hw_04

lisa liubovich

2024-02-14

Conceptual Exercises

1.

The standard error of B1^ would decrease. Mathematically, adding an observation would increase the value of the denominator in the standard error equation, thus decreasing the standard error. Intuitively, the standard error would decrease because an added observation that aligns with the mean of predictors helps to reduce the uncertainty in estimation by giving us more precise information about the slope of the regression line.

2.

The missing p-value for the intercept is less than 0.0001.

The formula for a t statistic is B1^/s(B1^). The Estimate is 0.997 and the standard error is 0.111. 0.997/0.111 is 8.981982.

```
0.997/0.111
```

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

3.

95% confidence interval: estimate ± multiplier * standard error = 0.997 ± 1.96 * 0.111 = (0.77944, 1.21456)

```
0.997 + 1.96 * 0.111
```

```
## [1] 1.21456
```

```
0.997 - 1.96 * 0.111
```

```
## [1] 0.77944
```

We estimate that the subjects have on average 0.997 units higher neuron activity for each additional year of experience. We have strong evidence that this association is not due to chance alone ($n = 15$, p-value $<$ 0.001).

We are 95% confident that the true population parameter falls between 1.21456 and 0.77944.

Because the confidence interval does not contain 0, we have even more evidence to conclude that the coefficient of B1 is statistically significant; that is, we have strong evidence to suggest that this association is not due to chance alone.

Wheatears

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(Sleuth3)
```

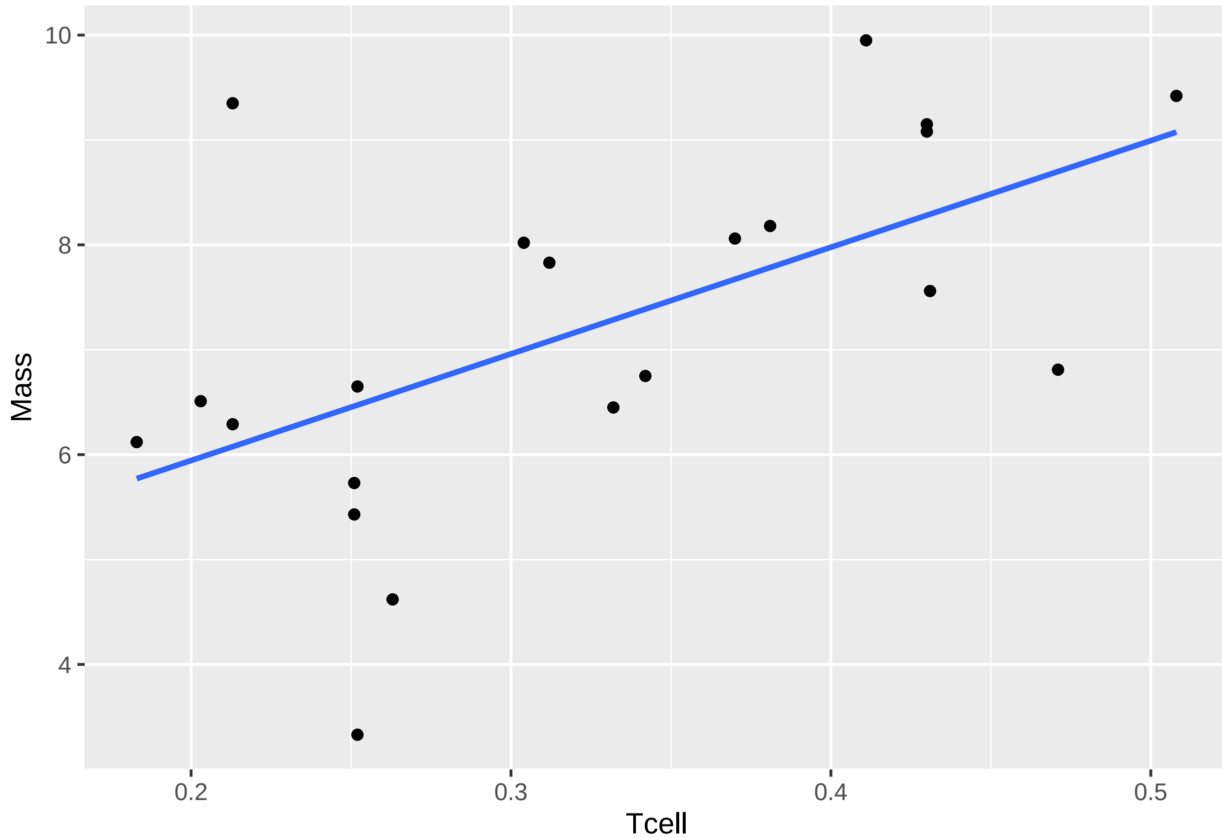
```
## Warning: package 'Sleuth3' was built under R version 4.3.2
```

```r
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~
```

```r
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.

sample size (n) = 21

H0: B1 = 0

Ha: B1 =/= 0

```
library(broom)
t_out <- tidy(lmout)
t_out
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic p.value
##   <chr>          <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)     3.91      1.11      3.52 0.00230
## 2 Tcell          10.2       3.30      3.08 0.00611
```

p-value of B1 = 0.006105022 –> 0.006105022 < 0.01

therefore, we reject the null hypothesis that the coefficient B1 is 0 at a 1% significance level.

estimated slope: 10.16512. We estimate that the average mass of stones carried by a bird on average are 10.16512 higher for each additional millimeter of a t-cell.

```
t_out <- tidy(lmout, conf.int = TRUE)
select(t_out, term, estimate, p.value, conf.low, conf.high)
```

```
## # A tibble: 2 x 5
##   term        estimate p.value conf.low conf.high
##   <chr>          <dbl>   <dbl>    <dbl>     <dbl>
## 1 (Intercept)     3.91 0.00230     1.58      6.24
## 2 Tcell          10.2  0.00611     3.27     17.1
```

The 95% confidence interval is 3.267001 to 17.06324. Because 0 is not in the interval, we have further strong evidence to indicate that the association between Mass and T-cell response measurement is not due to chance alone.

2.

```
3.91127 + (10.16512 * 0.35)
```

```
## [1] 7.469062
```

I would tell them that the estimated mass for such a measure would be approximately 7.469062 grams. I would emphasize that this is merely an estimate. We have a standard error of 3.295768, so a 95% confidence interval for this estimate is 1.009457 to 13.92877 grams. That is, we are 95% confident that the true mass falls between 1.009357 and 13.92877.

```
7.469062 + 1.96 * 3.295768
```

```
## [1] 13.92877
```

```
7.469062 - 1.96 * 3.295768
```

```
## [1] 1.009357
```

3.

sample-size (n) = 100

estimated slope = 10.16512

```
3.91127 + (10.16512 * 0.2)
```

```
## [1] 5.944294
```

Based on our prediction equation alone, the estimated average weight the birds can carry at the level of 0.2mm is 5.944294 grams. However, this is merely an estimate. Since we have no data for the 100 birds in this sample beyond their t-cell measurements, we cannot calculate the standard error and thus a confidence interval. I would emphasize that 5.944294 grams is merely a prediction and the true average weight might be higher or lower.

4.

```r
3.91127 + (10.16512 * 20)
```

```
## [1] 207.2137
```

Based on the prediction equation alone, the predicted mass is 207.2137. However, I would remind my colleague of a couple things:

First, we must consider the ethical considerations of "building" a bird like that. Birds' sizes are a natural result of evolutionary advantage and natural selection, and interfering with this (even just out of scientific curiosity) could have wide reaching consequences for ecosystems and species of bird.

The number of stones carried also depends on their individual weights. It is realistic for a bird to carry several stones of light weight, but to carry 207 grams of weight is unrealistic unless this is a very large bird.