p-variate Normal Distribution

04-02-2020

In the univariate normal distribution

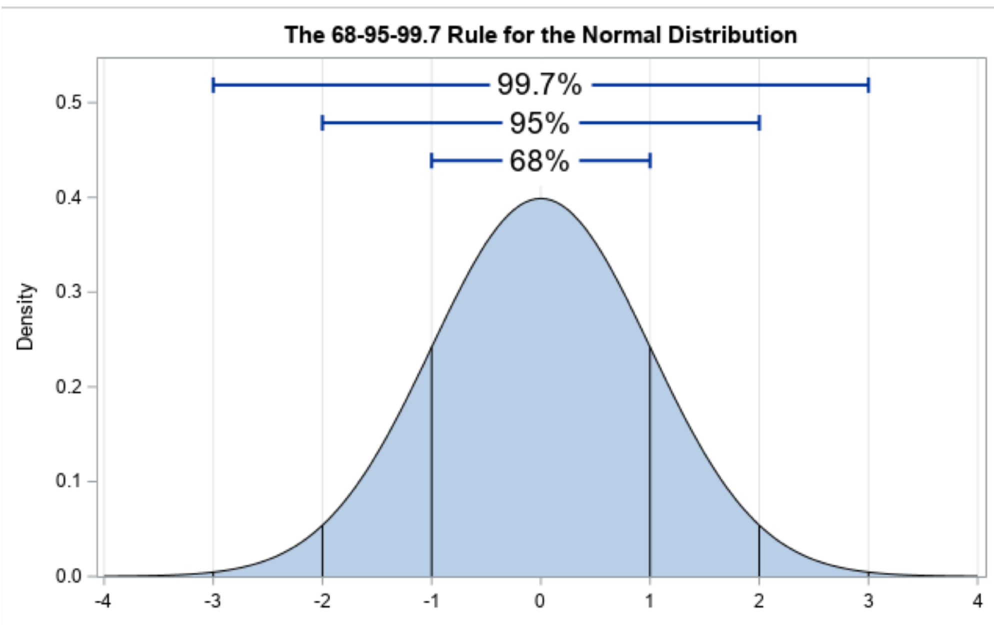
$$X \sim N(\mu, \sigma^2)$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{(-1/2\sigma^2)(x-\mu)^2} \quad (1) \quad , \; -\infty < x < \infty$$

we can rewrite (1) as

$$f(x) = \frac{1}{\sqrt{2\pi}(\sigma^2)^{1/2}} e^{(-1/2)(x-\mu)(\sigma^2)^{-1}(x-\mu)} \quad , \; -\infty < x < \infty$$

$$\mathbb{E}X = \mu, \quad var(x) = \sigma^2$$



$\mu-3\sigma \quad \mu-2\sigma \quad \mu-\sigma \quad \sigma \quad \mu+\sigma \quad \mu+2\sigma \quad \mu+3\sigma$

multivariate normal distribution:

$$\underset{\sim}{X} \sim N_p(\underset{p\times 1}{\mu}, \underset{p\times p}{\Sigma})$$

where: μ is a vector
Σ is a covariance matrix ≈ generalized population variance

$$f(\underline{x}) = \frac{1}{(\sqrt{2\pi})^p |\underline{\Sigma}|^{1/2}} \exp\left\{-1/2 \underbrace{(\underline{x}-\underline{\mu})' \underline{\Sigma}^{-1} (\underline{x}-\underline{\mu})}_{\text{mahalanobis distance}}\right\}$$

where $\underset{p\times 1}{\underline{X}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$, $E(\underline{X}) = \underset{p\times 1}{\underline{\mu}} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$

p is the number of variables

$$V(\underline{X}) = cov(\underline{X}) = \underline{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

$\underline{\Sigma}$ is:

(1) sym

(2) PD (positive definite)

$$\Delta^2 = (\underline{X}-\underline{\mu})' \underline{\Sigma}^{-1} (\underline{X}-\underline{\mu})$$

is $\boxed{\text{mahalanobis}}$ distance the square generalized distance from $\underline{X}$ to $\underline{\mu}$

recall the properties of this distance.

It can be shown that

$E(\Delta^2) = p$ see problem 4.4

therefore $p\uparrow \Rightarrow \Delta^2\uparrow$

as long as the number of variables increase, the mahalanobis distance increases.

Also, check the effect of $|\Sigma|$ on the density

Example:

$$X \sim N_2\left[\begin{pmatrix}5\\5\end{pmatrix}, \begin{pmatrix}\sigma^2 & \rho\sigma^2\\ \rho\sigma^2 & \sigma^2\end{pmatrix}\right]$$

$\downarrow$

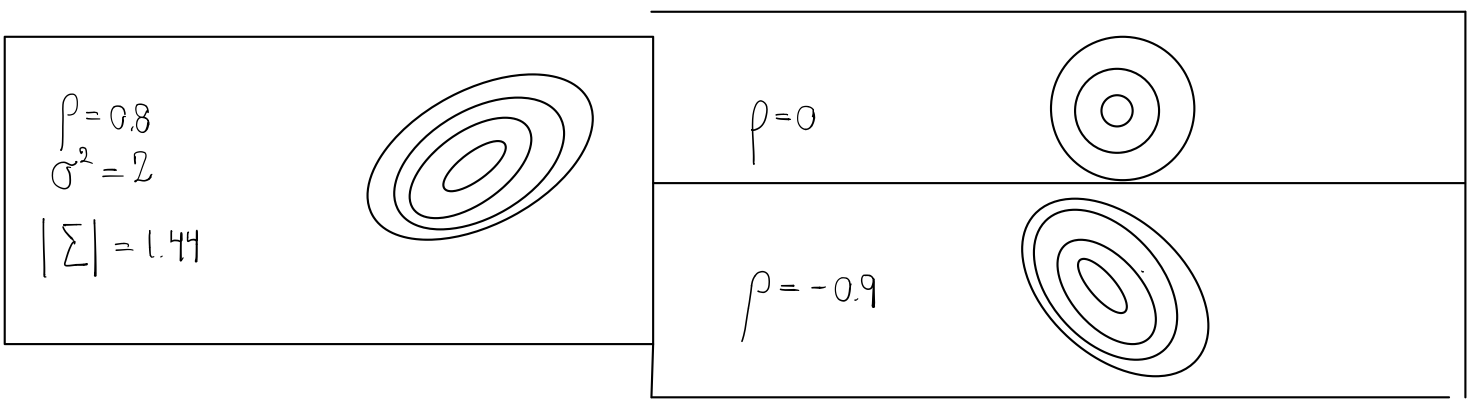
$\begin{pmatrix}X_1\\X_2\end{pmatrix}$

where

$$\rho = \frac{cov(X_1, X_2)}{\sigma x_1, \sigma x_2} = \frac{cov(X_1, X_2)}{\sigma^2}$$

where we asume X_1 and X_2 have the same variance, otherwise it would be $\rho_{\sigma_1\sigma_2}$

Portraying concentration of points with contour plots



If σ^2 is smaller, then the circules are closer to each other, ρ tells you the inclination of the circles.

Generalized population variance

In Section 3.11 we saw that the absolute value of S

$|S|$: generalized sample variance

$\Rightarrow |\underline{\Sigma}|$: generalized population variance

* In the univariate normal : If σ^2 is small $\rightarrow$ X values are concentrated near the mean.

* In the p-variate normal : If $|\Sigma|$ is small $\rightarrow$ $\underline{X}$'s are concentrated close to vector $\underline{\mu}$ in p-space or that there is multicollinearity among variables.

multicollinearity: indicates that the variables are highly intercorrelated, this implies that the effective dimensionality is less than p.
(see chapter 12 for finding a reduce number of new dimensions)

Also in the presence of the multicollinearity, one or some of the eigenvalues of Σ will be close to zero and hence $|\Sigma|$ would be small, since $|\Sigma| = \prod_{i=1}^{p} \lambda_i$
(see 2.108)

In a p-variate normal, reducing the intercorrelation between p-variates or increasing the variance of p-variates, leads to increase in $|\Sigma|$

SPSS { learn and download. (pendiente)

Some properties of p-variate distribution:

$$\underset{p\times 1}{\underset{\sim}{X}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \sim N_p(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$$

remember:

x, y independ $\Rightarrow$ $cov(x, y) = 0$

$cov(x, y) = 0$ $\cancel{\Rightarrow}$ x, y independ.

$\boxed{\text{only for normal distribution:}}$

x, y indepen $\Leftrightarrow$ $cov(x, y) = 0$

Ⓐ normality of linear combinations of the variables in y.

(1) any linear combination of $X_1 \dots X_p$ is normal

$z = \underset{\sim}{a}'\underset{\sim}{X}$ where $\underset{\sim}{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix}$ { vector of constants.

then $z \sim N(\underset{\sim}{a}'\mu, \underset{\sim}{a}' \underset{\sim}{\Sigma} a)$

$$\mathbb{E}\, a'X = a' \mathbb{E}\, X = a'\mu$$

$$V\, a'x = a' \operatorname{var} X a = a' \Sigma a$$

2) suppose $a_1 \dots a_q$ are q vectors with dimension of p ($q \le p$)

$$\underset{q\times q}{\underline{A}} = \begin{pmatrix} a_1' \\ \vdots \\ a_q' \end{pmatrix}$$

$$\underline{A}\,\underline{X} \sim N_q(\underline{A}\,\underline{\mu}, \underline{A}\,\underline{\Sigma}\,\underline{A}')$$

$$E(\underline{A}\,\underline{X}) = \underline{A}\,E\,\underline{X} = \underline{A}\,\underline{\mu}$$

$$V(\underline{A}\,x) = \underline{A}\,V(\underline{x})\,A' = \underline{A}\,\underline{\Sigma}\,\underline{A}'$$

(B) standardized variables

(3) Standardize p-variate normal distr.

(i) $\underline{z} = (T')^{-1}(\underline{X} - \underline{\mu})$ where $\underline{\Sigma} = T'T$ cholesky decomposition

then $\underline{z} \sim N_p(\underline{0}, I_p)$

(ii) $\underline{z} = (\Sigma^{1/2})^{-1}(\underline{x} - \underline{\mu})$ where $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$,

then $\underline{z} \sim N_p(\underline{0}, I_p)$

$$X_1 \dots X_n \xrightarrow{iid} N(\mu, \sigma^2)$$

theorem

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\bar{X} \perp\!\!\!\perp S^2$$

06-02-20

Moment Generating Function

univariate setting

$\mathbb{E}\, x^r = \mathbb{E}\, y^r \;\; \overset{\leftarrow}{\cancel{\rightarrow}} \;\; x \sim y$

$f(x) = f(y)$ pdf/pmf

$F(x) = F(y)$ cdf

$M_x(t) = M_y(t)$

$\boxed{M_x(t) = \mathbb{E}\, e^{tx}}$ for univariate X

moment generating function for univariate case

multivariate setting

$\underset{\sim}{x} = (x_1 \ldots x_p)'$

p×1 variables

Let $\underline{t} = (t_1 \ldots t_p)' \in \mathbb{R}^p$ $\quad$ t is a vector of constants

$\varphi(\cdot) = \mathbb{R}^p \to \mathbb{R}$, then $\boxed{\varphi_{\underline{x}}(t) = \mathbb{E}\, e^{\underline{t}'\underline{x}}}$ $\quad$ the result is a escalar

moment generating function for multivariate case

¿How can I find the moment generating function for a multivariate setting?

(1) $\varphi_{\underline{x}}((s, 0, \ldots 0)') = \mathbb{E}\, e^{\overset{t'}{\boxed{(s, 0 \ldots 0)}} \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}} = \mathbb{E}\, e^{s x_1} = M_{x_1}(s)$

$\varphi_{\underline{x}}((0, t_2, t_3, 0 \ldots 0)') = \varphi_{x_2, x_3}(t_2, t_3)$ $\quad$ $\left\{ \right.$ φ of x_2 and x_3 at point t_2 and t_3

if we had a linear combination:

(2) $\underset{\text{linear combination}}{y_{1\times1}} = \underset{1\times1}{\underline{a}'\underline{x}} \Rightarrow M_y(s) = \mathbb{E} e^{ys} = \mathbb{E}\, e^{s\underline{a}'\underline{x}} = \varphi_{\underline{x}}(s\,\underline{a})$

scalar, vector

(3) if X_i's are independent then

$$\varphi_{\underline{X}}(\underline{t}) = \prod_{1}^{n} M_{X_i}(t_i)$$

what is a moment generating function ?

$X \sim N(\mu, \sigma^2)$

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

$$M_x(t) = \mathbb{E} e^{tx} = \int e^{tx} f(x)\, dx$$

$$= \boxed{\exp\left\{ \mu t + t^2\sigma^2/2 \right\}}$$

→ moment generating function of a normal

$\boxed{X_i\text{'s independent}}$

$$M_{\bar{X}}(t) = \mathbb{E} e^{t\bar{X}} = \mathbb{E} e^{t/n \sum x_i}$$

r.v.

$$= \left(\mathbb{E} e^{t/n\, x_i}\right)^n = M_{x_1}(t/n)^n$$

moment generating function of one of this normals

$$m_x(t/n) = \exp\left\{ \frac{\mu t}{n} + \frac{\sigma^2}{2} \frac{t^2}{n^2} \right\}$$

$$(m_x(t/n))^n = \exp\left\{ \mu t + \frac{t^2}{n} \frac{\sigma^2}{2} \right\}$$

m.g.f of a normal distribution

mean ↑ variance ↑

Bivariate normal distribution

ʻ how can we get the bivariate normal distributions pdf ?

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

covariance

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$\sigma_{12} = cov(X_1, X_2)$

$f:10$

correlation

$$\rho = \frac{cov(X_1, X_2)}{\sigma_1 \sigma_2} \rightarrow cov(X_1, X_2) = \rho \sigma_1 \sigma_2$$

determinant of sigma

$$|\Sigma| = \sigma_1^2 \sigma_2^2 - \rho^2 \sigma_1^2 \sigma_2^2 = \sigma_1^2 \sigma_2^2 (1-\rho^2)$$

because I know this is a variance of a positive definite matrix w know this is positive

$$\rightsquigarrow \rho \in (0,1)$$

inverse of sigma

$$\Sigma^{-1} = \underbrace{\frac{1}{\sigma_1^2 \sigma_2^2 (1-\rho^2)}}_{\frac{1}{|\Sigma|}} \underbrace{\begin{pmatrix} \sigma_2^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}}_{\Sigma} = \underbrace{\frac{1}{1-\rho^2} \begin{bmatrix} 1/\sigma_1^2 & \rho/\sigma_1\sigma_2 \\ \rho/\sigma_1\sigma_2 & 1/\sigma_2^2 \end{bmatrix}}_{\Sigma^{-1}}$$

$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} e^{-1/2\, Q(x_1, x_2)}$$

where: $Q(x_1, x_2) = (\underline{x} - \underline{\mu})' \underline{\Sigma}^{-1} (\underline{x} - \mu)$ quadratic form

$$= \frac{(x_1-\mu_1)^2}{\sigma_1^2(1-\rho^2)} + \frac{(x_2-\mu_2)^2}{\sigma_2^2(1-\rho^2)} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2(1-\rho^2)}$$

therefore:

$$f_{x_1x_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} e^{-1/2(1-\rho^2)\left[\left(\frac{x_1-u_1}{\sigma_1}\right)^2 + \left(\frac{x_2-u_2}{\sigma_2}\right)^2 + \frac{2\rho(x_1-\mu_1)(x_2-u_2)}{\sigma_1\sigma_2}\right]}$$

X

pdf of a bivariate normal

conclusion: joint distributions are bivariate normals

Statement:

we want to show that if marginals are normals, then conditionals are normal:

$$(X, Y) \sim N_2(\underline{\mu}, \Sigma)$$

μ_x
μ_y

Let $Z_x = \dfrac{X - \mu_x}{\sigma_x}$

$Z_y = \dfrac{Y - \mu_y}{\sigma_y}$

then

lower case

(a) $Z_y \mid Z_x = z_x \sim N(\rho z_x, 1 - \rho^2)$

(b) $Y \mid X = \underbrace{x}_{\text{r.v.}} \sim N\left(\mu_y + \rho \sigma_y / \sigma_x (x - \mu_x), \sigma_y^2 (1 - \rho^2)\right)$

this shows that the conditional distributions are univariate normal

proof:

For this proof we use:

lema:

$Z_x \sim N(0,1)$

$Z_y \sim N(0,1)$

because whenever the joint distributions are bivariate normal the marginals are normal

$$\int_y f_{Z_x Z_y} \, dy = f_{Z_x}$$

$$\begin{pmatrix} Z_x \\ Z_y \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

N_2: bivariate normal

$$f_{Z_y \mid Z_x}(z_y, z_x) = \frac{f_{Z_x Z_y}(z_x, z_y)}{f_{Z_x}(z_x)}$$

$$= \frac{\frac{1}{2\pi\sqrt{1-\rho^2}} e^{(-1/2)(1-\rho^2)(z_x^2 + z_y^2 - 2\rho z_x z_y)}}{\frac{1}{\sqrt{2\pi}} e^{(-1/2) z_n^2}}$$

z_y : z of y

z_x : z of x

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1-\rho^2}} e^{-1/2} (1-\rho^2)(z_y^2 - 2\rho z_x z_y + \rho^2 z_x^2)$$

$$\Rightarrow z_y | z_x = z_x \sim N(\rho z_x, 1-\rho^2)$$

we showed that the conditional of $z_y | z_x$

$\boxed{\text{remark}}$: ρz_x : mean of the conditional

$1-\rho^2$: variance of the conditional

the conditional doesn't depend on x just on the variance of y

By an example, show that normality of marginals do not imply normality of joint distributions.

We will show it by an counterexample:

joint distribution

p-variate normal $\Rightarrow$ marginals are normal

$\cancel{\Leftarrow}$

$$F(x,y) = \Phi(x)\Phi(y)(1+\alpha(1-\Phi(x))(1-\Phi(y)))$$

bivariate joint cdf of x and y

$$F(x) = \lim_{y\to\infty} F(x,y) = \Phi(x) \Rightarrow X \sim N(0,1)$$

marginal of x

$$F(y) = \lim_{x\to\infty} F(x,y) = \Phi(y) \Rightarrow y \sim N(0,1)$$

marginal of y

$\rightsquigarrow \Phi(\infty) = 1$

$\therefore\ x \sim N(0,1)$; $y \sim N(0,1)$ but $F(x,y)$ is not $\sim N_2(0,1)$ why?

Show that $f(x,y) \neq f(x) f(y)$

$f(x,y) \overset{?}{=} \underbrace{\varphi(x)\varphi(y)(1+\alpha\varphi(y))(1-2\varphi(x))}$ } pdf

we get it we derivative — this is definitely not

copula

now f_x and f_y are correlated to each other to get $f_{x,y}$

this comes from: $\dfrac{\partial F(x,y)}{\partial x \partial y} = f(x,y)$

Theorem: if $\underline{x} \sim N_p(\mu, \Sigma)$, then $\underbrace{(\underline{x}-\underline{\mu})' \Sigma^{-1} (\underline{x}-\underline{\mu})}_{\text{mahalanobis distance}} \sim \chi_p^2$

p-variate

$E(\Delta) = p$

$V(\Delta) = 2p$

proof $y = \Sigma^{-1/2}(x-\mu) \sim N_p(0, \underline{I_p})$ } standardize

implies

y's iid $N(0,1)$

$\sum_1^p y_i^2 = y'y = \Delta \sim \chi_p^2$. chi-square with p-degrees of freedom.

we know: if z_i's indepe. $N(0,1)$

then: $\sum z_i^2 \sim \chi_p^2$. chi-square with p-degrees of freedom

remark:

$x \sim N_p(\mu, \sigma^2 I_p)$ → x_i's are inde. and $H_{p\times p}$ is an orthogonal matrix $\boxed{(H'H) = I}$

then $\underset{p\times1}{y} = \underset{p\times p}{H}\,\underset{p\times1}{X} \sim N_p(H\mu, \sigma^2 I)$ } variance hasn't change nor independence.

→ p-variate normal

→ because $\rho = 0$ and $\sigma^2 I$ covariance matrix

Note! x_i's independent and consequently y's are independent

new random variable

The variance doesn't change

$cov(y) = H \underbrace{cov(x)}_{\sigma^2 I_p} H' = \sigma^2 I$

$H I H' = H H' = I$

Exercise:

$x_1, x_2 \overset{iid}{\sim} N(0,1)$

show that $y_1 = \dfrac{x_1 - x_2}{\sqrt{2}}$; $y_2 = \dfrac{x_1 + x_2}{\sqrt{2}}$ are independent

---

$x_1, x_2 \overset{iid}{\sim} N(0,1)$

Show that $y_1 = \dfrac{x_1 - x_2}{\sqrt{2}}$; $y_2 = \dfrac{x_1 + x_2}{\sqrt{2}}$ are independent.

① I can define
$$H = \begin{array}{cc} x_1 & x_2 \\ \left(\begin{matrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{matrix}\right) & \begin{matrix} \to y_1 \\ \to y_2 \end{matrix} \end{array}$$
as orthogonal because:

$$HH' = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$$

$\sigma^2 I = 1(I) = I$

$$(y_1, y_2)' \sim N_2(H0, I)$$
$$N_2(0, I_2) \to$$ I here means they are independent

how can I check that my p-variate distribution comes from normal?

$$\underline{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \sim N_p(\mu, \Sigma) \, ?$$

$\downarrow$ how to estimate μ and Σ

all marginals must be normals, then $\underline{X}$ is normal.

univariate normal   qq plot $\to$ linear pattern
data is normal

next class

* mle

* manova

* generalization of the chi-squared distribution