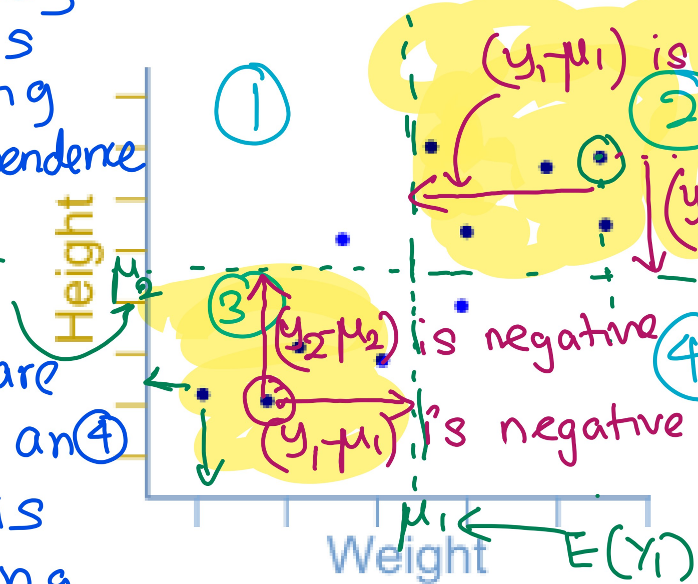
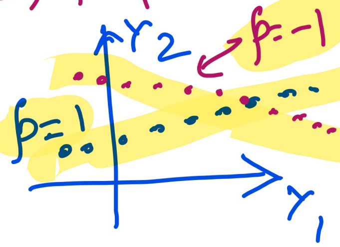
$(y_1-\mu_1)(y_2-\mu_2)$

(*) If more points are in quadrants ② and ③, then Covariance is positive indicating an increasing Linear dependence

$E(Y_2)$ = mean height

(*) If more points are in quadrants ① and ④ then Covariance is negative indicating a decreasing Linear dependence.





Pearson's Correlation:

"Rho"

$$Corr\,(Y_1,\ Y_2) = \rho_{Y_1Y_2} = \frac{Cov\,(Y_1,\ Y_2)}{\sigma_{Y_1}\,\sigma_{Y_2}} = \frac{E(Y_1Y_2)-E(Y_1)E(Y_2)}{\sqrt{Var(Y_1)\,Var(Y_2)}}$$

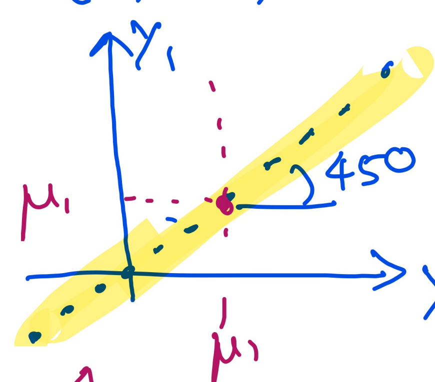
Note: $-1 \le \rho_{Y_1Y_2} \le 1$

Note: $Cov\,(Y_1,\ Y_2) = Cov\,(Y_2,\ Y_1)$. $= E(Y_2Y_1) - E(Y_2)E(Y_1)$

What is $Cov\,(Y_1,\ Y_1)$? By Definiton 5.10,

$$Cov(Y_1,Y_1) = E[(Y_1-\mu_1)(Y_1-\mu_1)]$$
$$= E[(Y_1-\mu_1)^2]$$
$$= Var(Y_1)$$

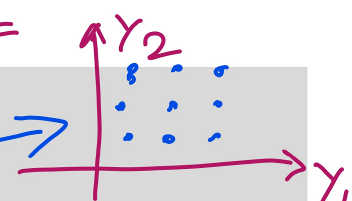
$$Corr(Y_1,Y_1) = \frac{Cov(Y_1,Y_1)}{\sqrt{Var(Y_1)Var(Y_1)}} = \frac{Var(Y_1)}{Var(Y_1)} = 1$$



Theorem 5.11:
If Y_1 **and** Y_2 are **independent** random variables, then

$$Cov\,(Y_1,\ Y_2) = 0$$

because $\quad E(Y_1Y_2) - [E(Y_1)\,E(Y_2)] \;=\; \underbrace{E(Y_1)\,E(Y_2)}_{\text{independence}} - [E(Y_1)\,E(Y_2)]$



Thus, **independent random variables must be uncorrelated.**

However, the **converse is not true**. That is, **if the covariance** between two random variables **is zero, this does necessarily imply they are independent**.

For independent variables, $E[g(Y_1)\,h(Y_2)] = E[g(Y_1)]\,E[h(Y_2)]$

If $cov(Y_1,Y_2)=0$, then Y_1 and Y_2 could be independent OR they have a NON-Linear dependence.

$\rho = Corr(Y_1, Y_2)$

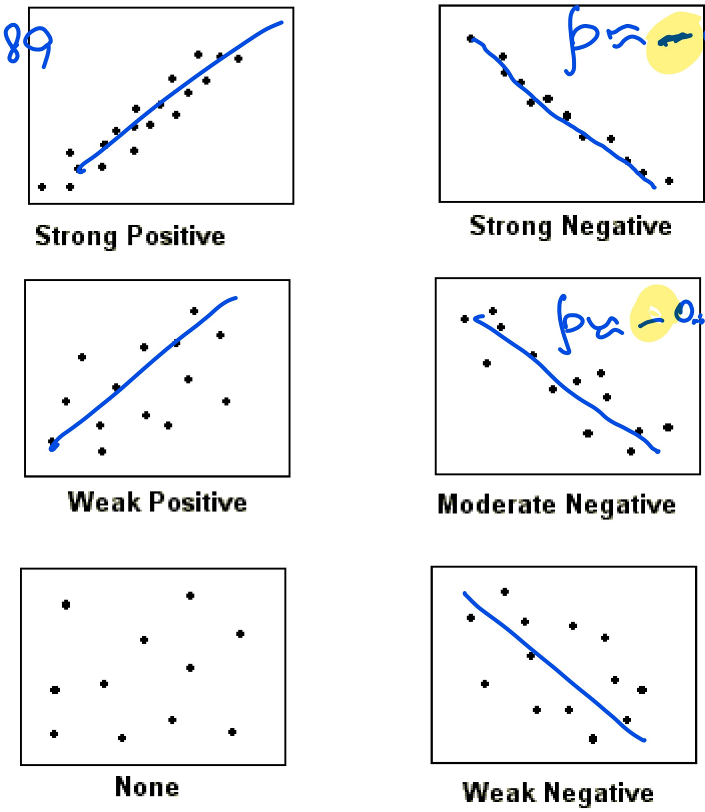
$\rho = 0$

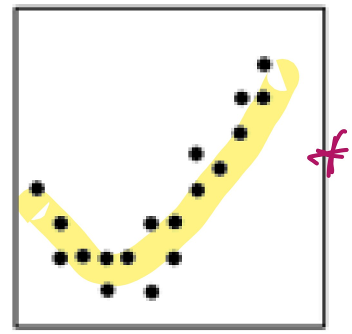
Degree of Correlation

$\rho \approx 0.89$ $\rho \approx -0.9$

$\rho = 0.7$ $\rho \approx -0.81$

ρ is almost zero





There is NO Linear dependence. However, there is a NON-Linear dependence. Therefore, covariance is almost zero.

Covariance is also useful when calculating the variance of linear functions of random variables.

Theorem 5.12:
If Y_1 and Y_2 are jointly distributed random variables, then

$$\boldsymbol{E(aY_1 + bY_2) = a\,E(Y_1) + b\,E(Y_2)}$$

$$\boldsymbol{Var(aY_1 + bY_2) = a^2\,Var(Y_1) + b^2\,Var(Y_2) + 2ab\,Cov\,(Y_1,\ Y_2)}$$

If Y_1 **and** Y_2 are **independent** random variables, then

$$\boldsymbol{Var(aY_1 + bY_2) = a^2\,Var(Y_1) + b^2\,Var(Y_2)}$$

Recall: $Var(Y) = E(Y^2) - [E(Y)]^2$

Proof:
Suppose Y_1 and Y_2 are jointly distributed random variables such that $E(Y_1) = \mu_{Y_1}$, $Var(Y_1) = \sigma^2_{Y_1}$, $E(Y_2) = \mu_{Y_2}$, and $Var(Y_2) = \sigma^2_{Y_2}$.

$$Var(aY_1 + bY_2) = E\left[(aY_1 + bY_2)^2\right] - \left[E(aY_1 + bY_2)\right]^2$$

$$= E\left[a^2Y_1^2 + b^2Y_2^2 + 2abY_1Y_2\right] + \left[aE(Y_1), bE(Y_2)\right]^2$$

$$= E\left[a^2Y_1^2\right] + E\left[b^2Y_2^2\right] + E\left[2abY_1Y_2\right] - \left[a^2[E(Y_1)]^2 + b^2[E(Y_2)]^2 + 2abE(Y_1)E(Y_2)\right]$$

$$Var(aY_1 + bY_2) = a^2E(Y_1^2) + b^2E(Y_2^2) + 2abE(Y_1Y_2) - a^2[E(Y_1)]^2 - b^2[E(Y_2)]^2 - 2abE(Y_1)E(Y_2)$$

$$= a^2\Big[\underbrace{E(Y_1^2) - [E(Y_1)]^2}_{Var(Y_1)}\Big] + b^2\Big[\underbrace{E(Y_2^2) - [E(Y_2)]^2}_{Var(Y_2)}\Big]$$
$$2ab\Big[\underbrace{E(Y_1Y_2) - E(Y_1)E(Y_2)}_{Cov(Y_1,Y_2)}\Big]$$

$$= a^2Var(Y_1) + b^2Var(Y_2) + 2ab\,Cov(Y_1,Y_2).$$

* Variance Can Never be Negative:

Example 5 Continued: Consider the joint pmf

$$p_{Y_1,Y_2}(y_1,y_2) = \binom{y_2}{y_1}p^{y_1}(1-p)^{(y_2-y_1)}\left(\frac{e^{-\lambda}\lambda^{y_2}}{y_2!}\right) \quad ; \quad y_2 = 0,1,\ldots; \quad y_1 = 0,1,\ldots,y_2$$

(a) Find $Corr(Y_1, Y_2)$.
(b) Find $E(3Y_1 - 4Y_2)$
(c) Find $Var(3Y_1 - 4Y_2)$

What do we know? From our previous work, we found Y_1 has a Poisson (λp) and therefore, $E(Y_1) = \lambda p$
$Var(Y_1) = \lambda p$.

Y_2 has a Poisson (λ) and therefore, $E(Y_2) = \lambda$
$Var(Y_2) = \lambda$.

(a) $$Corr(Y_1,Y_2) = \frac{Cov(Y_1,Y_2)}{\sqrt{Var(Y_1)Var(Y_2)}} = \frac{E(Y_1Y_2) - E(Y_1)E(Y_2)}{\sqrt{Var(Y_1)Var(Y_2)}}$$

$$E(Y_1Y_2) = \sum_{\text{all } y_1}\sum_{\text{all } y_2} y_1y_2\,p(y_1,y_2)$$

$$E(Y_1Y_2) = \sum_{y_2=0}^{\infty}\sum_{y_1=0}^{y_2}\left[y_1y_2\binom{y_2}{y_1}p^{y_1}(1-p)^{(y_2-y_1)}\frac{e^{-\lambda}\lambda^{y_2}}{y_2!}\right]$$

$$= \sum_{y_2=0}^{\infty}\left[y_2\frac{e^{-\lambda}\lambda^{y_2}}{y_2!}\sum_{y_1=0}^{y_2}\left[y_1\underbrace{\binom{y_2}{y_1}p^{y_1}(1-p)^{(y_2-y_1)}}_{Y_1\sim\text{Binomial pmf }(y_2,\,p)}\right]\right]$$

$$= \sum_{y_2=0}^{\infty}y_2\frac{e^{-\lambda}\lambda^{y_2}}{y_2!}\underbrace{\sum_{y_1=0}^{y_2}\left[y_1\,p_1(y_1)\right]}_{E(Y_1)=y_2p}$$

$$= \sum_{y_2=0}^{\infty}y_2\frac{e^{-\lambda}\lambda^{y_2}}{y_2!}(y_2p)$$

$$= p\sum_{y_2=0}^{\infty}\left[y_2^2\underbrace{\frac{e^{-\lambda}\lambda^{y_2}}{y_2!}}_{Y_2\text{ has Poisson pmf }(\lambda)}\right]$$

$$= pE(Y_2^2)$$

$$= p(\lambda+\lambda^2).$$

Recall: $Var(Y_2)=E(Y_2^2)-[E(Y_2)]^2$

$\lambda = E(Y_2^2)-\lambda^2$

$\Rightarrow E(Y_2^2)=\lambda+\lambda^2$

$$\Rightarrow Cov(Y_1,Y_2) = E(Y_1Y_2)-E(Y_1)E(Y_2)$$

$$= p(\lambda+\lambda^2)-(\lambda p)(\lambda) = \lambda p$$

$$Corr(Y_1,Y_2) = \frac{Cov(Y_1,Y_2)}{\sqrt{Var(Y_1)Var(Y_2)}} = \frac{\lambda p}{\sqrt{(\lambda p)(\lambda)}} = \frac{p}{\sqrt{p}} = \underline{\underline{\sqrt{p}}}$$

(b) $E(3Y_1-4Y_2) = 3E(Y_1)-4E(Y_2)$

$= 3(\lambda p)-4(\lambda) = \underline{\underline{\lambda(3p-4)}}$

(c) $Var(3Y_1-4Y_2)$ here $a=3$, $b=-4$ compared to aY_1+bY_2.

$$Var(aY_1+bY_2) = a^2Var(Y_1)+b^2Var(Y_2)+2abCov(Y_1,Y_2)$$

$$Var(3Y_1 - 4Y_2) = Var(3Y_1 + (-4)Y_2)$$
$$= 3^2 Var(Y_1) + (-4)^2 Var(Y_2) + 2(3)(-4)Cov(Y_1, Y_2)$$
$$= 9(\lambda p) + 16(\lambda) - 24(\lambda p)$$
$$= \underline{\lambda(16 - 15p)}$$ Is this non-zero?
$$= \frac{\lambda}{15}\left(\frac{16}{15} - p\right)$$ $\lambda > 0$ and $0 < p < 1$
Therefore, this is positive.

Example 11: (Problem 5.102) A firm purchases two types of chemicals. Type I chemical costs \$3 per gallon, whereas type II costs \$5 per gallon.
Let Y_1 be the number of gallons purchased from type I chemical.
The mean and variance for this are 40 and 4, respectively.
Let Y_2 be the number of gallons purchased from type II chemical.
The mean and variance for this are 65 and 8, respectively.
Assume that Y_1 and Y_2 are independent. $\Rightarrow Cov(Y_1, Y_2) = 0$
Find the mean and standard deviation of total amount of money spent on the two chemicals.

$E(Y_1) = 40$, $Var(Y_1) = 4$; $E(Y_2) = 65$, $Var(Y_2) = 8$

\$3 per gallon of Type I $\Rightarrow$ Cost is $3Y_1$
\$5 per gallon of Type II $\Rightarrow$ cost is $5Y_2$
Therefore, total cost is $3Y_1 + 5Y_2$.

Mean cost $= E[3Y_1 + 5Y_2]$
$$= 3E(Y_1) + 5E(Y_2)$$
$$= \boxed{(3 \times 40) + (5 \times 65)}$$
$$= \underline{445} \text{ dollars}$$

Standard Deviation = $\sqrt{\text{Variance}}$.
$$Var(3Y_1 + 5Y_2) = 3^2 Var(Y_1) + 5^2 Var(Y_2) + 2(3)(5)\underbrace{Cov(Y_1, Y_2)}_{0}$$
$$= (9 \times 4) + (25 \times 8)$$
$$= 236 \text{ dollars squared}$$

Standard deviation $= \sqrt{236}$
$$\approx \underline{15.36} \text{ dollars}$$