迴歸分析 ：Correlations

- Find 10 of your friends. Collect their information (Weight, Hight).
- We conjecture that hight is a good predictor of weight, i.e.

$$\text{weight}_i = \beta_0 + \beta_1 \text{hight}_i + \varepsilon_i$$

- Use the following way to examine the conjecture:
  - Tabulate your collected data.
  - Estimate β_0 and β_1 and their variances.
  - t-test for $H_0 : \beta_1 = 0$
  - ANOVA test for $H_0 : \beta_1 = 0$
  - t-test for zero Peason's correlation $H_0 : \rho = 0$
  - Confidence Interval for Peason's correlation
  - Calculate the Spearman's (rank) correlation coefficient (r_S)
  - Zero correlation test based on Spearman's correlation r_S

1. Tabulate collected data.

身高	體重
190	90
170	65
160	45
175	60
182	66
172	53
176	64
166	44
158	50
172	70

2. Estimate $\hat{\beta}_0$ and β_1 and their variances.

```
> n <- 10;
> x <- c(190, 170, 160, 175, 182, 172, 176, 166, 158, 172)
> y <- c(90, 65, 45, 60, 66, 53, 64, 44, 50, 70)
> lm.fit <- lm(y~x); summary(lm.fit)

Call:
lm(formula = y ~ x)

Residuals:
   Min     1Q Median     3Q    Max
-9.184 -6.242 -1.148  6.834  9.423

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -151.3470    44.5817  -3.395  0.00943 **
x              1.2321     0.2587   4.763  0.00142 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.448 on 8 degrees of freedom
Multiple R-squared:  0.7393,    Adjusted R-squared:  0.7067
F-statistic: 22.69 on 1 and 8 DF,  p-value: 0.001421
```

Beta0 的估計值

Beta1 的估計值

Beta0 的變異數

44.5817^2 = 1990.709

Beta1 的變異數

0.2587^2 = 0.0669

3. t-test for $H_0 : \beta_1 = 0$

```
> n <- 10;
> x <- c(190, 170, 160, 175, 182, 172, 176, 166, 158, 172)
> y <- c(90, 65, 45, 60, 66, 53, 64, 44, 50, 70)
> lm.fit <- lm(y~x); summary(lm.fit)

Call:
lm(formula = y ~ x)

Residuals:
   Min     1Q Median     3Q    Max
-9.184 -6.242 -1.148  6.834  9.423

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -151.3470    44.5817  -3.395  0.00943 **
x              1.2321     0.2587   4.763  0.00142 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.448 on 8 degrees of freedom
Multiple R-squared:  0.7393,    Adjusted R-squared:  0.7067
F-statistic: 22.69 on 1 and 8 DF,  p-value: 0.001421
```

由於 p-value（0.00142）小於顯著性水準（設定為 0.05），因此可以拒絕虛無假設，即斜率 beta1 顯著不等於 0。這也代表 x（自變量）對 y（因變量）的影響是顯著的。

4. ANOVA test for $H_0 : \beta_1 = 0$

```
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
x          1 1258.36 1258.36  22.686 0.001421 **
Residuals  8  443.74   55.47
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

F value = 22.686,

P 值 = 0.001421 <0.05(顯著水準),

因此可以拒絕虛無假設，即斜率 beta1 顯著不等於 0

5. *t*-test for zero Peason's correlation $H_0 : \rho = 0$

```
> cor.test(x, y, method = "pearson")

        Pearson's product-moment correlation

data:  x and y
t = 4.763, df = 8, p-value = 0.001421
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5019232 0.9663173
sample estimates:
      cor
0.8598243
```

p-value =0.001421，小於一般的顯著性水準（0.05）。因此可以拒絕虛無假設 H0: $\rho=0$, ρ 顯著不等於 0.

6. Confidence Interval for Peason's correlation

```
> cor.test(x, y, method = "pearson")$conf.int
[1] 0.5019232 0.9663173
attr(,"conf.level")
[1] 0.95
```

→信賴區間的下界是 0.5019232，上界是 0.9663173，95%的信心水準.

7. Calculate the Spearman's (rank) correlation coefficient (r_S)

```
> cor.test(x, y, method="spearman")

        Spearman's rank correlation rho

data:  x and y
S = 40.623, p-value = 0.01179
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.7538029
```

Spearman 的相關係數被稱為 rho（ρ, r_s）。它的估計值為 0.7538029，表示變數 x 和 y 之間存在一個強烈的單調相關關係。

8. Zero correlation test based on Spearman's correlation r_S

```
> cor.test(x, y, method="spearman")

        Spearman's rank correlation rho

data:  x and y
S = 40.623, p-value = 0.01179
alternative hypothesis: true rho
sample estimates:
      rho
0.7538029
```

虛無假設 $H_0 : \rho_s = 0$

p-value 為 0.01179 顯著小於顯著性水準（0.05）。這個結果表明在顯著性水準為 0.05 的情況下，有足夠的證據拒絕虛無假設，認為變數之間存在著顯著的 Spearman's rank 相關性。