

Module 2 Assignment 3 tidyverse

Ellen Bledsoe

2023-03-20

Curious how we would do all of this in `tidyverse`?

It takes fewer steps, but we haven't been using `tidyverse` very often, and not in this way before.

```
library(tidyverse)
grasshoppers <- read_csv("../data_raw/grasshoppers.csv")
```

Step 1: Calculate Population Parameter Estimates per Stratum

Use `group_by()`, `summarize()`, and `mutate()` to calculate parameter estimates

```
grasshoppers <- grasshoppers %>%
  # create groups for each stratum
  group_by(stratum) %>%
  # calculate the means, standard deviations, and variances for each stratum
  summarise(y_bar = mean(abund),
            sd = sd(abund),
            var = var(abund),
            n = n()) %>% # how many rows per stratum AKA how many units per stratum
  # create a new column to put in the total number of units per stratum
  mutate(N = if_else(stratum == "A", 200,
                    if_else(stratum == "B", 500, 300)),
         tau_hat = y_bar * N) # calculate totals per stratum
grasshoppers
```

```
## # A tibble: 3 x 7
##   stratum y_bar   sd   var     n     N tau_hat
##   <chr>   <dbl> <dbl> <dbl> <int> <dbl>   <dbl>
## 1 A      31.9  5.21  27.2   37   200   6373.
## 2 B      35.5  6.07  36.8   36   500  17736.
## 3 C      40    5.63  31.7   27   300  12000
```

Step 2: Calculate Measures of Uncertainty per Stratum

Use `mutate()` to calculate variance of the estimates for each stratum.

```
grasshoppers <- grasshoppers %>%
  # create new columns for:
  # 1) finite population correction factor,
  # 2) variance of the estimate of the mean, and
```

```
# 3) variance of the estimate of the total
mutate(pop_correction = (N-n)/N,
       var_y_bar = pop_correction * (var/n),
       var_tau_hat = N^2 * var_y_bar)
grasshoppers
```

```
## # A tibble: 3 x 10
##   stratum y_bar    sd   var     n     N tau_hat pop_correction var_y_bar var_t-1
##   <chr>   <dbl> <dbl> <dbl> <int> <dbl>   <dbl>         <dbl>    <dbl>    <dbl>
## 1 A      31.9  5.21  27.2   37   200   6373.         0.815    0.599  23944.
## 2 B      35.5  6.07  36.8   36   500  17736.         0.928    0.949 237335.
## 3 C      40    5.63  31.7   27   300  12000         0.91     1.07   96133.
## # ... with abbreviated variable name 1: var_tau_hat
```

Step 3: Calculate Overall Totals and Means

Use the `summarize()` function again to calculate the overall values.

```
grasshoppers_overall <- grasshoppers %>%
  summarize(overall_tau_hat = sum(tau_hat),
           overall_var_tau_hat = sum(var_tau_hat),
           overall_y_bar = overall_tau_hat/sum(N),
           overall_var_y_bar = overall_var_tau_hat / sum(N)^2) %>%
  slice(1) # still thinks we want groups, so just select the first row (all rows have the same values)
grasshoppers_overall
```

```
## # A tibble: 1 x 4
##   overall_tau_hat overall_var_tau_hat overall_y_bar overall_var_y_bar
##   <dbl>           <dbl>           <dbl>           <dbl>
## 1      36109.         357412.         36.1           0.357
```