- **Result 4.8.** Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ 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

C.f. AX in result 4.3, LNp.3-12

$$\mathbf{V}_1 = c_1\mathbf{X}_1 + c_2\mathbf{X}_2 + \cdots + c_n\mathbf{X}_n = [X_1 | X_2 | \cdots | X_n]\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = [c_1 \cdots c_n]\begin{bmatrix} X_1' \\ X_2' \\ \vdots \\ X_n' \end{bmatrix}'$$

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

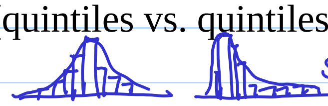
$$\begin{bmatrix} \left(\sum_{j=1}^n c_j^2\right)\boldsymbol{\Sigma} & (\mathbf{b}'\mathbf{c})\boldsymbol{\Sigma} \\ (\mathbf{b}'\mathbf{c})\boldsymbol{\Sigma} & \left(\sum_{j=1}^n b_j^2\right)\boldsymbol{\Sigma} \end{bmatrix}$$

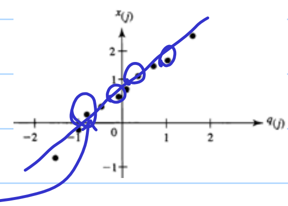
Consequently, $\mathbf{V}_1$ and $\mathbf{V}_2$ are independent if $\mathbf{b}'\mathbf{c} = \sum_{j=1}^n c_j b_j = 0$. $(b \perp c)$

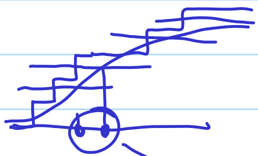
proof: $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$, $X \sim N_{np}\left(\underbrace{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}}_{\mu_X}, \underbrace{\begin{bmatrix} \Sigma & & & 0 \\ & \Sigma & & \\ & & \Sigma & \\ 0 & & & \Sigma \end{bmatrix}}_{=\Sigma_X}\right)$

$A = \begin{bmatrix} c_1 I & c_2 I & \cdots & c_n I \\ b_1 I & b_2 I & \cdots & b_n I \end{bmatrix}$, $AX = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$

$AX \sim N_{2p}(A\mu_X, \underline{A\Sigma_X A'})$

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





  - Let $\mathbf{X}$ be distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $|\boldsymbol{\Sigma}| > 0$. Then

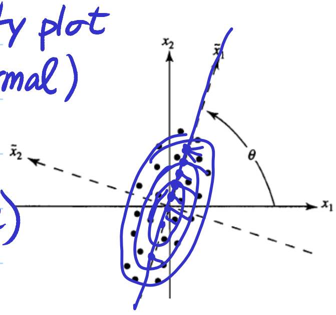
1-dim
    - the marginal distribution X_i is normal ← normal probability plot (QQ plot for normal)
    - linear combination of X_i is normal

2-dim
      - Many statisticians suggest plotting

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

← 1st principal component (future lecture)

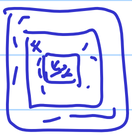
in which $\hat{\lambda}_1$ is the largest eigenvalue of $\mathbf{S}$.



2 variables: contour of pdf is ellipse

Scatter plot

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



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

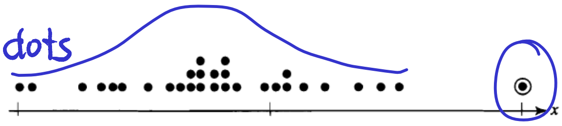
$d_1^2, d_2^2, \dots, d_n^2$ should behave like a chi-square random variable. → chi-square plot (QQ plot for chi-square)

  - 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)

Box-Cox Transformation

- detecting outliers

# of dots $=n$

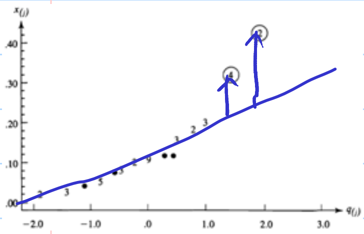


1-dim ➢ Make a dot plot for each variable.

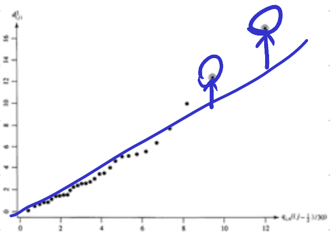
2-dim ➢ Make a scatter plot for each pair of variables.

1-dim ➢ 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. —— # of dots $= np$.

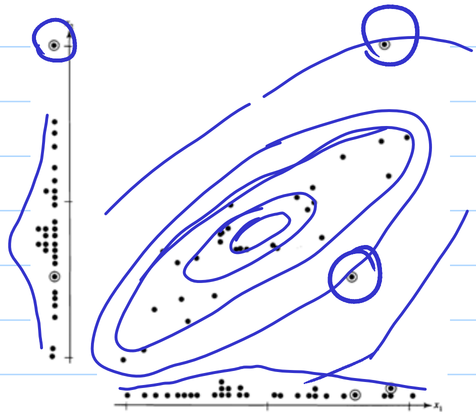
➢ Calculate the generalized squared distances $(\mathbf{x}_j - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$. Examine these distances for unusually large values. In a chi-square plot, these would be the points farthest from the origin.



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. (check LM, LNp.7-5~6)

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

Sample Mean Vector and Sample Covariance Matrix under Normality Assumption (cf. Result 3.1 LNp.3-3)

- 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

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

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 $\boldsymbol{\Sigma}$.

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

$tr\left(\sum_j (x_j-\mu)'\Sigma^{-1}(x_j-\mu)\right)$
$= \sum_j tr\left((x_j-\mu)'\Sigma^{-1}(x_j-\mu)\right)$
$= \sum_j tr\left(\Sigma^{-1}(x_j-\mu)(x_j-\mu)'\right)$
$= tr\left(\sum_j \left(\Sigma^{-1}(x_j-\mu)(x_j-\mu)'\right)\right)$

$-\bar{x}+\bar{x}$ $\quad -\bar{x}+\bar{x}$

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

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

$tr(n\,\Sigma^{-1}(\bar{x}-\mu)(\bar{x}-\mu)')$ $\;//$

$$\frac{1}{|\Sigma|^{n/2}}\;\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\right)\right] + n(\bar{\mathbf{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})$$

- 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

    dimension reduction $\bar{\mathbf{X}}$ and $\mathbf{S}$ are *sufficient statistics*

    If normality assumption is true, all information about parameters μ, Σ is kept in $\bar{X}$ & S.
- Distribution of $\bar{\mathbf{X}}$ and $\mathbf{S}$
  - 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. $\bar{\mathbf{X}}$ is distributed as $N_p(\boldsymbol{\mu}, (1/n)\boldsymbol{\Sigma})$. ← by result 4.8 (LN p.3-17)
    2. $(n-1)\mathbf{S}$ is distributed as a Wishart random matrix with $n-1$ d.f. $\sum_j (x_j-\bar{x})(x_j-\bar{x})'$
    3. $\bar{\mathbf{X}}$ and $\mathbf{S}$ are independent.
  - Wishart distribution
    - definition: $W_m(\cdot \mid \boldsymbol{\Sigma})$ = Wishart distribution with m d.f. ← Chi-square distribution.

      $$= \text{distribution of } \sum_{j=1}^{m} \underset{p\times 1}{\mathbf{Z}_j}\,\underset{1\times p}{\mathbf{Z}_j'}$$

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

      CAC' $\shortparallel$ $C\left(\sum_{j=1}^{m} Z_jZ_j'\right)C' = \sum_{j=1}^{m} (CZ_j)(CZ_j)'$, $CZ_j \sim N(0, C\Sigma C')$
    - 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 (when the normality assumption is dropped)
  - 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

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

    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

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

    and

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

    ($\mathbf{S}^{-1}$ → Σ^{-1})

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

    $(n-1)S \underset{\text{approximate}}{\sim} \text{Wishart}_{n-1}(\cdot \mid \Sigma)$

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