|  |
| --- |
| **Resampling from a Dataset**  rep\_sample\_n(<NAME OF DATASET>,  size = <SIZE OF SAMPLE>,  replace = TRUE,  reps = <NUMBER OF RESAMPLES>)  ***Note:*** The value of TRUE for replace means sampling is done *with replacement*. For sampling *without replacement*, you would need to set replace equal to FALSE. |
| **Working in the infer Package Workflow** |
| **Obtaining the Sample Slope**  obs\_slope <- <NAME OF DATASET> %>%  specify(response = <NAME OF VARIABLE>,  explanatory = <NAME OF VARIABLE>) %>%  calculate(stat = "slope")  ***Note:*** This step **must** be done **first**, before you find your confidence interval! |
| **Obtaining 1000 Bootstrap Slope Statistics**  bootstrap\_dist <- <NAME OF DATASET> %>%  specify(response = <NAME OF NUMERICAL VARIABLE>,  explanatory = <NAME OF CATEGORICAL VARIABLE>) %>%  generate(reps = 1000, type = "bootstrap") %>%  calculate(stat = "slope")  ***Note:*** You choose the number of reps. I recommend choosing at least 1000, to get a good idea of the shape of the bootstrap distribution – remember we need to verify it is approximately normal. |
| **Plotting the Bootstrap Distribution**  visualize(data = bootstrap,  method = “simulation”)  ***Note:*** You can add axis labels to this plot! All you need to do is connect the visualize() step to labs() using a **+** sign. |

|  |
| --- |
| **Obtaining a Percentile Confidence Interval from a Bootstrap Distribution**  get\_confidence\_interval(x = bootstrap,  level = 0.95,  type = “percentile”)  ***Note:*** You choose the confidence level of your interval! |
| **Obtaining an SE Confidence Interval from a Bootstrap Distribution**  get\_confidence\_interval(x = bootstrap,  level = 0.95,  type = “se”,  point\_estimate = obs\_slope)  ***Note:*** You choose the confidence level of your interval! |

|  |
| --- |
| **Obtaining the Sample Difference in Means**  obs\_diff\_in\_means <- <NAME OF DATASET> %>%  specify(response = <NAME OF NUMERICAL VARIABLE>,  explanatory = <NAME OF CATEGORICAL VARIABLE>) %>%  calculate(stat = "diff in means",  order = c(“<NAME OF FIRST GROUP>”, “<NAME OF FIRST GROUP>”)  )  ***Note:*** *This step* ***must*** *be done* ***before*** *you find your confidence interval and before finding your p-value!* |
| **Obtaining a Confidence Interval from a Bootstrap Distribution**  get\_confidence\_interval(x = bootstrap,  level = 0.95,  type = “percentile”,  point\_estimate = obs\_mean)  ***Note:*** *This step* ***must*** *come after you have obtained the bootstrapped differences in means* ***and*** *the observed difference in means!*  ***Note:*** If you want a 90% confidence interval, you change level to 0.90 |
| **Obtaining 1000 Permuted Differences in Means – Assuming the Null Hypothesis is True**  null\_dist <- <NAME OF DATASET> %>%  specify(response = <NAME OF NUMERICAL VARIABLE>,  explanatory = <NAME OF CATEGORICAL VARIABLE>) %>%  generate(reps = 1000, type = "permute") %>%  calculate(stat = "diff in means",  order = c(“<NAME OF FIRST GROUP>”, “<NAME OF FIRST GROUP>”)  ) |

|  |
| --- |
| **Plotting the Simulated Null Distribution**  visualize(data = null\_dist,  method = “simulation”)  ***Note:*** *This step* ***must*** *come after you have obtained the permuted differences in means!* |
| **Obtaining a p-value from a Null Distribution**  get\_pvalue(x = null\_dist,  obs\_stat = obs\_diff\_in\_means,  direction = “two-sided”)  ***Note:*** *This step* ***must*** *come after you have obtained the bootstrapped differences in means* ***and*** *the observed difference in means!*  ***Note:*** If you are doing a one-sided hypothesis test, you change alternative to either “greater” or “less” |
|  |
|  |

|  |
| --- |
| **Calculating Summary Statistics for One Numerical Variable and One Categorical Variable**  favstats(<NAME OF NUMERICAL VARIABLE> ~ <NAME OF CATEGORICAL VARIABLE>,  data = <NAME OF DATASET>)  ***Note:*** The ~ **must** be included! |
| **Obtaining an ANOVA Table**  aov(<NAME OF NUMERICAL VARIABLE> ~ <NAME OF CATEGORICAL VARIABLE>,  data = <NAME OF DATASE>) |

|  |
| --- |
| **Plotting the Bootstrap Distribution**  visualize(data = bootstrap,  method = “simulation”)  ***Note:*** *This is* ***the same*** *as plotting the bootstrap for one mean!* |
| **Obtaining the Sample Slope**  obs\_mean <- <NAME OF DATASET> %>%  specify(response = <NAME OF Y-VARIABLE>,  explanatory = <NAME OF X-VARIABLE>) %>%  calculate(stat = "mean")  ***Note:*** *This step* ***must*** *be done* ***before*** *you find your confidence interval!* |
| **Obtaining a Confidence Interval from a Bootstrap Distribution**  get\_confidence\_interval(x = bootstrap,  level = 0.95,  type = “percentile”,  point\_estimate = obs\_mean)  ***Note:*** *This is* ***the same*** *as how you found a confidence interval for one mean!* |