Rays R&D

Kenji Kajita

2024-10-13

```r
data <- read_csv("/Users/kenji/Documents/GitHub/Rays R&D/rays_questionnaire_dataset_2022.csv")
```

```
## Rows: 200 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (3): predictor_A, predictor_B, target_y
## 
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```r
sum(is.na(data$predictor_A))
```

```
## [1] 9
```

```r
summary(lm(target_y ~ predictor_A, data = data))
```

```
## 
## Call:
## lm(formula = target_y ~ predictor_A, data = data)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -50.890  -5.699   0.140   4.616  62.657 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  -10.752      1.049 -10.253   <2e-16 ***
## predictor_A    1.705      1.006   1.695   0.0918 .  
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
## 
## Residual standard error: 13.99 on 176 degrees of freedom
##   (22 observations deleted due to missingness)
## Multiple R-squared:  0.01607,    Adjusted R-squared:  0.01048 
## F-statistic: 2.874 on 1 and 176 DF,  p-value: 0.09179
```

```r
summary(lm(target_y ~ predictor_A + predictor_B, data = data))
```

```
##
## Call:
## lm(formula = target_y ~ predictor_A + predictor_B, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -47.597  -5.644   0.008   4.395  60.238
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -10.7136     1.1347  -9.442   <2e-16 ***
## predictor_A   1.7047     1.0466   1.629    0.105
## predictor_B   0.3679     0.3578   1.028    0.305
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 14.11 on 163 degrees of freedom
##   (34 observations deleted due to missingness)
## Multiple R-squared:  0.02188,    Adjusted R-squared:  0.009883
## F-statistic: 1.823 on 2 and 163 DF,  p-value: 0.1647
```

y = -10.7136 + 1.7047 x A + 0.3679 x B

```
result_1 <- -10.7136 + 1.7047 * 1.2 + 0.3679 * -2.1
result_1
```

```
## [1] -9.44055
```

To get Pr(>|t|) we need to get the t-value and p-value for each variable. To get the p-value we do:

```
## p-value = p-value = 2 * pt(abs(t value), residual df, lower.tail = FALSE)

P_1 <- 2 * pt(abs(0.377), 995, lower.tail = FALSE)
P_2 <- 2 * pt(abs(5.888), 995, lower.tail = FALSE)
P_3 <- 2 * pt(abs(9.923), 995, lower.tail = FALSE)
P_4 <- 2 * pt(abs(3.345), 995, lower.tail = FALSE)

P_1
```

```
## [1] 0.706254
```

```
P_2
```

```
## [1] 5.340641e-09
```

```
P_3
```

```
## [1] 3.409977e-22
```

```
P_4
```

```
## [1] 0.0008535592
```

```
#Winston_Walker <- P(H2/E) = [P(E/H2) x P(H2)]  / [P(E/H1) x P(H1)  + P(E/H2) x P(H2)]
# P(H1) = 0.5
# P(H2) = 0.5
# P(E/H1) = 0.29
# P(E/H2) = 0.35

Winston_Walker <- (0.35 * 0.5)/(0.29*0.5 + 0.35*0.5)*100
Winston_Walker
```

```
## [1] 54.6875
```

```
#Confidence Interval

S.E_mean <- 10/sqrt(10)
S.E_mean
```

```
## [1] 3.162278
```

```
alpha = 0.05
degrees.freedom = 10 - 1
t.score = qt(p = alpha/2, df = degrees.freedom, lower.tail = F)
print(t.score)
```

```
## [1] 2.262157
```

```
margin.error <- t.score * S.E_mean
margin.error
```

```
## [1] 7.153569
```

```
sample.mean <- 95

lower.bound <- sample.mean - margin.error
upper.bound <- sample.mean + margin.error
print(c(lower.bound,upper.bound))
```

```
## [1]  87.84643 102.15357
```