

Assessment: Tibbles, do, and broom, part 1

Question 1

1/1 point (graded)

As seen in the videos, what problem do we encounter when we try to run a linear model on our baseball data, grouping by home runs?

- ☐ There is not enough data in some levels to run the model.
- ☒ The `lm` function does not know how to handle grouped tibbles.
- ☐ The results of the `lm` function cannot be put into a tidy format.



Explanation

The `lm` function does not know how to handle grouped tibbles, so we can't simply run a linear model on the baseball data grouped by home runs. We need something to bridge between the grouped tibble and the `lm` function.

Submit

You have used 1 of 1 attempt

i Answers are displayed within the problem

Question 2

1/1 point (graded)

Tibbles are similar to what other class in R?

- ☐ Vectors
- ☐ Matrices

☒ Data frames

☐ Lists



Explanation

Tibbles are essentially modern versions of data frames.

Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem

Question 3

1/1 point (graded)

What are some advantages of tibbles compared to data frames?

Select ALL that apply.

☒ Tibbles display better.

☒ If you subset a tibble, you always get back a tibble.

☒ Tibbles can have complex entries.

☒ Tibbles can be grouped.



Answer

Correct:

Correct, this is one advantage of a tibble. There are several other correct answers listed here.

Correct, this is one advantage of a tibble. There are several other correct answers listed here.

Correct, this is one advantage of a tibble. There are several other correct answers listed here.

Correct, this is one advantage of a tibble. There are several other correct answers listed here.

Explanation

All of the listed answers are advantages of tibbles when compared to data frames: tibbles display better, they always return tibbles when subsetting, they can have complex entries, and they can be grouped.

Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem

Question 4

1/1 point (graded)

What are two advantages of the `do` command, when applied to the tidyverse?

Select TWO.

☐ It is faster than normal functions.

☐ It returns useful error messages.

☒ It understands grouped tibbles.

☒ It always returns a `data.frame`.



Answer

Correct:

Correct. The `do` function can understand grouped tibbles.

Correct. The `do` function always returns a `data.frame`.

Explanation

The `do` function serves as a useful bridge between base R functions and the tidyverse. It understands grouped tibbles and always returns a `data.frame`.

Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem

Question 5

1/1 point (graded)

You want to take the tibble `dat`, which we used in the video on the `do` function, and run the linear model $R \sim BB$ for each strata of HR. Then you want to add three new columns to your grouped tibble: the coefficient, standard error, and p-value for the BB term in the model.

You've already written the function `get_slope`, shown below.

```
get_slope <- function(data) {
  fit <- lm(R ~ BB, data = data)
  sum.fit <- summary(fit)

  data.frame(slope = sum.fit$coefficients[2, "Estimate"],
             se = sum.fit$coefficients[2, "Std. Error"],
             pvalue = sum.fit$coefficients[2, "Pr(>|t|)"])
}
```

What additional code could you write to accomplish your goal?



```
dat %>%
  group_by(HR) %>%
  do(get_slope)
```



```
dat %>%
  group_by(HR) %>%
  do(get_slope(.))
```



```
dat %>%
  group_by(HR) %>%
  do(slope = get_slope(.))
```



```
dat %>%
  do(get_slope(.))
```



Answer

Correct:

Correct. This will create a tibble with four columns: HR, slope, se, and pvalue for each level of HR.

Explanation

```
dat %>%
  group_by(HR) %>%
  do(get_slope(.))
```

This is the only command that correctly creates a tibble with four columns: HR, slope, se, and pvalue for each level of HR. The data frame must be passed to `get_slope` using `.`. If you name the results of the `do` command such as in the code `do(slope = get_slope(.))`, that will save all results in a single

column called `slope`. If you forget `group_by`, then the results will be a model on the data as a whole, rather than on the data stratified by home runs.

Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem

Question 6

1/1 point (graded)

The output of a `broom` function is always what?

☒ A data.frame

☐ A list

☐ A vector



Explanation

The `broom` function always outputs a data.frame.

Submit

You have used 1 of 1 attempt

i Answers are displayed within the problem

Question 7

1/1 point (graded)

You want to know whether the relationship between home runs and runs per game varies by baseball league. You create the following dataset:

```
dat <- Teams %>% filter(yearID %in% 1961:2001) %>%  
  mutate(HR = HR/G,  
         R = R/G) %>%  
  select(lgID, HR, BB, R)
```

What code would help you quickly answer this question?



```
dat %>%
  group_by(lgID) %>%
  do(tidy(lm(R ~ HR, data = .), conf.int = T)) %>%
  filter(term == "HR")
```



```
dat %>%
  group_by(lgID) %>%
  do(glance(lm(R ~ HR, data = .)))
```



```
dat %>%
  do(tidy(lm(R ~ HR, data = .), conf.int = T)) %>%
  filter(term == "HR")
```



```
dat %>%
  group_by(lgID) %>%
  do(mod = lm(R ~ HR, data = .))
```



Answer

Correct: Correct. This is a good application of the command `tidy`, from the `broom` package.

Explanation

```
dat %>%
  group_by(lgID) %>%
  do(tidy(lm(R ~ HR, data = .), conf.int = T)) %>%
  filter(term == "HR")
```

This code is a good application of the command `tidy`, from the `broom` package.

The `glance` function provides data on model fit rather than on effect estimates and confidence intervals. If you forget the line `group_by(lgID)`, your code will give you a single estimate for the entire dataset because you have not grouped the data by league ID.

```
dat %>%
  group_by(lgID) %>%
  do(mod = lm(R ~ HR, data = .))
```

This code gives get a data.frame with the column `mod`, which contains the linear model results. While it is possible to then extract effect estimates and confidence intervals from this model, it is not nearly as easy as using the tidy function.

Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem