Class Survey Project

All Members of Table 12

2025-10-09

Executive Summary

Data Description

Question of Interest

Basic Regression of the Class Survey Data

Here, we read in the standardized proficiency score dataset first.

```
csv_path <- "data/PSTAT197A_ Standardized_Proficiency_Score.csv"
df <- read_csv(csv_path, show_col_types = FALSE)

names(df)[names(df) == "Score_Avg (Difficulty of Courses)"] <- "Score_Avg"
```

```
# make sure columns are numeric
df <- df %>%
  mutate(
    Score_Avg        = as.numeric(Score_Avg),
    Prog_Proficiency = as.numeric(Prog_Proficiency),
    Stat_Proficiency = as.numeric(Stat_Proficiency),
    Math_Proficiency = as.numeric(Math_Proficiency)
  )

# define pairs and fit models
pairs <- tribble(
  ~y,                  ~x,
  "Prog_Proficiency",  "Score_Avg",
  "Stat_Proficiency",  "Score_Avg",
  "Math_Proficiency",  "Score_Avg"
)

fit_1 <- function(y, x, data) {
  sub <- data %>% select(all_of(c(y, x))) %>% drop_na()
  m   <- lm(reformulate(x, y), data = sub)
  list(
    glance = glance(m) %>% mutate(y = y, x = x, n = nobs(m)),
    tidy   = tidy(m)   %>% mutate(y = y, x = x),
```

```
    model  = m,
    data   = sub
  )
}

fits <- pairs %>%
  pmap(fit_1, data = df)
```

```
# summarize coefficients & fit stats
coef_tbl <- map_dfr(fits, "tidy") %>%
  select(y, x, term, estimate, std.error, statistic, p.value)

glance_tbl <- map_dfr(fits, "glance") %>%
  select(y, x, n, r.squared, adj.r.squared, AIC, BIC, sigma, p.value)

kable(coef_tbl, caption = "Regression Coefficients Summary")
```

Table 1: Regression Coefficients Summary

y	x	term	estimate	std.error	statistic	p.value
Prog_Proficiency	Score_Avg	(Intercept)	2.6153854	3.444774	0.7592329	0.4513503
Prog_Proficiency	Score_Avg	Score_Avg	1.8530853	1.613662	1.1483729	0.2563901
Stat_Proficiency	Score_Avg	(Intercept)	-3.6097496	3.662698	-0.9855440	0.3291985
Stat_Proficiency	Score_Avg	Score_Avg	5.1699736	1.715745	3.0132521	0.0040832
Math_Proficiency	Score_Avg	(Intercept)	5.3375240	3.900490	1.3684241	0.1774213
Math_Proficiency	Score_Avg	Score_Avg	0.7766173	1.827136	0.4250462	0.6726638

```
kable(glance_tbl, caption = "Model Fit Statistics Summary")
```

Table 2: Model Fit Statistics Summary

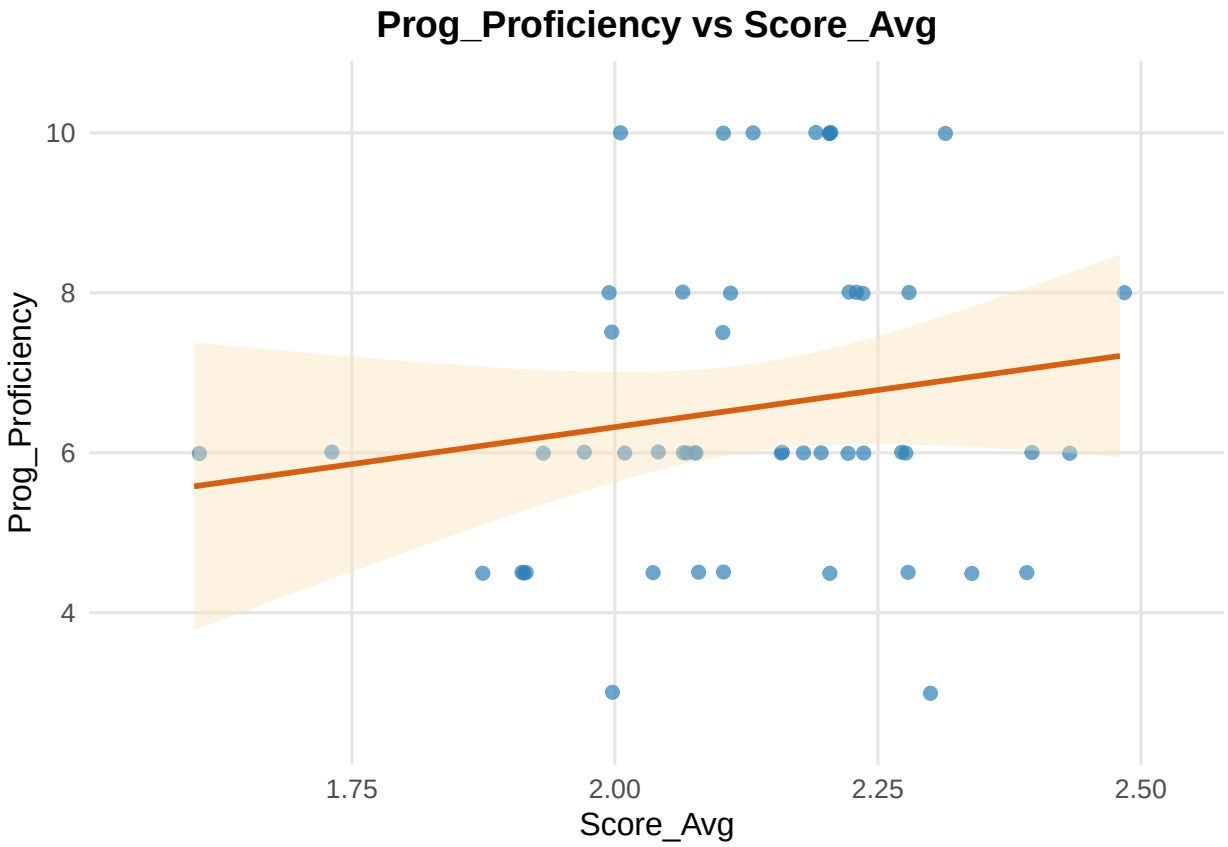
y	x	n	r.squared	adj.r.squared	AIC	BIC	sigma	p.value
Prog_Proficiency	Score_Avg	51	0.0262081	0.0063348	216.8064	222.6019	1.949931	0.2563901
Stat_Proficiency	Score_Avg	51	0.1563316	0.1391138	223.0633	228.8588	2.073288	0.0040832
Math_Proficiency	Score_Avg	51	0.0036735	-0.0166597	229.4793	235.2748	2.207891	0.6726638

```
# Plots with fitted line
plot_one <- function(y, x, data) {
  rngx <- range(data[[x]], na.rm = TRUE) + c(-0.05, 0.05)
  rngy <- range(data[[y]], na.rm = TRUE) + c(-0.5, 0.5)

  ggplot(data, aes(x = .data[[x]], y = .data[[y]])) +
    geom_point(
      alpha = 0.7,
      size = 2,
      color = "#2C7FB8",
      position = position_jitter(width = 0.01, height = 0.01)
    ) +
    geom_smooth(
```

```
      method = "lm",
      se = TRUE,
      color = "#D95F0E",
      fill = "#FEE0B6"
    ) +
    coord_cartesian(xlim = rngx, ylim = rngy) +
    labs(
      title = paste(y, "vs", x),
      x = x,
      y = y
    ) +
    theme_minimal(base_size = 12) +
    theme(
      plot.title = element_text(face = "bold", hjust = 0.5),
      panel.grid.minor = element_blank(),
      panel.grid.major = element_line(color = "grey90")
    )
}

plots <- pmap(
  pairs,
  \(y, x) plot_one(y, x, df %>% select(all_of(c(y, x))) %>% drop_na())
)
```

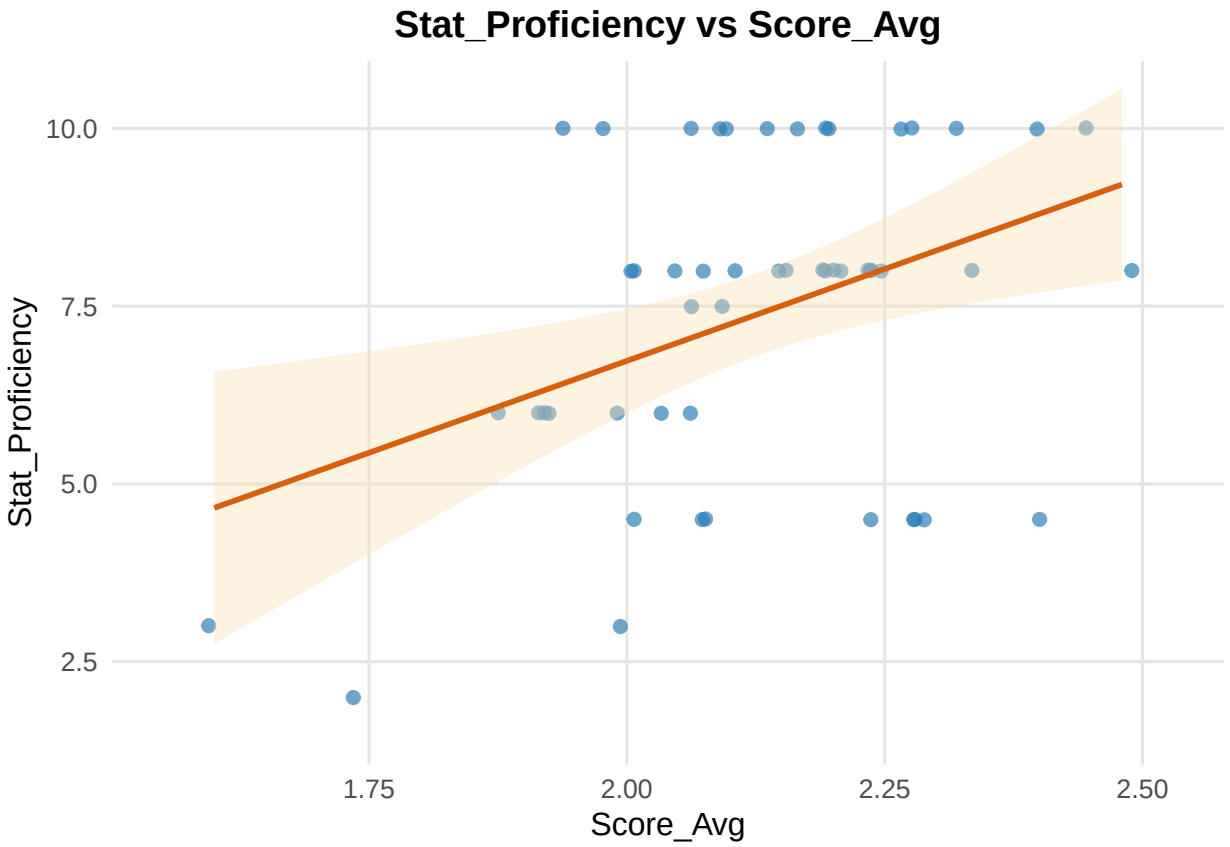
```
plots[[1]]
```

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



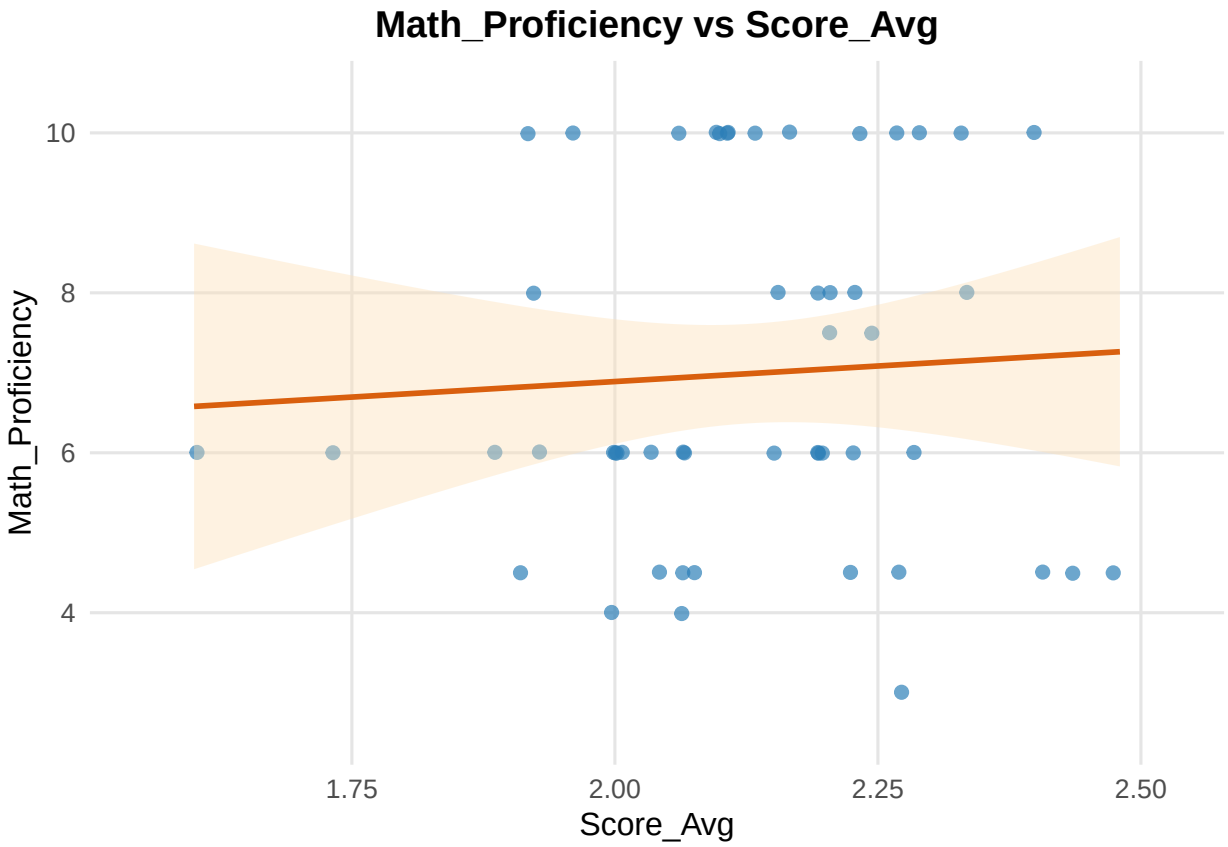

```
plots[[2]]
```

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




```
plots[[3]]
```

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


Math_Proficiency vs Score_Avg
Math_Proficiency
10
8
6
4
1.75
2.00
2.25
2.50
Score_Avg

Findings