Continuous probability distributions

A **continuous random variable** Y is real-valued. The possible values for Y form intervals of real numbers (uncountably infinite set of possible values).

Probability density function (pdf; probability curve): the probability that Y takes a value in the interval (a, b) is the area under the pdf between a and b. Area under the entire pdf is 1.

Expected value or **mean** of Y is a measure of the center of the distribution of Y. It is a constant indicating where the pdf would balance on a fulcrum:

$$\mathrm{E}(Y) = \mu = \text{calculus formula}$$

Variance of Y is a constant that measures the spread of the distribution. It is the expected value of a squared deviation of Y from the mean:

$$\mathrm{V}(Y) = \sigma^2 = \mathrm{E}\big[(Y-\mu)^2\big] = \text{calculus formula}$$

1. Uniform distribution on (0, 1)

Pdf is a horizontal line of height 1:

$$f(y) = 1, \quad 0 < y < 1.$$

1

$f(y)$

0

0 a b 1

y

$$P(a < Y \leq b) = (1) \times (b-a) = b-a$$

ex.

$$P(.1 < Y \leq .3) = .3 - .1 = .2$$

Notation: $Y \sim$ uniform(0, 1)

SAS:

The function RANDUNI(seed) generates a random variable from a uniform(0, 1) distribution.

2. Normal (or Gaussian) distribution

"bell-shaped curve"

$$f(y) = \frac{1}{\sqrt{\sigma^2 2\pi}} \mathrm{e}^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

$$-\infty < y < +\infty$$

$$-\infty < \mu < +\infty$$

$$0 < \sigma^2 < +\infty$$

Notation: $Y \sim \text{normal}(\mu, \sigma^2)$

ex. ACT scores
baseball batting averages
log(household income)
lengths of finch beaks

Finding probabilities with the normal model: the **standard normal distribution** $\mu = 0$ $\sigma^2 = 1$

$Y \sim \text{normal}(\mu, \sigma^2)$, $Z = \frac{Y-\mu}{\sigma}$: then $Z \sim \text{normal}(0, 1)$

Also, if $Z \sim \text{normal}(0, 1)$ then

$$Y = \sigma Z + \mu \sim \text{normal}\big(\mu, \sigma^2\big)$$

$P(Z \leq z)$ is the **cumulative distribution function** for a standard normal distribution (tabulated in Table 1 p. 1091):

ex. Proportion of ACT scores between 20 and 30
$\mu = 21.37 \qquad \sigma = 4.55$

$z_1 = \frac{20-\mu}{\sigma} \approx -0.30$ $\qquad$ $z_2 = \frac{30-\mu}{\sigma} \approx 1.90$

$\text{area}_1 = 0.3821$ $\qquad$ $\text{area}_2 = 0.9713$

$$\text{area} = \text{area}_2 - \text{area}_1 = 0.5892$$

SAS:

RANDNOR(seed) generates a standard normal random variable

PROBNORM(z) returns the area under a standard normal pdf to the left of z

Percentiles of the normal: z_α denotes the value of z in a standard normal distribution such that the area to the **right** is α

z_α is the $100(1-\alpha)$th **percentile** of the standard normal

A few useful percentiles:

α	$100(1-\alpha)$	z_α	CI % coverage
.40	60	0.25335	
.35	65	0.38532	
.30	70	0.52440	
.25	75	0.67449	
.20	80	0.84162	60
.15866	84.134	1.0000	68.269
.15	85	1.0364	70
.10	90	1.2816	80
.05	95	1.6449	90
.025	97.5	1.9600	95
.02275	97.725	2.0000	95.450
.02	98	2.0537	96
.01	99	2.3263	98
.005	99.5	2.5758	99
.00135	99.865	3.0000	99.730
.001	99.9	3.0902	
.0005	99.95	3.2906	99.9
.0001	99.99	3.7190	
.00005	99.995	3.8908	99.99