|  |
| --- |
| **Performing a t-test for a Difference in Means**  t\_test(x = <NAME OF DATASET>,  response = <NAME OF NUMERICAL VARIABLE>,  explanatory = <NAME OF CATEGORICAL VARIABLE>,  conf\_int = TRUE,  conf\_level = 0.90,  alternative = "two-sided")  ***Note:*** If you want a 95% confidence interval, you change conf\_level to 0.95  ***Note:*** If you are doing a one-sided hypothesis test, you change alternative to either “greater” or “less” |
| **Plotting a One-Variable Bar Plot with Counts**  ggplot(data = <NAME OF DATASET>,  mapping = aes(x = <NAME OF VARIABLE>)) +  geom\_bar() +  labs(x = "<TITLE FOR THE X-AXIS>")  ***Note:*** This bar plot has the variable names on the x-axis. If the names are squished, then you should use **y =** instead of **x =** . |
| **Plotting a One-Variable Bar Plot with Proportions**  ggplot(data = <NAME OF DATASET>,  mapping = aes(x = <NAME OF VARIABLE>**,** y = ..prop.., group = 1)) +  geom\_bar() +  labs(x = "<TITLE FOR THE X-AXIS>")  ***Note:*** This bar plot has the variable names on the x-axis. If the names are squished, then you should use **y =** instead of **x =** . |
| **Plotting a Two-Variable Bar Plot**  ggplot(data = <NAME OF DATASET>,  mapping = aes(x = <NAME OF VARIABLE 1>,  fill = <NAME OF VARIABLE 2>)) +  geom\_bar(position = “stack”) +  labs(x = "<TITLE FOR THE X-AXIS>",  fill = "<TITLE FOR THE COLOR LEGEND")  ***Note:*** You should fill by whichever variable has **fewer** values.  ***Note:*** If you want a side-by-side bar plot you need to change position to “dodge”. If you want a filled bar plot, you need change position to “fill”. |
| **Creating a Summary Table of Observations of One Variable**  <NAME OF DATASET> |>  count(<NAME OF VARIABLE>) |

|  |
| --- |
| **Creating a Summary Table of Observations from Two Variables**  <NAME OF DATASET> |>  count(<NAME OF VARIABLE 1>, <NAME OF VARIABLE 2>) |
| **Creating a Contingency Table of Observations from Two Variables**  <NAME OF DATASET> |>  count(<NAME OF RESPONSE VARIABLE>, <NAME OF EXPLANATORY VARIABLE>)|>  pivot\_wider(names\_from = <NAME OF EXPLANATORY VARIABLE>,  values\_from = n) |>  adorn\_totals(where = c(“row”, “col”))  ***Note:*** Your explanatory variable should be in the rows and your response variable should be in the columns. So, the variable you insert into names\_from should be the response variable you are interested in. |
| **Performing a Chi-Squared Goodness-of-Fit Test (One Variable)**  chisq\_test(x = <NAME OF DATASET>,  response = <NAME OF VARIABLE>) |
| **Performing a Chi-Squared Independence / Homogeneity Test (Two Variables)**  chisq\_test(x = <NAME OF DATASET>,  response = <NAME OF RESPONSE VARIABLE>,  explanatory = <NAME OF EXPLANATORY VARIABLE>) |
| **Obtaining the Sample X-Squared Statistic**  obs\_xsq <- <NAME OF DATASET> |>  specify(response = <NAME OF RESPONSE VARIABLE>,  explanatory = <NAME OF EXPLANATORY VARIABLE>) |>  calculate(stat = "Chisq")  ***Note:*** *This step* ***must*** *be done* ***before*** *you find your p-value!* |
| **Obtaining 1000 Permuted X-Squared Statistics – Assuming the Null Hypothesis is True**  null\_dist <- <NAME OF DATASET> |>  specify(response = <NAME OF RESPONSE VARIABLE>,  explanatory = <NAME OF EXPLANATORY VARIABLE>) |>  hypothesize(null = “independence”) |>  generate(reps = 1000, type = "permute") |>  calculate(stat = "Chisq") |
|  |

|  |
| --- |
| **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\_xsq,  direction = “greater”)  ***Note:*** *This step* ***must*** *come after you have obtained the bootstrapped differences in means* ***and*** *the observed difference in means!*  ***Note:*** In a Chi-Squared test we **always** use a greater than alternative, since we only look in the right tail! |