PO12Q - Introduction to Quantitative Political Analysis II: Worksheet Week 7 - Solutions

Dr Florian Reiche
F.Reiche@warwick.ac.uk

WARWICK
THE UNIVERSITY OF WARWICK
DEPARTMENT OF POLITICS & INTERNATIONAL STUDIES

1 | Regression – Hypothesis Testing & Confidence Intervals

```
## $coefficients
##                Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) -94306.813 10406.7267 -9.062101 3.328834e-16
## lifeexp       1503.211   144.3715 10.412105 7.027122e-20
```

1. Formulate the null and alternative hypotheses which is tested in this model.

- H_0: Life expectancy does not influence the level of per capita GDP.
- H_A: The higher the level of life expectancy, the higher the level of per capita GDP.

2. Build the regression function and interpret the coefficients.

$$\widehat{\text{gdp}}_i = -94306.813 + 1503.211\,\text{life}_i$$

3. Is the coefficient significant

- Yes, it is. The p-value is <0.05.

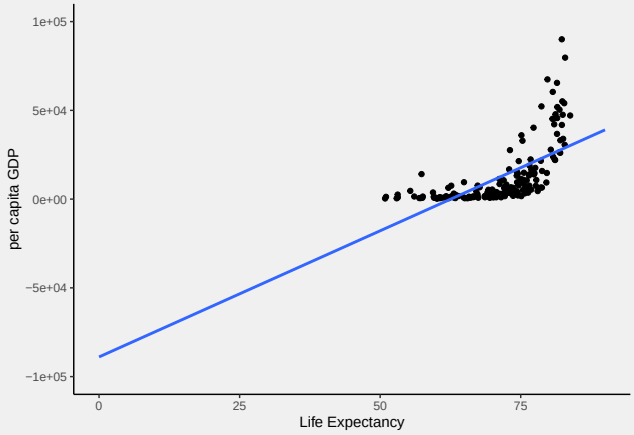
4. Interpret the coefficients.
   - **intercept:** if life expectancy was zero years, the per capita GDP would be US$ -94,306.81, on average (substantively, this makes zero sense.)
   - **slope:** For each additional year of life expectancy, per capita GDP increases by US$ 1,503.21, on average.

- If `life` was insignificant, you could not interpret it other than saying "there is no relationship". We have not rejected the null hypothesis, and so we are not confident enough that it is different from "no influence".

5. Plot the regression function in a suitable diagram using `ggplot`.

```
library(tidyverse)

ggplot(data=wdi, aes(x=lifeexp, y=gdppc)) +
  geom_point() +
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE) +
  theme_classic()+
  scale_x_continuous(name = "Life Expectancy", limits=c(0,90)) +
  scale_y_continuous(name = "per capita GDP", limits=c(-100000, 100000))+
  theme(
    panel.background = element_rect(fill='transparent'), #transparent panel bg
    plot.background = element_rect(fill='transparent', color=NA), #transparent plot bg
    panel.grid.major = element_blank(), #remove major gridlines
    panel.grid.minor = element_blank(), #remove minor gridlines
    legend.background = element_rect(fill='transparent'), #transparent legend bg
    legend.box.background = element_rect(fill='transparent') #transparent legend panel
  )
```



6. Explain how the t-value for `life` is obtained.

$$t = \frac{\hat{\beta}_2}{\hat{se}(\hat{\beta}_2)} = \frac{1503.211}{144.3715} = 10.412105$$

7. What do our results mean for the hypotheses?

- We have verified our alternative hypothesis and rejected the null hypothesis.

8. What does the value of "Multiple R-Squared" (this is equivalent to the R-Squared we calculated by hand last week) mean?

- R-Squared is 0.3936, which means that life expectancy is able to explain 39.36% of the variation in GDP.

9. Calculate the 95% confidence intervals for the coefficient `life` and the intercept. Compare your results to the R output below.

$$1503.211 - 1.96 \times 144.4 \leq \text{life} \leq 1503.211 + 1.96 \times 144.4$$

10. Find two explanations in the output for why the coefficient for `lifeexp` is statistically significant at the 5% level?
    1. p-value is 7.027122e-20
    2. $|t| > 2$ with df=167