- **Result 4.8.** Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be mutually independent with $\mathbf{X}_j$ distributed as $N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$. (Note that each $\mathbf{X}_j$ has the *same* covariance matrix $\boldsymbol{\Sigma}$.) Then

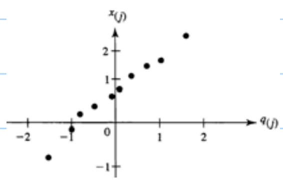
$$\mathbf{V}_1 = c_1\mathbf{X}_1 + c_2\mathbf{X}_2 + \cdots + c_n\mathbf{X}_n$$

is distributed as $N_p\left(\sum_{j=1}^{n} c_j\boldsymbol{\mu}_j, \left(\sum_{j=1}^{n} c_j^2\right)\boldsymbol{\Sigma}\right)$. Moreover, $\mathbf{V}_1$ and $\mathbf{V}_2 = b_1\mathbf{X}_1 + b_2\mathbf{X}_2 + \cdots + b_n\mathbf{X}_n$ are jointly multivariate normal with covariance matrix

$$\begin{bmatrix} \left(\sum_{j=1}^{n} c_j^2\right)\boldsymbol{\Sigma} & (\mathbf{b}'\mathbf{c})\boldsymbol{\Sigma} \\ (\mathbf{b}'\mathbf{c})\boldsymbol{\Sigma} & \left(\sum_{j=1}^{n} b_j^2\right)\boldsymbol{\Sigma} \end{bmatrix}$$

Consequently, $\mathbf{V}_1$ and $\mathbf{V}_2$ are independent if $\mathbf{b}'\mathbf{c} = \sum_{j=1}^{n} c_j b_j = 0$.

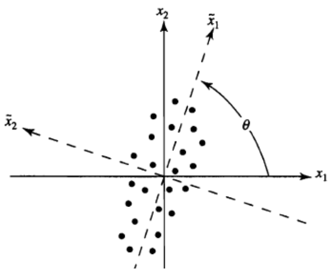
- Assessing the assumption of normality
  - Recall:
    - Q-Q plot (quintiles vs. quintiles plot)
    - histogram



  - Let $\mathbf{X}$ be distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $|\boldsymbol{\Sigma}| > 0$. Then
    - the marginal distribution X_i is normal
    - linear combination of X_i is normal
      - Many statisticians suggest plotting

$$\hat{\mathbf{e}}_1'\mathbf{x}_j \quad \text{where} \quad \mathbf{S}\hat{\mathbf{e}}_1 = \hat{\lambda}_1\hat{\mathbf{e}}_1$$

in which $\hat{\lambda}_1$ is the largest eigenvalue of $\mathbf{S}$.



    - $(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ is distributed as χ_p^2
      - Mahalanobis distance

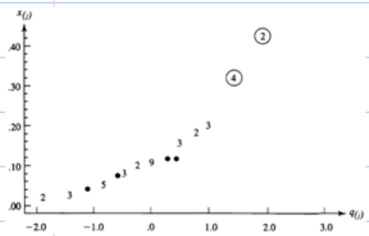
$$d_j^2 = (\mathbf{x}_j - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}), \qquad j = 1, 2, \ldots, n$$

$d_1^2, d_2^2, \ldots, d_n^2$ should behave like a chi-square random variable.

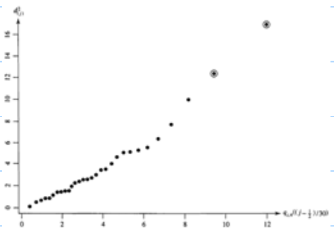
  - In some case, data is clearly non-normal but a transformation to approximate normality is possible. For example, for count data, consider the square root transform. For proportion data, the logit transform, and for correlations r, the $0.5\log[(1+r)/(1-r)]$ is worth a try. (more details in textbook, 4.8)

- detecting outliers
  - ➢ Make a dot plot for each variable.

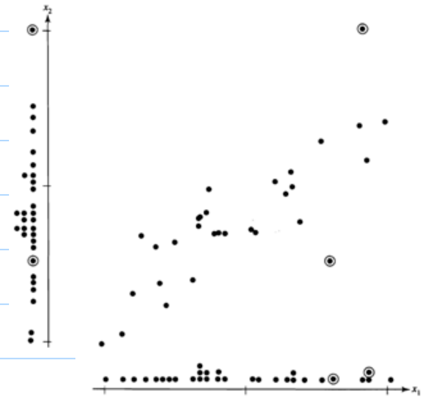
  - ➢ Make a scatter plot for each pair of variables.
  - ➢ Calculate the standardized values $z_{jk} = (x_{jk} - \bar{x}_k)/\sqrt{s_{kk}}$ for $j = 1, 2, \ldots, n$ and each column $k = 1, 2, \ldots, p$. Examine these standardized values for large or small values.
  - ➢ Calculate the generalized squared distances $(\mathbf{x}_j - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$. Examine these distances for unusually large values. In a chi-square plot, these would be the points farthest from the origin.



Normal probability plot



Chi-square plot



  - ➢ Note. When sample size is large, the appearance of few extreme values is reasonable
  - ➢ If outliers are identified, they should be examined for content. Depending upon the nature of the outliers and the objectives of the investigation, outliers may be deleted or appropriately "weighted" in a subsequent analysis.

❖ **Reading**: Textbook, 4.1, 4.2, 4.6, 4.7

Sample Mean Vector and Sample Covariance Matrix under Normality Assumption

- Maximum likelihood estimator
  - ➢ **Result 4.11.** Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from a normal population with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Then

$$\hat{\boldsymbol{\mu}} = \overline{\mathbf{X}} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})' = \frac{(n-1)}{n}\mathbf{S}$$

are the *maximum likelihood estimators* of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively. Their observed values, $\bar{\mathbf{x}}$ and $(1/n)\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'$, are called the *maximum likelihood estimates* of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

$$\left\{\begin{array}{c}\text{Joint density}\\ \text{of } \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\end{array}\right\} = \frac{1}{(2\pi)^{np/2}}\frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_j - \boldsymbol{\mu})/2}$$

$$\propto \exp\left\{-\text{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})'\right)\right]\Big/ 2\right\}$$

$$\text{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})'\right)\right]$$

$$= \text{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\right)\right] + n(\bar{\mathbf{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})$$

- Sufficient statistics
  - Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from a multivariate normal population with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Then

$$\overline{\mathbf{X}} \text{ and } \mathbf{S} \text{ are } \textit{sufficient statistics}$$

- Distribution of $\overline{X}$ and S
  - Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample of size n from a p-variate *normal* distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then

    1. $\overline{\mathbf{X}}$ is distributed as $N_p(\boldsymbol{\mu}, (1/n)\boldsymbol{\Sigma})$.
    2. $(n-1)\mathbf{S}$ is distributed as a Wishart random matrix with $n-1$ d.f.
    3. $\overline{\mathbf{X}}$ and $\mathbf{S}$ are independent.

  - Wishart distribution
    - definition:

$$W_m(\cdot \mid \boldsymbol{\Sigma}) = \text{Wishart distribution with } m \text{ d.f.}$$
$$= \text{distribution of } \sum_{j=1}^{m} \mathbf{Z}_j \mathbf{Z}_j'$$

where the $\mathbf{Z}_j$ are each independently distributed as $N_p(\mathbf{0}, \boldsymbol{\Sigma})$.

    - some properties

1. If $\mathbf{A}_1$ is distributed as $W_{m_1}(\mathbf{A}_1 \mid \boldsymbol{\Sigma})$ independently of $\mathbf{A}_2$, which is distributed as $W_{m_2}(\mathbf{A}_2 \mid \boldsymbol{\Sigma})$, then $\mathbf{A}_1 + \mathbf{A}_2$ is distributed as $W_{m_1+m_2}(\mathbf{A}_1 + \mathbf{A}_2 \mid \boldsymbol{\Sigma})$. That is, the degrees of freedom add.)
2. If $\mathbf{A}$ is distributed as $W_m(\mathbf{A} \mid \boldsymbol{\Sigma})$, then $\mathbf{CAC}'$ is distributed as $W_m(\mathbf{CAC}' \mid \mathbf{C\Sigma C}')$.

- Large-Sample
  - Law of Large Number

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be independent observations from a population with mean $\boldsymbol{\mu}$ and finite (nonsingular) covariance $\boldsymbol{\Sigma}$. Then

$$\overline{\mathbf{X}} \text{ converges in probability to } \boldsymbol{\mu}$$

and

$$\mathbf{S} \text{ (or } \hat{\boldsymbol{\Sigma}} = \mathbf{S}_n) \text{ converges in probability to } \boldsymbol{\Sigma}$$

  - **Result 4.13 (The central limit theorem).** Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be independent observations from any population with mean $\boldsymbol{\mu}$ and finite covariance $\boldsymbol{\Sigma}$. Then

$$\sqrt{n}\,(\overline{\mathbf{X}} - \boldsymbol{\mu}) \text{ has an approximate } N_p(\mathbf{0}, \boldsymbol{\Sigma}) \text{ distribution}$$

and

$$n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) \text{ is approximately } \chi_p^2$$

for large sample sizes. Here n should also be large relative to p.

❖ **Reading**: Textbook, 4.3, 4.4, 4.5