

# Hypothesis Testing

Felix Vo

2025-03-18

```
#Import libraries:
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.4      v tidyr     1.3.1
## v purrr      1.0.4
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(knitr)
```

## Hypothesis Testing:

- Statistical relationship between mental health disorder and number of commit suicide

```
df_canada <- mental_health_data %>%
  filter(Geography == "Canada", Gender == "Total, gender of person")
suicide_indicator <- "Commit Suicide"

mental_health_indicators <- c("Major depressive episode, life",
                              "Eating disorder, current diagnosed condition",
                              "Suicidal thoughts, life",
                              "Suicidal thoughts, 12 months",
                              "Social phobia, life",
                              "Generalized anxiety disorder, life",
                              "Bipolar disorder, life")

df_filtered <- df_canada %>%
  filter(Indicators %in% c(suicide_indicator, mental_health_indicators)) %>%
  select(Year, Indicators, Population) %>%
  group_by(Year, Indicators) %>%
  summarise(Population = sum(Population, na.rm = TRUE), .groups = "drop")

df_pivot <- df_filtered %>%
```

```

pivot_wider(names_from = Indicators, values_from = Population)

print(df_pivot)

```

```

## # A tibble: 3 x 9
##   Year 'Commit Suicide' Eating disorder, current diagn-1 Major depressive epi-2
##   <int>          <int>          <int>          <int>
## 1  2002          3650          243070         6074098
## 2  2012          4100          223445         6368844
## 3  2022          3567          457800         8625900
## # i abbreviated names: 1: 'Eating disorder, current diagnosed condition',
## #   2: 'Major depressive episode, life'
## # i 5 more variables: 'Suicidal thoughts, 12 months' <int>,
## #   'Suicidal thoughts, life' <int>, 'Bipolar disorder, life' <int>,
## #   'Generalized anxiety disorder, life' <int>, 'Social phobia, life' <int>

```

```

# Handle missing values (replace NA with median)
df_pivot_filled <- df_pivot %>%
  mutate(across(where(is.numeric), ~ ifelse(is.na(.), median(., na.rm = TRUE), .)))

# Correlation analysis
cor_results <- data.frame(
  Indicator = character(),
  Correlation = numeric(),
  t_value = numeric(),
  p_value = numeric(),
  stringsAsFactors = FALSE
)

for (indicator in mental_health_indicators) {
  if (indicator %in% colnames(df_pivot_filled)) {
    cor_test <- cor.test(df_pivot_filled[[indicator]], df_pivot_filled[["Commit Suicide"]])
    cor_results <- rbind(cor_results, data.frame(
      Indicator = indicator,
      Correlation = round(cor_test$estimate, 3),
      t_value = round(cor_test$statistic, 3),
      p_value = round(cor_test$p.value, 4)
    ))
  } else {
    cor_results <- rbind(cor_results, data.frame(
      Indicator = indicator,
      Correlation = NA,
      t_value = NA,
      p_value = NA
    ))
  }
}

```

```

## Warning in cor(x, y): the standard deviation is zero

```

```

# Display results in a table
kable(cor_results, caption = "Pearson Correlation Between Mental Health Indicators and Suicide Rates")

```

Table 1: Pearson Correlation Between Mental Health Indicators and Suicide Rates

	Indicator	Correlation	t_value	p_value
cor	Major depressive episode, life	-0.534	-0.631	0.6416
cor1	Eating disorder, current diagnosed condition	-0.678	-0.921	0.5261
cor2	Suicidal thoughts, life	-0.556	-0.669	0.6248
cor3	Suicidal thoughts, 12 months	-0.544	-0.648	0.6339
cor4	Social phobia, life	NA	NA	NA
cor5	Generalized anxiety disorder, life	-0.929	-2.515	0.2409
cor6	Bipolar disorder, life	-0.929	-2.515	0.2409

## Suicide Thoughts with other Indicators Pearson correlation test

```

suicide_indicators <- c("Suicidal thoughts, life", "Suicidal thoughts, 12 months")

mental_health_indicators <- c("Major depressive episode, life",
                              "Eating disorder, current diagnosed condition",
                              "Social phobia, life",
                              "Generalized anxiety disorder, life",
                              "Bipolar disorder, life")

df_filtered <- df_canada %>%
  filter(Indicators %in% c(suicide_indicators, mental_health_indicators)) %>%
  select(Year, Indicators, Population) %>%
  group_by(Year, Indicators) %>%
  summarise(Population = sum(Population, na.rm = TRUE), .groups = "drop")

df_pivot <- df_filtered %>%
  pivot_wider(names_from = Indicators, values_from = Population)

df_pivot_filled <- df_pivot %>%
  mutate(across(where(is.numeric), ~ ifelse(is.na(.), median(., na.rm = TRUE), .)))

cor_results <- data.frame(
  Indicator = character(),
  Suicide_Indicator = character(),
  Correlation = numeric(),
  t_value = numeric(),
  p_value = numeric(),
  stringsAsFactors = FALSE
)

for (suicide in suicide_indicators) {
  for (indicator in mental_health_indicators) {
    if (indicator %in% colnames(df_pivot_filled) && suicide %in% colnames(df_pivot_filled)) {
      cor_test <- cor.test(df_pivot_filled[[indicator]], df_pivot_filled[[suicide]])
      cor_results <- rbind(cor_results, data.frame(
        Indicator = indicator,
        Suicide_Indicator = suicide,
        Correlation = round(cor_test$estimate, 3),

```

```

    t_value = round(cor_test$statistic, 3),
    p_value = round(cor_test$p.value, 4)
  ))

  print(paste("Correlation between", indicator, "and", suicide, ":"))
  print(cor_test)
  cat("\n")
} else {
  cor_results <- rbind(cor_results, data.frame(
    Indicator = indicator,
    Suicide_Indicator = suicide,
    Correlation = NA,
    t_value = NA,
    p_value = NA
  ))
}
}
}

```

```

## [1] "Correlation between Major depressive episode, life and Suicidal thoughts, life :"
##
## Pearson's product-moment correlation
##
## data: df_pivot_filled[[indicator]] and df_pivot_filled[[suicide]]
## t = 38.093, df = 1, p-value = 0.01671
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
##      cor
## 0.9996556
##
## [1] "Correlation between Eating disorder, current diagnosed condition and Suicidal thoughts, life :"
##
## Pearson's product-moment correlation
##
## data: df_pivot_filled[[indicator]] and df_pivot_filled[[suicide]]
## t = 6.3971, df = 1, p-value = 0.09872
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
##      cor
## 0.9880013
##
## Warning in cor(x, y): the standard deviation is zero
##
## [1] "Correlation between Social phobia, life and Suicidal thoughts, life :"
##
## Pearson's product-moment correlation
##
## data: df_pivot_filled[[indicator]] and df_pivot_filled[[suicide]]
## t = NA, df = 1, p-value = NA
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
## cor

```

```

## NA
##
##
## [1] "Correlation between Generalized anxiety disorder, life and Suicidal thoughts, life :"
##
## Pearson's product-moment correlation
##
## data: df_pivot_filled[[indicator]] and df_pivot_filled[[suicide]]
## t = 1.452, df = 1, p-value = 0.384
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
##      cor
## 0.8235706
##
##
## [1] "Correlation between Bipolar disorder, life and Suicidal thoughts, life :"
##
## Pearson's product-moment correlation
##
## data: df_pivot_filled[[indicator]] and df_pivot_filled[[suicide]]
## t = 1.452, df = 1, p-value = 0.384
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
##      cor
## 0.8235706
##
##
## [1] "Correlation between Major depressive episode, life and Suicidal thoughts, 12 months :"
##
## Pearson's product-moment correlation
##
## data: df_pivot_filled[[indicator]] and df_pivot_filled[[suicide]]
## t = 83.325, df = 1, p-value = 0.00764
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
##      cor
## 0.999928
##
##
## [1] "Correlation between Eating disorder, current diagnosed condition and Suicidal thoughts, 12 months"
##
## Pearson's product-moment correlation
##
## data: df_pivot_filled[[indicator]] and df_pivot_filled[[suicide]]
## t = 5.8497, df = 1, p-value = 0.1078
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
##      cor
## 0.9857011

## Warning in cor(x, y): the standard deviation is zero

## [1] "Correlation between Social phobia, life and Suicidal thoughts, 12 months :"
##

```

```

## Pearson's product-moment correlation
##
## data: df_pivot_filled[[indicator]] and df_pivot_filled[[suicide]]
## t = NA, df = 1, p-value = NA
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
## cor
## NA
##
## [1] "Correlation between Generalized anxiety disorder, life and Suicidal thoughts, 12 months :"
```

```

## Pearson's product-moment correlation
##
## data: df_pivot_filled[[indicator]] and df_pivot_filled[[suicide]]
## t = 1.4086, df = 1, p-value = 0.393
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
## cor
## 0.8154076
##
## [1] "Correlation between Bipolar disorder, life and Suicidal thoughts, 12 months :"
```

```

## Pearson's product-moment correlation
##
## data: df_pivot_filled[[indicator]] and df_pivot_filled[[suicide]]
## t = 1.4086, df = 1, p-value = 0.393
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
## cor
## 0.8154076

```

```
kable(cor_results, caption = "Pearson Correlation Between Mental Health Indicators and Suicidal Thoughts")
```

Table 2: Pearson Correlation Between Mental Health Indicators and Suicidal Thoughts

	Indicator	Suicide_Indicator	Correlation	t_value	p_value
cor	Major depressive episode, life	Suicidal thoughts, life	1.000	38.093	0.0167
cor1	Eating disorder, current diagnosed condition	Suicidal thoughts, life	0.988	6.397	0.0987
cor2	Social phobia, life	Suicidal thoughts, life	NA	NA	NA
cor3	Generalized anxiety disorder, life	Suicidal thoughts, life	0.824	1.452	0.3840
cor4	Bipolar disorder, life	Suicidal thoughts, life	0.824	1.452	0.3840
cor5	Major depressive episode, life	Suicidal thoughts, 12 months	1.000	83.325	0.0076
cor6	Eating disorder, current diagnosed condition	Suicidal thoughts, 12 months	0.986	5.850	0.1078
cor7	Social phobia, life	Suicidal thoughts, 12 months	NA	NA	NA

	Indicator	Suicide_Indicator	Correlation	t_value	p_value
cor8	Generalized anxiety disorder, life	Suicidal thoughts, 12 months	0.815	1.409	0.3930
cor9	Bipolar disorder, life	Suicidal thoughts, 12 months	0.815	1.409	0.3930

## Summary Table

Mental Health Indicator	Suicide Indicator	Correlation (r)	t-value	p-value	Significance
Major depressive episode, life	Suicidal thoughts, life	<b>0.9997</b>	38.093	<b>0.0167</b>	<b>Significant</b>
Eating disorder, current diagnosed condition	Suicidal thoughts, life	0.9880	6.3971	0.0987	Not significant
Social phobia, life	Suicidal thoughts, life	NA	NA	NA	Not available
Generalized anxiety disorder, life	Suicidal thoughts, life	0.8236	1.452	0.384	Not significant
Bipolar disorder, life	Suicidal thoughts, life	0.8236	1.452	0.384	Not significant
Major depressive episode, life	Suicidal thoughts, 12 months	<b>0.9999</b>	83.325	<b>0.0076</b>	<b>Significant</b>
Eating disorder, current diagnosed condition	Suicidal thoughts, 12 months	0.9857	5.8497	0.1078	Not significant
Social phobia, life	Suicidal thoughts, 12 months	NA	NA	NA	Not available
Generalized anxiety disorder, life	Suicidal thoughts, 12 months	0.8154	1.4086	0.393	Not significant
Bipolar disorder, life	Suicidal thoughts, 12 months	0.8154	1.4086	0.393	Not significant

```
economic_data <-read.csv("C:/Users/felix/Desktop/CODING/felix's works/Mental-Health-On-Suicide-Rates-T
```

```

economic_canada <- economic_data %>%
  filter(Geo == "Canada") %>%
  select(Year, CPI, Inflation_rate)

suicide_indicators <- c("Suicidal thoughts, life", "Suicidal thoughts, 12 months")

suicide_canada <- df_canada %>%
  filter(Geography == "Canada", Gender == "Total, gender of person", Indicators %in% suicide_indicators)
  select(Year, Indicators, Population) %>%
  group_by(Year, Indicators) %>%
  summarise(Population = sum(Population, na.rm = TRUE), .groups = "drop")

# Convert suicide data to wide format
suicide_pivot <- suicide_canada %>%
  pivot_wider(names_from = Indicators, values_from = Population)

# Merge suicide and economic data
merged_data <- merge(suicide_pivot, economic_canada, by = "Year", all.x = TRUE)

# Display merged data
kable(merged_data, caption = "Merged Economic and Suicide Data in Canada")

```

Table 4: Merged Economic and Suicide Data in Canada

Year	Suicidal thoughts, 12 months	Suicidal thoughts, life	CPI	Inflation_rate
2002	1839590	6690688	100.0	2.3
2012	1885445	6730530	121.7	1.5
2022	2284800	7143600	141.9	6.8

```

# Correlation Analysis
cor_results <- data.frame(
  Suicide_Indicator = character(),
  Economic_Indicator = character(),
  Correlation = numeric(),
  p_value = numeric(),
  stringsAsFactors = FALSE
)

economic_indicators <- c("CPI", "Inflation_rate")

for (suicide in suicide_indicators) {
  for (economic in economic_indicators) {
    if (suicide %in% colnames(merged_data) && economic %in% colnames(merged_data)) {
      cor_test <- cor.test(merged_data[[suicide]], merged_data[[economic]])
      cor_results <- rbind(cor_results, data.frame(
        Suicide_Indicator = suicide,
        Economic_Indicator = economic,
        Correlation = round(cor_test$estimate, 3),
        p_value = round(cor_test$p.value, 4)
      ))
    }
  }

  print(paste("Correlation between", suicide, "and", economic, ":"))
}

```



```

    print(cor_test)
    cat("\n")
  }
}
}

```

```

## [1] "Correlation between Suicidal thoughts, life and CPI :"
##
## Pearson's product-moment correlation
##
## data: merged_data[[suicide]] and merged_data[[economic]]
## t = 1.9945, df = 1, p-value = 0.2959
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
##      cor
## 0.893936
##
##
## [1] "Correlation between Suicidal thoughts, life and Inflation_rate :"
##
## Pearson's product-moment correlation
##
## data: merged_data[[suicide]] and merged_data[[economic]]
## t = 4.4723, df = 1, p-value = 0.14
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
##      cor
## 0.9759016
##
##
## [1] "Correlation between Suicidal thoughts, 12 months and CPI :"
##
## Pearson's product-moment correlation
##
## data: merged_data[[suicide]] and merged_data[[economic]]
## t = 2.0675, df = 1, p-value = 0.2868
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
##      cor
## 0.9002295
##
##
## [1] "Correlation between Suicidal thoughts, 12 months and Inflation_rate :"
##
## Pearson's product-moment correlation
##
## data: merged_data[[suicide]] and merged_data[[economic]]
## t = 4.191, df = 1, p-value = 0.1491
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
##      cor
## 0.9726944

```

```
# Display correlation results
kable(cor_results, caption = "Correlation Between Economic Indicators and Suicidal Thoughts")
```

Table 5: Correlation Between Economic Indicators and Suicidal Thoughts

	Suicide_Indicator	Economic_Indicator	Correlation	p_value
cor	Suicidal thoughts, life	CPI	0.894	0.2959
cor1	Suicidal thoughts, life	Inflation_rate	0.976	0.1400
cor2	Suicidal thoughts, 12 months	CPI	0.900	0.2868
cor3	Suicidal thoughts, 12 months	Inflation_rate	0.973	0.1491

**Inflation Rate has a strong relationship with suicidal thoughts:** Inflation rate is highly correlated with: Suicidal thoughts, life ( $r = 0.9759$ ,  $p = 0.1400$ ) Suicidal thoughts, 12 months ( $r = 0.9727$ ,  $p = 0.1491$ ) The p-values are slightly above 0.05, meaning the results are not statistically significant.

**CPI is not a strong predictor:** CPI (Consumer Price Index) has a moderate correlation with suicidal thoughts, but the p-values are too high to conclude a meaningful relationship.