Statistics 5041
9. Multivariate Normal Distribution

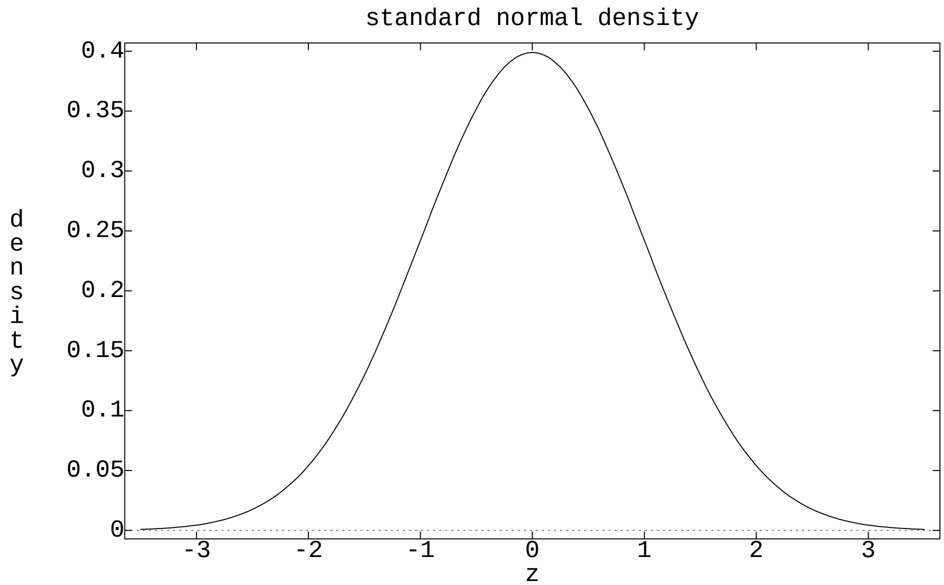
Gary W. Oehlert
School of Statistics
313B Ford Hall
612-625-1557
gary@stat.umn.edu

The *univariate normal* distribution plays a key role in univariate statistics.

$$x \sim \mathrm{N}(\mu, \sigma^2)$$

means that x has normal distribution with mean μ and variance σ^2. *Standardized* x is $z = (x - \mu)/\sigma$. The density of x is

$$\begin{aligned} f(x; \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2}} e^{-\frac{1}{2}(x-\mu)\sigma^{-2}(x-\mu)} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}z^2} \end{aligned}$$



$z \sim \mathrm{N}(0, 1)$ follows a standard normal distribution.
$z^2 = (x - \mu)^2/\sigma^2$ is χ^2_1 (chi squared with 1 degree of freedom).
$E(z^2) = 1$. $\mathrm{Var}(z^2) = 2$. $P(z^2 > k) = P(|z| > \sqrt{k})$, which we can get from a normal table. For example, $P(z^2 > 3.84) = P(|z| > 1.96) = .05$.
Suppose that $x_1, x_2, \ldots, x_p$ are *independent* normals with expectations μ_i and variances σ^2_i.
Let x be the vector with elements x_i; let μ be the vector with elements μ_i; and let Σ be the diagonal matrix with elements $\sigma^2_i = \sigma_{ii}$.
What is the density of x?

$$
\begin{aligned}
f(x;\mu,\Sigma) &= f(x_1,x_2,\ldots,x_p;\mu_1,\ldots,\mu_p,\sigma_1^2,\ldots,\sigma_p^2) \\
&= \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma_i^2}} e^{-\frac{1}{2}(x_i-\mu_i)\sigma_i^{-2}(x_i-\mu_i)} \\
&= \frac{1}{(2\pi)^{p/2}} \frac{1}{\sqrt{\prod_{i=1}^{p}\sigma_i^2}} e^{-\frac{1}{2}\sum_{i=1}^{p}(x_i-\mu_i)\sigma_i^{-2}(x_i-\mu_i)} \\
&= \frac{1}{(2\pi)^{p/2}} \frac{1}{|\Sigma|^{.5}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}
\end{aligned}
$$

Multivariate normal distribution. Let x and μ be p-vectors, and let Σ be a symmetric, positive definite matrix.

$$x \sim \mathrm{N}_p(\mu, \Sigma)$$

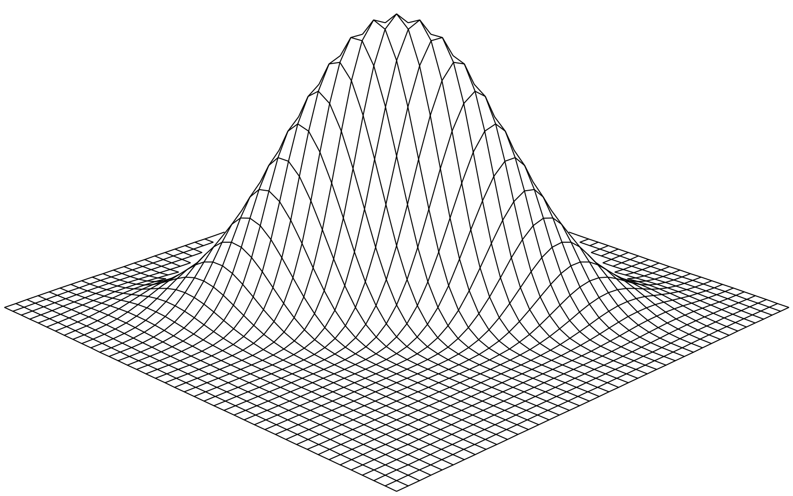
means that x follows the multivariate normal distribution with mean μ and variance Σ. The density is

$$f(x;\mu,\Sigma) = \frac{1}{(2\pi)^{p/2}} \frac{1}{|\Sigma|^{.5}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}$$

Standard multivariate normal has $\mu = 0$ and $\Sigma = \mathbf{I}_p$.
Some facts:
$E[x] = \mu$
$\mathrm{Var}[x] = \Sigma$
$(x-\mu)'\Sigma^{-1}(x-\mu) = \mathrm{trace}(\Sigma^{-1}(x-\mu)(x-\mu)')$ has a χ_p^2 distribution, with expected value p and variance $2p$.
Mode at μ; all level curves are ellipses centered at μ.
If q^2 is the upper α percent point of a χ_p^2, then the ellipse $(x-\mu)'\Sigma^{-1}(x-\mu) \le q^2$ describes a region with probability $1-\alpha$.

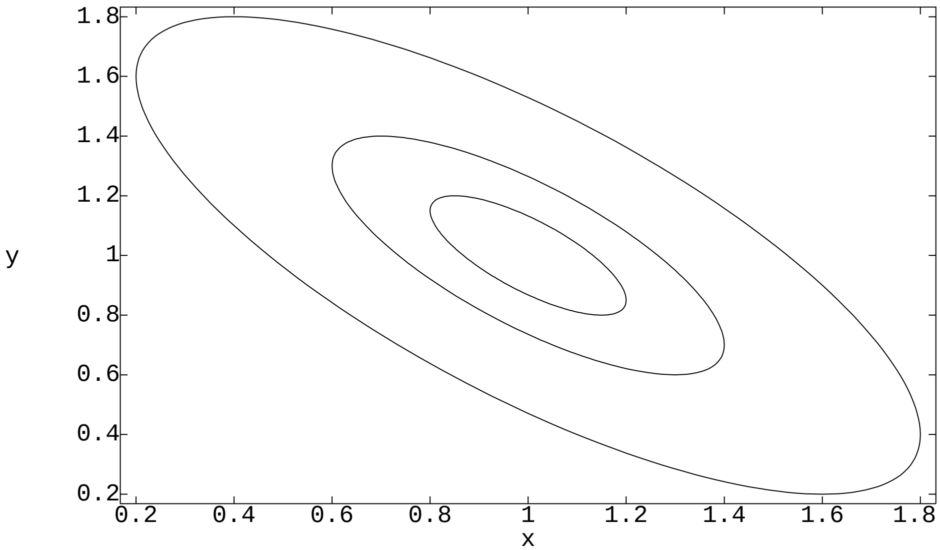
$$p = 2, \mu = (1,1)', \Sigma = \begin{bmatrix} .16 & -.12 \\ -.12 & .16 \end{bmatrix}$$

Shown from (4,4) direction.



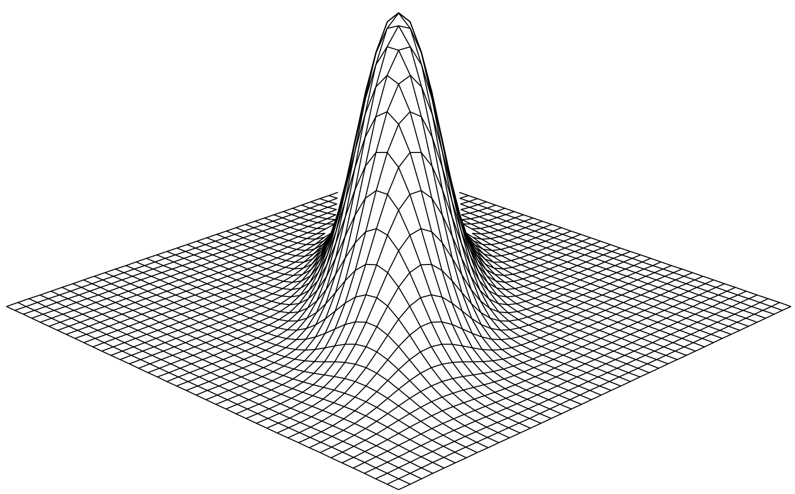
$$p = 2, \mu = (1,1)', \Sigma = \begin{bmatrix} .16 & -.12 \\ -.12 & .16 \end{bmatrix}$$

Contours at $2^2, 1^2, .5^2$ (probabilities .865, .393, and .117).



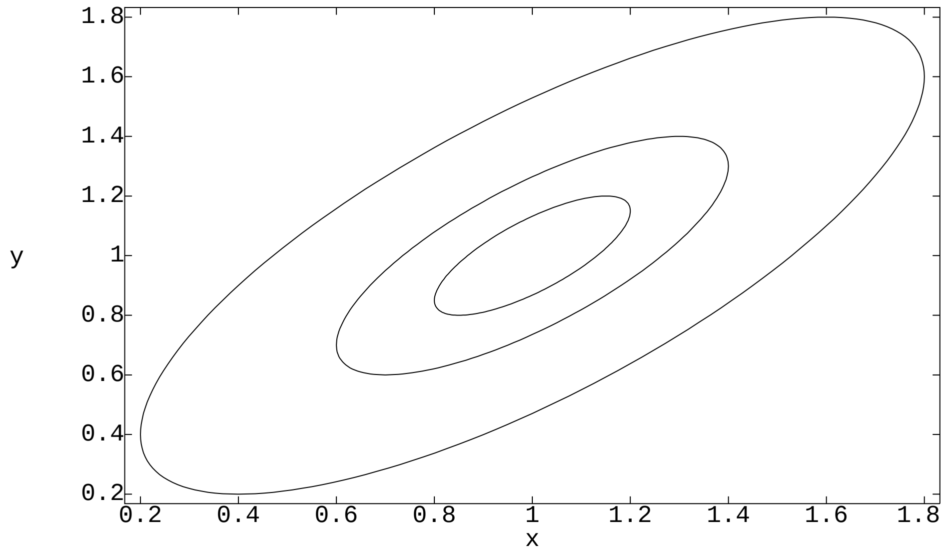
$$p = 2, \mu = (1,1)', \Sigma = \begin{bmatrix} .16 & .12 \\ .12 & .16 \end{bmatrix}$$

Shown from (4,4) direction.



$$p = 2, \mu = (1,1)', \Sigma = \begin{bmatrix} .16 & .12 \\ .12 & .16 \end{bmatrix}$$

Contours at $2^2, 1^2, .5^2$ (probabilities .865, .393, and .117).



Properties of the Multivariate Normal.
All *marginal* distributions are normal.
Divide x into its first p_1 elements and its remaining $p_2 = p - p_1$ elements: $x' = [x'_1, x'_2]$ Partition μ and Σ in the same way (subscripts on x and μ now indicate the partition instead of the individual element)

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

$x_i \sim \mathrm{N}_{p_i}(\mu_i, \Sigma_{ii})$
μ_1 and x_1 are p_1 vectors.
μ_2 and x_2 are p_2 vectors.
Σ_{11} is $p_1 \times p_1$.
Σ_{12} is $p_1 \times p_2$.
$\Sigma_{21} = \Sigma'_{12}$ is $p_2 \times p_1$.
Σ_{22} is $p_2 \times p_2$.
x_1 and x_2 are independent if $\Sigma_{12} = 0$.
All *linear combinations* are normal.
Let $\mathbf{B}$ be $q \times p$, then

$$\mathbf{B}x \sim \mathrm{N}_q(\mathbf{B}\mu, \mathbf{B}\Sigma\mathbf{B}')$$

If $\Sigma = \mathbf{U}\Lambda\mathbf{U}'$ is the spectral decomposition of Σ, then $v = \mathbf{U}'x$ has distribution

$$\mathbf{U}'x \sim \mathrm{N}_p(\mathbf{U}'\mu, \mathbf{U}'\Sigma\mathbf{U}) = \mathrm{N}_p(\mathbf{U}'\mu, \Lambda)$$

In particular, v has independent components.
All *conditional* distributions are normal.
$x_2|x_1$ is normal with mean

$$\mu_{2\bullet 1} = \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)$$

and variance

$$\Sigma_{22\bullet 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

This is a *linear* regression of x_2 on x_1
$\beta_{2\bullet 1} = \Sigma_{21}\Sigma_{11}^{-1}$ is a $p_2 \times p_1$ matrix of regression coefficients.
$\Sigma_{2\bullet 1}$ does not depend on x_1.
Try these out with $p_1 = p_2 = 1$ and compare with simple linear regression.

```
Cmd> Sigma <- .5^(abs(run(4)-run(4)'))

Cmd> Sigma
(1,1)         1       0.5      0.25     0.125
(2,1)       0.5         1       0.5      0.25
(3,1)      0.25       0.5         1       0.5
(4,1)     0.125      0.25       0.5         1

Cmd> g1 <- vector(1,2); g2 <- vector(3,4)

Cmd> Sigma11 <- Sigma[g1,g1]

Cmd> Sigma12 <- Sigma[g1,g2]


Cmd> Sigma21 <- Sigma[g2,g1]

Cmd> Sigma22 <- Sigma[g2,g2]

Cmd> Sigma21%*%solve(Sigma11)
(1,1)             0          0.5
(2,1)             0         0.25

Cmd> Sigma22-Sigma21%*%solve(Sigma11)%*%Sigma12
(1,1)          0.75        0.375
(2,1)         0.375       0.9375

Cmd> Sigma <- dmat(4,1)+rep(1,4)*rep(1,4)'

Cmd> Sigma
(1,1)         2         1         1         1
(2,1)         1         2         1         1
(3,1)         1         1         2         1
(4,1)         1         1         1         2

Cmd> Sigma11 <- Sigma[g1,g1]

Cmd> Sigma12 <- Sigma[g1,g2]

Cmd> Sigma21 <- Sigma[g2,g1]
```

```
Cmd> Sigma22 <- Sigma[g2,g2]

Cmd> Sigma21%*%solve(Sigma11)
(1,1)      0.33333      0.33333
(2,1)      0.33333      0.33333

Cmd> Sigma22-Sigma21%*%solve(Sigma11)%*%Sigma12
(1,1)       1.3333      0.33333
(2,1)      0.33333       1.3333
```

Let $\vec{\mathbf{X}}_1, \vec{\mathbf{X}}_2, \ldots, \vec{\mathbf{X}}_n$ be idependent with $\vec{\mathbf{X}}_i$ having distribution $\mathrm{N}_p(\mu_i, \Sigma)$. Then

$$V_1 = \sum_{i=1}^n c_i \vec{\mathbf{X}}_i \sim \mathrm{N}(\sum_{i=1}^n c_i \mu_i, (\sum_{i=1}^n c_i^2)\Sigma)$$

If $V_2 = \sum_{i=1}^n b_i \vec{\mathbf{X}}_i$, then V_1 and V_2 are jointly normal with covariance

$$\sum_{i=1}^n (b_i c_i)\Sigma$$

This is completely analogous to the univariate situation.
Sampling Distribution. Suppose that $\vec{\mathbf{X}}_i$ are iid $N_p(\mu, \Sigma)$. Then $\overline{\mathbf{x}}$ has distribution

$$\mathrm{N}(\mu, \frac{1}{n}\Sigma)$$

$\overline{\mathbf{x}}$ is an unbiased estimate of μ and is also the maximum likelihood estimate of μ.
This is completely analogous to the univariate situation.

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^n (\vec{\mathbf{X}}_i - \overline{\mathbf{x}})(\vec{\mathbf{X}}_i - \overline{\mathbf{x}})'$$

is an unbiased estimate of Σ, and $\frac{n-1}{n}\mathbf{S}$ is the maximum likelihood estimate of Σ.
$\overline{\mathbf{x}}$ and $\mathbf{S}$ are independent.
This is completely analogous to the univariate situation.
If z_i are iid $\mathrm{N}(0, \sigma^2)$ (univariate), then

$$\sum_{i=1}^n z_i^2 \sim \sigma^2 \chi_n^2$$

If z_i are iid $\mathrm{N}(0, \Sigma)$ (p-variate), the

$$\sum_{i=1}^n z_i z_i' \sim W_n(\Sigma)$$

which is a *Wishart* distribution with n degrees of freedom and parameter Σ.
$(n-1)\mathbf{S} \sim W_{n-1}(\Sigma)$
Wishart density only exists if degrees of freedom greater than dimension.
Let $\mathbf{V}_1 \sim W_n(\Sigma)$ and $\mathbf{V}_2 \sim W_m(\Sigma)$, then

$$\mathbf{V}_1 + \mathbf{V}_2 \sim W_{n+m}(\Sigma)$$

(df add if Σ matches).

$$\mathbf{C}\mathbf{V}_1\mathbf{C}' \sim W_n(\mathbf{C}\Sigma\mathbf{C}')$$

Law of large numbers
$x_1, x_2, \ldots, x_n$ are p-variate iid from a population with mean μ.
Then $\overline{\mathbf{x}}$ converges (in probability) to μ as n tends to infinity.
If Σ exists, then $\mathbf{S}$ converges to Σ in probability as n tends to infinity.
$x_1, x_2, \ldots, x_n$ are p-variate iid from a population with mean μ and nonsingular variance Σ. Then

$$\sqrt{n}(\overline{\mathbf{x}} - \mu) \rightarrow \mathrm{N}(0, \Sigma)$$

and

$$\sqrt{n}(\overline{\mathbf{x}} - \mu)'\mathbf{S}^{-1}(\overline{\mathbf{x}} - \mu) \rightarrow \chi^2_p$$

as $n - p$ goes to infinity.
Multivariate Standardization.
x has mean μ and variance Σ.
We want $\mathbf{C}$ so that

$$z = \mathbf{C}(x - \mu)$$

has mean 0 and variance $\mathbf{I}_p$ and is standardized.
$\mathbf{C}(x - \mu)$ has mean 0 and variance $\mathbf{C}\Sigma\mathbf{C}'$, so we need $\mathbf{C}$ such that $\mathbf{C}\Sigma\mathbf{C}' = \mathbf{I}_p$.
We want to write $\Sigma = \mathbf{B}\mathbf{B}'$ for some nonsingular $\mathbf{B}$. Then

$$\mathbf{B}^{-1}\Sigma(\mathbf{B}')^{-1} = \mathbf{I}_p$$

so $\mathbf{C} = \mathbf{B}^{-1}$ is what we need.
One choice derived from the spectral decomposition of Σ is

$$\mathbf{B} = \mathbf{U}\Lambda^{.5}\mathbf{U}'$$

This is a symmetric square root.
There are other choices, so the multivariate standardization *is not unique*.
Another common choice:

$$\Sigma = \mathbf{L}\mathbf{U} = \mathbf{U}'\mathbf{U}$$

where $\mathbf{U}$ is *upper triangular*.
This is called the *Cholesky Decomposition* of Σ.
z is not unique, but

$$\begin{aligned} ||z||^2 &= z'z = (x - \mu)'\mathbf{C}'\mathbf{C}(x - \mu) \\ &= (x - \mu)'(\mathbf{B}')^{-1}\mathbf{B}^{-1}(x - \mu) \\ &= (x - \mu)'\Sigma^{-1}(x - \mu) \end{aligned}$$

is unique.
$||z||^2 = \sum_{i=1}^{p} z_i^2 \sim \chi^2_p$, showing that $(x - \mu)'\Sigma^{-1}(x - \mu) \sim \chi^2_p$.
Standarizing, or at least diagonalizing the covariance matrix, is often the beginning of understanding in multivariate.

```
Cmd> Sigma <- matrix(vector(16,-12,-12,16),2)

Cmd> Sigma
(1,1)            16           -12
(2,1)           -12            16

Cmd> c <- cholesky(Sigma);c
(1,1)             4            -3
(2,1)             0        2.6458

Cmd> c'%*%c
(1,1)            16           -12
(2,1)           -12            16

Cmd> eigout <- eigen(Sigma);\
evec <- eigout$vectors;\
eval <- eigout$values

Cmd> d <- evec%*%dmat(eval^.5)%*%evec'

Cmd> d
(1,1)        3.6458       -1.6458
(2,1)       -1.6458        3.6458

Cmd> d%*%d
(1,1)            16           -12
(2,1)           -12            16
```