Correlation bewteen Item difficulty and Point Biserial Correlation

Chang Lu

2025-02

In first part of the analysis, we will find whether there is correlation between r_pb and item difficulty.

```
## New names:
## * `` -> `...1`
## * `E.` -> `E....10`
## * `E.` -> `E....12`
```

```
# Compute item difficulty as the mean proportion of correct responses
item_difficulty <- df_students %>%
  summarise(across(Q1:Q20, ~ mean(as.numeric(.), na.rm = TRUE))) %>%
  pivot_longer(cols = everything(), names_to = "Item", values_to = "Item_Difficulty")

print(item_difficulty)
```

```
## # A tibble: 20 x 2
##    Item  Item_Difficulty
##    <chr>           <dbl>
##  1 Q1              0.548
##  2 Q2              0.323
##  3 Q3              0.290
##  4 Q4              0.355
##  5 Q5              0.419
##  6 Q6              0.710
##  7 Q7              0.323
##  8 Q8              0.484
##  9 Q9              0.484
## 10 Q10             0.258
## 11 Q11             0.161
## 12 Q12             0.484
## 13 Q13             0.194
## 14 Q14             0.581
## 15 Q15             0.484
## 16 Q16             0.613
## 17 Q17             0.645
## 18 Q18             0.419
## 19 Q19             0.581
## 20 Q20             0.645
```

```
# Compute point biserial correlation for each question using Pearson correlation
point_biserial <- map_df(paste0("Q", 1:20), function(q) {
  cor_value <- cor.test(as.numeric(df_students[[q]]), as.numeric(df_students$`Total Score`), method = "
```

```
  tibble(Item = q, Point_Biserial_Correlation = cor_value)
})

print(point_biserial)
```

```
## # A tibble: 20 x 2
##    Item  Point_Biserial_Correlation
##    <chr>                      <dbl>
##  1 Q1                         0.448
##  2 Q2                         0.439
##  3 Q3                        -0.236
##  4 Q4                         0.261
##  5 Q5                         0.307
##  6 Q6                         0.491
##  7 Q7                         0.706
##  8 Q8                         0.339
##  9 Q9                         0.464
## 10 Q10                       -0.163
## 11 Q11                        0.267
## 12 Q12                        0.250
## 13 Q13                        0.361
## 14 Q14                        0.614
## 15 Q15                        0.589
## 16 Q16                        0.604
## 17 Q17                        0.298
## 18 Q18                        0.614
## 19 Q19                        0.542
## 20 Q20                        0.354
```

The r_pb I calculated is equal to the original data client provided. Also,we can notice that there are some negative rpb, which suggests that students who performed well overall tended to get that specific question wrong more often than those who performed poorly.

Now we move to figure out correlation between them:

```
# Combine item difficulty and point biserial correlation data
combined_IDPB <- left_join(item_difficulty, point_biserial, by = "Item")

print(combined_IDPB)
```

```
## # A tibble: 20 x 3
##    Item  Item_Difficulty Point_Biserial_Correlation
##    <chr>           <dbl>                      <dbl>
##  1 Q1              0.548                      0.448
##  2 Q2              0.323                      0.439
##  3 Q3              0.290                     -0.236
##  4 Q4              0.355                      0.261
##  5 Q5              0.419                      0.307
##  6 Q6              0.710                      0.491
##  7 Q7              0.323                      0.706
##  8 Q8              0.484                      0.339
##  9 Q9              0.484                      0.464
## 10 Q10             0.258                     -0.163
```

```
## 11 Q11             0.161                     0.267
## 12 Q12             0.484                     0.250
## 13 Q13             0.194                     0.361
## 14 Q14             0.581                     0.614
## 15 Q15             0.484                     0.589
## 16 Q16             0.613                     0.604
## 17 Q17             0.645                     0.298
## 18 Q18             0.419                     0.614
## 19 Q19             0.581                     0.542
## 20 Q20             0.645                     0.354
```

1. Pearson Correlation

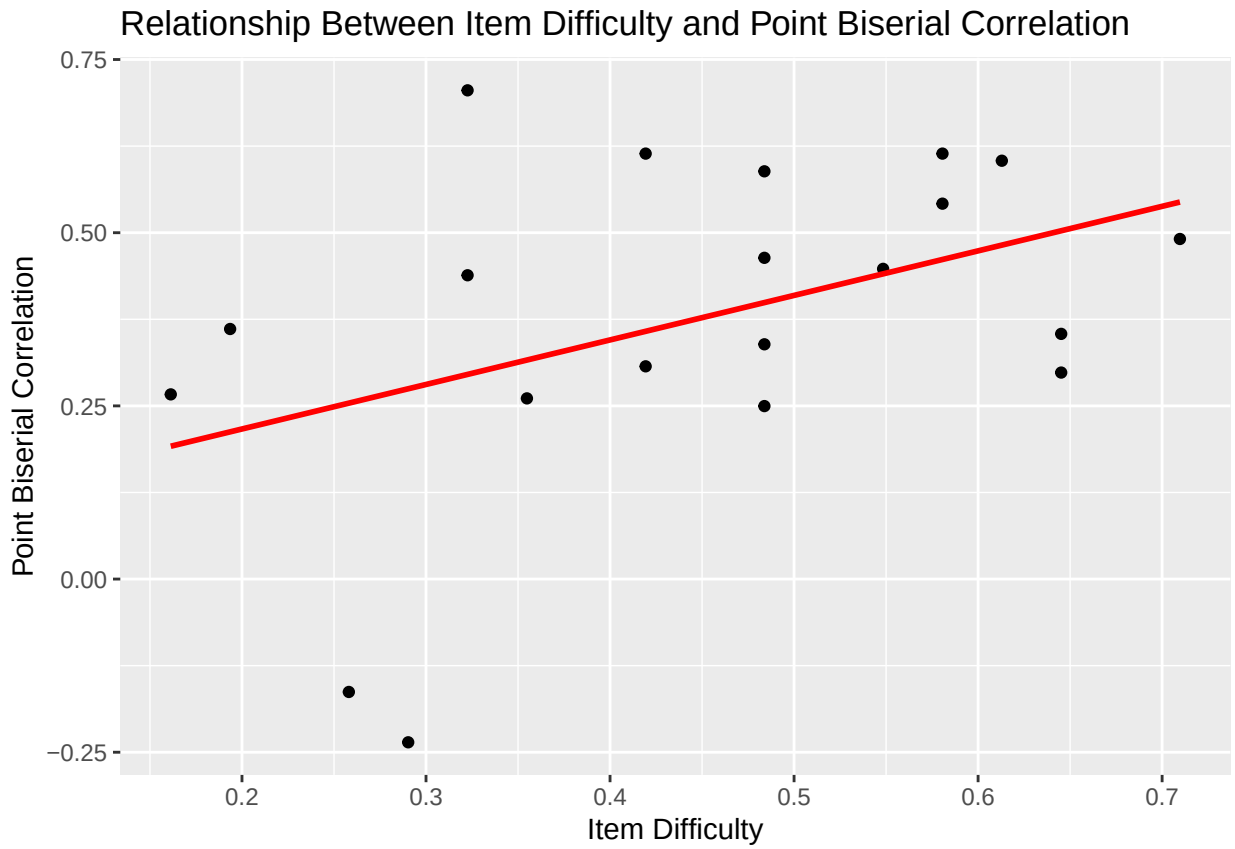
```r
correlation_result <- cor(combined_IDPB$Item_Difficulty,
                          combined_IDPB$Point_Biserial_Correlation,
                          use = "complete.obs")

print(paste("Pearson Correlation coef between Item Difficulty and Point Biserial Correlation:", round(c
```

```
## [1] "Pearson Correlation coef between Item Difficulty and Point Biserial Correlation: 0.423"
```

```r
ggplot(combined_IDPB, aes(x = Item_Difficulty, y = Point_Biserial_Correlation)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Relationship Between Item Difficulty and Point Biserial Correlation",
       x = "Item Difficulty",
       y = "Point Biserial Correlation")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



A Pearson correlation of 0.423 indicates a moderate positive relationship between item difficulty and point biserial correlation, which suggests that easier questions (higher item difficulty, >=0.8) tend to better discriminate between high and low performance students(higher point biserial correlation). But some outliers exist.

2. Spearman's Rank correlation

```
spearman_corr <- cor(combined_IDPB$Item_Difficulty,
                     combined_IDPB$Point_Biserial_Correlation,
                     method = "spearman", use = "complete.obs")
print(paste("Spearman's Rank Correlation coef:", round(spearman_corr, 3)))
```

```
## [1] "Spearman's Rank Correlation coef: 0.361"
```

3. Kendall Tau correlation

```
kendall_corr <- cor(combined_IDPB$Item_Difficulty,
                    combined_IDPB$Point_Biserial_Correlation,
                     method = "kendall")

print(paste("Kendall Correlation coef:", round(kendall_corr, 3)))
```

```
## [1] "Kendall Correlation coef: 0.244"
```

The relationship is weak to moderate (0.423 Pearson, 0.361 Spearman, 0.244 Kendall). In a word, some easy questions do have higher discrimination power. However, not all easy questions discriminate well, and medium-difficulty questions (0.4-0.6) tend to be the best discriminators.