村上　明
11/6/08
pp. 58-65

3.3.1 The normal distribution

<u>正規分布について</u>

- p.58 の Figure 3.6.の左上は、最も基本的な正規分布（μ （ミュー：平均）= 0, σ（シグマ：標準偏差） = 1）で、標準正規分布と呼ばれる
- 平均は縦線が入っているところで、density が最も高い箇所でもある
- 横線は標準偏差で、分散の程度を示す
- 平均を変化させることで、グラフを左右に動かすことができる
- 標準偏差を変化させると、グラフの広がりを変えることができる
- x 軸と曲線に囲まれた面積は 1 である
- 左右対称なので、平均以下となる確率が 0.5、平均以上となる確率も 0.5

<u>正規分布の描き方</u>

1. x の範囲を指定する

x = seq(-4, 4, 0.1)

2. y の値を dnorm( )関数で求める

y = dnorm(x)

- dnorm( )関数に引数をほかに付けなければ、平均が 0、SD が 1 の正規分布となる
- それぞれ指定したい場合は、mean = ○○, sd = ○○と指定する

3. プロットする

plot(x, y, xlab = "x", ylab = "density", ylim = c(0, 0.8), type = "l")
mtext("normal(0,1)", 3, 1)

- plot( )関数の引数で type = “l”と指定することにより、実線を引くことができる

4. 線を引く

abline(v = 0, lty = 2)
lines(c(-1, 0), rep(dnorm(-1), 2), lty = 2)

- 直線を引くには abline( )を用いる。縦線は v = ○○で、横線は h = ○○で位置を指定。lty は「ラインタイプ」の略で、どのような線を用いるか。
- 部分的に線を引くには、lines ( )関数を用いる。lines(c(x1, x2), c(y1, y2), lty = 1)で、(x1, y1)から(x2, y2)までの線を引くことができる。

<u>他の正規分布の描き方</u>

Figure 3.6 の右下のグラフ（μ = 4, σ = 0.5）

x = seq(0, 8, 0.1)
y = dnorm(x, mean = 4, sd = 0.5)
plot(x, y, xlab = "x", ylab = "density", ylim = c(0, 0.8), type = "l")

その他正規分布関連のグラフ

Figure 3.7 の左上のグラフ

x = seq(-3, 3, by = 0.01)
y = pnorm(seq(-3, 3, by = 0.01))
plot(x,y, ylab = "pnorm(x)", type = "l")
abline(v = 0, lty = 2)

右上のグラフ

plot(y,x, xlab = "p", ylab = "qnorm(x)", type = "l")
abline(h = 0, lty = 2)

右下のグラフ

shadenormal.fnc(qnts = c(0.025, 0.975))
※左右 2.5%ずつを shade する

標準化について

正規分布であれば、μ = 0、σ = 1 の標準正規分布に標準化することができる。
標準化：平均を引いて標準偏差で割る

x = rnorm(10, 3, 0.1)
(x - mean(x)) / sd(x)

scale( )関数を用いても同じ。元のベクトルの平均値と標準偏差も出力する。
scale(x)

正規分布で確率を求める

μ = 1、σ = 3 の正規分布で、x の値が 0 から-1 の間の面積
pnorm(0, 1, 3) - pnorm(-1, 1, 3)

標準化されていれば、平均や標準偏差を指定しなくても良い
pnorm(-1/3) - pnorm(-2/3)

分散を求める

分散は var( )関数で求められる

v = rnorm(20, 4, 2)
sd(v)
sqrt(var(v))

各値の平均からの距離を平均してしまうとゼロになる
mean(v - mean(v))

距離を二乗して、データ数から 1 を引いた数で割ったのが分散
sum((v - mean(v))^2)/(length(v) - 1)

3.3.2 The *t*, *F*, and χ^2 distributions

正規分布のほかによく出てくる分布が t 分布、F 分布、カイ二乗分布

<u>t 分布について</u>

- *t* 分布は正規分布に似ていて、自由度（df: degree of freedom）が 1 つある
- 自由度によって分布の形（特に両端）が異なる
- 自由度が上がれば上がるほど、標準正規分布に近くなる
- *t* 分布関連の関数として dt( )、pt( )、 qt( )、rt( )が用意されている
- 下記の値を比較してみると、標準正規分布では 1%水準で有意な値も *t* 分布上では有意でなくなる

pnorm(-3, 0, 1)
pt(-3, 2)

<u>F 分布について</u>

- F 分布は自由度が 2 つ
- 分子となる自由度が分母となる自由度よりどの程度大きいかが検定の際に重要となる
- 下記を比較するとその意味がわかる

1 - pf(6, 1, 1)
1 - pf(6, 20, 8)

<u>カイ二乗分布について</u>

- カイ二乗分布は自由度が 1 つ
- contingency table 内の値が同一かどうかはカイ二乗分布を利用したカイ二乗検定で確かめられる
- 下記を比較すると、p 値は自由度によることがわかる

1 - pchisq(4, 1)
1 - pchisq(4, 5)
1 - pchisq(4, 10)