Air passengers 1949-1960

22 marks

Airline passengers.
Here we will examine the monthly total of the number of air passengers in the US from 1949 to 1960.

The data are available as the data set `AirPassengers` from the `datasets` package in the standard `R` distribution.

```
AirPassengers
```

```
##      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 1949 112 118 132 129 121 135 148 148 136 119 104 118
## 1950 115 126 141 135 125 149 170 170 158 133 114 140
## 1951 145 150 178 163 172 178 199 199 184 162 146 166
## 1952 171 180 193 181 183 218 230 242 209 191 172 194
## 1953 196 196 236 235 229 243 264 272 237 211 180 201
## 1954 204 188 235 227 234 264 302 293 259 229 203 229
## 1955 242 233 267 269 270 315 364 347 312 274 237 278
## 1956 284 277 317 313 318 374 413 405 355 306 271 306
## 1957 315 301 356 348 355 422 465 467 404 347 305 336
## 1958 340 318 362 348 363 435 491 505 404 359 310 337
## 1959 360 342 406 396 420 472 548 559 463 407 362 405
## 1960 417 391 419 461 472 535 622 606 508 461 390 432
```

a. **(7 marks)** Use the function `decompose()` on this data and plot the results.

   i. *(2 marks)* Show your plot.

   as shown

   ii. *(1 mark)* Describe the trend.

   we observe the number of airpassengers is clearly increasing in terms of time in a approximately linear fashion.

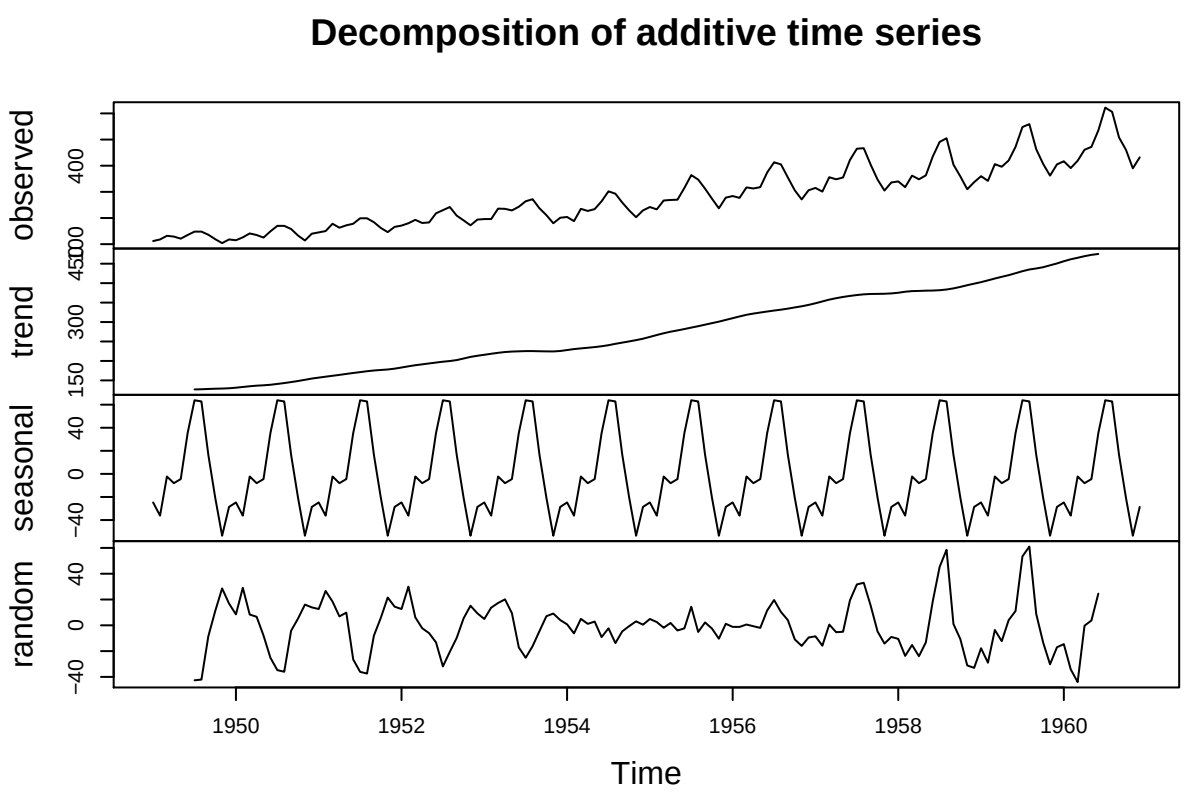
   iii. *(2 marks)* Describe the seasonal pattern.

   we observe there is a clear seasonal pattern, some months clearly have more airpassengers than others. The overall seasonal pattern follows roughly a sin or cos wave shape each year (with period of one year).

   iv. *(2 marks)* What do you conclude from the residuals?

   We observe there is a pattern in residuals, the variance is smaller as time approaches year 1955, and larger otherwise. the shape of residual is rough opposite (symmric with respect to the time axis) with seasonal pattern before year 1955, and is roughly same shape with seasonal pattern after year 1955 so this is probably a bad model with large variance. And the model didn't explain some property of the data.

```
plot(decompose(AirPassengers))
```



b. **(9 marks)** Use the function `stl()` on this data with seasonality loess span `s.window = 7` and the local polynomial for the seasonal loess being a line.

i. *(2 marks)* Show your plot.

ii. *(2 marks)* Describe the trend. Is it significant? Why or why not?

we observe the number of airpassengers is clearly increasing in terms of time in a approximately linear fashion.

it is significant, as its shape looks strongly linearly monotonicly increasing, its y value has a very large span (almost same as span of data),

iii. *(3 marks)* Describe the seasonal pattern. Is it significant? Why or why not?

we observe there is a clear seasonal pattern, some months clearly have more airpassengers than others. The overall seasonal pattern follows roughly a sin or cos wave shape each year(with period of one year). Also the pattern grows stronger with respect to time, having increasing variance.

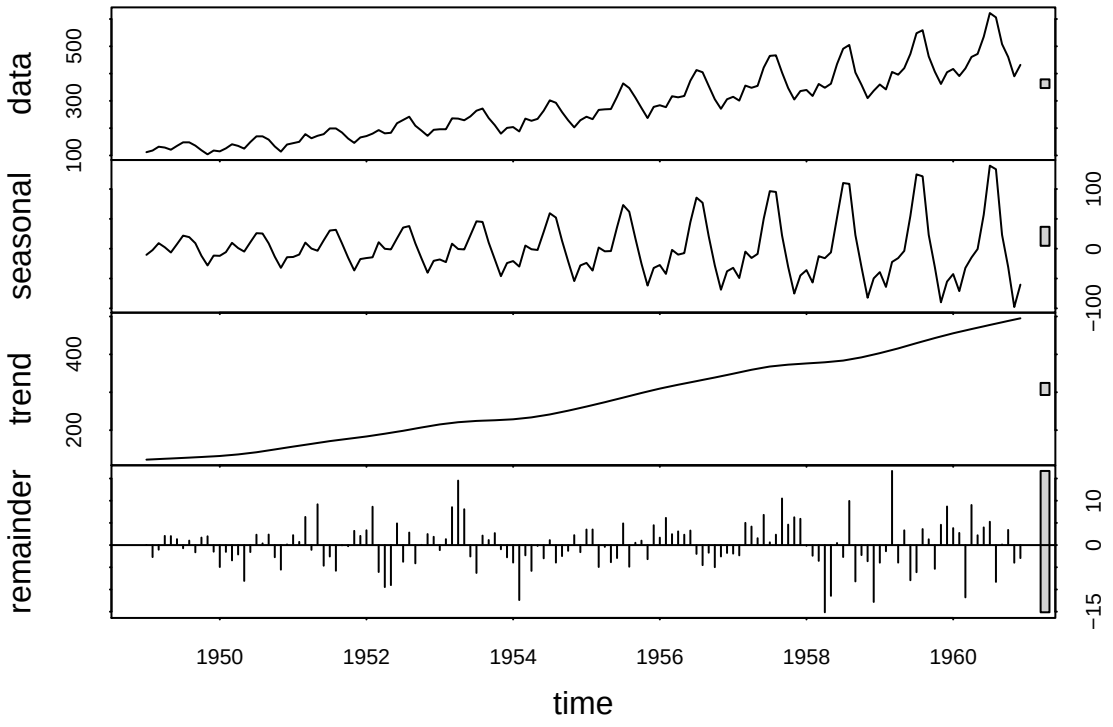
it is significant in shape, as there is clearly a sin wave shape, it is also significant in terms of absolute value. as its span is 1/2 of the span of the data

iv. *(2 marks)* What do you conclude from the residuals?

there is little pattern in residual, it has no clear shape, it has small span (there isn't even extreme value)

so this is probably a good model with small variance.

```
plot(stl(AirPassengers, s.window = 7, s.degree = 1))
```



c. **(6 marks)** Compare the plots from parts (a) and (b).

i. *(2 marks)* What are the major differences in the plots?

plot in (a) has constant seasonal pattern each year, (b) has varying seasonal pattern, with more variation each year.

plot in (a) has a clear pattern in residuals, (b) has very little pattern in resudual.

ii. *(2 marks)* What characteristics of the two methods caused these differences?

model in (a) uses the same seasonal trend for all years by calculating average of air passengers in that month for all years. Model in (b) uses varying seasonal trend by smoothing the number of air passengers for that month in all years.

model b also smooth the overall trend, but the overall trend is already very smooth, the effect on trend is not large

iii. *(2 marks)* Which fitted model do you prefer and **why**?

I prefer the second model, the residual has less variance and less pattern. And the seasonal pattern is more significant (has larger absolute value)