STAT 420 Fall 2015

Examples for 10/8/2015

Correlation Coefficient:

$$r = \frac{1}{n-1} \cdot \sum \left(\frac{x-\overline{x}}{s_x} \right) \cdot \left(\frac{y-\overline{y}}{s_y} \right) \qquad r = \frac{\sum (x-\overline{x})(y-\overline{y})}{\sqrt{\sum (x-\overline{x})^2} \cdot \sqrt{\sum (y-\overline{y})^2}}$$

$$r = \frac{n \sum x\,y \;-\; \left(\sum x\right)\left(\sum y\right)}{\sqrt{n \sum x^2 - \left(\sum x\right)^2} \cdot \sqrt{n \sum y^2 - \left(\sum y\right)^2}}$$

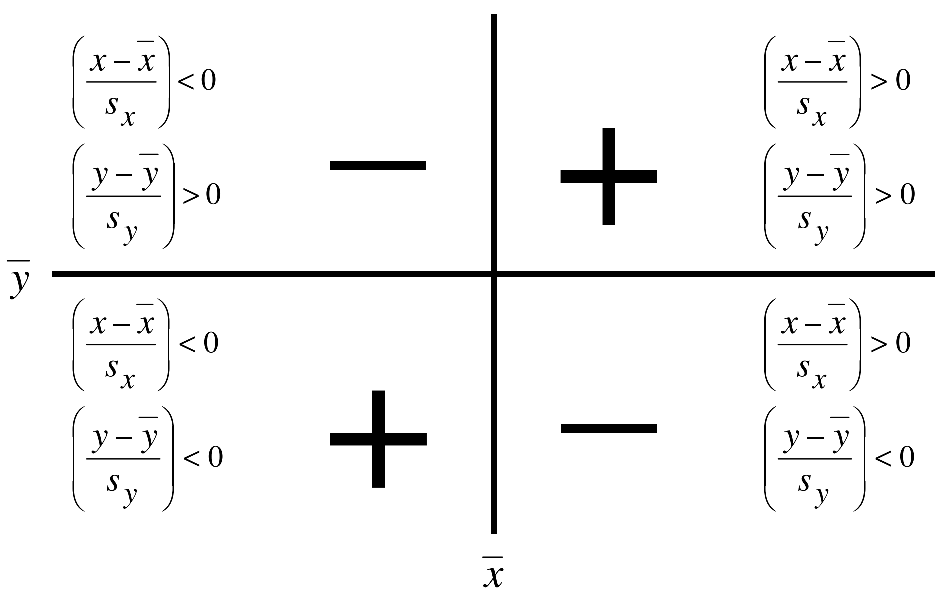
1. The value of r is always between -1 and $+1$.

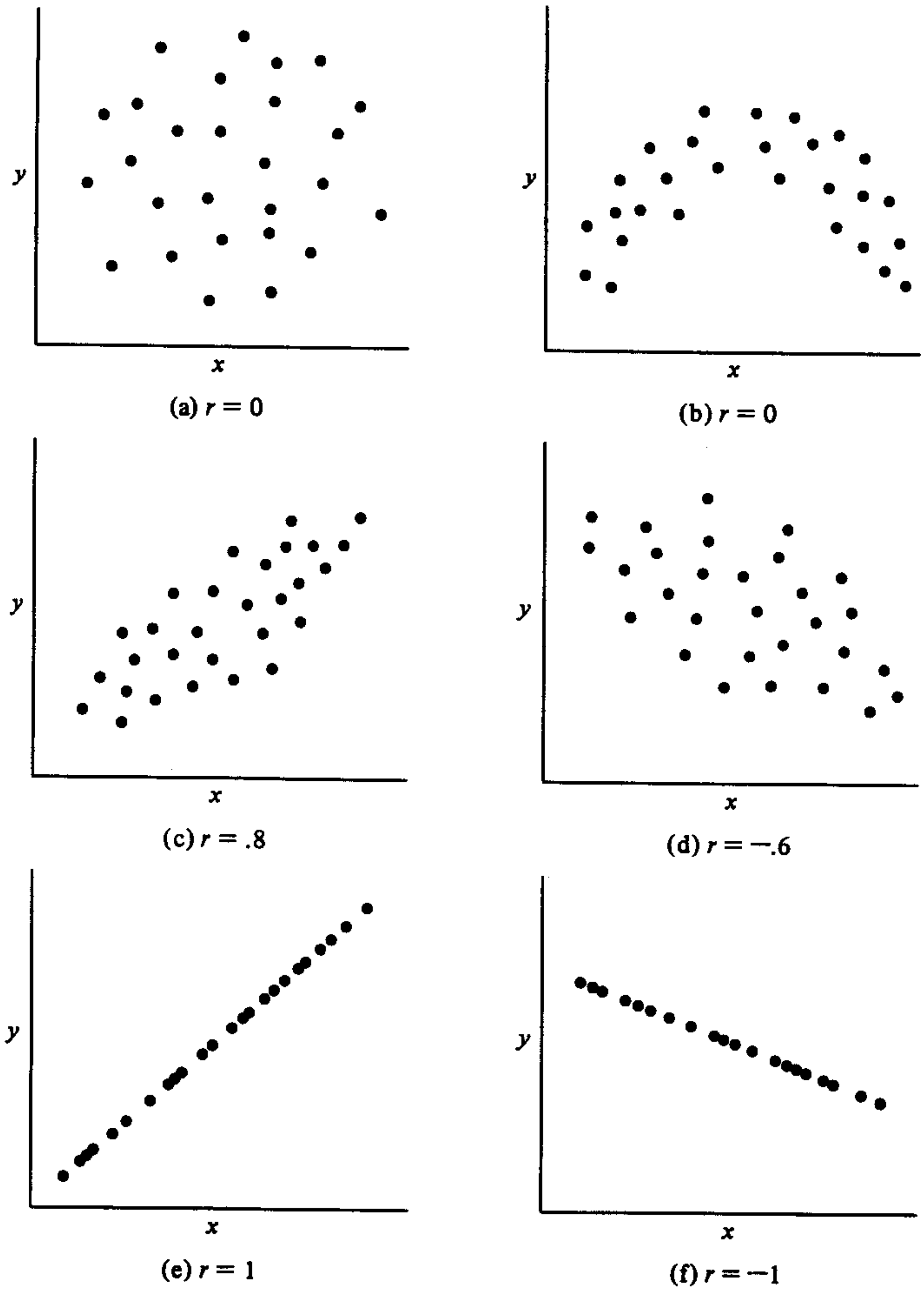
2. The magnitude of r indicates the strength of a <u>linear</u> relation, whereas its sign indicates the direction.

3. A value of r close to zero means that the <u>linear</u> association is very weak.

4. The value of r does not depend on which of the two variables under study is labeled x and which is labeled y, and is independent of the units in which x and y are measured.

5. r is invariant under *linear* transformations of x and y. That is, if $v = a + b\,x$ and $w = c + d\,y$, then $r_{xy} = r_{vw}$ if b and d are of the same sign, $r_{xy} = -\,r_{vw}$ if b and d are of the opposite sign.





Slope of the least-squares regression line $\hat{\beta}_1 = r \cdot \frac{s_y}{s_x}.$

1. The owner of *Momma Leona's Pizza* restaurant chain believes that if a restaurant is located near a college campus, then there is a linear relationship between sales and the size of the student population. Suppose data were collected from a sample of 10 *Momma Leona's Pizza* restaurants located near college campuses. For the ith restaurant in the sample, x_i is the size of the student population (in thousands) and y_i is the quarterly sales (in thousands of dollars). The values of x_i and y_i for the 10 restaurants in the sample are summarized in the following table:

Restaurant	Student Population (1000s)	Quarterly Sales ($1000s)
i	x_i	y_i
1	2	58
2	6	105
3	8	88
4	8	118
5	12	117
6	16	137
7	20	157
8	20	169
9	22	149
10	26	202

$\bar{x} = 14, \quad \bar{y} = 130$

$SXX = 568$

$SXY = 2{,}840$

$SYY = 15{,}730$

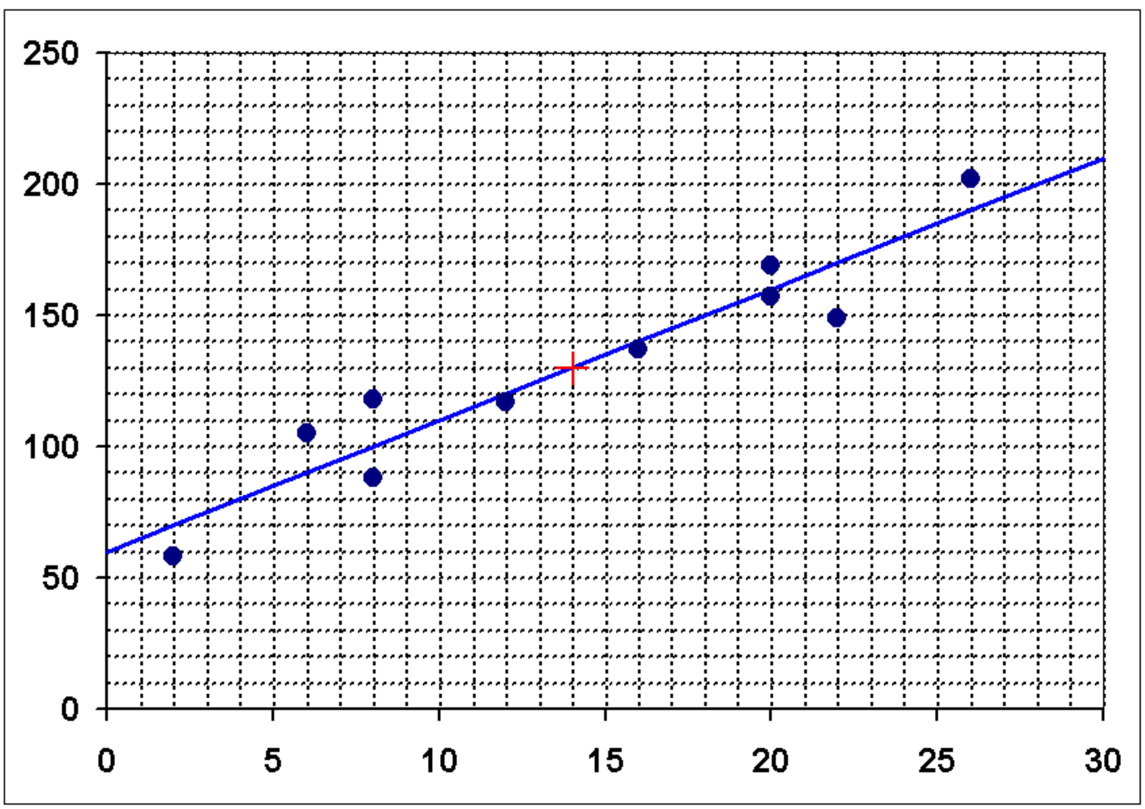
$\hat{\beta}_1 = 5, \quad \hat{\beta}_0 = 60$

$\hat{y} = 60 + 5 \cdot x$

$RSS = 1{,}530$

$R^2 = 0.9027$

$s_e^2 = 191.25$



a) Compute the sample correlation coefficient.

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2} \cdot \sqrt{\sum(y-\bar{y})^2}} = \frac{2840}{\sqrt{568} \cdot \sqrt{15730}} = \mathbf{0.950123}.$$

OR

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2} \cdot \sqrt{n\sum y^2 - (\sum y)^2}}$$

$$= \frac{10 \cdot 21040 - 140 \cdot 1300}{\sqrt{10 \cdot 2528 - 140^2} \cdot \sqrt{10 \cdot 184730 - 1300^2}} = \frac{28400}{\sqrt{5680} \cdot \sqrt{157300}}$$

$$= \mathbf{0.950123}.$$

Note that $(r)^2 = R^2$. (since SSRegr $= \sum(\hat{y}-\bar{y})^2 = \hat{\beta}_1^2$ SXX.)

$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ – a random sample from a Bivariate Normal distribution with parameters $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho$

To test $H_0: \rho = 0$, use $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ ($n-2$ degrees of freedom)

To test $H_0: \rho = \rho_0$, use $W = \dfrac{1}{2}\ln\dfrac{1+r}{1-r}$

W is approximately normal with mean $\mu_W = \dfrac{1}{2}\ln\dfrac{1+\rho}{1-\rho}$

and variance $\sigma_W^2 = \dfrac{1}{n-3}$.

b) Assume (X, Y) have a Bivariate Normal distribution. Test $H_0: \rho = 0$ vs. $H_1: \rho \neq 0$ at a 1% level of significance.

Test Statistic: $$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.950123 \cdot \sqrt{10-2}}{\sqrt{1-0.950123^2}} = \mathbf{8.617}.$$

Rejection Region: Rejects H_0 if $t < -t_{0.005}(8\text{ df})$ or $t > t_{0.005}(8\text{ df})$.

$\pm t_{0.005}(8\text{ df}) = \pm 3.355$. **Reject H_0.**

Since $1 - r^2 = 1 - R^2 = \dfrac{\text{SSResid}}{\text{SYY}}$,

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{\dfrac{\text{SXY}}{\sqrt{\text{SXX}}\sqrt{\text{SYY}}}\sqrt{n-2}}{\sqrt{\dfrac{\text{SSResid}}{\text{SYY}}}} = \frac{\dfrac{\text{SXY}}{\sqrt{\text{SXX}}}}{\sqrt{\dfrac{\text{SSResid}}{n-2}}} = \frac{\dfrac{\text{SXY}}{\sqrt{\text{SXX}}}}{s_e} = \frac{\dfrac{\text{SXY}}{\text{SXX}}}{s_e \big/ \sqrt{\text{SXX}}} = \frac{\hat{\beta}_1 - 0}{\hat{\sigma}_{\hat{\beta}_1}}.$$

OR

$$W = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\cdot\ln\left(\frac{1+0.950123}{1-0.950123}\right) = 1.833.$$

Under H_0, $$\mu_W = \frac{1}{2}\ln\frac{1+\rho_0}{1-\rho_0} = \frac{1}{2}\cdot\ln\left(\frac{1+0}{1-0}\right) = 0,$$

$$\sigma_W^2 = \frac{1}{n-3} = \frac{1}{7}.$$

Test Statistic: $$z = \frac{W-\mu_W}{\sigma_W} = \frac{1.833-0}{\sqrt{1/7}} = \mathbf{4.85}.$$

Rejection Region: Rejects H_0 if $z < -z_{0.005}$ or $z > z_{0.005}$.

$\pm z_{0.005} = \pm 2.576$. **Reject H_0.**

c) Find the p-value of the test $H_0 : \rho = 0.8$ vs. $H_1 : \rho > 0.8$.

$$W = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\cdot\ln\left(\frac{1+0.950123}{1-0.950123}\right) = 1.833.$$

Under H_0, $\mu_W = \frac{1}{2}\ln\frac{1+\rho_0}{1-\rho_0} = \frac{1}{2}\cdot\ln\left(\frac{1+0.80}{1-0.80}\right) = 1.0986,$

$$\sigma_W^2 = \frac{1}{n-3} = \frac{1}{7}.$$

Test Statistic: $z = \frac{W-\mu_W}{\sigma_W} = \frac{1.833-1.0986}{\sqrt{1/7}} = \mathbf{1.943}.$

P-value = right tail = $P(Z > 1.943)$ = **0.026**.

2. The following data are indexed prices of gold and copper over a 10-year period. Assume that the indexed values constitute a random sample from a bivariate normal distribution.

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})\cdot(y-\bar{y})$	$(y-\bar{y})^2$
76	80	16	12	256	192	144
62	68	2	0	4	0	0
70	73	10	5	100	50	25
59	60	−1	−8	1	8	64
53	64	−7	−4	49	28	16
54	68	−6	0	36	0	0
55	65	−5	−3	25	15	9
58	62	−2	−6	4	12	36
57	67	−3	−1	9	3	1
56	73	−4	5	16	−20	25
600	680	0	0	500	288	320

a) Test for the existence of linear relationship between the indexed prices of the two metals. That is, test $H_0: \rho = 0$ vs. $H_1: \rho \neq 0$. Use a 5% level of significance.

b) Is there enough evidence to conclude $\rho > 0.40$. That is, test $H_0: \rho = 0.40$ vs. $H_1: \rho > 0.40$. Use a 5% level of significance. What is the p-value of this test?