Life Expectancy Analysis with Simple Linear Regression Models

Conducting two linear regression analyses between

- Predicted variable (DV): Life Expectancy (LE)
- Explanatory variables (IVs): People per TV (PPTV), People per Physician (PPP)

Creating Models

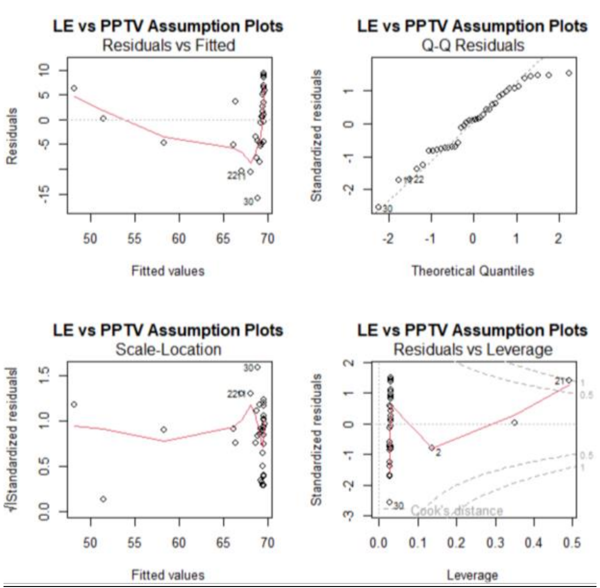
```
> data = LifeExpectancy
>
> model_pptv = lm(data$LE ~ data$PPTV)
> model_pptv

Call:
lm(formula = data$LE ~ data$PPTV)

Coefficients:
(Intercept)    data$PPTV
   69.64813     -0.03626
```

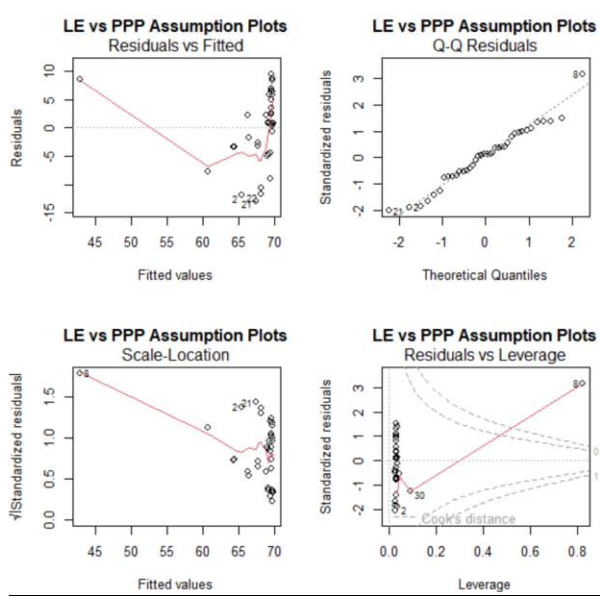
```
> model_ppp = lm(data$LE ~ data$PPP)
> model_ppp

Call:
lm(formula = data$LE ~ data$PPP)

Coefficients:
(Intercept)     data$PPP
 69.9264143   -0.0007374
```

a. Checking Linear Regression Assumptions

```
windows()
par(mfrow = c(2,2))
plot(model_pptv, main="LE vs PPTV Assumption Plots")
```

```
windows()
par(mfrow = c(2,2))
plot(model_ppp, main="LE vs PPP Assumption Plots")
```





The plots show:

1. Residuals vs Fitted: Data points are **randomly spread** around 0 with **no clear pattern**.

2. Q-Q Residuals: Most data points follow along the line, suggesting **normal** residuals.

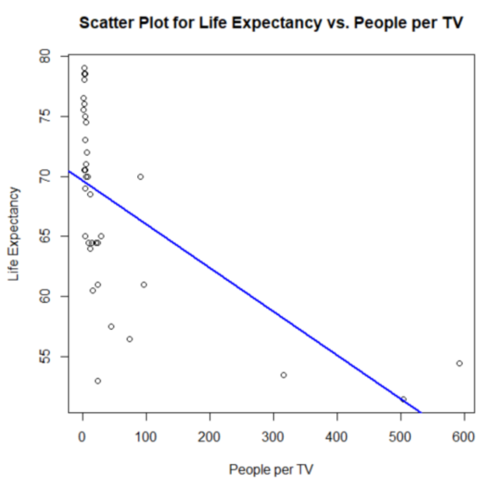
3. Scale-Location: Random spread with approximately the same number of data points above and below the line, suggesting **equal variance**.

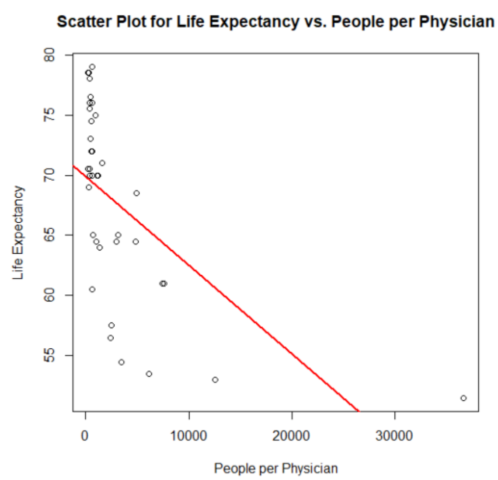
4. Residuals vs Leverage: **No extreme outliers** with high leverage.

The conditions are reasonable enough to proceed with linear regression.

b. Constructing Scatterplots (Linearity Assumption)

```
windows()
plot(data$PPTV, data$LE, main="Scatter Plot for Life Expectancy vs.
People per TV", xlab="People per TV", ylab="Life Expectancy")
abline(model_pptv, col="blue", lwd=2)

windows()
plot(data$PPP, data$LE, main="Scatter Plot for Life Expectancy vs.
People per Physician", xlab="People per Physician", ylab="Life
Expectancy")
abline(model_ppp, col="red", lwd=2)
```





The scatterplots suggest:

- **Moderate negative linear** relationship between **LE** and **PPTV** (as PPTV increases, LE decreases)
- **Moderate negative linear** relationship between **LE** and **PPP** (as PPP increases, LE decreases)

c. Regression Line Equations

```
> # LE vs PPTV model
> summary(model_pptv)

Call:
lm(formula = data$LE ~ data$PPTV)

Residuals:
     Min       1Q   Median       3Q      Max 
-15.8141  -4.6061   0.5876   5.3647   9.4171 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 69.648132   1.101058  63.256  < 2e-16 ***
data$PPTV   -0.036264   0.007937  -4.569 5.56e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.293 on 36 degrees of freedom
Multiple R-squared:  0.3671,	Adjusted R-squared:  0.3495 
F-statistic: 20.88 on 1 and 36 DF,  p-value: 5.561e-05
```

```
> # LE vs PPP model
> summary(model_ppp)

Call:
lm(formula = data$LE ~ data$PPP)

Residuals:
     Min       1Q   Median       3Q      Max 
-12.8567  -4.1791   0.7933   5.5884   9.5226 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 69.9264143  1.1510293  60.751  < 2e-16 ***
data$PPP    -0.0007374  0.0001693  -4.356 0.000105 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.401 on 36 degrees of freedom
Multiple R-squared:  0.3451,	Adjusted R-squared:  0.3269 
F-statistic: 18.97 on 1 and 36 DF,  p-value: 0.0001054
```

i. Linear equation for predicting LE vs. PPTV

- $\hat{LE}_{PPTV}$ = 69.6481 - 0.0363(PPTV)

ii. Linear equation for predicting LE vs. PPP

- $\hat{LE}_{PPP}$ = 69.9264 - 0.0007374(PPP)

iii. Slope for LE vs. PPTV Regression Line = -0.0363

- For an **increase** of each additional person per TV, life expectancy **decreases** by 0.0363 years on average.

iv. Slope for LE vs. PPP Regression Line = -0.0007

- For an **increase** of each additional person per physician, life expectancy **decreases** by 0.0007374 years on average.

d. Correlation Coefficient (r)

```
> cor(data$PPTV, data$LE)
[1] -0.6058468
> cor(data$PPP, data$LE)
[1] -0.5874798
```

i. LE vs. PPTV(*r*): -0.6058

ii. LE vs. PPP(*r*): -0.5875

Both relationships are **moderate negative** because r is between -0.5 and -0.8.

The relationship is slightly stronger for people per TV.

e. Significance Test for an Alpha Level of 5%

```
> # LE vs PPTV model
> summary(model_pptv)

Call:
lm(formula = data$LE ~ data$PPTV)

Residuals:
     Min       1Q   Median       3Q      Max 
-15.8141  -4.6061   0.5876   5.3647   9.4171 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 69.648132   1.101058  63.256  < 2e-16 ***
data$PPTV   -0.036264   0.007937  -4.569 5.56e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.293 on 36 degrees of freedom
Multiple R-squared:  0.3671,	Adjusted R-squared:  0.3495 
F-statistic: 20.88 on 1 and 36 DF,  p-value: 5.561e-05
```

```
> # LE vs PPP model
> summary(model_ppp)

Call:
lm(formula = data$LE ~ data$PPP)

Residuals:
     Min       1Q   Median       3Q      Max 
-12.8567  -4.1791   0.7933   5.5884   9.5226 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 69.9264143  1.1510293  60.751  < 2e-16 ***
data$PPP    -0.0007374  0.0001693  -4.356 0.000105 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.401 on 36 degrees of freedom
Multiple R-squared:  0.3451,	Adjusted R-squared:  0.3269 
F-statistic: 18.97 on 1 and 36 DF,  p-value: 0.0001054
```

i. LE vs PPTV **p-value**: 5.56e-05 < **alpha**

So, the linear relationship between life expectancy and people per TV is **statistically significant**.

ii. LE vs PPTV **p-value**: 0.000105 < **alpha**

So, the linear relationship between life expectancy and people per physician is **statistically significant**.

f. Coefficient of Determination (r^2)

```
> # LE vs PPTV model
> summary(model_pptv)

Call:
lm(formula = data$LE ~ data$PPTV)

Residuals:
     Min       1Q   Median       3Q      Max 
-15.8141  -4.6061   0.5876   5.3647   9.4171 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 69.648132   1.101058  63.256  < 2e-16 ***
data$PPTV   -0.036264   0.007937  -4.569 5.56e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.293 on 36 degrees of freedom
Multiple R-squared:  0.3671,	Adjusted R-squared:  0.3495 
F-statistic: 20.88 on 1 and 36 DF,  p-value: 5.561e-05
```

```
> # LE vs PPP model
> summary(model_ppp)

Call:
lm(formula = data$LE ~ data$PPP)

Residuals:
     Min       1Q   Median       3Q      Max 
-12.8567  -4.1791   0.7933   5.5884   9.5226 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 69.9264143  1.1510293  60.751  < 2e-16 ***
data$PPP    -0.0007374  0.0001693  -4.356 0.000105 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.401 on 36 degrees of freedom
Multiple R-squared:  0.3451,	Adjusted R-squared:  0.3269 
F-statistic: 18.97 on 1 and 36 DF,  p-value: 0.0001054
```

i. LE vs. PPTV($r2$): 0.3671

- **36.71% of variation** in life expectancy is explained by people per TV.

ii. LE vs. PPP($r2$): 0.3451

- **34.51% of the variation** in life expectancy is explained by people per physician.

g. Predicting LE for New Countries on PPTV and PPP Data on Web Search

Equation:
• LE_{PPTV}^ = 69.6481 - 0.0363(PPTV)
 	
Range of PPTV in the Dataset:
`> range(data$PPTV)`
`[1]   1.3 592.0`
	Equation:
	• LE_{PPP}^ = 69.9264 - 0.0007374(PPP)
	
	Range of PPP in the Dataset:
	`> range(data$PPP)`
	`[1]   226 36660`

Country: Ghana

TV Market Volume (by 2029): 223,600 pieces

Population (2025 est.): 34.92 million (https://www.worldometers.info/world-population/ghana-population/)

People per TV (PPTV): 34,920,000 / 223,600 = 156.1717 people (est.)

LE^ = 69.6481 - 0.0363(156.1717) = 63.97907 ~ 64 years

URL: https://www.statista.com/outlook/cmo/consumer-electronics/tv-radio-multimedia/televisions/ghana

Country: Bhutan

People per Physician (PPP) (in 2022): 1,000

LE^ = 69.9264 - 0.0007374(1000) = 69.189 ~ 69.2 years

URL: https://data.worldbank.org/indicator/SH.MED.PHYS.ZS

h. Predicting PPTV and PPP for Life Expectancy of 66.41 Years in Djibouti

```
> model_pptv_inv = lm(PPTV ~ LE, data=data)
> model_pptv_inv

Call:
lm(formula = PPTV ~ LE, data = data)

Coefficients:
(Intercept)           LE
     737.85       -10.12
```

```
> model_ppp_inv = lm(PPP ~ LE, data=data)
> model_ppp_inv

Call:
lm(formula = PPP ~ LE, data = data)

Coefficients:
(Intercept)           LE
    34651.4       -468.1
```

Predicted PPTV

- PPTV^ = 737.85 – 10.12(LE)
- 737.85 - (10.12*66.41) = 65.7808
- **65 people per TV**

Predicted PPP

- PPP^ = 34651.4 – 468.1(LE)
- 34651.4 - (468.1*66.41) = 3564.879
- **3564 people per physician**