SET_8

Oluwanifemi

2024-11-22

```
model_1 <- lm(formula = lifeExpF ~ I(log(ppgdp)) + pctUrban + fertility,
data = UN11)

summary(model_1)
```

```
##
## Call:
## lm(formula = lifeExpF ~ I(log(ppgdp)) + pctUrban + fertility,
##     data = UN11)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.6165  -1.6683   0.5406   2.6425  11.2263
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   64.21764    3.84281  16.711  < 2e-16 ***
## I(log(ppgdp))  2.17842    0.42888   5.079 8.83e-07 ***
## pctUrban       0.02116    0.02353   0.900    0.369
## fertility     -4.19652    0.39396 -10.652  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.145 on 195 degrees of freedom
## Multiple R-squared:  0.7456, Adjusted R-squared:  0.7417
## F-statistic: 190.5 on 3 and 195 DF,  p-value: < 2.2e-16
```

Question 1(b)

Conclusion for the Overall F-Test

From the R output, the **overall F-test** evaluates whether at least one of the regression coefficients (excluding the intercept) is significantly different from zero.

- **F-statistic**: 190.5
- **Degrees of Freedom (DF)**: 3, 195
- **p-value**: $< 2.2 \times$ 10-16
- **Significance Level** (α): 0.01

Conclusion: The p-value for the F-test ($< 2.2 \times$ 10-16) is much smaller than the significance level of 0.01. **We reject the null hypothesis**, concluding that the model as a whole is statistically significant.

Reasoning: At least one predictor in the model significantly contributes to explaining the variability in life expectancy.

```
reduced_model <- lm(lifeExpF ~ fertility, data = UN11)

# Perform the nested F-test
anova_result <- anova(reduced_model, model_1)
anova_result
```

```
## Analysis of Variance Table
##
## Model 1: lifeExpF ~ fertility
## Model 2: lifeExpF ~ I(log(ppgdp)) + pctUrban + fertility
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    197 6528.4
## 2    195 5161.7  2    1366.6 25.814 1.133e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testing Whether the Parameters for `log(ppgdp)` and `pctUrban` Are Both Zero

1. Hypotheses

- **Null Hypothesis (H_0)**: The parameters for `log(ppgdp)` and `pctUrban` are both 0. These variables do not contribute to explaining the variability in `lifeExpF`.
- **Alternative Hypothesis (H_a)**: At least one of the parameters for `log(ppgdp)` or `pctUrban` is not 0. Including these variables significantly improves the model.

2. Test Statistic The test statistic is the F-statistic from the ANOVA output:

$$F = 25.814$$

- **Degrees of Freedom**: - Numerator (df_1): 2 (number of predictors added: `log(ppgdp)` and `pctUrban`). - Denominator (df_2): 195 (residual degrees of freedom from the full model).

3. Significance Level The significance level (α) is 0.05.

4. p-value From the ANOVA output:

$$p = 1.133 \times 10^{-10}$$

5. Decision Compare the p-value to the significance level ($\alpha = 0.05$):

$$p = 1.133 \times 10^{-10} < 0.05$$

Since the p-value is much smaller than 0.05, we **reject the null hypothesis**.

6. Conclusion At the 0.05 significance level, there is sufficient evidence to conclude that the parameters for `log(ppgdp)` and/or `pctUrban` are not 0. Adding these predictors significantly improves the model's ability to explain `lifeExpF`.