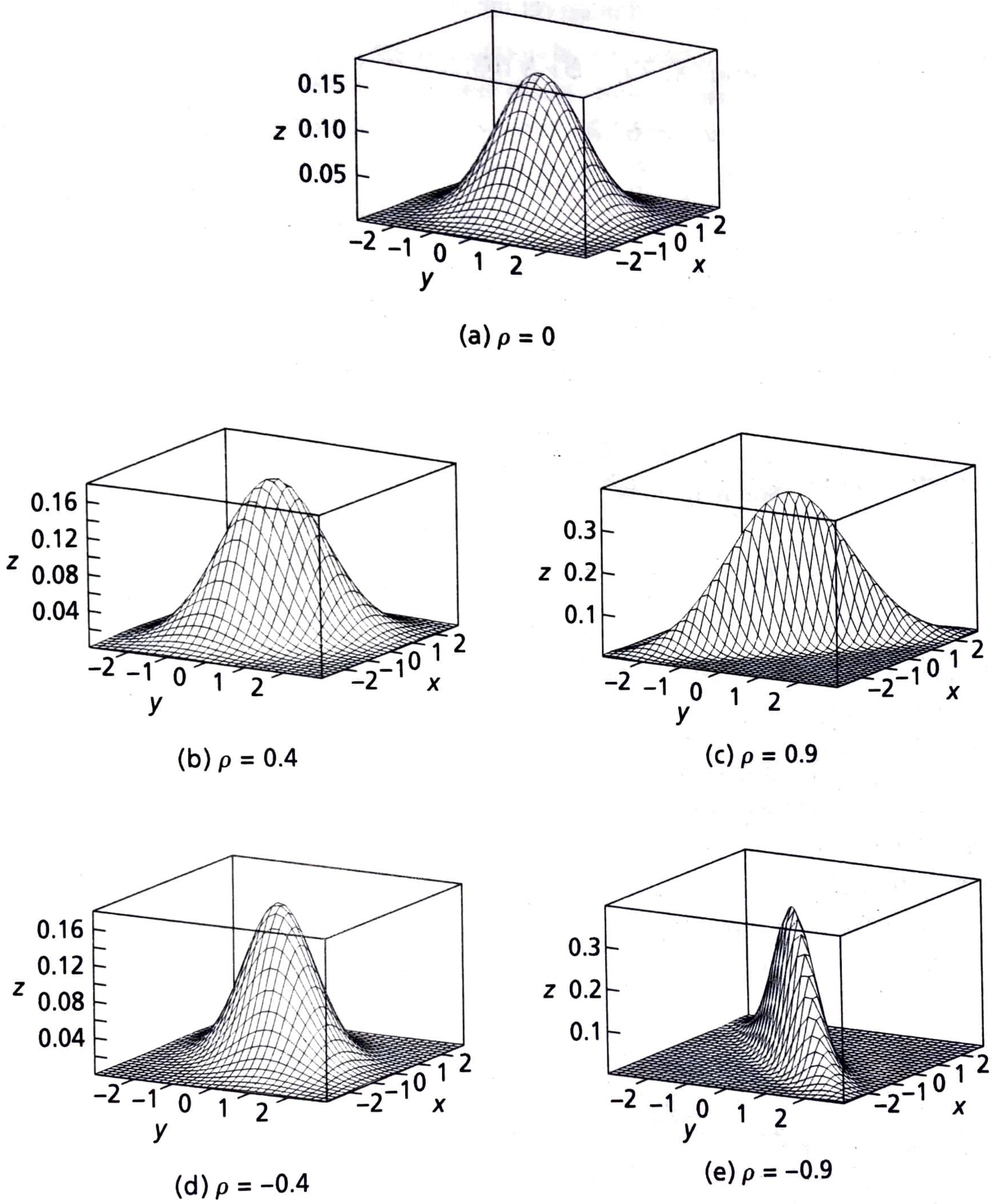
Figure 10.2 Selected standard bivariate normal distributions



EXAMPLE 10.21 ***The Bivariate Normal Distribution***

Heights of Mothers and Sons For a family with sons, let X and Y denote the heights, in inches, of the mother and eldest son, respectively. Data from the *National Health and Nutrition Survey* and research by K. Pearson and A. Lee ("On the Laws of Inheritance in Man. I. Inheritance of Physical Characters," *Biometrika,* 1903, Vol. 2, pp. 357–462) show that it's reasonable to presume that X and Y have a bivariate normal distribution. Specifically, we can assume that $(X, Y) \sim \mathcal{BVN}(63.7, 2.7^2, 69.1, 2.9^2, 0.5)$.

a) Determine and graph a joint PDF of this bivariate normal distribution.

b) Identify and interpret the marginal distribution of X.

c) Identify and interpret the marginal distribution of Y.
d) Identify and interpret the correlation coefficient of X and Y.

Solution Because $(X, Y) \sim \mathcal{BVN}(63.7, 2.7^2, 69.1, 2.9^2, 0.5)$, we have

$$\mu_X = 63.7, \quad \sigma_X = 2.7, \quad \mu_Y = 69.1, \quad \sigma_Y = 2.9, \quad \rho = 0.5. \tag{10.59}$$

a) From Equations (10.59),

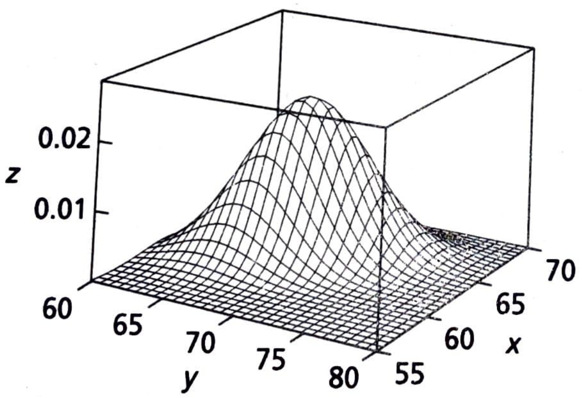
$$2\sigma_X\sigma_Y\sqrt{1-\rho^2} = 2 \cdot 2.7 \cdot 2.9\sqrt{1-0.5^2} = 7.83\sqrt{3},$$
$$1-\rho^2 = 1-0.5^2 = 3/4,$$
$$2\rho = 2 \cdot 0.5 = 1.$$

Therefore, by Definition 10.4 and Equations (10.59), a joint PDF of mother and eldest son heights is

$$f_{X,Y}(x,y) = \frac{1}{7.83\sqrt{3}\pi} e^{-\frac{2}{3}\left\{\left(\frac{x-63.7}{2.7}\right)^2 - \left(\frac{x-63.7}{2.7}\right)\left(\frac{y-69.1}{2.9}\right) + \left(\frac{y-69.1}{2.9}\right)^2\right\}}.$$

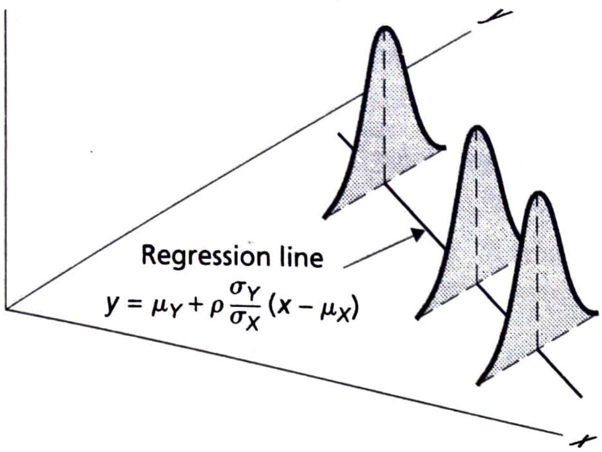
A graph of this bivariate normal PDF is shown in Figure 10.3.

Figure 10.3 Graph of joint PDF of a bivariate normal distribution with parameters $\mu_X = 63.7$, $\sigma_X^2 = 2.7^2$, $\mu_Y = 69.1$, $\sigma_Y^2 = 2.9^2$, and $\rho = 0.5$



b) Because (X, Y) is bivariate normal, we know that the marginal distribution of X is $\mathcal{N}(\mu_X, \sigma_X^2) = \mathcal{N}(63.7, 2.7^2)$. Hence the heights of mothers (with sons) are normally distributed with mean 63.7 inches and standard deviation 2.7 inches.
c) Because (X, Y) is bivariate normal, we know that the marginal distribution of Y is $\mathcal{N}(\mu_Y, \sigma_Y^2) = \mathcal{N}(69.1, 2.9^2)$. Hence the heights of eldest sons are normally distributed with mean 69.1 inches and standard deviation 2.9 inches.
d) We know that $\rho(X, Y) = \rho = 0.5$. Hence the correlation between the heights of mothers and their eldest sons is 0.5. We see that a moderate positive correlation exists between the height of a mother and that of her eldest son. In particular, eldest sons of tall mothers tend to be taller than eldest sons of short mothers. ∎

Figure 10.4 Conditional distributions of Y given $X = x$ for bivariate normal random variables



Prediction in the Bivariate Normal Setting

Referring to Equation (10.64) and applying the prediction theorem (Proposition 10.8 on page 602), we conclude that the best possible predictor of Y, given that X equals x, is the quantity $\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$. The equation $y = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$ is called the **regression equation,** and the straight line it represents is called the **regression line,** as illustrated in Figure 10.4.

EXAMPLE 10.22 ***Prediction in the Bivariate Normal Setting***

Heights of Mothers and Sons Refer to Example 10.21 on page 611.

a) Determine the conditional distribution of Y given $X = x$.
b) Obtain the regression equation for predicting the height of an eldest son based on the height of his mother.
c) Apply part (b) to predict the height of an eldest son whose mother is 62 inches tall.

Solution Because $(X, Y) \sim \mathcal{BVN}(63.7, 2.7^2, 69.1, 2.9^2, 0.5)$, we have

$$\mu_X = 63.7, \quad \sigma_X = 2.7, \quad \mu_Y = 69.1, \quad \sigma_Y = 2.9, \quad \rho = 0.5. \tag{10.65}$$

a) From Equations (10.63) and (10.65), we conclude that the conditional distribution of Y given $X = x$ is the normal distribution with mean

$$\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X) = 69.1 + 0.5 \cdot \frac{2.9}{2.7}(x - 63.7) \approx 34.9 + 0.537x \tag{10.66}$$

and variance $\sigma_Y^2(1 - \rho^2) = 2.9^2(1 - 0.5^2) \approx 6.31$.

b) From Relation (10.66), the regression equation for predicting the height of an eldest son, based on the height of his mother, is given approximately by $y = 34.9 + 0.537x$.
c) To use a mother's height to predict the height of her eldest son, we use the regression equation found in part (b). For a mother who is 62 inches tall, we predict the height of her eldest son to be roughly $34.9 + 0.537 \cdot 62$, or about 68.2 inches. ■