

## 相関と回帰式

### 相関係数

```
> cor(Type, Token)
[1] 0.889026
```

詳しくは、[Type と Token](#) 参照

### 相関関係の検定

- ・ p 値は、相関が無いのに有ると間違える確率。
  - ・ p 値が高ければ、間違っている確率が高くなる。

### 相関関係の散布図と回帰直線

- ・ Token が x 軸、Type が y 軸
- ・  $y=ax+b$  の回帰直線

```
> lm(Type ~ Token)
```

```
Call:
lm(formula = Type ~ Token)
```

```
Coefficients:
(Intercept)      Token
    43.587         0.298
```

```
> plot(Token, Type)
```

- ・ x と y が入れ替わる

```
> abline(lm(Type ~ Token))
```

### 95%の予測区間

- ・ 予測 predict() する。区間 interval を「予測」に

```
> pred.95 <- predict(lm(Type ~ Token), interval = "prediction")
```

- ・ 結果を保存した pred.95 のデータをデータフレーム型に変更

```
> pred.95 <- as.data.frame(pred.95)
```

- ・ データをプロット

```
> plot(Token, Type)
```

- ・ フィット（回帰直線）を黒で描く

```
> lines(Token, pred.95$fit, col = "black")
```

- 上限値を赤で描く

```
> lines(Token, pred.95$upr, col = "red")
```

- 下限値を青で描く

```
> lines(Token, pred.95$lwr, col = "blue")
```