村上　明
pp. 114-116

4.6 A note on statistical significance

有意差とは
○　差が偶然ではない
×　意味のある差

以下、それを例証する
100 個の y = 0.3x 上データに大きなノイズを発生させる

```
n = 100
(x = seq(1, 100, length = n))
y = 0.3 * x + rnorm(n, 0, 80)
```

有意差はない

```
model100 = lm(y ~ x)
summary(model100)$coef
```

→noise が大きすぎて差がなくなってしまった

今度はデータ数を増やしてみる

```
n = 1000
(x = seq(1, 100, length = n))
y = 0.3 * x + rnorm(n, 0, 80)
model1000 = lm(y ~ x)
summary(model1000) $coef
```

→すると有意差ありとなった

しかし、x から y を予測するのは不可能であることが、散布図を描いてみるとわかる

```
plot(x,y)
abline(lm(y ~ x))
```

また、r^2 の値が小さい（1%）ことからもそれがわかる

⇒よって差の大きさを見るには p 値は不適切

少なくとも信頼区間を見るべき
上記の例では信頼区間は confint 関数で求められる

```
confint(model100)
confint(model1000)
```

後者の方が 0.3 に近いので情報量が多いと言える

0.3 の傾きが意味あることなのかどうかも分野の背景知識等による
r^2の値が低くても理論的に意味があることもある