

# 4.1 Scatter Diagrams and Linear Correlation with work

## 4.1 Scatter Diagrams and Linear Correlation

### Essential Questions:

How does correlation help me analyze the relationship I am comparing?

### Focus Points:

- Make a scatter diagram.
- Visually estimate the location of the "best-fitting" line for a scatter diagram.
- Use sample data to compute the sample correlation coefficient,  $r$ .
- Investigate the meaning of the correlation coefficient,  $r$ .

Oct 16-8:27 AM

A **scatter diagram** is a graph in which data pairs  $(x, y)$  are plotted as individual points on a grid with horizontal axis  $x$  and vertical axis  $y$ . We call  $x$  the **explanatory variables** and  $y$  the **response variable**.

Oct 16-8:33 AM

**Scatterplot for quality characteristics, 2001**  
 Scattered but clearly negative correlation  
 NO SLOPE!!  
 Correlation or association.  
 negative association  
 Draw the "best-fitting" line.  
 Describe the correlation (moderate linear, perfectly linear, etc) and association (positive, negative, none).  
 almost linear; positive association  
 weak correlation no clear associative

Oct 16-8:35 AM

### WHAT DOES A SCATTER PLOT TELL US?

A scatter diagram shows the relationship between two paired variables  $x$  and  $y$ . For each pair  $(x, y)$  of the data set, we plot the point on a grid with horizontal axis  $x$  and vertical axis  $y$ .

- To the extent the data points on the graph fall closer to a straight line, we can say that the **relationship between  $x$  and  $y$  is more linear**.
- To the extent the data pairs are scattered over the graph, we can say **no linear relationship between  $x$  and  $y$  is apparent**.

Oct 16-8:40 AM

L1 X L2 Y L3 Z  
 20 1.0 1.0  
 42 1.0 1.0  
 47 1.0 1.0  
 50 1.0 1.0  
 60 1.0 1.0  
 80 1.0 1.0  
 L2(8) =

2000 MEMORY  
 6:Standard  
 7:ZTrig  
 8:ZInteger  
 9:ZoomStat  
 0:ZoomFit  
 1:ZQuadrant1  
 2:ZFrac1/2

Plot1 Plot2 Plot3  
 On Off  
 Type: [ ] [ ] [ ]  
 Xlist:L1  
 Ylist:L2  
 Mark: [ ] [ ] [ ]

Oct 16-8:46 AM

The **sample correlation coefficient,  $r$** , is a numerical measurement that assesses the strength of a **linear relationship** between two variables  $x$  and  $y$ .

1.  $r$  is a **unitless measurement** between  $-1$  and  $1$ . In symbols,  $-1 \leq r \leq 1$ . If  $r = 1$ , there is a **perfect positive linear correlation**. If  $r = -1$ , there is **perfect negative linear correlation**. If  $r = 0$ , there is **no linear correlation**. The closer  $r$  is to  $1$  or  $-1$ , the better a **line describes the relationship** between the two variables  $x$  and  $y$ .
2. **Positive** values of  $r$  imply as  $x$  is increasing,  $y$  tends to **increase**; **Negative** values of  $r$  imply that as  $x$  increases,  $y$  tends to **decrease**.
3.  $r$  will **not** change if variables are converted to **different units**.

Oct 16-8:58 AM

# 4.1 Scatter Diagrams and Linear Correlation with work

**Finding "r"**

$$r = \frac{1}{n-1} \sum \frac{(x - \bar{x})}{s_x} \cdot \frac{(y - \bar{y})}{s_y}$$

degrees of freedom "df" = n - 1

---

```

EDIT  [M-LOC] TESTS
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7:QuartReg
8:LinReg(a+bx)
    
```

```

LinReg(a+bx)
Xlist:L1
Ylist:L2
FreqList:
Store RegEQ:
r=0.9189898338
    
```

Oct 16-9:09 AM

p. 149

**Then**

There is no linear relation among the points of the scatter diagram.

There is a perfect linear relation between x and y values; all points lie on the least-squares line.

The x and y values have a positive correlation. By this, we mean that large x values are associated with large y values, and small x values are associated with small y values.

The x and y values have a negative correlation. By this, we mean that large x values are associated with small y values, and small x values are associated with large y values.

Oct 16-9:17 AM

**Example 1: Sand Dunes**

Sand driven by wind creates large, beautiful dunes at the Great Sand Dunes National Monument, Colorado. Of course, the same natural forces also create large dunes in the Great Sahara Desert and Saudi Arabia. Is there a linear correlation between wind velocity and sand drift rate? Let x be a random variable representing wind velocity (in 10 cm/sec) and let y be a random variable representing drift rate of sand (in 100 gm/cm/sec). A test site at the Great Sand Dunes National Monument gave the following information about x and y. Graph a scatter diagram, determine r, and describe what r tells you about the relationship between the variables.

x	70	115	105	82	93	125	88
y	3	45	21	7	16	62	12

*Great Sand Dunes*

```

LinReg
y=a+bx
a=-79.97763496
b=1.070565553
r^2=.8997719968
r=.9485631222
    
```

*r = .949*

Oct 16-9:18 AM

What does the correlation coefficient, r, is a sample statistic from a data set of ordered pairs (x, y). It is a measurement indicating the strength of a linear relationship between x and y.

- r is a **unitless** measurement from **-1 to 1**.
- r close to **1** means **positive linear** relationship increasing from left to right.
- r close to **-1** means **negative linear** relationship decreasing from left to right.
- r close to **0** means **no linear relationship** between the variables.

Oct 16-9:31 AM

p is used for population correlation coefficient

Oct 16-9:34 AM

**REMINDER!!!**

Lurking Variables and Causation  
z is the **lurking variable**

(a) Causation    (b) Common response    (c) Confounding

Oct 16-9:35 AM

# 4.1 Scatter Diagrams and Linear Correlation with work

HW: pg. 154: 1 - 7(o), 13, 15

1. Explanatory variables is placed along the horizontal axis. Response variables is placed along the vertical axis.

3. Decreases.

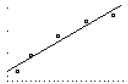
5. a) Moderate    b) None    c) High

7. a) None    b) Increasing population might be a lurking variable causing both variables to increase.

13. a) Lines slopes upward.    b) Strong, positive    c)  $r = 0.972$ ; increase.

L1	L2	L3	Z
3	80		
8	85		
12	120		
18	125		
24	130		
30	135		
36	140		
42	145		
48	150		
54	155		
60	160		
66	165		
72	170		
78	175		
84	180		
90	185		
96	190		
102	195		
108	200		
114	205		
120	210		
126	215		
132	220		
138	225		
144	230		
150	235		
156	240		
162	245		
168	250		
174	255		
180	260		
186	265		
192	270		
198	275		
204	280		
210	285		
216	290		
222	295		
228	300		
234	305		
240	310		
246	315		
252	320		
258	325		
264	330		
270	335		
276	340		
282	345		
288	350		
294	355		
300	360		
306	365		
312	370		
318	375		
324	380		
330	385		
336	390		
342	395		
348	400		
354	405		
360	410		
366	415		
372	420		
378	425		
384	430		
390	435		
396	440		
402	445		
408	450		
414	455		
420	460		
426	465		
432	470		
438	475		
444	480		
450	485		
456	490		
462	495		
468	500		
474	505		
480	510		
486	515		
492	520		
498	525		
504	530		
510	535		
516	540		
522	545		
528	550		
534	555		
540	560		
546	565		
552	570		
558	575		
564	580		
570	585		
576	590		
582	595		
588	600		
594	605		
600	610		
606	615		
612	620		
618	625		
624	630		
630	635		
636	640		
642	645		
648	650		
654	655		
660	660		
666	665		
672	670		
678	675		
684	680		
690	685		
696	690		
702	695		
708	700		
714	705		
720	710		
726	715		
732	720		
738	725		
744	730		
750	735		
756	740		
762	745		
768	750		
774	755		
780	760		
786	765		
792	770		
798	775		
804	780		
810	785		
816	790		
822	795		
828	800		
834	805		
840	810		
846	815		
852	820		
858	825		
864	830		
870	835		
876	840		
882	845		
888	850		
894	855		
900	860		
906	865		
912	870		
918	875		
924	880		
930	885		
936	890		
942	895		
948	900		
954	905		
960	910		
966	915		
972	920		
978	925		
984	930		
990	935		
996	940		
1002	945		
1008	950		
1014	955		
1020	960		
1026	965		
1032	970		
1038	975		
1044	980		
1050	985		
1056	990		
1062	995		
1068	1000		
1074	1005		
1080	1010		
1086	1015		
1092	1020		
1098	1025		
1104	1030		
1110	1035		
1116	1040		
1122	1045		
1128	1050		
1134	1055		
1140	1060		
1146	1065		
1152	1070		
1158	1075		
1164	1080		
1170	1085		
1176	1090		
1182	1095		
1188	1100		
1194	1105		
1200	1110		
1206	1115		
1212	1120		
1218	1125		
1224	1130		
1230	1135		
1236	1140		
1242	1145		
1248	1150		
1254	1155		
1260	1160		
1266	1165		
1272	1170		
1278	1175		
1284	1180		
1290	1185		
1296	1190		
1302	1195		
1308	1200		
1314	1205		
1320	1210		
1326	1215		
1332	1220		
1338	1225		
1344	1230		
1350	1235		
1356	1240		
1362	1245		
1368	1250		
1374	1255		
1380	1260		
1386	1265		
1392	1270		
1398	1275		
1404	1280		
1410	1285		
1416	1290		
1422	1295		
1428	1300		
1434	1305		
1440	1310		
1446	1315		
1452	1320		
1458	1325		
1464	1330		
1470	1335		
1476	1340		
1482	1345		
1488	1350		
1494	1355		
1500	1360		

$y = a + bx$   
 $a = 55.73170732$   
 $b = 0.894308943$   
 $r = 0.94522448$   
 $r^2 = 0.9722468841$



15. a) line slopes downward    b) strong, negative    c)  $r = -0.990$ ; decrease.

L1	L2	L3	Z
1000	90		
875	100		
850	85		
825	105		
800	95		
775	110		
750	100		
725	115		
700	105		
675	120		
650	110		
625	125		
600	115		
575	130		
550	120		
525	135		
500	125		
475	140		
450	130		
425	145		
400	135		
375	150		
350	140		
325	155		
300	145		
275	160		
250	150		
225	165		
200	155		
175	170		
150	160		
125	175		
100	165		
75	180		
50	170		
25	185		
0	175		
-25	190		
-50	180		
-75	195		
-100	185		
-125	200		
-150	190		
-175	205		
-200	195		
-225	210		
-250	200		
-275	215		
-300	205		
-325	220		
-350	210		
-375	225		
-400	215		
-425	230		
-450	220		
-475	235		
-500	225		
-525	240		
-550	230		
-575	245		
-600	235		
-625	250		
-650	240		
-675	255		
-700	245		
-725	260		
-750	250		
-775	265		
-800	255		
-825	270		
-850	260		
-875	275		
-900	265		
-925	280		
-950	270		
-975	285		