• Assessing the assumption of normality
  - Recall:
    - Q-Q plot (quintiles vs. quintiles plot)
    - histogram
    - Q-Q plot (quintiles vs. quintiles plot)
  - Let \( \mathbf{X} \) be distributed as \( N_p(\mathbf{\mu}, \Sigma) \) with \( |\Sigma| > 0 \). Then
    - the marginal distribution \( X_i \) is normal
    - linear combination of \( X_i \) is normal
  - Many statisticians suggest plotting
    - Mahalanobis distance
      \[ d_j^2 = (\mathbf{x}_j - \bar{\mathbf{x}})^\top \Sigma^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}), \quad j = 1, 2, \ldots, n \]
      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 $(x_j - \bar{x})'S^{-1}(x_j - \bar{x})$. Examine these distances for unusually large values. In a chi-square plot, these would be the points farthest from the origin.

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 $X_1, X_2, \ldots, X_n$ be a random sample from a normal population with mean $\mu$ and covariance $\Sigma$. Then

$$\hat{\mu} = \bar{X} \quad \text{and} \quad \hat{\Sigma} = \frac{1}{n} \sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})' = \frac{(n - 1)}{n} S$$

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

$$\exp \left\{ -\text{tr} \left[ \Sigma^{-1} \left( \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})' + n(\bar{x} - \mu)(\bar{x} - \mu)' \right) \right] / 2 \right\} \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} e^{-\frac{1}{2} \sum_{j=1}^{n} (x_j - \mu)(x_j - \mu)'}$$

tr $\left[ \Sigma^{-1} \left( \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})' + n(\bar{x} - \mu)(\bar{x} - \mu)' \right) \right] / 2$
• Sufficient statistics
  ➢ Let \( X_1, X_2, \ldots, X_n \) be a random sample from a multivariate normal population with mean \( \mu \) and covariance \( \Sigma \). Then \( \bar{X} \) and \( S \) are sufficient statistics.

• Distribution of \( \bar{X} \) and \( S \)
  ➢ Let \( X_1, X_2, \ldots, X_n \) be a random sample of size \( n \) from a \( p \)-variate normal distribution with mean \( \mu \) and covariance matrix \( \Sigma \). Then
  1. \( \bar{X} \) is distributed as \( N_p(\mu, \frac{1}{n} \Sigma) \).
  2. \( (n - 1)S \) is distributed as a Wishart random matrix with \( n - 1 \) d.f.
  3. \( \bar{X} \) and \( S \) are independent.

Wishart distribution
  ➢ definition: \( W_m(\cdot \mid \Sigma) = \text{Wishart distribution with } m \text{ d.f.} \)
  \[ = \text{distribution of } \sum_{i=1}^{m} Z_iZ_i' \]  
  where the \( Z_i \) are each independently distributed as \( N_p(0, \Sigma) \).

• some properties
  1. If \( A_1 \) is distributed as \( W_{m_1}(A_1 \mid \Sigma) \) independently of \( A_2 \), which is distributed as \( W_{m_2}(A_2 \mid \Sigma) \), then \( A_1 + A_2 \) is distributed as \( W_{m_1+m_2}(A_1 + A_2 \mid \Sigma) \). That is, the degrees of freedom add.
  2. If \( A \) is distributed as \( W_m(A \mid \Sigma) \), then \( CA \Sigma C' \) is distributed as \( W_m(CA \Sigma C' \mid \Sigma) \).

• Large-Sample (when the normality assumption is dropped)
  ➢ Law of Large Number
    Let \( X_1, X_2, \ldots, X_n \) be independent observations from a population with mean \( \mu \) and finite (nonsingular) covariance \( \Sigma \). Then
    \[ \bar{X} \] converges in probability to \( \mu \)
    \[ S \text{ (or } \hat{\Sigma} = S_n \) converges in probability to \( \Sigma \)

  ➢ Result 4.13 (The central limit theorem). Let \( X_1, X_2, \ldots, X_n \) be independent observations from any population with mean \( \mu \) and finite covariance \( \Sigma \). Then
    \[ \sqrt{n} (\bar{X} - \mu) \] has an approximate \( N_p(0, \Sigma) \) distribution
    and
    \[ n(\bar{X} - \mu) \Sigma^{-1}(\bar{X} - \mu) \] is approximately \( \chi^2 \)
    for large sample sizes. Here \( n \) should also be large relative to \( p \).
    \[ (n-1)S \approx \text{Wishart}_{n-1}(\cdot \mid \Sigma) \]

• Reading: Textbook, 4.3, 4.4, 4.5