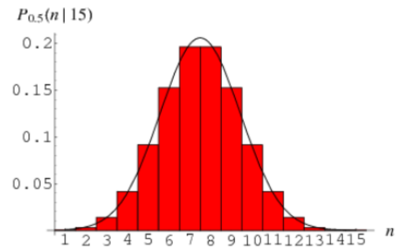
•The basic Practice of Statistics

•The normal or Gaussian distribution (or density, histogram).
•Density curve is a curve that: a) is always on or above the horizontal axis, and b) has area exactly 1 underneath it.
•A density cure describes the overall pattern of a distribution.
•The area under the curve and above any range of values is the proportion of all observation that fall in that range.

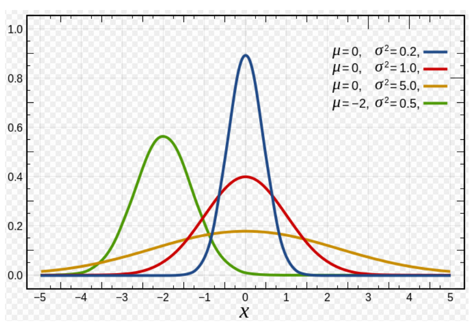


Review of Statistics and Probability Concepts

1

Normal or Gaussian or univariate Distribution

$N(\mu, \sigma^2)$

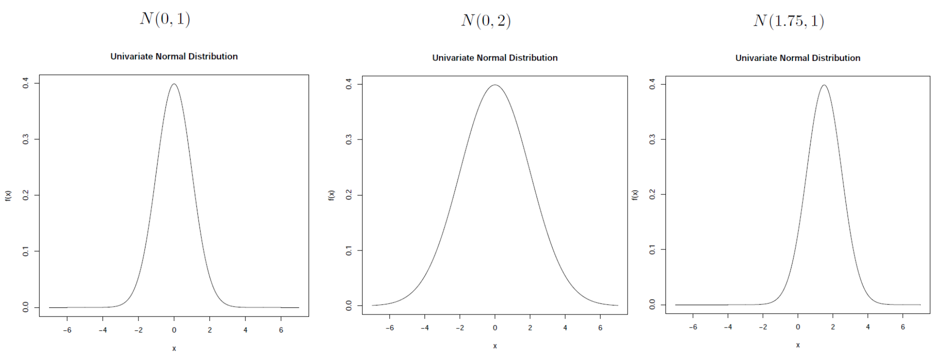


$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

where μ is the mean of the distribution and σ^2 is the variance which is the expectation of the squared deviation of random variable from its mean.
The mean of a density curve is the balance point, at which the curve would balance if made of solid materials.
The median of a density curve is the equal-area point, the point that divides the area under the curve in half.

Review of Statistics and Probability Concepts

2

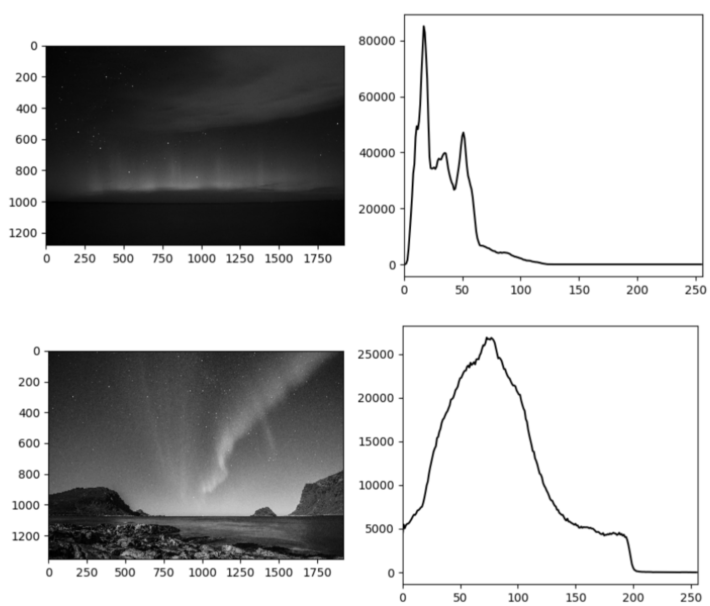


The variance σ^2 of a set of observation is an avarage of the squares of the deviation of the observations from their mean:

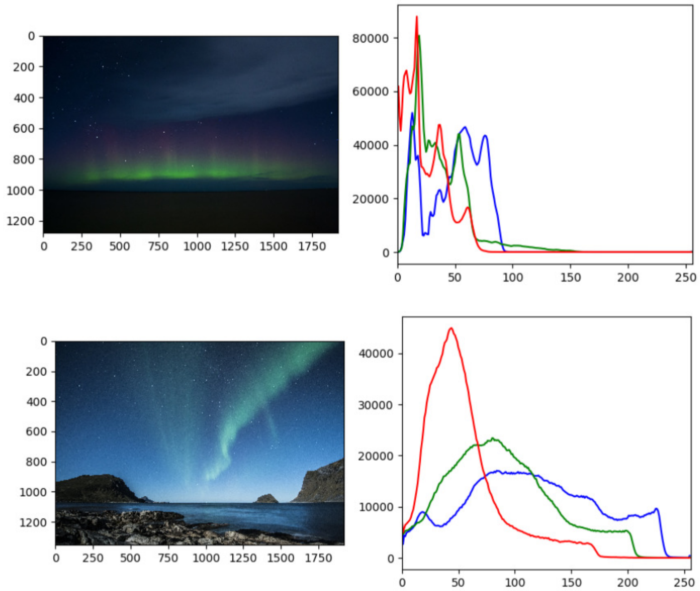
$$\sigma^2 = \frac{1}{n-1}\sum\left(x_i - \mu\right)^2$$

Standard deviation is the square root of the variance σ^2.

Review of Statistics and Probability Concepts 3



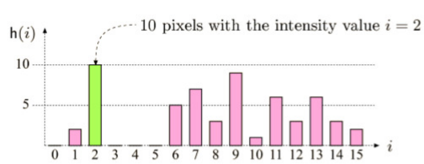
Review of Statistics and Probability Concepts 4



Review of Statistics and Probability Concepts 5

$\mathsf{h}(i) =$ the *number* of pixels in I with the intensity value i

Or, formally, $\mathsf{h}(i) = \mathrm{card}\{(u,v) \mid I(u,v) = i\}$



$\mathsf{h}(i)$	0	2	10	0	0	0	5	7	3	9	1	6	3	6	3	2
i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Cardinalities for Neighbours Lists
The function tallies the numbers of neighbours of regions in the neighbours list.

Review of Statistics and Probability Concepts 6

- If x is an observation from a distribution that has a mean μ and standard deviation σ, the standardized value of x is:

$$z = \frac{x-\mu}{\sigma}$$

- A standardized value is often called a z-score.
- The area under the curve for the univariate normal distribution is a function of the variance/standard deviation.
- In particular:

$$P(\mu-\sigma \le x \le \mu+\sigma) = 0.683$$
$$P(\mu-2\sigma \le x \le \mu+2\sigma) = 0.954$$

Review of Statistics and Probability Concepts 7

- Also note the term in the exponent:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

can be written as: $\left(\frac{(x-\mu)}{\sigma}\right)^2 = 1/2(x-\mu)(\sigma^2)^{-1}(x-\mu)$

- This is the square of the distance from x to μ in standard deviation units, and will be generalized for the Multivariate case.

Review of Statistics and Probability Concepts 8

The Multivariate Normal Distribution

Recall the univariate Normal distribution:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

The bivariate normal distribution:

$$\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}} = \mathcal{N}(\mu_1, \sigma_1^2)\mathcal{N}(\mu_2, \sigma_2^2)$$

Multivariate Normal Distribution- K-variate Normal distribution

The multivariate normal distribution function is defined as:

$$f(\mathrm{X}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-(\mathrm{X}-\mu)^T\Sigma^{-1}(\mathrm{X}-\mu)/2}$$

X is a n dimensional vector
μ is a n dimensional vector of mean values
Σ is the matrix of covariance
$|\Sigma|$ is the determinent of matrix Σ

The descriptive statistics $\sqrt{(\mathrm{X}-\mu)^T\Sigma^{-1}(\mathrm{X}-\mu)}$ is known as the Mahalanobis distance which represents the distance of the test point X from the mean μ.

Covariance matrix –

In probability theory and statistics, a covariance matrix is a matrix whose element at position i, j is the covariance between the ith and jth elements of a random vector. This expresses how two variable behave as a PAIR.

Intuitively, the covariance matrix generalizes the notion of variance to multiple dimensions. A positive value indicates a direct or increasing linear relationship. A negative value indicates a decreasing relationship.

Because the covariance of the ith random variable with itself is simply that random variable's variance, each element of the principle of the covariance matrix is variance of one of the random variables.

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$

$$\Sigma_{ij} = \mathrm{cov}\left(X_i, X_j\right) = E\left[\left(X_i - \mu_i\right)\left(X_j - \mu_j\right)\right]$$

where $\mu_i = E(X_i)$ is the expected value of the ith entry in the vector X, in other words:

$$\Sigma = \begin{bmatrix} E\left[(X_1-\mu_1)(X_1-\mu_1)\right] & E\left[(X_1-\mu_1)(X_2-\mu_2)\right] & \cdots & E\left[(X_1-\mu_1)(X_n-\mu_n)\right] \\ E\left[(X_2-\mu_2)(X_1-\mu_1)\right] & E\left[(X_2-\mu_2)(X_2-\mu_2)\right] & \cdots & E\left[(X_2-\mu_2)(X_n-\mu_n)\right] \\ \vdots & \vdots & \ddots & \vdots \\ E\left[(X_n-\mu_n)(X_1-\mu_1)\right] & E\left[(X_n-\mu_n)(X_2-\mu_2)\right] & \cdots & E\left[(X_n-\mu_n)(X_n-\mu_n)\right] \end{bmatrix}$$

$$= \begin{vmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{vmatrix}$$

The invesre Σ^{-1} if exists, is the invesre covariance matrix (concentration matrix).

Example:

Two variables (x and y) (along the columns) and two of their observations (along the rows) :

$$\begin{bmatrix} 3 & 7 \\ 2 & 4 \end{bmatrix}$$

$$\begin{bmatrix} \mathrm{var}(x) & \mathrm{cov}(x,y) \\ \mathrm{cov}(x,y) & \mathrm{var}(y) \end{bmatrix}$$

$$\mathrm{var}(x) = \frac{1}{n-1}\sum\left(x_i - \mu_x\right)^2$$

$$\mathrm{cov}(x,y) = \frac{1}{n-1}\sum\left(x_i - \mu_x\right)\left(y_i - \mu_y\right).$$

$$\mu_x = \frac{(3+2)}{2} = \frac{5}{2}$$

$$\mu_y = \frac{(7+4)}{2} = \frac{11}{2}$$

$$var(x) = (3 - \tfrac{5}{2})^2 + (2 - \tfrac{5}{2})^2$$

$$var(y) = (7 - \tfrac{11}{2})^2 + (4 - \tfrac{11}{2})^2$$

$$cov(x,y) = (3 - \tfrac{5}{2})(7 - \tfrac{11}{2}) + (2 - \tfrac{5}{2})(4 - \tfrac{11}{2})$$

Example:

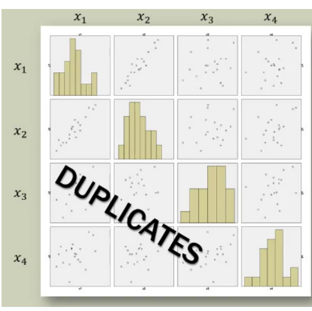
4 variables; x_1, x_2, x_3, x_4

Sample; $n = 20$

Statistics of Interest:
- Mean
- Variance
- Standard Deviation

Descriptive Statistics

	N	Mean		Std. Deviation	Variance
	Statistic	Statistic	Std. Error	Statistic	Statistic
x1	20	9.9550	.22448	1.00393	1.008
x2	20	20.0000	.21423	.95807	.918
x3	20	14.6800	.68528	3.06467	9.392
x4	20	15.7650	.33782	1.51076	2.282
Valid N (listwise)	20				



Matrix of Scatter Plots

Descriptive Statistics

	N	Mean		Std. Deviation	Variance
	Statistic	Statistic	Std. Error	Statistic	Statistic
x1	20	9.9550	.22448	1.00393	1.008
x2	20	20.0000	.21423	.95807	.918
x3	20	14.6800	.68528	3.06467	9.392
x4	20	15.7650	.33782	1.51076	2.282
Valid N (listwise)	20				

Inter-Item Covariance Matrix

	x1	x2	x3	x4
x1	1.008	.895	.634	.545
x2	.895	.918	.490	.652
x3	.634	.490	9.392	1.592
x4	.545	.652	1.592	2.282

The diagonal of a covariance matrix provides the variance of each individual variable; covariance with itself.

The off-diagonal entries in the matrix provide the covariance between each variable pair.

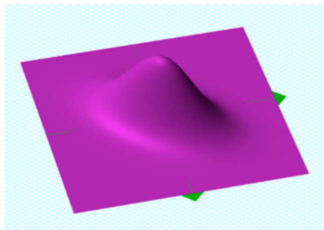
	x_1	x_2	x_3	x_4
x_1	$Var(x_1)$	$Cov(x_1, x_2)$	$Cov(x_1, x_3)$	$Cov(x_1, x_4)$
x_2		$Var(x_2)$	$Cov(x_2, x_3)$	$Cov(x_2, x_4)$
x_3			$Var(x_3)$	$Cov(x_3, x_4)$
x_4				$Var(x_4)$

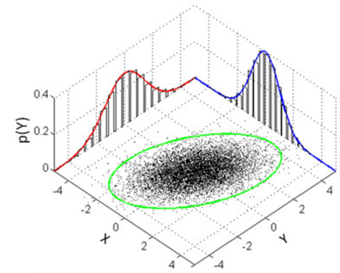
The diagonal of a covariance matrix provides the variance of each individual variable.

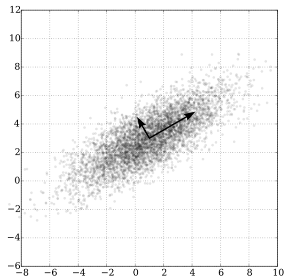
The off-diagonal entries in the matrix provide the **covariance** between each variable pair.

The bivariate normal distribution

$$\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}} = \mathcal{N}(\mu_1, \sigma_1^2)\mathcal{N}(\mu_2, \sigma_2^2)$$





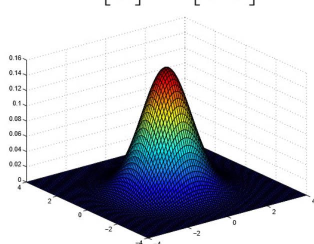


The directions of the arrows correspond to the eigenvectors of this bivariate covariance matrix and their lenghts to the square roots of the eigenvalues.

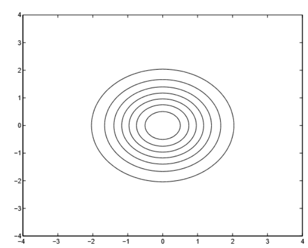
Review of Statistics and Probability Concepts

17

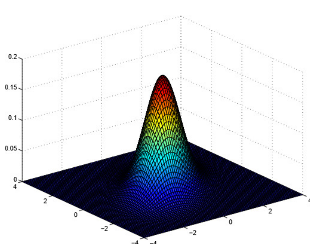
$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



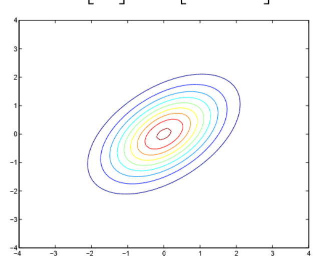
$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$



$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$



Review of Statistics and Probability Concepts

18

MVN Contours

- The lines of the contour plots denote places of equal probability, the lines represent points of both variables that lead to the same height on the for example, the z-axis (the height of the surface).
- The set all the vectors which has the same probability is also referred to as a level set.
- These contours can be constructed from the eigenvalues and eigenvectors of the covariance matrix.
- For example,

$$(\mathrm{X}-\mu)^T \Sigma^{-1} (\mathrm{X}-\mu) = c^2 = \text{ Level Set of X}$$

has ellipsoid centered at μ, and has axes $\pm c\sqrt{\lambda_i}\mathrm{e}_i$.

Review of Statistics and Probability Concepts

19

$$(\mathrm{X}-\mu)^T \Sigma^{-1} (\mathrm{X}-\mu) = c^2$$

Writing Σ in terms of eigenvector (using eigne value decomposition)

$$\Sigma = \mathrm{U}\Lambda\mathrm{U}^T$$

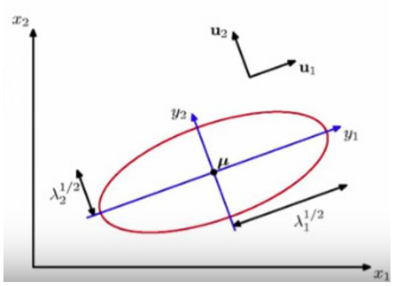
$$\Sigma = \begin{bmatrix} u_1 & u_2 \\ \downarrow & \downarrow \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} u_1 & \rightarrow \\ u_2 & \rightarrow \end{bmatrix}$$

Using the following transformation:

$$y_i = u_i^T (x-\mu)$$

We have:

$$c^2 = \frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2}$$



Review of Statistics and Probability Concepts

20

- Imagine we had a bivariate normal distribution with:

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

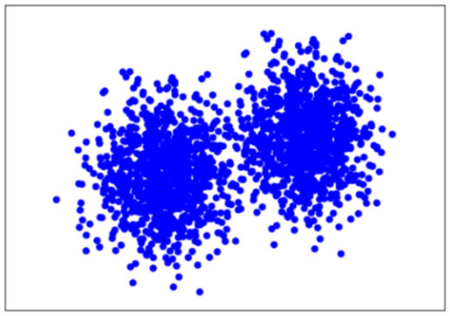
- The covariance matrix has eigenvalues and eigenvectors:

$$\boldsymbol{\lambda} = \begin{bmatrix} 1.5 \\ 0.5 \end{bmatrix}, \mathbf{E} = \begin{bmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{bmatrix}$$

Gaussian Mixture Model

- In statistics a **mixture model** is a probabilistic model for representing the presence of subpopulations within an overall population, without requiring that an observed data set should identify the sub-population to which an individual observation belongs.

- Formally a mixture model corresponds to the mixture distribution that represents the probability distribution of observations in the overall population.

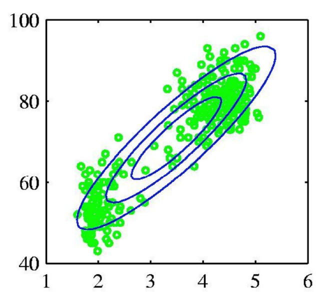


Single Guassian : $P_{\mu,\sigma^2}(x) = N(\mu,\sigma^2) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

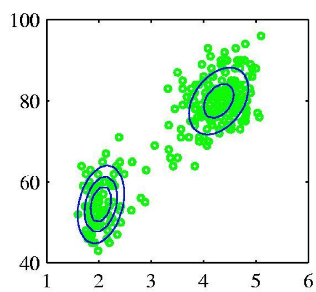
Two Gaussian: $P(\mathrm{X}) = \sum_j p_j \dfrac{1}{(2\pi)^{1/2}\left|\Sigma_j\right|^{1/2}} e^{-\frac{1}{2}(\mathrm{X}-\mu_j)^T \Sigma_j^{-1}(\mathrm{X}-\mu_j)}$

Where $\sum_j p_j = 1$

Review of Statistics and Probability Concepts 23



Single Gaussian



Mixture of two Gaussians

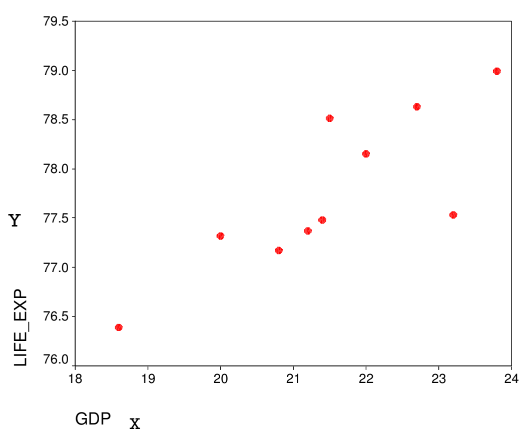
Review of Statistics and Probability Concepts 24

Scatter Plots and Correlation

- Response Variables, Explanatory variable
- A response variable measures an outcome of a measure.
- An explanatory variable explains or influences changes in a response variable.
- You will often find explanatory variables called independent variables, and response variables called dependent variables.

Country	Per Capita GDP (*X*)	Life Expectancy (*Y*)
Austria	21.4	77.48
Belgium	23.2	77.53
Finland	20.0	77.32
France	22.7	78.63
Germany	20.8	77.17
Ireland	18.6	76.39
Italy	21.5	78.51
Netherlands	22.0	78.15
Switzerland	23.8	78.99
United Kingdom	21.2	77.37

Scatter plot

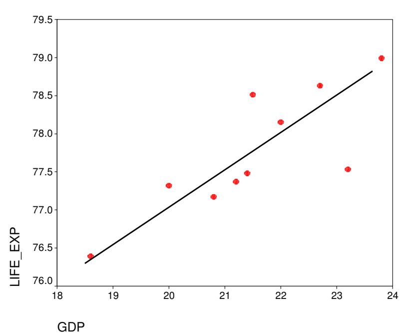


Interpreting scatter plots

- **Form:** Can relationship be described by straight line (linear)? ..by a curved line? etc.
- **Outliers?**: Any deviations from overall pattern?
- **Direction** of the relationship either:
  - Positive association (upward slope)
  - Negative association (downward slope)
  - No association (flat)
- ***Strength*:** Extent to which points *adhere* to imaginary trend line

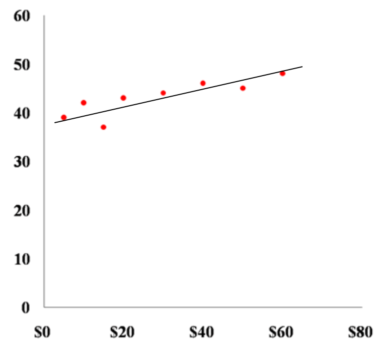
Review of Statistics and Probability Concepts 27

Example: Interpretation



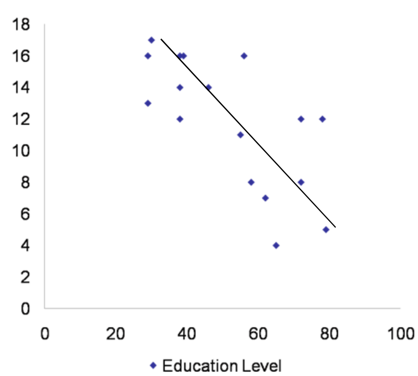
Interpretation:
- Form: linear (straight)
- Outliers: none
- Direction: positive
- Strength: difficult to judge by eye

Review of Statistics and Probability Concepts 28



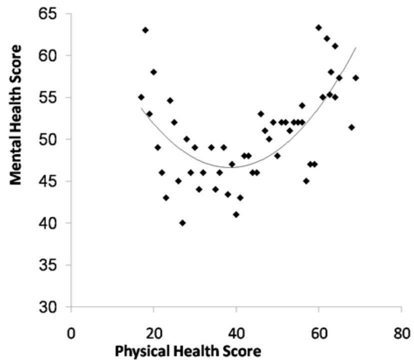
Interpretation
- Form: linear
- Outliers: none
- Direction: positive
- Strength: difficult to judge by eye (looks strong)

Review of Statistics and Probability Concepts 29



- Form: linear
- Outliers: none
- Direction: negative
- Strength: difficult to judge by eye (looks moderate)

Review of Statistics and Probability Concepts 30

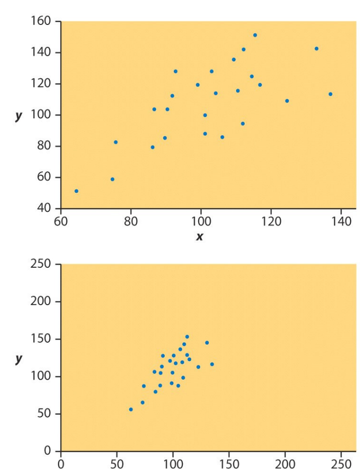


- Form: curved
- Outliers: none
- Direction: parabolic
- Strength: difficult to judge by eye (looks moderate)

Review of Statistics and Probability Concepts 31

Measuring linear Association: Correlation

- It is difficult to judge correlational *strength* visually alone.
- Here are identical data plotted on different scales.
- First relationship *seems* weaker than second.
- This is an artifact of the axis scaling.
- We use a statistical called the *correlation coefficient* to judge strength objectively.



Review of Statistics and Probability Concepts 32

Correlation

•The correlation measures the direction and strength of the linear relationship between two quantitative variables.

•Correlation is usually refers to by *r*.

•Suppose that we have data on variable *x* and *y* for *n* individuals.

•The values for the first individual are x_1 and y_1 , the values for the second individual are x_2 and y_2 , and so on.

•The means and standard deviations of the two variables are μ_x and σ_x for the x-values, and μ_y and σ_y , for the y-values.

•The correlation *r* between *x* and *y* is:

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \mu_x}{\sigma_x}\right)\left(\frac{y_i - \mu_y}{\sigma_y}\right)$$

- *r* ≡ Pearson's correlation coefficient
- *Always* between −1 and +1 (inclusive)
  - *r* = +1 ⇒ all points on upward sloping line
  - *r* = -1 ⇒ all points on downward line
  - *r* = 0 ⇒ no line or horizontal line
- The closer *r* is to +1 or −1, the **stronger** the correlation

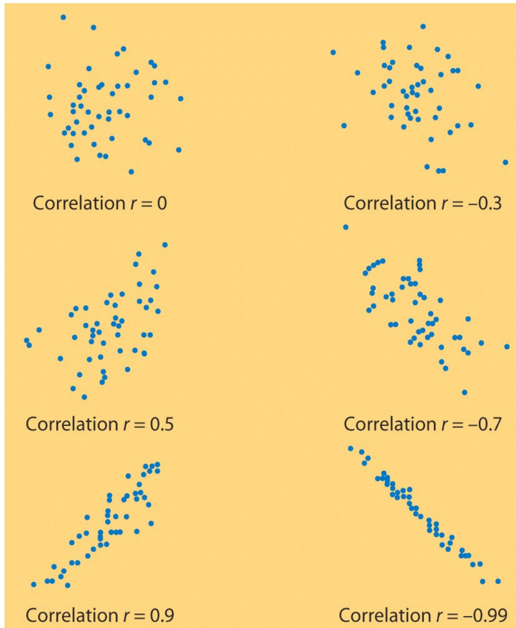
- Direction: positive, negative, ≈0
- Strength: the closer |*r*| is to 1, the stronger the correlation

  0.0 ≤ |*r*| < 0.3 ⇒ weak correlation

  0.3 ≤ |*r*| < 0.7 ⇒ moderate correlation

  0.7 ≤ |*r*| < 1.0 ⇒ strong correlation

  |*r*| = 1.0 ⇒ perfect correlation



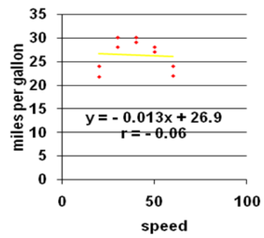
x	y	$\left(\frac{x_i-\mu_x}{\sigma_x}\right)$	$\left(\frac{y_i-\mu_y}{\sigma_y}\right)$	$\left(\frac{x_i-\mu_x}{\sigma_x}\right)\left(\frac{y_i-\mu_y}{\sigma_y}\right)$
21.4	77.48	-0.078	-0.345	0.027
23.2	77.53	1.097	-0.282	-0.309
20.0	77.32	-0.992	-0.546	0.542
22.7	78.63	0.770	1.102	0.849
20.8	77.17	-0.470	-0.735	0.345
18.6	76.39	-1.906	-1.716	3.271
21.5	78.51	-0.013	0.951	-0.012
22.0	78.15	0.313	0.498	0.156
23.8	78.99	1.489	1.555	2.315
21.2	77.37	-0.209	-0.483	0.101
				7.285

$\mu_x = 21.52; \sigma_x = 1.53$

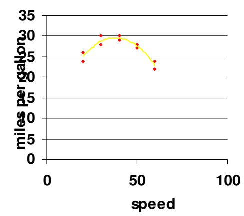
$\mu_y = 77.75; \sigma_y = 0.79$

- *r* applies to *linear* relations only
- Outliers have large influences on *r*
- Association does *not* imply causation

- Figure shows :miles per gallon" versus "speed" ("car data" *n* = 10)



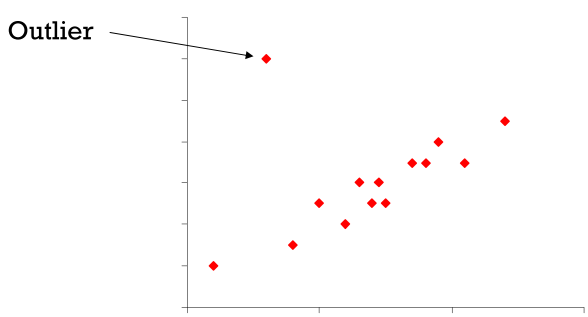
- *r* ≈ 0; but this is *misleading* because there *is* a strong non-linear relationship



Review of Statistics and Probability Concepts

37

Outliers Can Have a Large Influence



With the outlier, r ≅ 0
Without the outlier, r ≅ .8

Review of Statistics and Probability Concepts

38