

```
#####  
##### (5週目) 正規分布とt検定 #####  
#####
```

```
#-----  
# ∞ 5.1 正規分布  
#-----
```

```
par(cex=2) #フォントサイズを挙げる (デフォルトの2倍)
```

```
# [関数] curve(xを含む関数式, xmin,xmax)  
# [仕様] xを含む関数式のグラフをxminからxmaxの範囲で描画  
curve(x*x,-3,3)
```

```
# [関数] dnorm(x,a,b)  
# [仕様] 平均a,標準偏差bの正規分布の(x=x)における確率密度の値
```

```
# 平均170、標準偏差6の正規分布の概形 (17歳男性の身長近似)  
curve(dnorm(x,170,6),140,200)
```

```
xarray = x=seq(152,188,by=6)  
#[1] 152 158 164 170 176 182 188
```

```
# x=vの縦線を引く  
abline(v=xarray,lty=2) #lty=2 (点線)
```

```
options(digits=5) #5桁表示 (デフォルトは7桁)
```

```
# 確率密度 (平均値170cm付近に6.6%、10cm離れると1%を切る)  
data.frame(x=xarray,y=dnorm(xarray,170,6))
```

```
#   x           y  
#1 152 0.00073864  
#2 158 0.00899849  
#3 164 0.04032845  
#4 170 0.06649038  
#5 176 0.04032845  
#6 182 0.00899849  
#7 188 0.00073864
```

```
# 164cmから176cmの間に約68% (a-b~a+b)  
sum(dnorm(seq(164,176,by=0.1),170,6)/10) #[1] 0.68671
```

```
# 158cmから182cmの間に約95% (a-2b~a+2b)  
sum(dnorm(seq(158,182,by=0.1),170,6)/10) #[1] 0.95539
```

```
# 152cmから188cmの間に約99% (a-3b~a+3b)  
sum(dnorm(seq(152,188,by=0.1),170,6)/10) #[1] 0.99737
```

```
# 正規分布の性質  
# 平均値±標準偏差の範囲に約68% (偏差値40~偏差値60)  
# 平均値±2*標準偏差の範囲に約95% (偏差値30~偏差値70)  
# 平均値±3*標準偏差の範囲に約99.7% (偏差値20~偏差値80)
```

```
# a (平均値) と b (標準偏差) をどのような値に変えても成立します。
```

```
a = 70.0; b = 10.0;
sum(dnorm(seq(a-b,a+b,by=0.1),a,b)/10) #[1] 0.68511
sum(dnorm(seq(a-2*b,a+2*b,by=0.1),a,b)/10) #[1] 0.95594
sum(dnorm(seq(a-3*b,a+3*b,by=0.1),a,b)/10) #[1] 0.99734
```

これらの確率は、pnorm関数を使うと正確に確認することができます。

```
# [関数] pnorm(z得点)
# [仕様] : 正規分布における指定したz値
#         (平均値からのズレ/ 標準偏差)
#         の下側確率
```

```
pnorm(1) - pnorm(-1) #[1] 0.68269 (平均値±標準偏差)
pnorm(2) - pnorm(-2) #[1] 0.9549 (平均値±2*標準偏差)
pnorm(3) - pnorm(-3) #[1] 0.9973 (平均値±3*標準偏差)
```

とくに「平均値±2*標準偏差の範囲に約95%」は非常に重要な性質です。
(ひとまず偏差値30~70の間に95%の人が含まれる、と考えてください)

```
# (正確には、2ではなく1.96が係数となります。)
pnorm(1.96) - pnorm(-1.96) #[1] 0.95
```

次に、正規分布の確率分布から、有限個の標本を生成します。

```
# [関数] rnorm(n,a,b)
# [仕様] 平均a,標準偏差bの正規分布のサンプルをn個生成します。
```

```
options(digits=4) #4桁表示 (デフォルトは5桁)
a = 170; b = 6 # 平均170・標準偏差7
sum = 10 # 10個のサンプル
rnorm(sum,a,b)
#[1] 170.0 174.6 167.9 162.6 171.7 163.0 170.0 165.2 164.6 166.2
```

```
sum = 500000 # 500000個の標本のヒストグラム
h = hist(rnorm(sum,a,b),breaks=seq(140,200,by=1))
```

```
sum(h$counts[h$breaks>=164 & h$breaks<=176]) / sum #[1] 0.7188 (実行例)
sum(h$counts[h$breaks>=158 & h$breaks<=182]) / sum #[1] 0.9616 (実行例)
```

#おおよそ近似できていることがわかる。

```
#-----
# ∞ 5.2 標本平均と正規分布の関係：母集団がサイコロ
#-----
```

```
# サイコロベクトルの定義
dice=1:6
```

```
# 母集団である(無限)サイコロの平均値
mean(dice) #[1] 3.5
```

```

# 「サイコロをn回振って出た目の平均値」を求める関数
getMeanDice = function(n){
  sn = sample(dice,n,replace=T)
  mean(sn)
}

# 標本平均の分布：
# 「サイコロをn回振って出た目の平均値」 (1000試行)
h.2 = replicate(1000,getMeanDice(2))
h.4 = replicate(1000,getMeanDice(4))
h.6 = replicate(1000,getMeanDice(6))
h.8 = replicate(1000,getMeanDice(8))
h.10 = replicate(1000,getMeanDice(10))
h.100 = replicate(1000,getMeanDice(100))
h.1000 = replicate(1000,getMeanDice(1000))
h.10000 = replicate(1000,getMeanDice(10000))

# 「サイコロをn回振って出た目の平均値」の分布 (1000試行)！！
# ピークは移動せずも、幅が徐々に先鋭となることに注意！！

hist(h.2,breaks=seq(1,6,by=0.1))
hist(h.4,breaks=seq(1,6,by=0.1))
hist(h.6,breaks=seq(1,6,by=0.1))
hist(h.8,breaks=seq(1,6,by=0.1))
hist(h.10,breaks=seq(1,6,by=0.1))
hist(h.100,breaks=seq(1,6,by=0.05))
hist(h.1000,breaks=seq(1,6,by=0.01))
hist(h.10000,breaks=seq(1,6,by=0.001))

# [大数の法則]
# サンプルサイズ n を大きくしていくと、
# 標本平均 (h.nの平均) は、母集団 (mean(dice) = 3.5) に限りなく近づく

mean(h.2);mean(h.4);mean(h.6);mean(h.8);
#[1] 3.4445 [1] 3.48675 [1] 3.483833 [1] 3.5215
mean(h.10);mean(h.100);mean(h.1000);mean(h.10000);
#[1] 3.4775 [1] 3.49783 [1] 3.500009 [1] 3.500087

# [中心極限定理]
# サンプルサイズ n が大きいとき
# 標本平均分布の標準偏差は以下で近似できる
# 「母集団の標準偏差」 / sqrt(n)

# 要するに、サンプルサイズが n 倍となると
# グラフの幅は (1/√n) だけ縮小する。

# 標準偏差は√不偏分散
# 不偏分散は「平均からの差の2乗和」 / 「要素数-1」
# 不偏分散はvar(ベクトル)で一発で計算できます。

# [関数] var(標本ベクトル)
# [仕様] 不偏分散 (「平均からの差の2乗和」 / 「要素数-1」) を返す

# サンプルサイズが4倍になると、標準偏差が1/√4 (半分) となることが確認できる

```

```

sqrt(var(h.2));sqrt(var(h.4));sqrt(var(h.6));sqrt(var(h.8));
#[1] 1.214761 [1] 0.8135101 [1] 0.6645694 [1] 0.6132965

# サンプルサイズが100倍になると、標準偏差が $1/\sqrt{100}$  (1/10) となることが確認できる
sqrt(var(h.10));sqrt(var(h.100));sqrt(var(h.1000));sqrt(var(h.10000));
#[1] 0.5599603 [1] 0.1672452 [1] 0.05178183 [1] 0.0167836

# このサンプルサイズnを大きくしていくと
# 標本平均分布は「正規分布」に収束していく！！

hist(h.100,breaks=seq(1,6,by=0.05),xlim = c(3.0,4.0))
hist(h.1000,breaks=seq(1,6,by=0.01),xlim = c(3.3,3.7))
hist(h.10000,breaks=seq(1,6,by=0.001),xlim = c(3.4,3.6))

# 「正規分布」とは、任意の分布の標本平均分布と関係しています。
# もともとの分布が正規分布である必要はありません。
# あらゆる分布の（標本平均としての）メタ分布が「正規分布」なのです。
# この意味をよくよく噛み締めてください。

# 正規分布には、
# 平均値から±標準偏差の区域におよそ68%、
# 平均値から±2*標準偏差の区域におよそ95%
# の値が含まれるという特徴がありました。
# これを確認しましょう。

# サンプルサイズ10000の標本誤差平均「h.10000」を取り上げます。

# h.10000自体は1000のベクトルであることに注意してください。
h.10000

hlist = hist(h.10000,breaks=seq(1,6,by=0.001),xlim = c(3.4,3.6))

# ここでは母集団の平均値と標準偏差は未知であるとして、
# 10000個の標本から直接に推定します。

# h.10000の平均値および標準偏差
a = mean(h.10000); b = sqrt(var(h.10000));
#[1] 3.500438 #[1] 0.0167836

#サンプル数の総和は1000です。
sum(hlist$counts) #[1] 1000

# 平均値±標準偏差の区域の出現数のベクトル
hlist$counts[hlist$breaks>a-b & hlist$breaks<a+b]

# 平均値±標準偏差の区域の頻度総和
sum(hlist$counts[hlist$breaks>a-b & hlist$breaks<a+b])
#[1] 686 (68.7%)
sum(hlist$counts[hlist$breaks>a-2*b & hlist$breaks<a+2*b])
#[1] 952 (95.2%)

#およそ正規分布の性質を引き継いでいることがわかります。

#-----
# ∞ 5.3 標本平均と正規分布の関係：母集団が正規分布（成人男性の身長）
#-----

```

```

# 今度は、母集団が男性の身長の場合を検討します。
# a: 平均170cm、b: 標準偏差6cmとします。
a = 170; b = 6

# [関数] rnorm(n,a,b)
# [仕様] 平均a, 標準偏差bの正規分布のサンプルをn個生成します。

# n=100の標本平均の分布 (10000個)
h.100 = replicate(10000, mean(rnorm(100, a, b)))
hist(h.100, breaks=seq(165, 175, by=0.1))

# n=10000の標本平均の分布 (10000個)
h.10000 = replicate(10000, mean(rnorm(10000, a, b)))
hist(h.10000, breaks=seq(165, 175, by=0.01))

# ところで、h.100, h.10000の標準偏差の理論値はそれぞれ
#  $b/\sqrt{100} = b/10 = 0.6$ 
#  $b/\sqrt{10000} = b/100 = 0.06$ 

# 実測値はこちら (理論値とほぼ一致していることがわかります)
sqrt(var(h.100)) #0.6066
sqrt(var(h.10000)) #0.06066

#-----
# 〇〇〇 5.4 標本誤差と信頼区間
#-----

# [標本誤差分布]-----
# 標本誤差とは、母集団の平均値と標本の平均値の誤差を指し、
# 標本誤差分布とは、この標本誤差の分布を指します。

# 平均170、標準偏差6の正規分布から、サイズ100の標本を抽出
a = 170; b = 6; n = 100
h.100 = replicate(10000, mean(rnorm(n, a, b))) #10000の標本平均を集めます

# 標本平均の分布 (n=100)
hist(h.100, breaks=seq(165, 175, by=0.1))

# 標本誤差は標本平均から、母集団の平均を引いたものです。
hist(h.100-a, breaks=seq(-5, 5, by=0.1))

# 標本誤差分布の平均はゼロとなり、
# グラフの幅は母平均の標準偏差より狭くなっていることがわかります。

# 標本誤差分布は、標本平均分布と同様に、
# 標本数を増やしていくと、正規分布に近づきます。

# また、標本誤差分布の標準偏差は、
# 標本平均分布の標準偏差と同様に以下で計算されます。
#  $b / \sqrt{n} \mid (\text{母集団の標準偏差}) / \sqrt{\text{標本サイズ}}$ 

b / sqrt(100) #[1] 0.6

```

```
# 母集団が未知の場合、標本誤差分布の標準偏差は、
# bを（特定の）標本の不偏標準偏差sに変えて計算します。
```

```
s = sqrt(var(rnorm(100,a,b))); s/sqrt(100) #[1] 0.6403
s = sqrt(var(rnorm(100,a,b))); s/sqrt(100) #[1] 0.5449
s = sqrt(var(rnorm(100,a,b))); s/sqrt(100) #[1] 0.5801
```

```
# ばらつきはありますが、おおよそいい推定ができています。
```

```
# [信頼区間]-----
```

```
# 標本サイズ n の標本誤差分布（標本平均の母平均からのズレ）は、
# 平均ゼロ、標準偏差  $s/\sqrt{n}$ の正規分布で（大まかに）近似される。
# つまり、標本誤差の絶対値の約95%は、「 $2*s/\sqrt{n}$ 」以下に含まれる。
```

```
# これを逆から言えば、標本平均から±「 $2*s/\sqrt{n}$ 」の範囲内に、
# 95%の確率で母平均が存在することになる。
```

```
# 95%信頼区間とは
# 「95%の確率で、この区間が母集団平均を含んでいる」と同義。
```

```
# 統計の重要な仕事の一つは、未知の母平均を推定することです。
# すなわち「95%信頼区間」を母平均の推定にとって非常に重要な情報。
```

```
# [信頼区間の具体例]-----
```

```
# 以下に、成人男性の10人分の身長データがあります。
# 179,176,166,167,170,164,170,154,169,164
# このとき、成人男性の身長の母平均が、
# 95%の確率で含まれる領域（95%信頼区間）を特定してください。
```

```
n = 10 # サンプルサイズ
x = c(179,176,166,167,170,164,170,154,169,164) # 標本
xmean = mean(x) #[1] 167.9 （標本平均）
s = sqrt(var(x)) #[1] 6.887 （不偏標準偏差 | 標本の散らばり具合）
se = s / sqrt(n) #[1] 2.178 （標準誤差 | 標本平均の真値からの散らばり具合）
```

```
#95%信頼区間は（およそ）以下の163.5~172.3の領域にある。
xmin = xmean - 2.0*se #[1] 163.5
xmax = xmean + 2.0*se #[1] 172.3
```

```
# つまり
# 成人男性の身長の母平均は、95%の確率で
# 163.5cmから172.3cmに含まれる（と近似されます）
```

```
# （注意）この範囲は厳密に言うと正しくありません。
# なぜなら、（上記のケースでの）標本誤差の分布は、
# 正規分布よりも少し広がるからです。
# 実は、この際に使われるべき分布は
# 正規分布ではなく<t分布>となります。
```

```

#-----
# ooooo 5.5 t分布
#-----

# (復習)
# 母集団の分布の形状にかかわらず、
# 標本平均分布は (標本サイズ n が増えると!!) 正規分布に近づきます。
# 逆に言えば、母集団の分布の形状によっては、
# 標本サイズ n が小さいと、
# 標本平均分布の形は正規分布とは大きく異なることがあります。

# 母集団の分布が正規分布のとき、
# <標本誤差>を<標準誤差>で割った値の分布としての
# 標本平均分布は (標本サイズ n にかかわらず!)
# 「正規分布と良く似た形の分布」 (平均0・標準偏差1) となります。
# この「正規分布と良く似た形の分布」のことを<t分布>と呼びます。
# (当然、t分布は、nが大きくなると正規分布に近づきます。 )

# t分布は正規分布よりも裾野がわずかに広がります。
# この裾野の広がり は n が小さいときにより大きくなります。

# 例えばn=10のときのt分布の95%信頼区間は
# 標準誤差の±2.26個分の区間になります。

# 95%信頼区間の下限と上限の正確な値は以下の通りです。
xmin = xmean - 2.26*se #[1] 163
xmax = xmean + 2.26*se #[1] 172.8

# [関数] dt(x,df=m)
# [仕様] 自由度m (=n-1) のt分布の、xにおける確率密度を返します。
m = 10-1
curve(dt(x,df=m),-4,4,lty=1)

# t分布と標準正規分布の形状を比べてみます。
curve(dt(x,df=3),-4,4,lty=1) #ltyは線のタイプ (1は実線)
#add=Tとすると重ね書きします。
curve(dt(x,df=6),-4,4,lty=1,add=T)
curve(dt(x,df=9),-4,4,lty=1,add=T)
curve(dt(x,df=12),-4,4,lty=1,add=T)
curve(dnorm(x,0,1),-4,4,lty=2,add=T) #lty=2:点線

# nが小さいとピークがつぶれ裾野が広がることがわかります。

# x=-∞からの累積確率がpの確率点を求めるには[qt]を使います。
# [関数] qt(p,df=m)
# [仕様] 自由度m (=n-1) のt分布における、累積確率がpの時のt値を返します。

# 95%信頼区間の場合、左側の累積確率は0.025となります。
# 実際、先ほどの理論値と一致しています。
qt(0.025,df=9) #[1] -2.262
abline(v=qt(0.025,df=9))
abline(v=qt(0.975,df=9))

# nが大きくなると正規分布の95%区間 (±1.96) とほぼ一致します。
qt(0.025,df=1000) #[1] -1.962

```

```

# 自由度が小さくなると、95%信頼区間は広がります。
qt(0.025,df=3) #[1] -3.182

# これは、標本数が少なくなると、
# 母平均の推定範囲がぼやけることと対応します。

# p=0.5となるt値は常にゼロです（グラフの対称性より明らか）。
qt(0.5,df=9) #[1] 0

# 逆に検定量のt値がわかっている場合、
# 下側確率は以下で求めることができます。

# [関数] pt(t0, df=m, lower.tail=TRUE/FALSE)
# [仕様] 自由度m (=n-1) のt分布における、検定量t0に対する
#         下側確率P(t<t0) : lower.tail=T (デフォルト) または
#         上側確率P(t>t0) : lower.tail=F

pt(-2.262,df=9) #[1] 0.02501
pt(2.262,df=9)  #[1] 0.975
pt(2.262,df=9,lower.tail=F) #[1] 0.02501 (上側確率)

#-----
# ○○○○○ 5.6 1標本のt検定
#-----

# アメリカ人の成人男性の平均身長は175cm（2001年調べ）です。
# 日本人の成人男子の身長が、
# アメリカ人の平均身長と比較して違いがあるかを検定します。

# [帰無仮説] 日本人の成人男性の母集団平均は175cmである。
# [対立仮説1] 日本人の成人男性の母集団平均は175cmではない。（両側検定）
# [対立仮説2] 日本人の成人男性の母集団平均は175cmより低い。（下側片側検定）

# これを以下の2つのサンプルで考えましょう。

n = 10 # サンプルサイズ
x1 = c(179,176,166,167,170,164,170,154,169,164) # 標本1
x2 = c(153,176,168,167,150,153,170,154,195,191) # 標本2

#箱ひげ図のplot
boxplot(x1,x2)

# 標本平均 (xmean1 > xmean2)
xmean1 = mean(x1) #167.9
xmean2 = mean(x2) #167.7

# 標本の不偏標準偏差 (s1 < s2)
s1 = sqrt(var(x1)) #6.887
s2 = sqrt(var(x2)) #15.97

# サンプル1のt値は

```



```

t1 = (xmean1 - 175) / (s1/sqrt(n)) #[1] -3.26
# サンプル2のt値は
t2 = (xmean2 - 175) / (s2/sqrt(n)) #[1] -1.445

# [サンプル1のt検定]-----

# t<t1となる確率p1は (p値) ?
p1 = pt(t1,df=9) #[1] 0.004919 (下側約0.5%の領域)

# 母平均が175cmのときに、
# サンプル1以上に下側に偏る確率は0.492% (片側検定)
# サンプル1以上のズレが生じる確率は0.984% (両側検定)

# これらはいずれについても5%を下回る。よって、

# 帰無仮説
# 「(日本人成人男性の) 母集団平均は175cm」である
# は棄却され、
# 対立仮説1 「(日本人成人男性の) 母集団平均は175cmではない」または
# 対立仮説2 「(日本人成人男性の) 母集団平均は175cmより低い」が
# 採用される。

# 統計的には、日本人の成人男性の身長は、
# アメリカ人よりも有意に低いと結論できる。

# [サンプル2のt検定]-----

# t<t2となる確率p2 (p値) は?
p2 = pt(t2,df=9) #[1] 0.091

# 母平均が175cmのときに、

# サンプル1ほどに下側に偏る確率は9.1% (片側検定)
# サンプル1ほどにズレが生じる確率は18.2% (両側検定)

# いずれについても有意水準の5%を上回る。よって
# 帰無仮説
# 「(日本人成人男性の) 母集団平均は175cm」は支持される。

# 統計的には、日本人の成人男性とアメリカ人の成人男性との間に
# 有意な差は存在しない。

# 自由度9のt分布上で、2つのサンプルのt値がどこにあるかを確認します。

#まずt分布を描画したのちに
curve(dt(x,df=9),-4,4,lty=1)
#t=t1とt=t2に線を引きます。| abline(縦線v=x、横線h=y)
abline(v=t1,lty=2);abline(v=t2,lty=2);

#! 標本2と標本1では平均値が標本1の方が高いにもかかわらず
#! 標本1でのみ有意な差が検出されていることに注意。
#! 標本2はデータの散らばりが大きいため、
#! 標準偏差で割って標準得点化されるt値では、
#! 元々の平均差が低く査定される。

```

```
# [Rの関数を用いる]-----
```

```
# t.test関数を用いれば、  
# 以上の計算を自動的に行ってくれます。
```

```
# [関数] t.test(標本ベクトル, mu=比較対象の母平均, alternative=片側or両側)  
# [仕様] 標本ベクトルと比較対象の母平均とを比較する1標本t検定  
#       alternative="two.sided" (両側検定)  
#       alternative="less" (片側検定：下側確率)  
#       alternative="greater" (片側検定：上側確率)
```

```
# サンプル1のt検定 (下側確率)  
t.test(x1,mu=175,alternative="less")  
#data: x1  
#t = -3.3, df = 9, p-value = 0.005  
#alternative hypothesis: true mean is less than 175  
#95 percent confidence interval:  
# -Inf 171.9  
#sample estimates:  
#mean of x  
#167.9
```

```
# サンプル1のt検定 (両側確率)  
t.test(x1,mu=175,alternative="two.sided")  
#t = -3.3, df = 9, p-value = 0.01  
#alternative hypothesis: true mean is not equal to 175  
#95 percent confidence interval:  
# 163.0 172.8  
#sample estimates:  
#mean of x  
# 167.9
```

```
# サンプル2のt検定 (下側確率)  
t.test(x2,mu=175,alternative="less")  
#data: x2  
#t = -1.4, df = 9, p-value = 0.09  
#alternative hypothesis: true mean is less than 175  
#95 percent confidence interval:  
# -Inf 177  
#sample estimates:  
# mean of x  
# 167.7
```

```
# サンプル2のt検定 (両側確率)  
t.test(x2,mu=175,alternative="two.sided")  
#data: x2  
#t = -1.4, df = 9, p-value = 0.2  
#alternative hypothesis: true mean is not equal to 175  
#95 percent confidence interval:  
# 156.3 179.1  
#sample estimates:  
# mean of x
```

```
# 167.7
```

```
#-----
```

```
# ○○○○○○ 5.7 2標本のt検定 (対応あり)
```

```
#-----
```

```
# 同じ被験者のサプリメント
```

```
# 接種前 (pre) と接種後 (post) の体重の変化
```

```
subj = LETTERS[c(1:8)] #LETTERSは大文字のアルファベットのベクトル
```

```
pre = c(95,80,80,85,75,75,80,85)
```

```
post = c(90,75,75,75,80,65,75,80)
```

```
data = data.frame(subj=subj,pre=pre,post=post,change=post-pre)
```

```
#  subj pre post change
```

```
#1  A  95  90    -5
```

```
#2  B  80  75    -5
```

```
#3  C  80  75    -5
```

```
#4  D  85  75   -10
```

```
#5  E  75  80     5
```

```
#6  F  75  65   -10
```

```
#7  G  80  75    -5
```

```
#8  H  85  80    -5
```

```
#箱ひげ図
```

```
boxplot(data$pre,data$post)
```

```
# サプリメント接種前と接種後で体重に変化があるか (両側検定)
```

```
# あるいは体重は減少したか (片側検定: 下側)
```

```
# 実は、対応のあるt検定は、
```

```
# 少しの操作で1標本t検定に変換できる。
```

```
# なぜなら、
```

```
# 帰無仮説「標本Aと標本Bに差はない」は
```

```
# 帰無仮説「標本 (A-B) の母集団平均は0と等しい」
```

```
# と同義であるためである。
```

```
#片側検定
```

```
t.test(change,dm=0,alternative="less")
```

```
#data:  change
```

```
#t = -3.1, df = 7, p-value = 0.009
```

```
#alternative hypothesis: true mean is less than 0
```

```
#95 percent confidence interval:
```

```
# -Inf -1.899
```

```
#sample estimates:
```

```
#mean of x
```

```
#      -5
```

```
# change (接種後-接種前) の母平均が0であると仮定したときに
```

```
# changeが、標本以上の減少を示す確率は0.9%
```

```
# 母平均を0とする仮説は棄却され
```

```
# 対立仮説 (片側検定) である「changeは0より小さい」が採用される。
```

```

# ちなみに、t.testに
# 比較対象の2つのベクトルを引数として入れることもできます。
# この場合、paired=TRUEを指定する必要があります。

t.test(data$post,data$pre,paired=TRUE,alternative="less")
#data: data$post and data$pre
#t = -3.1, df = 7, p-value = 0.009
#alternative hypothesis: true mean is less than 0
#95 percent confidence interval:
# -Inf -1.899
#sample estimates:
#mean of x
#      -5

#-----
# ○○○○○○○ 5.8 宿題
#-----

# 以下から、奇数偶数の好み実験結果の一部のデータをインポートしてください。
# http://lab.kenrikodaka.com/\_download/class/2022\_AppliedMedia/oddeven.csv

#Environmentパネルの「Import Dataset」から
#oddeven.csvのデータを、expdatという変数に読み込みます。

# 数値をFactorに変換します。

expdat$gender = factor(expdat$sex,levels=0:1,labels=c("FEMALE","MALE"))
expdat$domhand =
factor(expdat$domhand,levels=0:1,labels=c("RIGHT","LEFT"))
expdat$preference =
factor(expdat$preference,levels=0:1,labels=c("EVEN","ODD"))

#女性の奇数と偶数の好みのベクトル (573人分)
female.preference = expdat$preference[expdat$gender=="FEMALE"]

# 偶数好きは370人 : 64.5% (370/573) が偶数好き
sum(female.preference == "EVEN")/ 573

# 573人からランダムに100人を取り出して、偶数の数をカウントする関数
countFemaleEven = function(){
  count = 0
  for(i in 1:100){
    r = sample(573,1)
    if(female.preference[r] == "EVEN"){
      count = count+1
    }
  }
  count
}

# サイズ10 (n=10) の標本
sample = replicate(10,countFemaleEven())
#[1] 72 58 59 65 53 68 68 71 72 64

```

```
# 標本平均  
mean(sample) #[1] 65.8
```

```
# countFemaleEvenについて  
# 2つ以上のサイズ（例えば、n=5, n=10）の標本平均を  
# 繰り返し（1000回から10000回程度）計算し、  
# サイズ毎の標本平均の分布を可視化してください。
```

```
# このとき、標本平均のピークが、  
# 母集団の平均64.5付近にあるかを確認してください。
```

```
# さらに、標本サイズ n が異なると、  
# 標本平均の分布がどのように変わるかを考察してください。
```

```
#作成したプログラムは、指定のリンク先に提出してください。  
#ファイル名は「Class1027_Work_1950XX.R」をお願いします。  
#（注）このファイルを上書き保存したものを提出すればよいです。
```