Expected value of linear combination of random variables

1. The expected value of the sum of n random variables is the sum of n respective expected values.

$$E(L) = E(c_1X_1 + \ldots + c_nX_n) = c_1E(X_1) + c_2E(X_2) + \ldots c_nE(X_n)$$

2. Ex. (Renal Disease) Suppose the expected values of serum creatinine for the white and the black individuals are 1.3 and 1.5 respectively. What is the expected value of the average serum-creatinine level of a single white and a single black individual? The expected value of the average serum-creatinine level $= E(0.5X_1 + 0.5X_2) = 0.5E(X_1) + 0.5E(X_2) = 0.65 + 0.75 = 1.4$

3. Ex. (Renal Disease) Suppose X_1, X_2 are defined as before. If we know that $Var(X_1) = Var(X_2) = 0.25$, then what is the variance of the average serum-creatinine level over a single white and a single black individual? $Var(0.5X_1 + 0.5X_2) = (0.5)^2Var(X_1) + (0.5)^2Var(X_2) = 0.25(0.25) + 0.25(0.25) = 0.125$

4. If X_1, X_2 are defined as in previous examples and are each normally distributed, then what is the distribution of the average $L = 0.5X_1 + 0.5X_2$? Since $E(L) = 1.4, Var(L) = .125$, then $(X_1 + X_2)/2 \sim N(1.4, 0.125)$

Normal Approximation to the Binomial distribution

Ex. Suppose a binomial distribution has parameters $n = 25, p = .4$. How can $P(7 \leq X \leq 12)$ be approximated? $np = 25(.4) = 10, np(1-p) = 25(.4)(.6) = 6.0$. This distribution can be approximated by $N(10, 6)$. $P(7 \leq X \leq 12)$ equals the area under the normal curve from 6.5 to 12.5.

Normal Approximation to the Poisson Distribution

Ex. Consider again the distribution of the number of bacteria in a petri plate of area A. Assume the probability of observing X bacteria is given exactly by a Poisson distribution with parameter $\mu = \lambda A$, where $\lambda = 0.1\, bacteria/cm^2$ and $A = 100\, cm^2$. Suppose 20 bacteria are observed in this area. How unusual is this event? The exact distribution of the number of bacteria observed in $100\, cm^2$ is Poisson with parameter $\mu = 10$. We can approximate this distribution by a normal distribution with mean = 10 and variance = 10

Association between two variables

1. Covariance between two variables X and Y is denoted by Cov(X, Y) and defined by

$$Cov(X, Y) = E(X - E(X))(Y - E(Y))$$

2. Covariance is not convenient for expressing the strength of association between two variables.

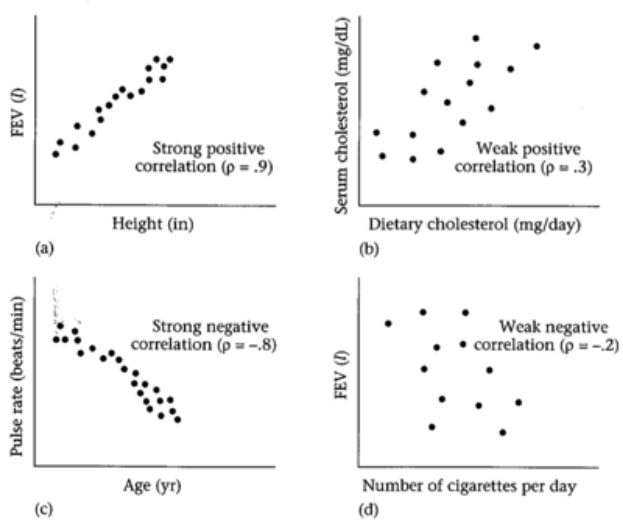
3. The correlation coefficient between 2 random variables X and Y is denoted by $Corr(X, Y)$ and is defined by

$$\rho = corr(X, Y) = \frac{Cov(X, Y)}{\sigma_x \sigma_y}$$

4. ρ is a dimensionless quantity between -1 and 1, for linearly related random variables, 0 implies independence In general, correlation zero does not necessarily imply independence

5. 1 implies nearly perfect positive dependence, -1 implies nearly perfect negative dependence.

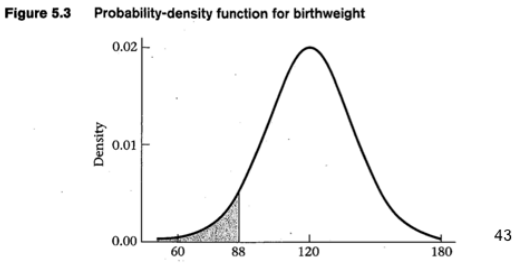
Figure 5.16 Interpretation of various degrees of correlation



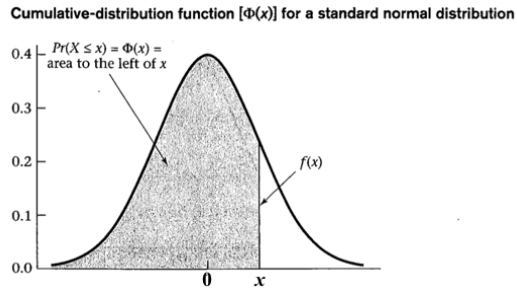
40

Cumulative distribution function

Cumulative distribution function (cdf) for the random variable X evaluated at the point a is defined as the probability that X will take on values $\leq$ a. It is represented by the area under the pdf to the left of a.



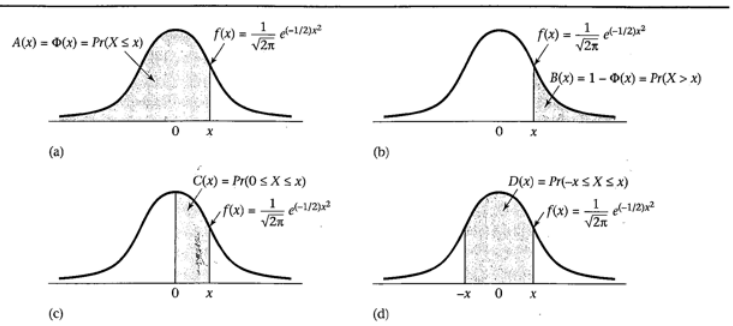
(a) cdf



(b) cdf of normal

Normal Table

Table 3 The normal distribution



x	A[a]	B[b]	C[c]	D[d]	x	A	B	C	D
0.0	.5000	.5000	.0	.0	0.32	.6255	.3745	.1255	.2510
0.01	.5040	.4960	.0040	.0080	0.33	.6293	.3707	.1293	.2586
0.02	.5080	.4920	.0080	.0160	0.34	.6331	.3669	.1331	.2661
0.03	.5120	.4880	.0120	.0239	0.35	.6368	.3632	.1368	.2737
0.04	.5160	.4840	.0160	.0319	0.36	.6406	.3594	.1406	.2812
0.05	.5199	.4801	.0199	.0399	0.37	.6443	.3557	.1443	.2886
0.06	.5239	.4761	.0239	.0478	0.38	.6480	.3520	.1480	.2961
0.07	.5279	.4721	.0279	.0558	0.39	.6517	.3483	.1517	.3035
0.08	.5319	.4681	.0319	.0638	0.40	.6554	.3446	.1554	.3108
0.09	.5359	.4641	.0359	.0717	0.41	.6591	.3409	.1591	.3182
0.10	.5398	.4602	.0398	.0797	0.42	.6628	.3372	.1628	.3255
0.11	.5438	.4562	.0438	.0876	0.43	.6664	.3336	.1664	.3328
0.12	.5478	.4522	.0478	.0955	0.44	.6700	.3300	.1700	.3401