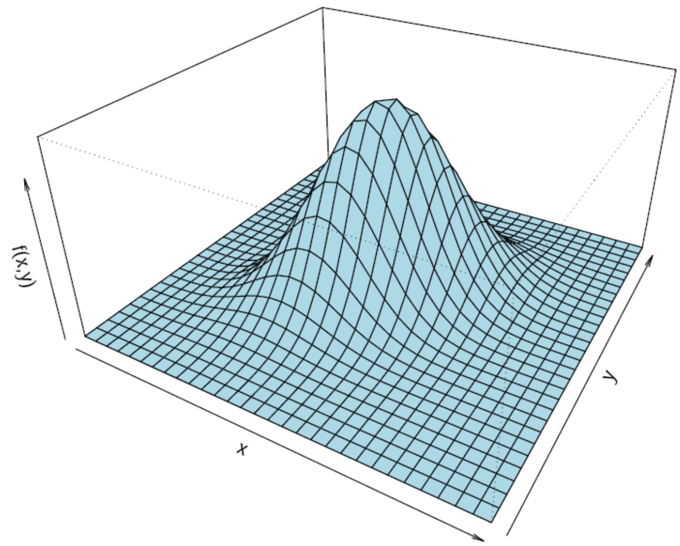
- **Simple Linear Regression**
  - Model: $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$
  - <u>*Pearson Correlation Coefficient*</u> between 2 Variables (r_{yx})
    - $r_{yx} = \dfrac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2 \sum (y_i - \overline{y})^2}} = \hat{\beta}_1 \dfrac{s_x}{s_y}$
    - $R^2 = r_{yx}^2 = \frac{SS(Regression)}{SS(Total)}$

- **Multiple Linear Regression**
  - Model: $Y_i = \beta_0 + \beta_1 \; X_{i1} + ... + \beta_k \; X_{ik} + \epsilon_i$
  - Lots of *Pearson Correlation Coefficients*: $r_{yx_1}, r_{yx_2}, ..., r_{yx_k}, r_{x_i x_j}$
  - <u>*Multiple Correlation Coefficient*</u> ($R_{y|x_1,x_2,...,x_k}$)
    - Measures overall linear association between Y and the entire set of variables $X_1, X_2, ..., X_k$
    - $R_{y|x_1,x_2,...,x_k} = r_{y\hat{y}} = \dfrac{\sum (y_i - \overline{y})(\hat{y}_i - \overline{\hat{y}})}{\sqrt{\sum (y_i - \overline{y})^2 \sum (\hat{y}_i - \overline{\hat{y}})^2}} = \sqrt{\dfrac{SS(Regression)}{SS(Total)}}$
    - $R^2 = R^2_{y|x_1,x_2,...,x_k}$
  - <u>*Partial Correlation Coefficients*</u> (e.g., $r_{yx_1|z}$)
    - Measures the linear association between 2 variables after controlling for the effects of other variables (called *covariates*)
    - $r_{yx_1|z} = \dfrac{r_{yx_1} - r_{yz} r_{x_1 z}}{\sqrt{(1 - r_{yz}^2)(1 - r_{x_1 z}^2)}}$

Bivariate Normal Distribution

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}e^{\frac{-1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right)}$$



$$E(Y \mid X=x)=\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)$$

$$\hat{\mu}_{Y|X=x}=\bar{y}+r\frac{s_y}{s_x}(x-\bar{x})$$