Correlations & Bivariate Distribution

Monday, 28 February 2022 11:43

Correlations:

$$Cor(X,Y) = \rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

σ_X : variance of X $\quad Cov(X,Y) \equiv \sigma_{XY}$

σ_Y : variance of Y

Properties:

1. ρ is the covariance of the standardization of X and Y.

$X \to (X-\mu_X)/\sigma_X$ $\to \tilde{\mu}_X = 0, \tilde{\sigma}_X = 1$

$Y \to (Y-\mu_Y)/\sigma_Y$ $\to \tilde{\mu}_Y = 0, \tilde{\sigma}_Y = 1$

$Cov(\tilde{X},\tilde{Y}) \equiv Corr(\tilde{X},\tilde{Y})$

2. ρ is dimensionless.

3. $-1 \le \rho \le +1$

$\rho = +1$ if and only if $Y = aX + b$ with $a > 0$

$\rho = -1$ if and only if $Y = aX + b$ with $a < 0$

Proof

Example:

We flip a fair coin 3 times.

X: # of heads in the first 2 flips.

Y: # of heads in the last 2 flips.

$Cov(X,Y)$ $\quad p(x,y)$

X \ Y	0	1	2	$P(x_i)$
0	1/8	1/8	0	1/4
1	1/8	2/8	1/8	1/2
2	0	1/8	1/8	1/4
$P(y_i)$	1/4	1/2	1/4	$\boxed{1}$ ← Total Prob.

$Cov(x,y) = 1/4.$ $\qquad Var(x) = \sum x_i M(x_i)$

$Var(x) = 1/2 \rightarrow \sigma_x = 1/\sqrt{2}.$

$Var(y) = 1/2 \rightarrow \sigma_y = 1/\sqrt{2}.$

$$Cor(x,y) \equiv \rho = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{1/4}{1/\sqrt{2} \cdot 1/\sqrt{2}} = \frac{1}{2}.$$

Two-Dimensional (Bivariate) Distributions

2-dim dist: $f(x,y)$

$x_i, y_j \rightarrow$ drawn from $f(x,y)$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y) = 1$$

$$\sigma_x = \sqrt{V_x} \leftarrow V_x = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\mu_x)^2 f(x,y)\,dx\,dy.$$

$$\sigma_y = \sqrt{V_y} \leftarrow V_y = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (y-\mu_y)^2 f(x,y)\,dx\,dy.$$

$$\mu_x = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x \cdot f(x,y)\,dx\,dy.$$

$$\mu_y = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y \cdot f(x,y)\,dx,dy.$$

$$\underbrace{Cov(x,y)}_{\sigma_{xy}} = V_{xy} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\mu_x)(y-\mu_y) f(x,y)\,dx\,dy. \leftarrow$$

$Z = x+y \rightarrow Var(Z) \equiv V_Z \text{ (or } \sigma_Z) = \sigma_x + \sigma_y + 2\sigma_{xy}$

$Z = x-y \rightarrow Var(Z) \equiv \sigma_Z = \sigma_x + \sigma_y - 2\sigma_{xy}$

Bivariate Gaussian Distribution

$$p(x,y \mid \underbrace{\mu_x, \mu_y, \sigma_x, \sigma_y, \sigma_{xy}}_{\text{parameters of the Gaussian distribution.}}) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(\frac{-z^2}{2(1-\rho^2)}\right)$$

$$z^2 = \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}.$$

$$\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

Contours in the (x,y) plane.

- $p(x,y \mid \mu_x, \mu_y, \sigma_x, \sigma_y, \sigma_{xy}) = \text{Constant}.$
- ellipses centered on $(x=\mu_x, y=\mu_y)$
- angle α $\quad (-\pi/2 \le \alpha \le \pi/2)$

$$\tan(2\alpha) = 2\rho\frac{\sigma_x\sigma_y}{\sigma_x^2-\sigma_y^2} = 2\frac{\sigma_{xy}}{\sigma_x^2-\sigma_y^2}.$$

$$\left.\begin{aligned} P_1 &= (x-\mu_x)\cos\alpha + (y-\mu_y)\sin\alpha \\ P_2 &= -(x-\mu_x)\sin\alpha + (y-\mu_y)\cos\alpha. \end{aligned}\right\}$$

$$P_2 = -(x-\mu_x)\sin\alpha + (y-\mu_y)\cos\alpha$$

There will be No correlation b/w P_1

$$\sigma_{1,2}^2 = \frac{\sigma_x^2+\sigma_y^2}{2} \sqrt{\left(\frac{\sigma_x^2-\sigma_y^2}{2}\right)^2+\sigma_{xy}^2}$$

Robust way to Estimate from Data: