Multivariate Normal Distribution

- Researchers often assume that the joint distribution of the $p \times 1$ vectors of measurements on each sample unit is a p-dimensional *multivariate normal distribution*.

- The multivariate normal (MVN) assumption is often a good approximate model

- This is generally the result of a fundamental result called the Multivariate Central Limit Theorem that tells us that the distribution of the sum, or mean, of many multivariate random vectors approximately has a multivariate normal distribution regardless of the distributions the random vectors in the sum.

Multivariate Normal Distribution

- The MVN is a generalization of the univariate normal distribution.

- Recall that if X has a normal distribution with mean μ and variance σ^2, then the formula for its density function is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right), \quad -\infty < x < \infty.$$

- The density can be re-expressed as:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)\right).$$

Matrix Terminology

- If it exists, the inverse of a $p \times p$ matrix A, denoted by A^{-1}, is the unique matrix such that

$$AA^{-1} = A^{-1}A = I = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

- A must be a positive definite matrix for the inverse to exist, i.e.,

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_p \end{bmatrix} A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} > 0 \text{ for any } \begin{bmatrix} x_1 & x_2 & \cdots & x_p \end{bmatrix} \neq \mathbf{0}$$

Eigenvalues and Eigenvectors

- Recall that if $(\lambda_j, \mathbf{e}_j)$ is an eigenvalue-eigenvector pair for a $p \times p$ matrix A, then by definition

$$A\mathbf{e}_j = \lambda_j \mathbf{e}_j$$

- Eigenvectors are usually scaled to have length one, i.e., $1 = \mathbf{e}_j'\mathbf{e}_j = e_{j1}^2 + e_{j2}^2 + \cdots + e_{jp}^2$.

- Eigenvectors are mutually orthogonal, i.e., $\mathbf{e}_j'\mathbf{e}_k = 0$ for any $j \neq k$

- Also $(\lambda_j^{-1}, \mathbf{e}_j)$ is an eigenvalue-eigenvector pair of A^{-1} when A is positive definite and A^{-1} exists.

Eigenvalues and Eigenvectors

- A $p \times p$ matrix is symmetric if $A^{'} = A$, where $A^{'}$ is the transpose of A.

- **Spectral decomposition** of any $p \times p$ symmetric matrix:

$$A = \lambda_1 \mathbf{e}_1 \mathbf{e}_1^{'} + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^{'} + \cdots + \lambda_p \mathbf{e}_p \mathbf{e}_p^{'}$$

- Then $A^{-1} = \lambda_1^{-1} \mathbf{e}_1 \mathbf{e}_1^{'} + \lambda_2^{-1} \mathbf{e}_2 \mathbf{e}_2^{'} + \cdots + \lambda_p^{-1} \mathbf{e}_p \mathbf{e}_p^{'}$, if A^{-1} exists.

- $A^{1/2} = \lambda_1^{1/2} \mathbf{e}_1 \mathbf{e}_1^{'} + \lambda_2^{1/2} \mathbf{e}_2 \mathbf{e}_2^{'} + \cdots + \lambda_p^{1/2} \mathbf{e}_p \mathbf{e}_p^{'}$, and $A^{1/2} A^{1/2} = A$

- $A^{-1/2} = \lambda_1^{-1/2} \mathbf{e}_1 \mathbf{e}_1^{'} + \lambda_2^{-1/2} \mathbf{e}_2 \mathbf{e}_2^{'} + \cdots + \lambda_p^{-1/2} \mathbf{e}_p \mathbf{e}_p^{'}$,

  and $A^{-1/2} A^{-1/2} = A^{-1}$ and $A^{-1/2} A A^{-1/2} = I$

Determinant of a $p \times p$ Matrix

- The determinant of the matrix A is denoted by $|A|$ or $\det(A)$.
- The determinant of a $p \times p$ matrix is the product of its eigenvalues, i.e., $|A| = \det(A) = \lambda_1 \lambda_2 \cdots \lambda_p$.
- All of the eigenvalues of a positive definite matrix are positive.
- The determinant of a positive definite matrix must be larger than zero.
- If at least one eigenvalue is zero and the inverse of the matrix does not exist, then the determinant of the matrix is zero.

Multivariate Normal Distribution

- Consider a $p \times 1$ random vector $\mathbf{x} = [x_1, x_2, \ldots, x_p]'$.

- $\mathbf{x}$ has a multivariate normal distribution with mean vector μ and covariance matrix Σ if it has a probability density function (pdf) of the form

$$f(\mathbf{x}) = (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right],$$

where $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$.

- The density only exists if Σ is positive definite.

Multivariate Normal Distribution

- The quadratic form $(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})$ in the density formula is the squared statistical distance of $\mathbf{x}$ from $\boldsymbol{\mu}$.

- This quadratic form is often referred to as the square of the *Mahalanobis distance*.

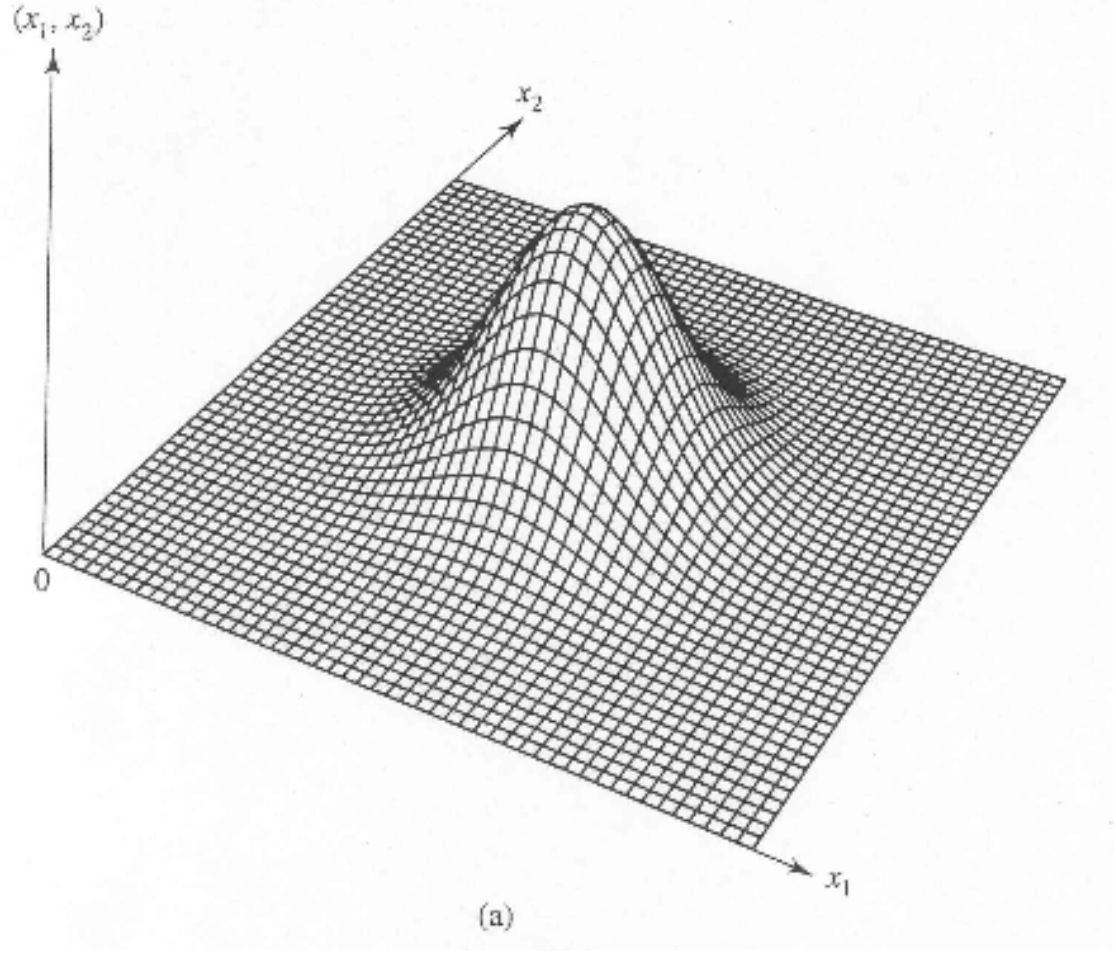
- We use the notation

$$\mathbf{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma) \text{ or } \mathbf{x} \sim \mathrm{MVN}(\boldsymbol{\mu}, \Sigma)$$

  to indicate that the p-dimensional random vector $\mathbf{X}$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.

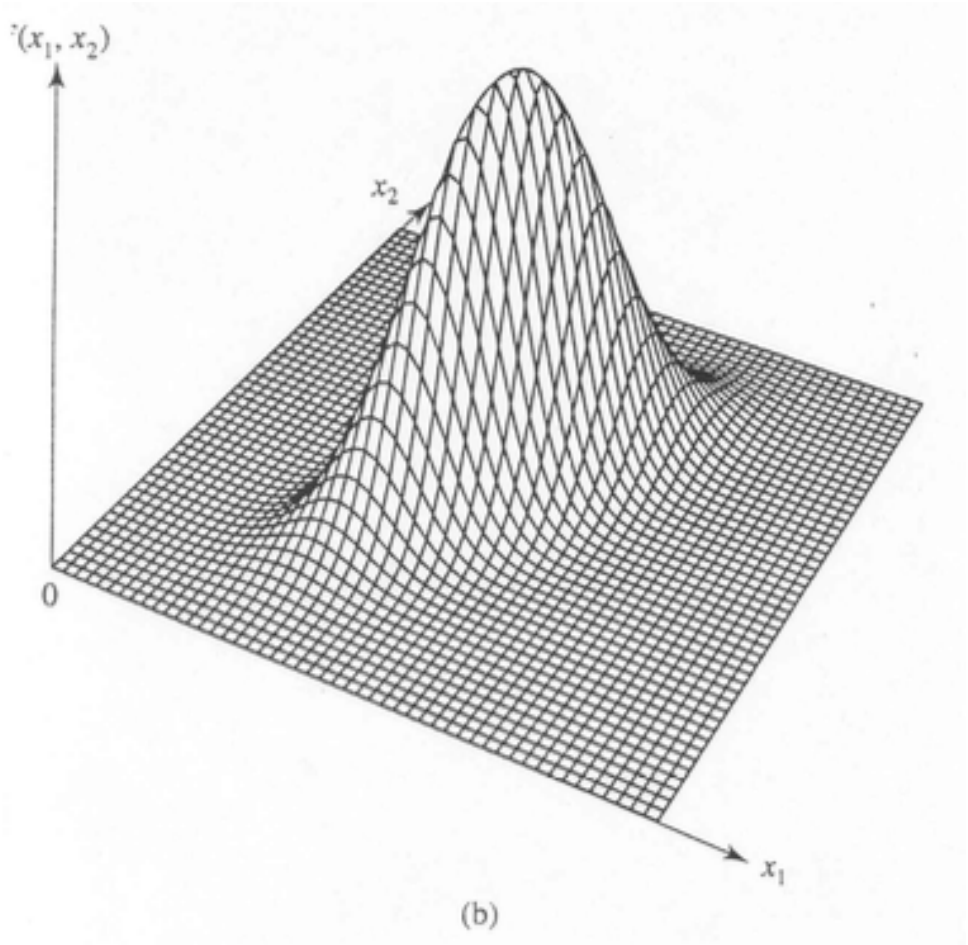
- If $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$, then $(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) \sim \chi^2_p$.

Example: Bivariate Normal Density



Example: Bivariate Normal Density

$\sigma_{11} = \sigma_{22}$, and $\rho_{12} > 0$



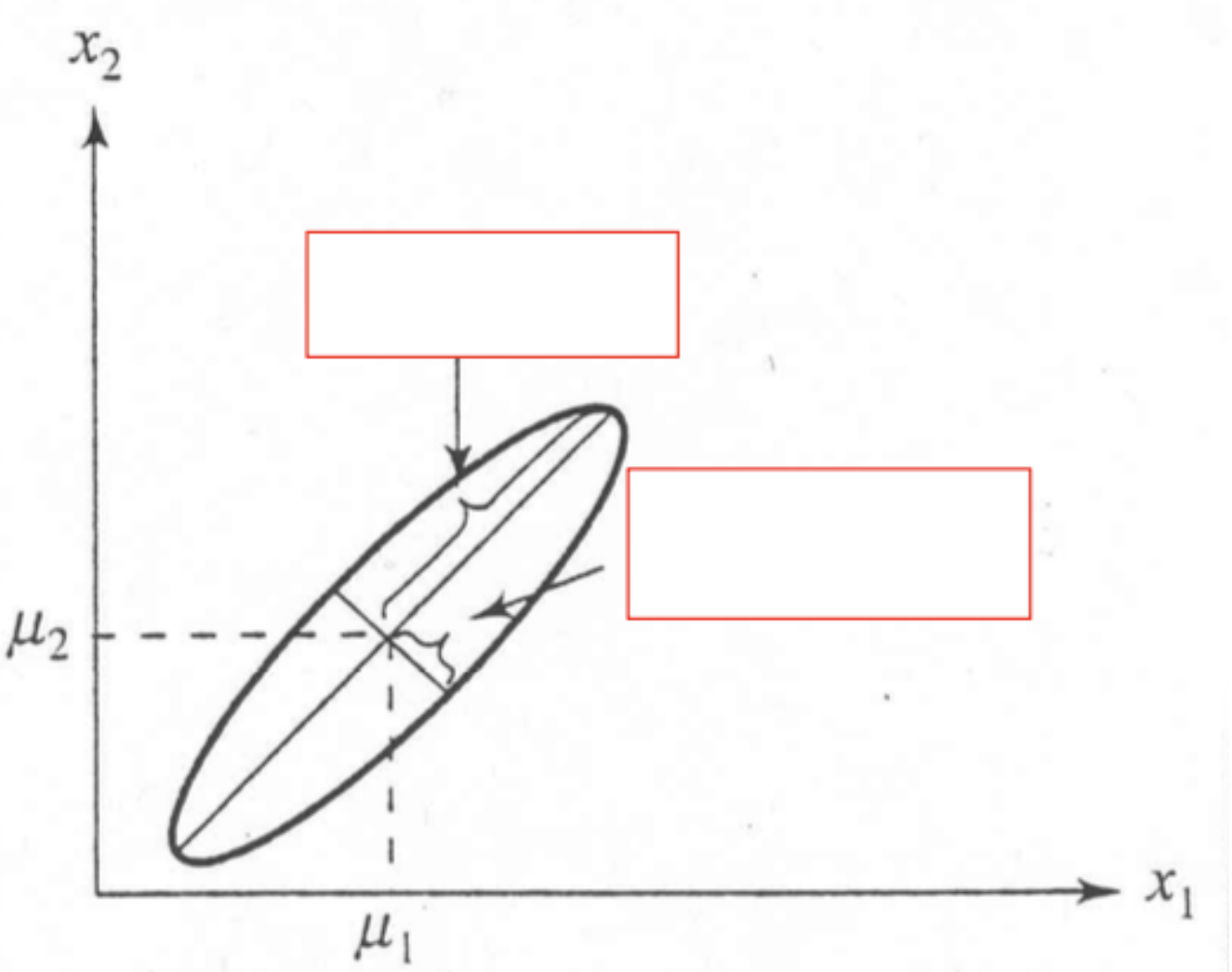
Density Contours

- A set of all data vectors $\mathbf{x}$ that correspond to a constant height of the density function forms an ellipsoid centered at $\boldsymbol{\mu}$.

- The MVN density is constant for all $\mathbf{x}$'s that are the same statistical distance from the population mean vector, i.e., all $\mathbf{x}$'s that satisfy

$$\sqrt{(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} = constant.$$

- The axes of the ellipses are in the directions of the eigenvectors of $\boldsymbol{\Sigma}$ and the length of the j-th longest axis is proportional to $\sqrt{\lambda_j}$, where λ_j is the eigenvalue associated with the j-th eigenvector of $\boldsymbol{\Sigma}$.


x_2
μ_2
μ_1
x_1

Eigenvalues and Eigenvectors

- The probability is $1-\alpha$ that the value of a random vector will be inside the ellipsoid defined by

$$(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \leq \chi^2_{(p),1-\alpha}$$

  where $\chi^2_{(p),1-\alpha}$ is the upper $(100 \times \alpha)$ percentile of a chi-square distribution with p degrees of freedom.

- This is smallest region that has probability $(1-\alpha)$ of containing a vector of observations randomly selected from the population.

- The jth axis of this ellipsoid is determined by the eigenvector associated with the j-th largest eigenvalue of $\boldsymbol{\Sigma}$, for $j = 1, ..., p$.

- The distance along the j-th axis from the center to the boundary of the ellipsoid is $\sqrt{\lambda_j}\left(\frac{2}{p\Gamma(p/2)}\right)^{1/p}\sqrt{\chi^2_{(p),1-\alpha}}$

Gamma Function

$$\Gamma\left(\frac{p}{2}\right) = \left(\frac{p}{2} - 1\right)\left(\frac{p}{2} - 2\right)\cdots(2)(1)$$

when p is an even integer, and

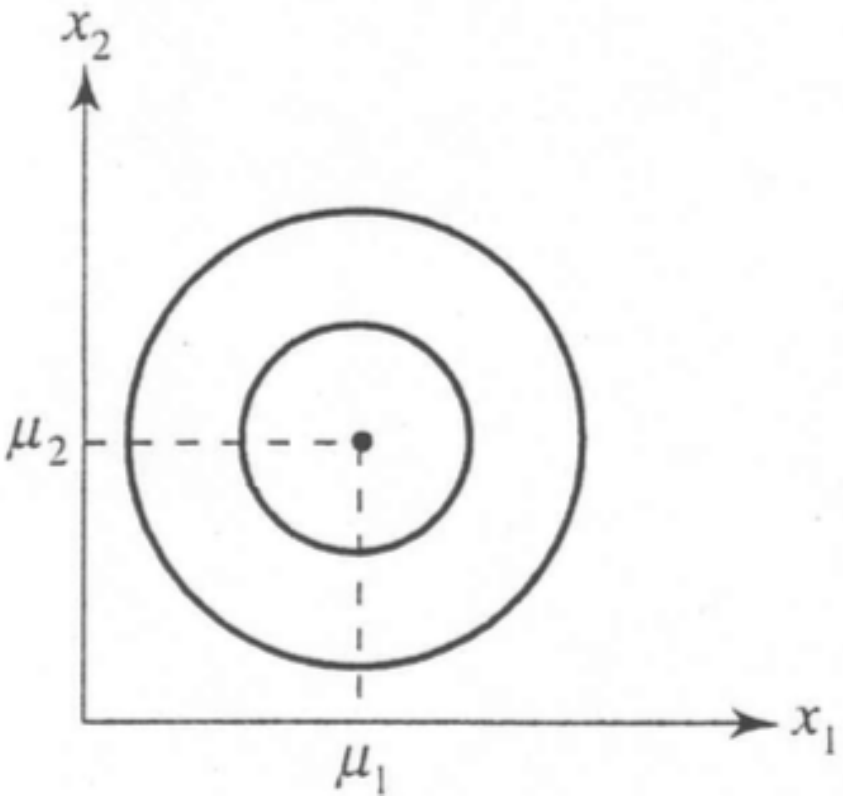
$$\Gamma(\frac{p}{2}) = \frac{(p-2)(p-4)\cdots(3)(1)}{2^{(p-1)/2}}\sqrt{\pi}$$

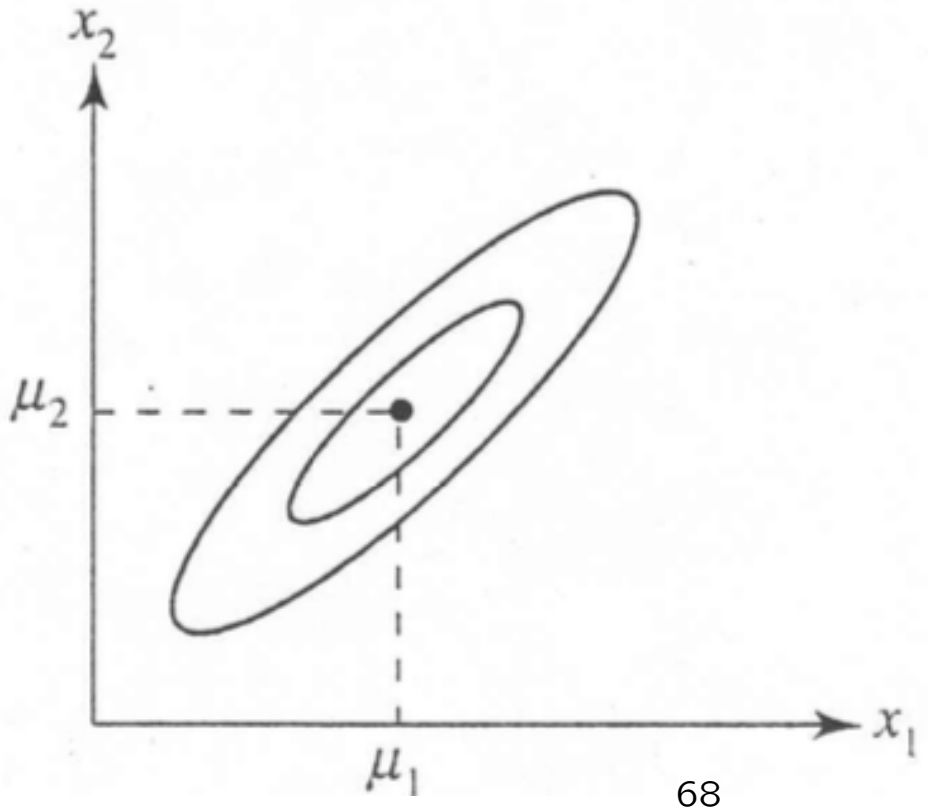
when p is an odd integer

and $\Gamma(1)$ is defined to be 1.0

More Bivariate Normal Examples

50% and 90% contours of two bivariate normal densities. Density is the highest when $\mathbf{x} = \boldsymbol{\mu}$.





Central $(1-\alpha)\times 100\%$ Region of a Bivariate Normal Distribution

- The ratio of the lengths of the major and minor axes is

$$\frac{\text{Length of major axis}}{\text{Length of minor axis}} = \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}}$$

- The area of the ellipse containing the central $(1-\alpha)\times 100\%$ of a bivariate normal population is

$$area = \pi\chi^2_{(2),1-\alpha}\sqrt{\lambda_1}\sqrt{\lambda_2} = \pi\chi^2_{(2),1-\alpha}|\mathbf{\Sigma}|^{1/2}$$

- For p-dimensional normal distributions the hypervolume of the p-dimensional ellipsoid is

$$\frac{2\pi^{p/2}}{p\Gamma(p/2)}\left[\chi^2_{(p),1-\alpha}\right]^{p/2}|\mathbf{\Sigma}|^{1/2}$$

where $\Gamma(\cdot)$ is the gamma function

Overall Measures of Variability

- Generalized variance:

$$|\Sigma| = \lambda_1 \lambda_2 \cdots \lambda_p$$

- Generalized standard deviation

$$|\Sigma|^{1/2} = \sqrt{\lambda_1 \lambda_2 \cdots \lambda_p}$$

- Total variance

$$\begin{aligned} trace(\Sigma) \equiv tr(\Sigma) &:= \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} \\ &= \lambda_1 + \lambda_2 + \cdots + \lambda_p \end{aligned}$$

Sample Estimates: Air Samples

$$\mathbf{x}_i = \begin{bmatrix} x_{i1} \leftarrow \ CO \text{ concentration} \\ x_{i2} \leftarrow \ N_2O \text{ concentration} \end{bmatrix}$$

$$\mathbf{x}_1 = \begin{bmatrix} 7 \\ 12 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 4 \\ 9 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} 4 \\ 5 \end{bmatrix} \quad \mathbf{x}_4 = \begin{bmatrix} 5 \\ 8 \end{bmatrix} \quad \mathbf{x}_5 = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$$

Sample Estimates: Air Samples

- Sample mean vector:

$$\bar{\mathbf{x}} = \begin{bmatrix} 4.8 \\ 8.4 \end{bmatrix}$$

- Sample covariance matrix:

$$S = \begin{bmatrix} 1.7 & 2.6 \\ 2.6 & 6.3 \end{bmatrix}$$

- Sample correlation: r_{12}=0.7945
- Generalized variance: $|S| = 3.95$
- Total variance; $trace(S) = 1.7 + 6.3 = 8.0$

Examples

The overall measures of variability can be the same for samples with different covariance matrices.

$$S = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \qquad S = \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix} \qquad S = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

$$r_{12} = 0.8 \qquad r_{12} = -0.8 \qquad r_{12} = 0$$

$$|S| = 9 \qquad |S| = 9 \qquad |S| = 9$$

$$tr(S) = 10 \qquad tr(S) = 10 \qquad tr(S) = 6$$

Assessing Normality

- Assessing multivariate normality is difficult.
- If any single variable does not have a normal distribution, then the joint distribution of p random variables cannot have a normal distribution.
- We can check normality for each variable individually. If we reject normality for any variable then the joint distribution is not multivariate normal.
- Also look at scatter plots of pairs of variables.
- Look for outliers.

Normal Q-Q plots

- A quantile-quantile (Q-Q) plot can also be constructed for each of the p variables.

- In a Q-Q plot, we plot the ordered data (sample quantiles) against the quantiles that would be expected if the sample came from a standard normal distribution.

- If the hypothesis of normality holds, the points in the plot will fall closely along a straight line.

Normal Q-Q plots

- The slope of the line passing through the points is an estimate of the population standard deviation.

- The intercept of the estimated line is an estimate of the population mean.

- The sample quantiles are just the sample order statistics. For a sample $x_1, x_2, ..., x_n$, quantiles are obtained by ordering sample observations

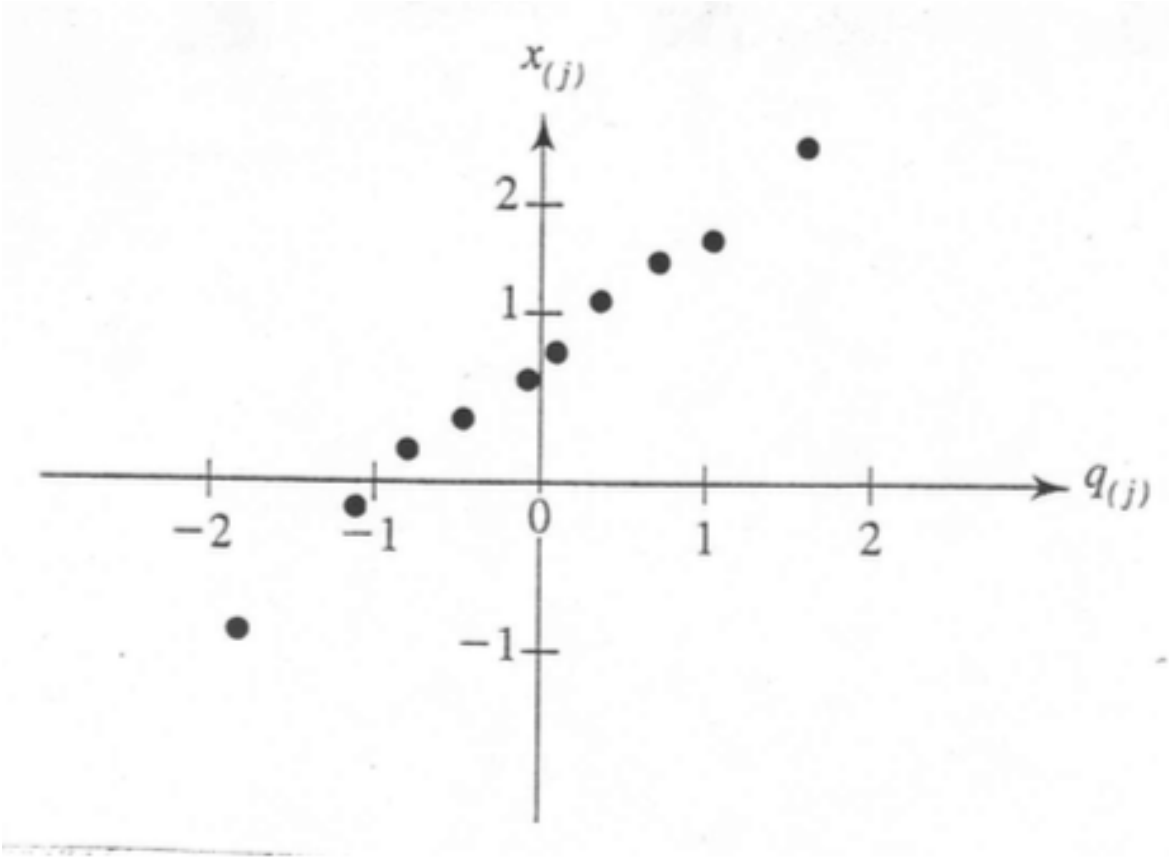
$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)},$$

where $x_{(j)}$ is the jth smallest sample observation.

Example

Ordered Observations $x_{(j)}$	Probability Level $(j-0.5)/n$	Standard Normal Quantiles $q_{(j)}$
-1.00	0.05	-1.645
-0.10	0.15	-1.036
0.16	0.25	-0.674
0.41	0.35	-0.385
0.62	0.45	-0.125
0.80	0.55	0.125
1.26	0.65	0.385
1.54	0.75	0.674
1.71	0.85	1.036
2.30	0.95	1.645

Q-Q Plot Example



Shapiro-Wilk Test

- A weighted correlation between the $x_{(j)}$ and the $q_{(j)}$:

$$W = \left(\frac{\sum_{i=1}^{n} a_j (x_{(i)} - \bar{x})(q_{(i)} - \bar{q})}{\sqrt{\sum_{i=1}^{n} a_j^2 (x_{(i)} - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (q_{(i)} - \bar{q})^2}} \right)^2 .$$

- We expect values of W to be close to one if the sample arises from a normal population.
- Reject the null hypothesis that data were sampled from a normal distribution if W is too small.
- This test has been extended to p-dimensional normal distributions
  - mshapiro.test function in the R mvnormtest package.
  - SAS has other tests in the MODEL procedure in SAS/ETS

Stiffness of boards (J&W Example 4.14)

Four measures of stiffness on each of n = 30 boards were obtained. The data (shown below) are posted in Canvas as board_stiffness.dat

```
 1 1889 1651 1561 1778
 2 2493 2048 2087 2197
 3 2119 1700 1815 2222
 .    .    .    .    .
 .    .    .    .    .
 .    .    .    .    .
28 1655 1675 1414 1597
29 2326 2301 2065 2234
30 1490 1382 1214 1284
```

See R code for illustrations

Transformations to Near Normality

- If observations show gross departures from normality, it might be necessary to transform some of the variables to near normality.

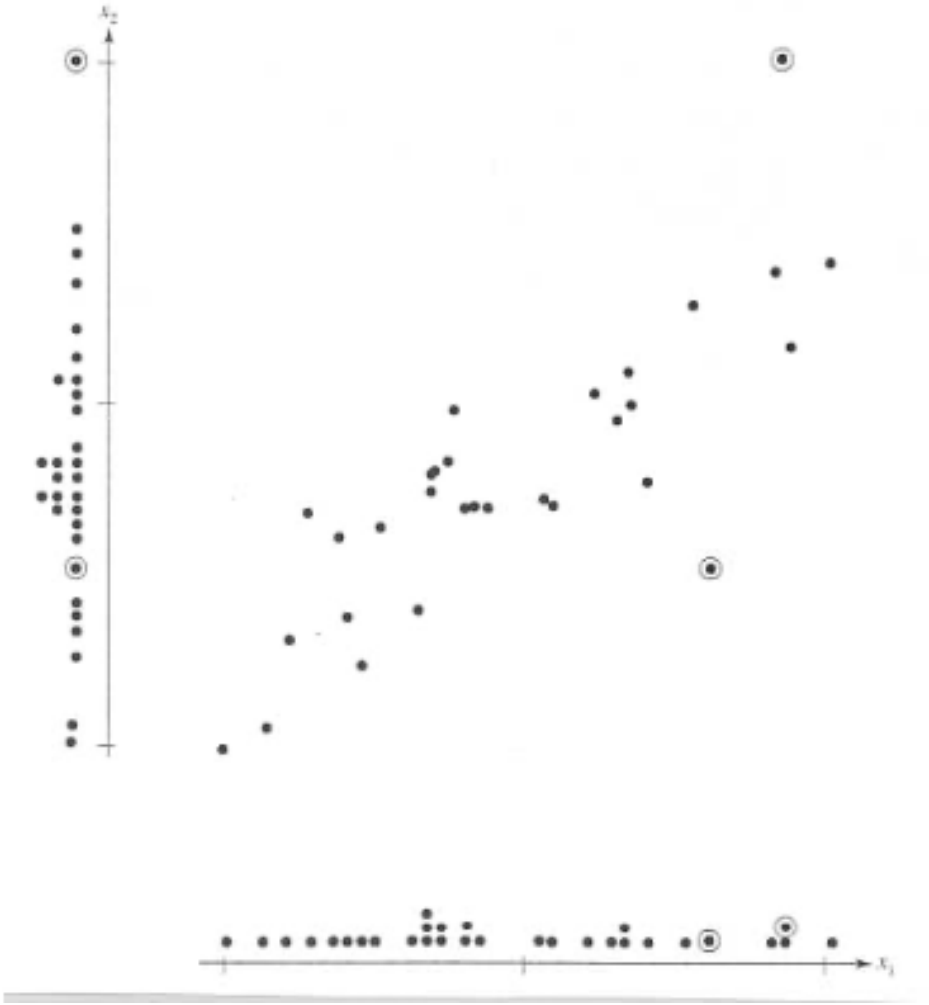
- Some suggestions:

Original scale	Transformed scale
Right skewed data	$\sqrt{x} \quad log(x) \quad 1/\sqrt{x} \quad 1/x$
x are counts	$\sqrt{x}$
x are proportions $\widehat{p}$	$\text{logit}(\widehat{p}) = 1/2 \log[(\widehat{p})/(1-\widehat{p})]$
x are correlations r	Fisher's $z(r) = 1/2 \log[(1+r)/(1-r)]$

Detecting Outliers

- An outlier is a measurement that appears to be much different from neighboring observations.
- In the univariate case with adequate sample sizes, and assuming that normality holds, an outlier can be detected by:
  1. Standardizing the n measurements so that they are approximately $N(0,\ 1)$.
  2. Flagging observations with standardized values below or above 3.5 or thereabouts.
- In p dimensions, detecting outliers is not so easy. A sample unit which may not appear to be an outlier in each of the marginal distributions can still be an outlier relative to the multivariate distribution.

Detecting Outliers



Steps for Detecting Outliers

1. Investigate all univariate marginal distributions by computing the standardized values $z_{ji} = (x_{ji} - \bar{x}_i)/\sqrt{\sigma_{ii}}$ for the j-th sample unit and the i-th variable.

2. If p is moderate, construct all bivariate scatter plots. There are $p(p-1)/2$ of them.

3. For each sample unit, calculate the squared distance $d_j^2 = (\mathbf{x}_j - \bar{\mathbf{x}})' S^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})$, where $\mathbf{x}_j$ is the $p \times 1$ vector of measurements on the j-th sample unit.

4. To decide if d_j^2 is 'extreme', recall that the d_j^2 are approximately χ_p^2. For example, if $n = 100$, we would expect to observe about 5 squared distances larger than the 0.95 percentile of the χ_p^2 distribution.

Example: Stiffness of Boards

- Recall that p=4 measurements were taken on each of 30 boards.
- Data are shown on the next slide along with the four columns of standardized values and a column of squared distances.
- Note that boards 9 and 16 have unusually large d^2, of 11.96 and 16.94, respectively. The probability of observing values larger than 11.96 and 16.94 are 0.01768 and 0.00198, respectively.
- In a sample of size 30, we would expect about 0.53 boards (30 x 0.01768) with $d^2 > 11.96$ and about 0.0594 boards (30 x 0.00198) with $d^2 > 16.94$.
- Board 16, with $d^2 = 16.85$ is not flagged as an outlier when we only consider the univariate standardized measurements.

Obs	Board	X1	X2	X3	X4	z1	z2	z3	z4	d_squared	prob
1	1	1889	1651	1561	1778	-0.06089	-0.30926	0.17108	0.16427	0.5936	0.96377
2	2	2493	2048	2087	2197	1.76882	0.93679	1.90603	1.46211	7.2337	0.12404
3	3	2119	1700	1815	2222	0.63585	-0.15547	1.00887	1.53955	7.5891	0.10784
4	4	1645	1627	1110	1533	-0.80004	-0.38459	-1.31650	-0.59461	5.0517	0.28203
5	5	1976	1916	1614	1883	0.20266	0.52248	0.34589	0.48950	1.3888	0.84614
6	6	1712	1712	1439	1546	-0.59708	-0.11780	-0.23133	-0.55434	2.1898	0.70090
7	7	1943	1685	1271	1671	0.10269	-0.20255	-0.78546	-0.16716	4.9996	0.28734
8	8	2104	1820	1717	1874	0.59041	0.22117	0.68563	0.46163	1.2979	0.86173
9	9	2983	2794	2412	2581	3.25319	3.27823	2.97801	2.65154	11.9557	0.01768
10	10	1745	1600	1384	1508	-0.49711	-0.46934	-0.41274	-0.67205	0.7581	0.94397
11	11	1710	1591	1518	1667	-0.60314	-0.49758	0.02925	-0.17955	1.9812	0.73922
12	12	2046	1907	1627	1898	0.41471	0.49424	0.38877	0.53597	0.4794	0.97548
13	13	1840	1841	1595	1741	-0.20933	0.28708	0.28322	0.04966	2.6897	0.61102
14	14	1867	1685	1493	1678	-0.12753	-0.20255	-0.05321	-0.14548	0.1179	0.99833
15	15	1859	1649	1389	1714	-0.15177	-0.31554	-0.39625	-0.03397	1.0760	0.89806
16	16	1954	2149	1180	1281	0.13602	1.25379	-1.08561	-1.37518	16.9426	0.00198
17	17	1325	1170	1002	1176	-1.76943	-1.81896	-1.67272	-1.70041	3.5050	0.47711
18	18	1419	1371	1252	1308	-1.48467	-1.18809	-0.84813	-1.29154	3.9476	0.41315
19	19	1828	1634	1602	1755	-0.24568	-0.36262	0.30631	0.09303	1.4325	0.83852
20	20	1725	1594	1313	1646	-0.55770	-0.48817	-0.64692	-0.24460	1.4320	0.83862
21	21	2276	2189	1547	2111	1.11146	1.37934	0.12490	1.19573	9.9061	0.04204
22	22	1899	1614	1422	1477	-0.03060	-0.42539	-0.28740	-0.76807	4.6411	0.32614
23	23	1633	1513	1290	1516	-0.83640	-0.74240	-0.72279	-0.64727	0.7763	0.94159
24	24	2061	1867	1646	2037	0.46015	0.36869	0.45144	0.96652	2.5522	0.63532
25	25	1856	1493	1356	1533	-0.16086	-0.80517	-0.50509	-0.59461	4.1892	0.38100
26	26	1727	1412	1238	1469	-0.55164	-1.05940	-0.89430	-0.79285	3.2583	0.51556
27	27	2168	1896	1701	1834	0.78429	0.45971	0.63285	0.33773	2.0625	0.72427
28	28	1655	1675	1414	1597	-0.76975	-0.23394	-0.31379	-0.39637	2.8309	0.58651
29	29	2326	2301	2065	2234	1.26292	1.73087	1.83346	1.57672	6.5374	0.16245
30	30	1490	1382	1214	1284	-1.26959	-1.15356	-0.97346	-1.36588	2.5841	0.62964