Homework_4

Emmenta Janneh

2024-02-11

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.2
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✓ dplyr     1.1.2     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.3     ✓ tibble    3.2.1
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────
tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```
library(broom)
```

```
## Warning: package 'broom' was built under R version 4.3.2
```

Conceptual Exercises

1. The standard error of β1 would decrease when a new observational unit is obtained with Xnew = Xmean. When Xnew = Xmean, the distance of each data point from the mean, Xi - Xmean, remains the same for all data points, including the new one. Therefore, the denominator of the formula remains constant.

Since the denominator remains constant and Var is also assumed to be constant, the standard error SE would decrease as n (the sample size) increases with the addition of the new observational unit.

2.

```
2 * pt(-7.52, df = 13)
```

```
## [1] 4.373129e-06
```

```
0.997 / 0.111
```

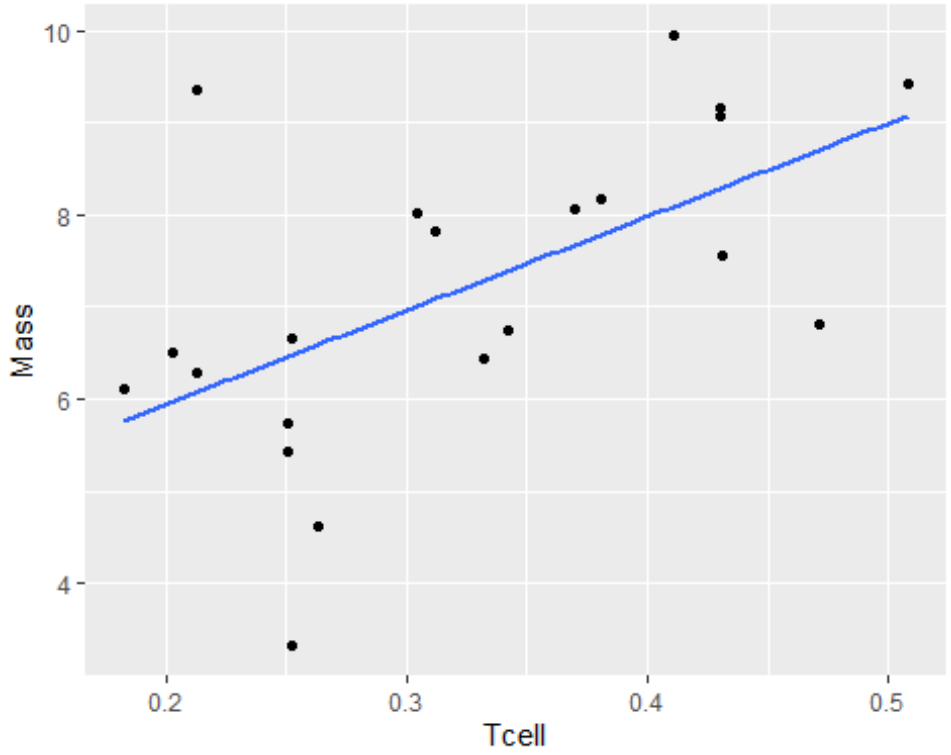
```
## [1] 8.981982
```

A tibble: 2 x 5
term estimate std.error statistic p.value
<chr> <dbl> <dbl> <dbl> <dbl>
1 (Intercept) 8.39 1.11 7.52 0.00000437
2 Years 0.997 0.111 8.982 0.000000618

3.
```
estimate <- c(8.39, 0.997)
std_error <- c(1.11, 0.111)
statistic <- c(7.52, 8.982)
df <- 15 - 2
t_critical <- qt(0.975, df)

lower_ci <- estimate - t_critical * std_error
upper_ci <- estimate + t_critical * std_error

ci_data <- data.frame(term = c("(Intercept)", "Years"),
                      lower_ci = lower_ci,
                      upper_ci = upper_ci
                      )
ci_data

##          term  lower_ci  upper_ci
## 1 (Intercept) 5.9919908 10.788009
## 2       Years 0.7571991  1.236801
```

Wheatears

```
library(Sleuth3)

## Warning: package 'Sleuth3' was built under R version 4.3.2

data("ex0727")
data("ex0727")
wheat <- ex0727
glimpse(wheat)

## Rows: 21
## Columns: 2
## $ Mass  <dbl> 3.33, 4.62, 5.43, 5.73, 6.12, 6.29, 6.45, 6.51, 6.65, 6.75,
6.81…
## $ Tcell <dbl> 0.252, 0.263, 0.251, 0.251, 0.183, 0.213, 0.332, 0.203,
0.252, 0…

lmout <- lm(Mass ~ Tcell, data = wheat)
ggplot(wheat, aes(x = Tcell, y = Mass)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE)

## `geom_smooth()` using formula = 'y ~ x'
```



1.

```
n <- nrow(wheat)
n

## [1] 21

tidy(lmout, conf.int = TRUE)

## # A tibble: 2 × 7
##   term        estimate std.error statistic p.value conf.low conf.high
##   <chr>          <dbl>     <dbl>     <dbl>   <dbl>    <dbl>     <dbl>
## 1 (Intercept)     3.91      1.11      3.52 0.00230     1.58      6.24
## 2 Tcell          10.2       3.30      3.08 0.00611     3.27     17.1
```

Form the above information, we have a positive linear association between Tcell and Mass. `3.91` is the y-intercept of the regression line. Birds with `1mm` higher Tcell are estimated to carry on average `10.2g` higher stone mass (95% confidence interval of `3.3g` lower to `17.1g` higher). With a strong evidence that this association is not due to chance alone *(p = 0.006, n = 21)*.

2.

```
newbird <- data.frame(Tcell = c(0.35))
predict(object = lmout, newdata = newbird, interval = "confidence") %>%
cbind(newbird)
```

```
##        fit      lwr      upr Tcell
## 1 7.469063 6.793505 8.144621  0.35
```

We estimate that a bird with Tcell of 0.35mm can carry on average a stone with mass `7.5g` (95% confidence interval of 6.8g lower to 8.1g higher).

3.

```
newbird <- data.frame(Tcell = c(0.2))
predict(object = lmout, newdata = newbird, interval = "confidence")

##        fit      lwr     upr
## 1 5.944295 4.869499 7.01909
```

Birds with 0.2mm Tcell are estimated to carry on average `5.9g` mass of stone (95% confidence interval of 4.9g lower to 7.0g higher mass).

4.

```
advbird <- data.frame(Tcell = c(20))
predict(object = lmout, newdata = advbird, interval = "predict")

##        fit      lwr      upr
## 1 207.2137 71.45158 342.9758
```

We estimate that a "super" bird with Tcell of 20mm can carry on average `207.2g` (95% prediction interval of 71.5g lower to 342.9g higher).