

CS_08

steppe

2/28/2022

```
library(kableExtra)
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5      v purrr 0.3.4
## v tibble 3.1.4       v dplyr 1.0.7
## v tidyr 1.1.3        v stringr 1.4.0
## v readr 2.0.1        v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter()     masks stats::filter()
## x dplyr::group_rows() masks kableExtra::group_rows()
## x dplyr::lag()         masks stats::lag()

data(iris)
head(iris)

##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4          0.2  setosa
## 2          4.9         3.0          1.4          0.2  setosa
## 3          4.7         3.2          1.3          0.2  setosa
## 4          4.6         3.1          1.5          0.2  setosa
## 5          5.0         3.6          1.4          0.2  setosa
## 6          5.4         3.9          1.7          0.4  setosa

a <- iris %>%
  split(.$Species) %>%
  map(summary) %>%
  map(as.data.frame) %>%
  map(drop_na) %>%
  bind_rows() %>%
  separate(Freq, c('Statistic', 'Value'), ":") %>%
  filter(!str_detect(Var2, 'Species')) %>%
  mutate(Species = c(rep('setosa', times = 24),
                     rep('versicolor', times = 24),
                     rep('virginica', times = 24))) %>%
  select(-Var1) %>%
  rename(Variable = Var2) %>%
  filter(!str_detect(Statistic, 'Qu'))

iris_table <- iris %>%
  group_by(Species) %>%
  summarise(across(everything(), list(min = min,
                                     max = max,
```

```

        mean = mean,
        median = median))) %>%
pivot_longer(-Species, values_to = 'Value') %>%
separate(name, c('Tissue', 'Statistic'), sep = "_") %>%

# style down here
mutate(Tissue = str_replace(Tissue, '[.]', ' ')) %>%
mutate(Species = paste0('I. ', Species)) %>%
mutate(Value = round(Value, digits = 1)) %>%

pivot_wider(names_from = Tissue, values_from = Value) %>%
mutate(Statistic=fct_relevel(Statistic, "min", "mean", "median", 'max')) %>%
arrange(Species, Statistic)

iris_table[c(2:4, 6:8, 10:12),1] <- ""
knitr::kable(iris_table)

```

Species	Statistic	Sepal Length	Sepal Width	Petal Length	Petal Width
I. setosa	min	4.3	2.3	1.0	0.1
	mean	5.0	3.4	1.5	0.2
	median	5.0	3.4	1.5	0.2
	max	5.8	4.4	1.9	0.6
I. versicolor	min	4.9	2.0	3.0	1.0
	mean	5.9	2.8	4.3	1.3
	median	5.9	2.8	4.3	1.3
	max	7.0	3.4	5.1	1.8
I. virginica	min	4.9	2.2	4.5	1.4
	mean	6.6	3.0	5.6	2.0
	median	6.5	3.0	5.5	2.0
	max	7.9	3.8	6.9	2.5

```

model_summary <- summary(lm(Sepal.Length ~ Sepal.Width, data = iris))
model_summary_statistics <- tibble("intercept" = model_summary$coefficients[1,1],
                                   "slope"      = model_summary$coefficients[2,1],
                                   "p_value"     = model_summary$coefficients[2,4],
                                   "r_squared"    = model_summary$r.squared)

model_summary_statistics

```

```

## # A tibble: 1 x 4
##   intercept slope p_value r_squared
##   <dbl>   <dbl>   <dbl>   <dbl>
## 1     6.53 -0.223   0.152   0.0138

```

```

mssT <- model_summary_statistics %>%
  mutate(across(where(is.numeric), ~ round(.x, digits = 3)))

knitr::kable(mssT,
  col.names = c('Intercept', 'Slope', 'p-Value', 'R2'),
  align = "cccc",
  caption = "Results of a linear model with one continuous predictor and response"
)

```

```

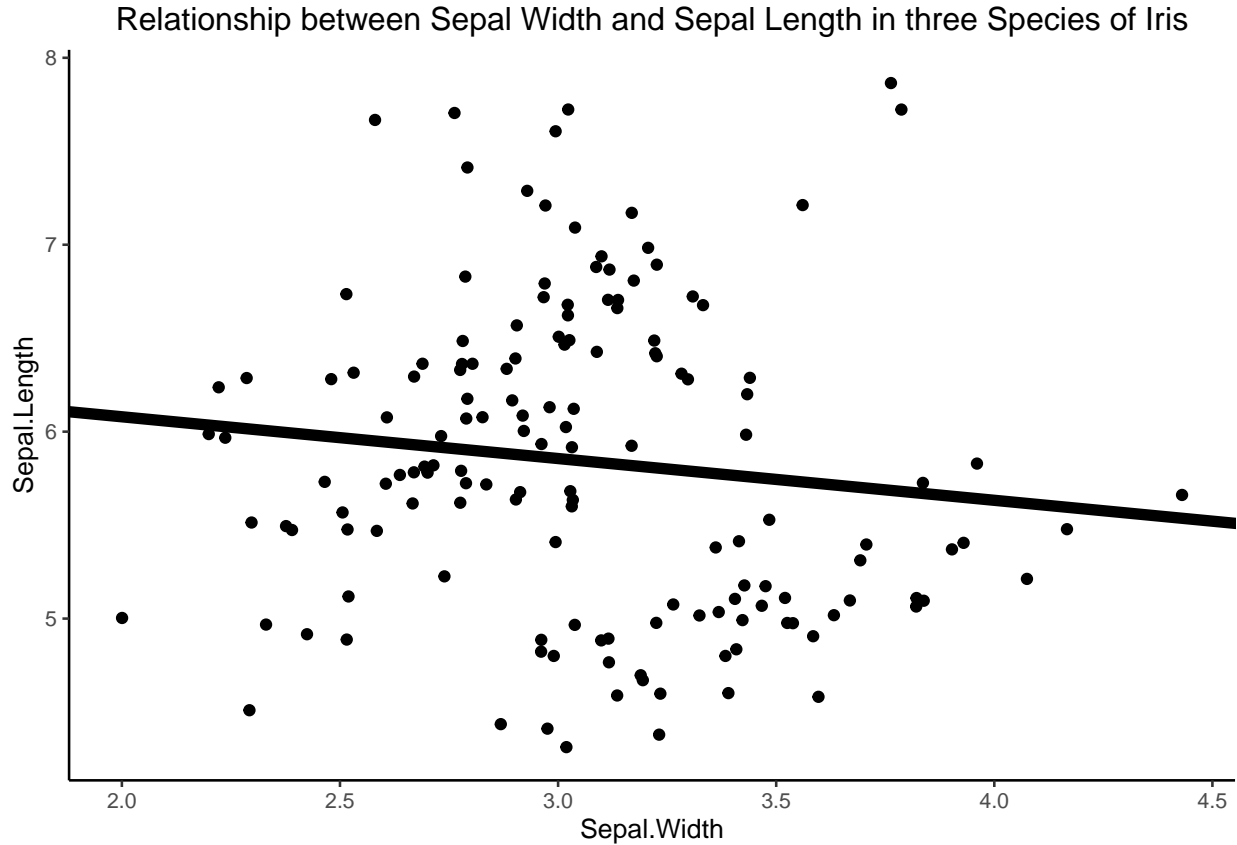
ggplot(iris, aes(x = Sepal.Width, y = Sepal.Length))+
  geom_jitter()+
  geom_abline(intercept = model_summary$coefficients[1,1],

```

Table 1: Results of a linear model with one continuous predictor and response

Intercept	Slope	p-Value	R ²
6.526	-0.223	0.152	0.014

```
slope = model_summary$coefficients[2,1], size = 2) +
theme_classic(base_size = 10) +
labs(title = 'Relationship between Sepal Width and Sepal Length in three Species of Iris') +
theme(plot.title = element_text(hjust = 0.5))
```



```
model <- lm(Sepal.Length ~ Sepal.Width, data = iris)
conf_int <- as.data.frame(predict(model, interval = "confidence"))
data <- cbind(conf_int, iris)

ggplot(data, aes(x = Sepal.Width, y = Sepal.Length)) +
  geom_jitter() +
  geom_abline(intercept = model_summary$coefficients[1,1],
              slope = model_summary$coefficients[2,1], size = 2) +
  theme_classic(base_size = 10) +
  geom_line(aes(y = lwr), color = "black", linetype = "dashed") +
  geom_line(aes(y = upr), color = "black", linetype = "dashed") +
  geom_ribbon(aes(ymin=lwr, ymax=upr), alpha=0.1) +
  labs(title = 'Relationship between Sepal Width and Sepal Length in three Species of Iris') +
  theme(plot.title = element_text(hjust = 0.5))
```

Relationship between Sepal Width and Sepal Length in three Species of Iris

