.018	-142564	7447537
32	NJ	69	-8.24	67.8976	-559.476	4610.084
33	NM	131	53.76	2890.138	155373.8	8352895
34	NY	127	49.76	2476.058	123208.6	6130861
35	NC	106	28.76	827.1376	23788.48	684156.6
36	ND	12	-65.24	4256.258	-277678	18115729
37	OH	81	3.76	14.1376	53.15738	199.8717
38	OK	51	-26.24	688.5376	-18067.2	474084
39	OR	51	-26.24	688.5376	-18067.2	474084
40	PA	68	-9.24	85.3776	-788.889	7289.335
41	RI	44	-33.24	1104.898	-36726.8	1220799
42	SC	114	36.76	1351.298	49673.7	1826005
43	SD	7	-70.24	4933.658	-346540	24340977
44	TN	108	30.76	946.1776	29104.42	895252.1
45	TX	169	91.76	8419.898	772609.8	70894676
46	UT	38	-39.24	1539.778	-60420.9	2370915
47	VT	22	-55.24	3051.458	-168563	9311393
48	VA	86	8.76	76.7376	672.2214	5888.659
49	WA	55	-22.24	494.6176	-11000.3	244646.6
50	WV	71	-6.24	38.9376	-242.971	1516.137
51	WI	29	-48.24	2327.098	-112259	5415383
52	WY	62	-15.24	232.2576	-3539.61	53943.59
53						
54	Mean =	77.24	Sum =	88755.12	1000676	3.28E+08
55	N =	50	Sum =	88755.12		
56						
57	Variance =	1811.329				
58	Std Dev =	42.55971				
59	Skewness =	0.27595				
60	Kurtosis =	-0.88334				

Descriptive Statistics Shortcut:

19. Enter the following labels.

A63: **Shortcuts:**

A64: **Min.** =

A65: **Max.** =

A66: **Mode** =

A67: **Median** =

A68: **Mean** =

A69: **Variance**=

A70: **Std Dev.**=

A71: **Skewness**=

A72: **Kurtosis**=

20. Enter the following formulas.

B64: **=min(b3:b52)**

B65: **=max(b3:b52)**

B66: **=mode(b3:b52)**

B67: **=median(b3:b52)**

B68: **=average(b3:b52)**

B69: **=var(b3:b52)**

B70: **=stdev(b3:b52)**

B71: **=skew(b3:b52)**

B72: **=kurt(b3:b52)**

This part of the worksheet should now look like below.

	A	B
63	Shortcut:	
64	Min =	7
65	Max =	169
66	Mode =	69
67	Median =	70
68	Mean =	77.24
69	Variance=	1811.329
70	Std Dev =	42.55971
71	Skewness=	0.27595
72	Kurtosis=	-0.88334

Frequency Distribution

21. Create a frequency distribution for the variable MURDER. Enter the following into the cells indicated below.

H3: **50**
H4: **100**
H5: **150**
H6: **200**
I3: **0-50**
I4: **51-100**
I5: **101-150**
I6: **151-200**
I7: **Total**
J2: **Frequency**

22: To create a frequency distribution you must enter an array formula (a formula that is entered into an array of cells). The intervals for the distribution are in cells I3 to I6. The upper value in the range for each interval is listed in cells H3 to H6. the format of the frequency function is: =FREQUENCY(*data_range*,*interval_range*)

To enter an array formula, highlight the cells into which the frequency distribution will be located. In this case highlight cells J3 to J6. Type the following formula (but do not press ENTER when you finish).

=frequency(b3:b52,h3:h6)

After you finish typing in the formula press the following three keys simultaneously: CONTROL, SHIFT, and ENTER.

The frequency for each interval should appear in the cells J3 to J6.

23. To complete the frequency distribution enter the following formula.

J7: **=sum(j3:j6)**

Your frequency distribution should now look like the one below.

	H	I	J
2			Frequency
3	50	0-50	16
4	100	51-100	18
5	150	101-150	14
6	200	151-200	2
7		Total	50
8			

Bivariate Statistics

24. Return to the Data sheet and copy cells A1:A52, D1:E52. Paste this data for the variables STATE, POLICE, and ROBBERY into cells A1:C52 on Sheet3. Label this sheet BIVARIATE.

25. Calculate a Pearson's product moment correlation coefficient (r) for the variables ROBBERY and POLICE by breaking down the formula for Pearson's r which is:

$$r = (\sum(Y_i - Y_{\text{bar}})(X_i - X_{\text{bar}}) / (N-1)) / (\sqrt{(\sum(X_i - X_{\text{bar}})^2 / (N-1))(\sum(Y_i - Y_{\text{bar}})^2 / (N-1))})$$

26. Enter the following labels

- A54: **N =**
- A55: **Mean =**
- A56: **Std Dev =**
- A58: **Covar XY=**
- A59: **Var X=**
- A60: **Var Y=**
- A61: **Pearson r=**
- E2: $(X_i - X_{\text{bar}})$
- F2: $(Y_i - Y_{\text{bar}})$
- G2: $(Y_i - Y_{\text{bar}})(X_i - X_{\text{bar}})$
- H2: $(X_i - X_{\text{bar}})^2$
- I2: $(Y_i - Y_{\text{bar}})^2$

27. Enter the following formulas.

- B54: **=count(b3:b52)**
- B55: **=average(b3:b52)**
- B56: **=stdev(b3:b52)**
- C54: **=count(c3:c52)**
- C55: **=average(c3:c52)**
- C56: **=stdev(c3:c52)**
- E3: **=b3-\$b\$55**

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

29. Enter the following formula.

- F3: **=c3-\$c\$55**

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

31. Enter the following formula.

- G3: **=e3*f3**

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

33. Enter the following formula.

H3: **=e3^2**

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

35. Enter the following formula.

I3: **=f3^2**

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

37. Enter the following formulas.

B58: **=sum(g3:g52)/(b54-1)**

B59: **=sum(h3:h52)/(b54-1)**

B60: **=sum(i3:i54)/(b54-1)**

B61: **=sqrt(b58^2/(b59*b60))**

38. Your completed worksheet should look like the one below.

	A	B	C	D	E	F	G	H	I
1	State Crime Data Set								
2	STATE	POLICE	ROBBERY		$(X_i - \bar{X}_{bar})$	$(Y_i - \bar{Y}_{bar})$	$(Y_i - \bar{Y}_{bar})(X_i - \bar{X}_{bar})$	$(X_i - \bar{X}_{bar})^2$	$(Y_i - \bar{Y}_{bar})^2$
3	AL	230	1321		-21.2	-266.14	5642.168	449.44	70830.5
4	AK	350	900		98.8	-687.14	-67889.432	9761.44	472161.4
5	AZ	310	1936		58.8	348.86	20512.968	3457.44	121703.3
6	AR	190	809		-61.2	-778.14	47622.168	3745.44	605501.9
7	CA	270	3842		18.8	2254.86	42391.368	353.44	5084394
8	CO	270	1601		18.8	13.86	260.568	353.44	192.0996
9	CT	260	2180		8.8	592.86	5217.168	77.44	351483
10	DE	280	1370		28.8	-217.14	-6253.632	829.44	47149.78
11	FL	300	3555		48.8	1967.86	96031.568	2381.44	3872473
12	GA	240	1976		-11.2	388.86	-4355.232	125.44	151212.1
13	HI	290	1902		38.8	314.86	12216.568	1505.44	99136.82
14	ID	240	468		-11.2	-1119.14	12534.368	125.44	1252474
15	IL	320	2170		68.8	582.86	40100.768	4733.44	339725.8
16	IN	210	1414		-41.2	-173.14	7133.368	1697.44	29977.46
17	IA	200	549		-51.2	-1038.14	53152.768	2621.44	1077735
18	KS	230	1131		-21.2	-456.14	9670.168	449.44	208063.7
19	KY	200	952		-51.2	-635.14	32519.168	2621.44	403402.8
20	LA	290	1970		38.8	382.86	14854.968	1505.44	146581.8
21	ME	200	308		-51.2	-1279.14	65491.968	2621.44	1636199
22	MD	310	3927		58.8	2339.86	137583.768	3457.44	5474945
23	MA	290	2355		38.8	767.86	29792.968	1505.44	589609
24	MI	250	2440		-1.2	852.86	-1023.432	1.44	727370.2
25	MN	190	991		-61.2	-596.14	36483.768	3745.44	355382.9
26	MS	200	810		-51.2	-777.14	39789.568	2621.44	603946.6
27	MO	280	2236		28.8	648.86	18687.168	829.44	421019.3
28	MT	240	340		-11.2	-1247.14	13967.968	125.44	1555358
29	NE	220	822		-31.2	-765.14	23872.368	973.44	585439.2
30	NV	360	4606		108.8	3018.86	328451.968	11837.44	9113516
31	NH	240	420		-11.2	-1167.14	13071.968	125.44	1362216
32	NJ	350	3037		98.8	1449.86	143246.168	9761.44	2102094
33	NM	280	1279		28.8	-308.14	-8874.432	829.44	94950.26
34	NY	370	6413		118.8	4825.86	573312.168	14113.44	23288925
35	NC	220	823		-31.2	-764.14	23841.168	973.44	583909.9
36	ND	180	77		-71.2	-1510.14	107521.968	5069.44	2280523
37	OH	210	2237		-41.2	649.86	-26774.232	1697.44	422318
38	OK	230	1049		-21.2	-538.14	11408.568	449.44	289594.7
39	OR	240	1524		-11.2	-63.14	707.168	125.44	3986.66
40	PA	240	1779		-11.2	191.86	-2148.832	125.44	36810.26
41	RI	280	1186		28.8	-401.14	-11552.832	829.44	160913.3
42	SC	240	1181		-11.2	-406.14	4548.768	125.44	164949.7
43	SD	200	201		-51.2	-1386.14	70970.368	2621.44	1921384
44	TN	210	1806		-41.2	218.86	-9017.032	1697.44	47899.7
45	TX	240	2085		-11.2	497.86	-5576.032	125.44	247864.6
46	UT	240	802		-11.2	-785.14	8793.568	125.44	616444.8
47	VT	200	389		-51.2	-1198.14	61344.768	2621.44	1435539
48	VA	230	1201		-21.2	-386.14	8186.168	449.44	149104.1
49	WA	210	1351		-41.2	-236.14	9728.968	1697.44	55762.1
50	WV	180	485		-71.2	-1102.14	78472.368	5069.44	1214713
51	WI	240	707		-11.2	-880.14	9857.568	125.44	774646.4
52	WY	310	444		58.8	-1143.14	-67216.632	3457.44	1306769
53									
54	N =	50	50						
55	Mean =	251.2	1587.14						
56	Std Dev =	48.80783	1228.557						
57									
58	Covar XY =	40985.95							
59	Var X =	2382.204							
60	Var Y =	1509353							
61	Pearson r =	0.683518							

Shortcut to Pearson r

39. A much simpler method of calculating this correlation coefficient is to use the PEARSON function in Excel. Enter the following formula. It should give you the same result as in cell B61.

B62: **=pearson(b3:b52,c3:c52)**

Scatterplot

40. Create a scatterplot for ROBBERY (as Y) and POLICE (as X). Open the INSERT menu and select the CHART option.

41. Respond as follows to the steps in ChartWizard as prompted.

Step 1: Chart Type

Select XY (Scatter) and the chart sub-type at the top of the left hand column. Then click on NEXT.

Step 2: Chart Source Data

If the cell range for the data and the series source are correct, then click on NEXT.

Step 3: Chart Options

a. With the tab "Titles" highlighted, type in the following information next to each of the title options.

Chart title: **Scattergram of Robbery and Police**

Axis title--Category (X): **Police**

Axis title--Value (Y): **Robbery**

b. Click on the "Legend" tab. Since only one variable is graphed, turn off the legend.

c. Click on the "Gridlines" tab. Under the "Value (y) axis" click on the check mark to turn off the major gridlines.

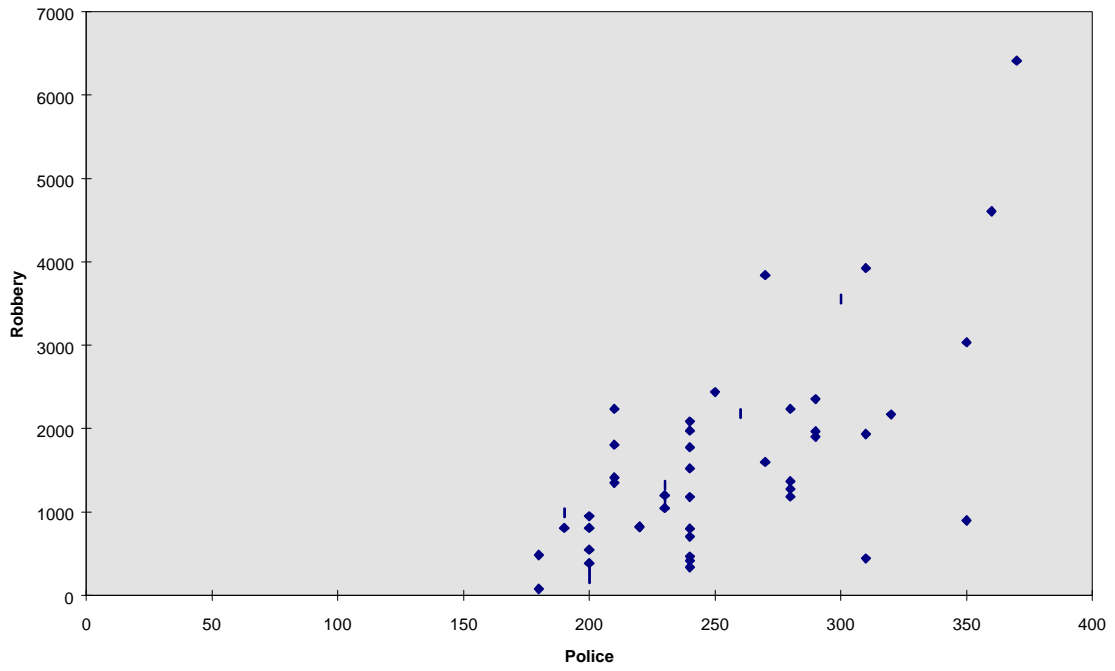
d. After you have selected the chart options you want, click on NEXT.

Step 4: Chart Location

Click on the empty circle to the left of the "As new sheet" option. The text bar to the right of this option with the word "Chart1" should now be highlighted. Give the sheet that the chart will appear on a more descriptive label. Type "Scatter" and click on FINISH.

Your scattergram should look like the one below.

Scattergram of Robbery and Police



	A	B	C	D	E	F
1	State Crime Data Set					
2	STATE	URBAN	MURDER	POLICE	ROBBERY	REGION
3	AL	60	132	230	1321	3
4	AK	64.5	97	350	900	4
5	AZ	83.8	103	310	1936	4
6	AR	51.6	92	190	809	3
7	CA	91.3	143	270	3842	4
8	CO	80.6	69	270	1601	4
9	CT	78.8	47	260	2180	1
10	DE	70.7	69	280	1370	3
11	FL	84.3	145	300	3555	3
12	GA	62.3	138	240	1976	3
13	HI	86.5	87	290	1902	4
14	ID	54	31	240	468	4
15	IL	83	106	320	2170	2
16	IN	64.2	89	210	1414	2
17	IA	58.6	22	200	549	2
18	KS	66.7	69	230	1131	2
19	KY	50.8	88	200	952	3
20	LA	68.6	157	290	1970	3
21	ME	47.5	28	200	308	1
22	MD	80.3	95	310	3927	3
23	MA	83.8	41	290	2355	1
24	MI	70.7	102	250	2440	2
25	MN	66.8	26	190	991	2
26	MS	47.3	145	200	810	3
27	MO	68.1	111	280	2236	2
28	MT	52.9	40	240	340	4
29	NE	62.7	44	220	822	2
30	NV	85.3	20	360	4606	4
31	NH	52.2	25	240	420	1
32	NJ	89	69	350	3037	1
33	NM	72.2	131	280	1279	4
34	NY	84.6	127	370	6413	1
35	NC	48	106	220	823	3
36	ND	48.8	12	180	77	2
37	OH	73.3	81	210	2237	2
38	OK	67.3	51	230	1049	3
39	OR	67.9	51	240	1524	4
40	PA	69.3	68	240	1779	1
41	RI	87	44	280	1186	1
42	SC	54.1	114	240	1181	3
43	SD	46.4	7	200	201	2
44	TN	60.4	108	210	1806	3
45	TX	79.6	169	240	2085	3
46	UT	84.4	38	240	802	4
47	VT	33.8	22	200	389	1
48	VA	66	86	230	1201	3
49	WA	73.6	55	210	1351	4
50	WV	36.2	71	180	485	3
51	WI	64.2	29	240	707	2
52	WY	62.8	62	310	444	4

