FDS::CHEATSHEETS["project 4"]

2 sample t-test

Case use:
- When you have a numerical variable and two groups and you want to test if the average of this numerical variable is different in the two groups

Underlying assumption:
- The observations in the two groups are independent and normally distributed ()

To perform a t-test on tidy data, we use **infer::t_test().** **e.g.:**

```
library(infer)
?gss # for more info on the dataset

gss %>%
  tidyr::drop_na(college) %>%
  t_test(formula = hours ~ college,
    order = c("degree", "no degree"),
    alternative = "two-sided")
```

Example using the gss dataset, which is included in the {infer} package

Variables:
- *college*: categorical variable with values degree and no degree
- *hours*: average number of hours worked per week

T-test is carried to respond to the following question: Is there a statistically significant difference at 95% confidence level in the average number of hours worked per week, by college degree status?

```
#> # A tibble: 1 × 7
#>   statistic t_df p_value alternative estimate lower_ci upper_ci
#>       <dbl> <dbl>   <dbl> <chr>          <dbl>    <dbl>    <dbl>
#> 1      1.12  366.   0.264 two.sided       1.54    -1.16     4.24
```

Interpretation of results:

As the p-value is greater than 0.05, we cannot reject the null hypothesis that there is no difference in the average number of hours works, between student with and without a college degree. In other words, we do not have enough evidence to state that there is a difference.

X^2 Chi-square test of independence:

Case use:
- When you want to test if two categorical variables are associated (not independent)

Underlying assumption:
-

To perform a chi-square test on tidy data, we use **infer::chisq_test().** **e.g.:**

```
library(infer)

gss %>%
  chisq_test(college ~ finrela)
```

Example using the gss dataset, which is included in the {infer} package

Variables:
- *college*: categorical variable with values degree and no degree
- *finrela*: categorical variable that gives the respondent's self-identification of family income

This chi-squared test is carried to respond to the following question: Is there an association between income (self identified, in this case) and educational attainment, in the studied sample? Or, in other words, are the differences in educational attainment between the different income classes only due to random noise?

```
#> # A tibble: 1 × 3
#>   statistic chisq_df p_value
#>       <dbl>    <dbl>   <dbl>
#> 1      30.7       5. 0.0000108
```

Interpretation of results:

Since the p-value is < 0.05, we can reject the null hypothesis that there is no association between income and educational attainment.