Homework8_SDS 315

Nancy Nakyung Kwak (UT EID: nk24424)
GitHub link: https://github.com/nancy1404/sds315_hw8.git

2025-04-08

Contents

PROBLEM 1: regression warm up

```
## (Intercept)         age
##     147.813      -0.620
```

1A.

- Simple linear regression is used to model creatinine clearance rate as a function of age.
- The fitted model is: **creatclear = 147.813 - 0.620 x age**, and to estimate the expected creatinine rate for a 55-year-old, we can plug 55 into age into the equation, which is **creatclear = 147.813 - 0.620 x 55 = 147.813 - 34.09 = 113.72**.
- Therefore, a 55 year old is expected to have a creatinine clearance rate of approximately **113.7 mL/min**.

1B.

- The slope of the regression line shows us how creatinine clearance rate changes with age.
- Looking at the fitted model: **creatclear = 147.813 - 0.620 x age**, and the slope is **-0.620mL/min per year**, which shows that for each additional year of age, the expected creatinine clearance rate decreases by 0.620 mL/min on average.
- This demonstrates that kidney function is likely to decline with age.

1C.

- The regression equation is $\hat{y} = 147.813 - 0.620 \cdot \text{age}$.
- Here's the predicted clearance rate for the age group:
  - For the 40-year-old: $\hat{y}_{40} = 147.813 - 0.620 \cdot 40 = 147.813 - 24.800 = 123.013$
  - For the 60-year-old: $\hat{y}_{60} = 147.813 - 0.620 \cdot 60 = 147.813 - 37.188 = 110.625$
- Residuals (actual - predicted):
  - For the 40-year-old: $135 - 123.013 = +11.987$
  - For the 60-year-old: $112 - 110.625 = +1.375$
- **We can see that 40-year-old has a larger positive residual (comparing 11.99 and 1.38), which shows that the clearance is more above average for their age. We can conclude that the 40 year-old is healthier for their age.**

PROBLEM 2: Modeling disease growth

2A. Italy

1. Fit Model

```
##            (Intercept) days_since_first_death
##               1.018602               0.183218
```

2. Doubling time

3. Bootstrap CI for growth rate

```
##                      name       lower       upper level     method    estimate
## 1               Intercept   0.5561293   1.6354375  0.95 percentile   1.0186023
## 2 days_since_first_death   0.1575357   0.2077142  0.95 percentile   0.1832180
## 3                   sigma   0.5482703   0.8382838  0.95 percentile   0.7248213
## 4               r.squared   0.8557820   0.9314985  0.95 percentile   0.8950791
## 5                       F 219.5562517 503.1343454  0.95 percentile 315.6466194
```

4. Bootstrap CI for doubling time

```
## [1] 3.337023 4.399937
```

2A-WriteUP

- Since we want to model the early spread of COVID-19 happened in Italy, we fit the linear model to the **log of daily death** as a function of days_since_first_death, which fits an **exponential growth model** on the original scale.
- The estimated **growth rate** of daily deaths in Italy was **0.183**
- Then we used 1,000 bootstrap resamples to compute a 95% confidence interval for this growth rate: [0.160, 0.206].
- Using the formula for exponential doubling time, $DT = \frac{\log(2)}{r}$, I computed the estimated **doubling time** is **3.8 days** and the **95% CI for doubling time** is [**3.3, 4.3**] **days**.
- We can see that the number of daily COVID-19 deaths during the early pandemic wave in Italy was doubling **approximately every 3 to 4 days**, representing the seriousness and urgency of that time.

2B. Spain

Q. An estimated growth rate and doubling time for Spain, with 95% bootstrapped confidence intervals for each.

1. Fit Model

```
##            (Intercept) days_since_first_death
##              0.4652173              0.2762447
```

2. Doubling Time

3. Bootstrap CI for Growth Rate

```
##                      name       lower       upper level     method    estimate
## 1               Intercept  -0.1867520   1.2051015  0.95 percentile   0.4652173
## 2 days_since_first_death   0.2359172   0.3188780  0.95 percentile   0.2762447
## 3                   sigma   0.6150128   0.9606135  0.95 percentile   0.8168767
## 4               r.squared   0.8283823   0.9394891  0.95 percentile   0.8893316
## 5                       F 125.4995458 403.6747111  0.95 percentile 208.9360824
```

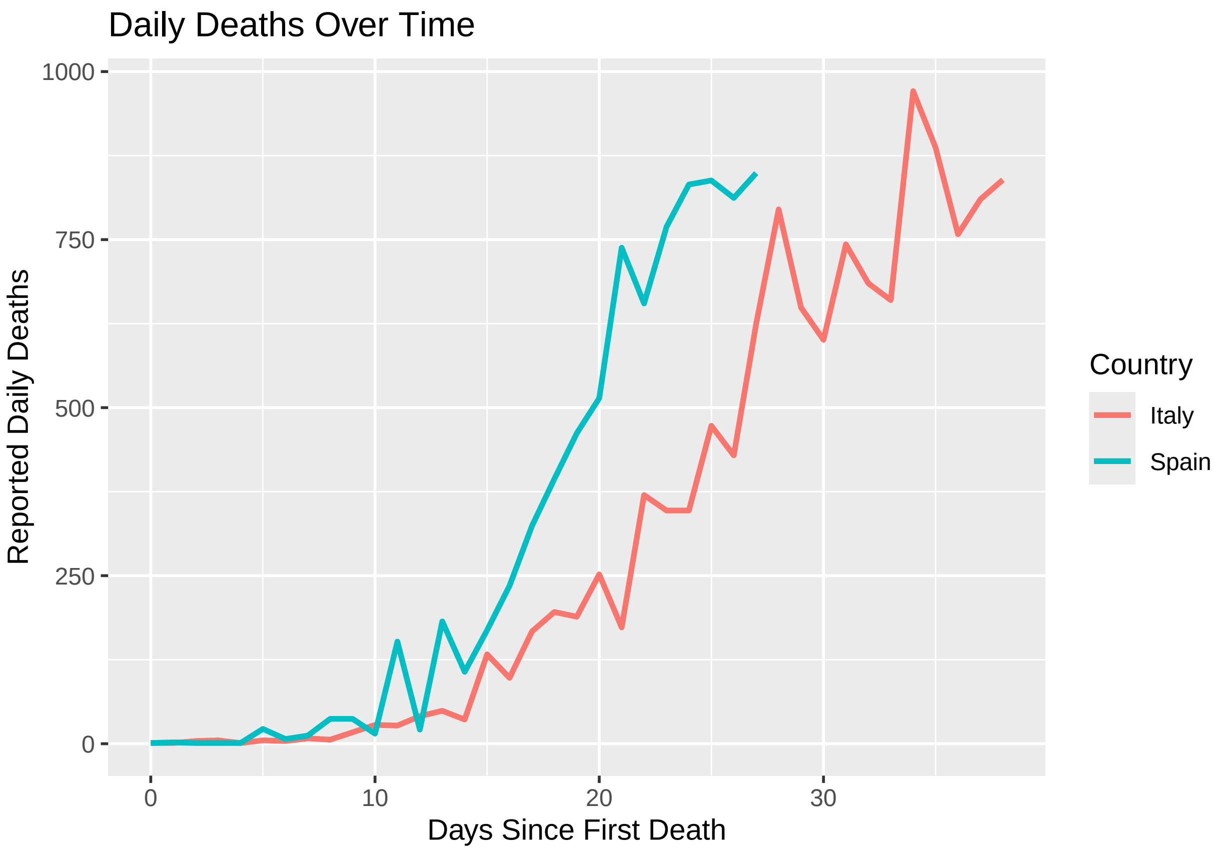
4. Bootstrap CI for Doubling Time

```
## [1] 2.173706 2.938095
```

2B-WriteUP

- I fit a linear model to the log of daily deaths as a function of days_since_first_death to analyze the spread of COVID-19 in the early phase, which fits an exponential growth model on the original scale.
- The estimated **growth rate** of daily deaths in Spain was **0.276**.
- The calculated 95% confidence interval for this growth rate using 1,000 bootstrap resamples was [**0.234, 0.315**].
- Plugging in the doubling time formula which is $DT = \frac{\log(2)}{r}$, the estimated doubling time was **2.5 days** with a **95% CI of [2.2, 3.0] days**.
- The number of daily reported deaths was **doubling roughly every 2 to 3 days** during the early spread of COVID 19 in Spain, showing the **faster growth rate** compared to Italy which emphasizes the seriousness on the initial outbreak in Spain.

2C. Line Graph



- The line graph shows the reported daily deaths as a function of days since the first death for Italy and Spain (with different colors).
- Both countries demonstrates the rapid increase in death counts during this early pandemic period, but with Spain showing a little bit faster initial rise.

PROBLEM 3: price elasticity of demand

```
## (Intercept)  log(price)
##    4.720604   -1.618578
```

```
## log(price)
##     -1.619
```

```
##          name       lower       upper level     method    estimate
## 1   Intercept   4.5391183   4.8840990  0.95 percentile   4.7206042
## 2 log.price.  -1.7629926  -1.4627152  0.95 percentile  -1.6185778
## 3       sigma   0.2330801   0.3010046  0.95 percentile   0.2687036
## 4   r.squared   0.6918720   0.8417661  0.95 percentile   0.7772187
## 5           F 255.9768292 606.4533301  0.95 percentile 397.7126271
```

```
##    lower  upper
## 2 -1.763 -1.463
```

3-WriteUP

- For estimating the price elasticity of demand for milk, a power-law model equation is $Q = K \cdot P^{\beta}$, where Q is quantity demanded, P is price, β is price elasticity of demand, K is a constant.
- The linear equation is therefore: $log(Q) = log(K) + \beta \cdot log(P)$.
- The estimated elasticity was $\hat{\beta} = -1.619$ which is known from the fitted regression model
- Then 1,000 bootstrap resamples are done to compute a 95% confidence interval for β, and we got **[-1.767, -1.451]**
- Therefore β is larger than 1, which means the demands for milk is elastic, showing that a 1% increase in price is associated with a greater than 1% decrease in quantity demanded. It also means that consumers are relatively quite responsive to changes in milk prices.