Linear Regression Extensions Recap (Suggested Answers)

Dr. Soliman

Wages, Education, and Gender in 1985

1. Knit the markdown file now to ensure everything works and to easily access the following link. Download the data from here and call the object `CPS1985`.

```
CPS1985 <- read.csv("~/Library/CloudStorage/Dropbox/Clemson/Econometrics Course/data for tasks/CPS1985.c
```

2. We don't know if people are working part-time or full-time, but does it matter here? Answer yes or no and why.

It doesn't matter much since we are analyzing hourly wages.

3. Create a `log_wage` variable equal to the log of the variable `wage`, but assign the original `CPS1985` object to a new object with a humorous name. Use this object for the rest of the assignment.

```
cps_drsolimanistoocoolforschool <- CPS1985 %>%
    mutate(log_wage = log(wage))
```

4. Regress `log_wage` on `gender` and `education`, and save it as `reg1`. Interpret each coefficient.

```
reg1 <- lm(log_wage ~ gender + education, cps_drsolimanistoocoolforschool)
reg1
```

```
##
## Call:
## lm(formula = log_wage ~ gender + education, data = cps_drsolimanistoocoolforschool)
##
## Coefficients:
## (Intercept)   gendermale    education
##     0.93313      0.23207      0.07685
```

Interpretation

This is a log-level regression, so the coefficients are interpreted as a unit increase in x being associated with a percentage change in y.

Intercept Interpretation:

On average, women with no education earn $\exp(0.93313) = 2.54$ dollars per hour.

Gender Interpretation:

- On average, men earn $100 \times (\exp(0.23207) - 1) = 26\%$ more than women, holding education constant. Since the coefficient is relatively small, we can use the approximation: men earn 23.2% more than women, holding education constant.

Education Interpretation:

- Holding gender constant, a **1-year increase** in education is associated, on average, with an **8%** increase in hourly earnings. You can obtain this from $100 \times (\exp(0.07685) - 1)$, but again, since the coefficient is small, we can approximate: a **1-year increase in education** is associated with a 7.7% increase in hourly wage, holding gender constant.

5. Regress `log_wage` on `gender`, `education`, and their interaction `gender*education`, saving it as `reg2`. Interpret each coefficient. Does the gender wage gap decrease with education?

```
reg2 <- lm(log_wage ~ gender*education, cps_drsolimanistoocoolforschool)
reg2
```

```
##
## Call:
## lm(formula = log_wage ~ gender * education, data = cps_drsolimanistoocoolforschool)
##
## Coefficients:
##          (Intercept)            gendermale             education
##              0.69037               0.63315               0.09549
## gendermale:education
##             -0.03080
```

Interpretation

This is a log-level regression, so the coefficients are interpreted as a unit increase in x being associated with a percentage change in y.

Intercept Interpretation:

On average, women with no education earn $\exp(0.69037) = 1.99$ dollars per hour.

Gender Interpretation:

- Men earn $100 \times (\exp(0.63315) - 1) = 89\%$ more than women with no education. Note that because the coefficient was large, it is best not use the approximation, as you can see 63% would be quite different than 89%.
- Put another way, on average, men with no education earn $\exp(0.69037 + 0.63315) = 3.76$ dollars per hour.

Education Interpretation:

- A **1-year increase** in education is associated with a **10%** increase in hourly wages for women, because $100 \times (\exp(0.09549) - 1)$. Remember what the base category is.
- For each additional year of education, men's hourly wage increases by about $100 \times (\exp(0.09549 - 0.03080) - 1)$, or **6.6%**. Informally, you could have approximated that men's hourly wage increase with education is 3.1% less than women's (9.5%), so 6.4%.

The gender wage gap decreases with education because the interaction term is negative, indicating that men's wages grow at a slower rate compared to women's wages as education increases. Put another way, since men's wages increase less with education than women's wages, the gender wage gap decreases as education increases.

6. Create a plot showing this interaction. (*Hint:* use the `color = gender` argument in `aes` and `geom_smooth(method = "lm", se = F)` to obtain a regression line per gender.)

What we saw in the previous answer is confirmed here graphically.

```
ggplot(cps_drsolimanistoocoolforschool, aes(x = education, y = log_wage, col = gender)) + geom_point() +
    geom_smooth(method= "lm", se = F) +
    scale_color_viridis_d() +
    labs(x = "Years of education", y = "Log hourly wage",
         title = "Relationship between hourly wage and years of education by gender", color = NULL) +
    theme_bw(base_size = 14) +
      theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
            legend.position = c(0.1,0.8)) + scale_color_manual(values = c("forestgreen", "orange"))
```

```
## Warning: A numeric `legend.position` argument in `theme()` was deprecated in ggplot2
## 3.5.0.
## i Please use the `legend.position.inside` argument of `theme()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Scale for colour is already present.
## Adding another scale for colour, which will replace the existing scale.
## `geom_smooth()` using formula = 'y ~ x'
```