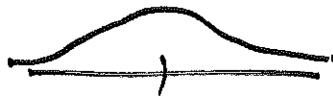
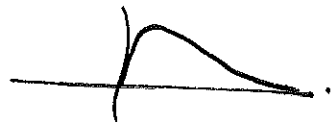
$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\,\mathrm{cov}(X_1, X_2)$$

$$Z \sim N(0,1)$$



$$X = Z$$
$$Y = Z^2$$



$$\cancel{E\{(X-E(X))\}\{(Y-E(Y))\}}$$

$$\mathrm{Cov}(X, Y)$$

$$= E\big((X - E(X))(Y - E(Y))\big)$$

$$= E\left[XY - E(X)Y - X\,E(Y) + E(X)E(Y)\right]$$

$$= E[XY] - E(X)E(Y)$$

$$= E[Z^3] - E[Z]\,E[Z^2]$$

$$K\int_{-\infty}^{\infty} z^3 \exp\left(-\frac{z^2}{2}\right) dz = 0$$

$$= 0$$

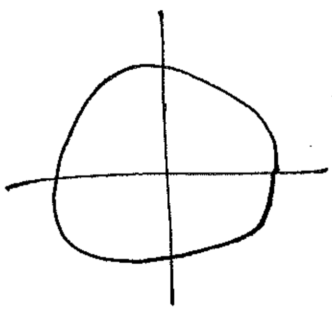
X, Y are uncorrelated

however, they are <u>not</u> independent

- if you know the value of X, you also know the value of Y

covariance / correlation measures
<u>linear</u> association.

If the relationship is nonlinear, we can still have $\mathrm{cov}(X,Y) = 0$



$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2\pi} & x^2+y^2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

X, Y are <u>dependent</u>

$$\mathrm{cov}(X,Y) = E[XY] - E[X]\,E[Y]$$

$$E[XY] = \iint xy\, f_{X,Y}(x,y)\, dx\, dy$$

$$= \frac{1}{2\pi} \iint_{\substack{\text{interior} \\ \text{of } \bigcirc}} xy \; dx\, dy.$$

$$= \frac{1}{2\pi} \iint xy \; dx\, dy + \frac{1}{2\pi} \iint xy \; dx\, dy$$

$= 0$ the two integrals have the same value but opposite signs.

Again - dependent RVs with covariance $= 0$

Correlation.

Covariance - the size of the covariance depends on the scale of the variation in X, Y.

Depends on the units of measurement.

Correlation - standardized version of covariance.

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{SD(X)\,SD(Y)}$$

$$= \text{Cov}\left(\frac{X - E(X)}{SD(X)} , \frac{Y - E(Y)}{SD(Y)} \right)$$

Correlation is the covariance between standardized versions of the RVs.

theorem: $-1 \leq \text{Corr}(X,Y) \leq 1.$ - we now know what "large" values are.

Proof: Assume X, Y are standardized.

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X,Y)$$

$$\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y) - 2\,\text{Cov}(X,Y)$$

let $\rho = \text{cov}(X, Y)$

Greek letter rho ρ

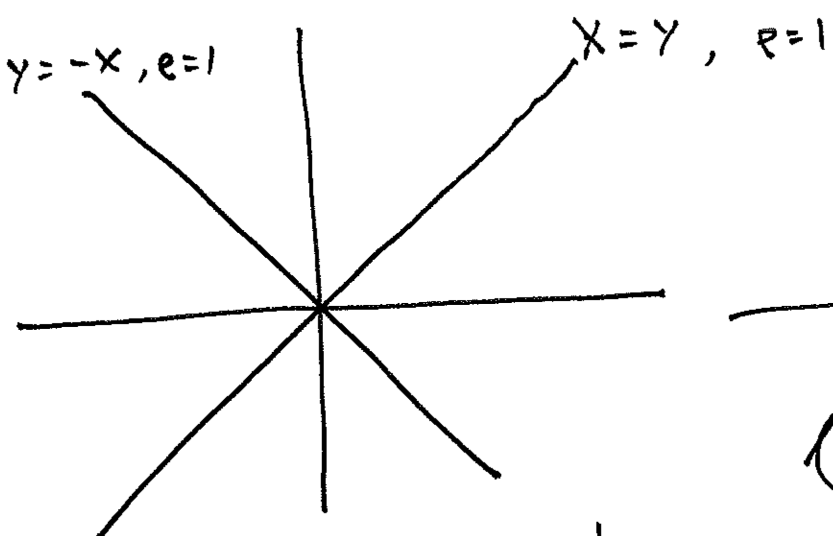
$\text{Var}() \geq 0$

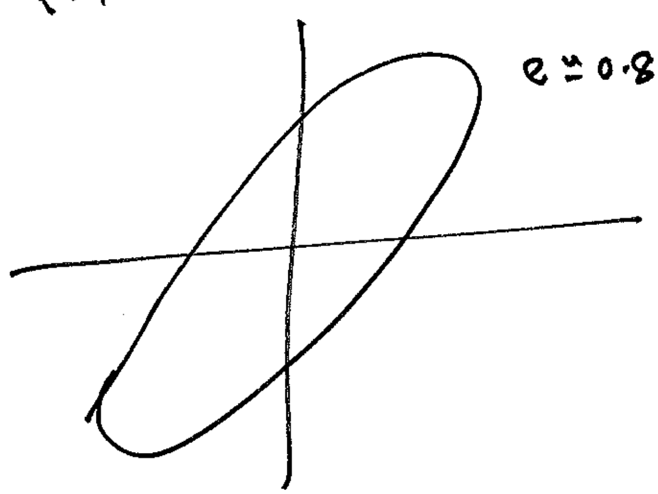
$$\text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X,Y) \geq 0$$
$$\text{Var}(X) + \text{Var}(Y) - 2\,\text{cov}(X,Y) \geq 0$$

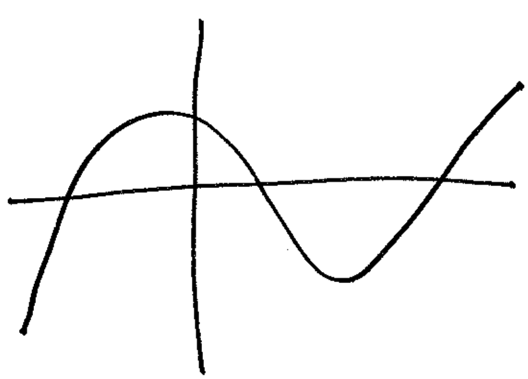
$$2 + 2\rho \geq 0$$
$$2 - 2\rho \geq 0$$

hence $-1 \leq \rho \leq 1$.

Correlation is between -1 and +1.







- nonlinear dependence <u>may</u> have $\rho = 0$ even though X, Y are dependent.

Variance of Binomial.

$X \sim \text{Bin}(n, p)$

$X = X_1 + X_2 + \dots X_n \qquad X_i \sim \text{iid Bernoulli}(p)$

$$\begin{aligned} \text{Var}(X_i) &= E(X_i^2) - (E(X_i))^2 \\ &= p - p^2 \\ &= p(1-p) \\ &= pq \end{aligned}$$

X_i can be considered an indicator variable for success on the ith trial.

X_i^2 is an indicator variable for the same event.

$\text{Var}(X) = npq.$