Full_fact_check

May 20, 2022

[1]:
```
library(tidyverse)
library(dtplyr)

load("/data/dbir/2022dbir/data/verisr/dbir_22_full.Rda")
```

```
Warning message in system("timedatectl", intern = TRUE):
"running command 'timedatectl' had status 1"
  Attaching packages                      tidyverse
1.3.1

 ggplot2 3.3.5      purrr   0.3.4
 tibble  3.1.5      dplyr   1.0.7
 tidyr   1.1.4      stringr 1.4.0
 readr   2.0.2      forcats 0.5.1

  Conflicts
tidyverse_conflicts()
 dplyr::filter() masks stats::filter()
 dplyr::lag()    masks stats::lag()
```

[11]:
```
#This year alone we analyzed 5,212 breaches and 23,896 incidents and
#over the years have amassed 234,638 breaches, 914,547 incidents, and 8.9␣
 ↪terabytes of data
#'dfb50d2d-1e27-47f3-86f3-d7712397d2ec'

vz %>%
    filter(plus.dbir_year==2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    nrow()
vz %>%
    filter(plus.dbir_year==2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    nrow()

vz %>%
    #filter(subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
```

```
    nrow()

vz %>%
    #filter(subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    nrow()
```

5212

23896

234691

914572

0.1 Actions

[16]:
```
#'a57a6d7d-fc54-4551-9eac-9f6102327e50
## Errors are around 14% of breaches

vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("action") %>%
    verisr::test_veris_proportion(Ea = "Error", prop = .14, direction = "equal")
```

```
The hypothesis is TRUE because the proportion of 14% is in the 95% range between
13.28% and 15.23%.
```

TRUE

[7]:
```
#058c10ee-6771-4ed3-9e54-34d1fed79ed4
## DoS Represents 45% of the incidents

vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022('action.*.variety', ci.method = "bootstrap") %>%
    verisr::test_veris_proportion(Ea="DoS", prop = .46, direction = "equal")
```

```
The hypothesis is TRUE because the proportion of 46% is in the 95% range between
44.99% and 46.46%.
```

TRUE

[11]:
```
#'0cae4a45-3c52-4e6d-83aa-963798b44808'
## Backdoor or c2 @ 17% of incidents

vz %>%
```

```
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022('action.*.variety', ci.method="bootstrap") %>%
    verisr::test_veris_proportion(Ea="Backdoor or C2", prop=.16, direction =␣
 ↪"equal")
```

The hypothesis is TRUE because the proportion of 17% is in the 95% range between 15.96% and 17.07%.

TRUE

[15]:
```
#b25e5143-4121-487e-a183-5041fc2cf874
## Other is in the top 3 of incidents, which means that 73% of incidents are␣
 ↪captured in the top 10
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022('action.*.variety', top=10, ci.method="bootstrap") %>%
    filter(enum=='Other') %>%
    pull(freq) - 1
```

-0.72955

[7]:
```
#'76982226-0774-4ee0-b58d-4217e901acdf'
#Breaches due to misconfiguration errors appear to have peaked in 2019 at 11%␣
 ↪of breaches and potentially settled to a new steady state of 6% of breaches

chunk <- vz %>%
    filter(plus.dbir_year%in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("action.*.variety", by="timeline.incident.year", ci.
 ↪method = "bayes") %>%
    filter(enum=="Misconfiguration") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_proportion(Ea="2019", prop = .11, direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea="2021", prop = .06, direction = "equal")
```

```
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"7 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
```

```
"2 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"10 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"8 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"4 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"8 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
Warning message in verisr::getenumCI2022(., "action.*.variety", by =
"timeline.incident.year", :
"Removing all rows with n < 5.  5 is the smallest samples size appropriate for
the DBIR. Use force = TRUE to avoid this removal.  If this leaves n othing but
columns with no sample size, those will be removed as well."
The hypothesis is TRUE because the proportion of 11% is in the 95% range between
9.94% and 12.28%.
```

TRUE

```
The hypothesis is TRUE because the proportion of 6% is in the 95% range between
4.78% and 6.40%.
```

TRUE

[3]:
```
#'106c1a06-a5df-4761-8bdf-274156fc1a9c'
#ransomware rose to to the total to now 25%
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022('action.*.variety', top=10, ci.method="bootstrap") %>%
    verisr::test_veris_proportion("Ransomware", prop = .25, direction = "equal")
```

```
The hypothesis is TRUE because the proportion of 25% is in the 95% range between
23.33% and 25.94%.
```

TRUE

[5]:
```
#'b4160efc-b9c2-40d2-8d42-e5b3deb1938b'
#61% of incidents in System INtrusion

vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(`pattern.System Intrusion`) %>%
    verisr::getenumCI2022("action.*.vector", ci.method = "bootstrap") %>%
    verisr::test_veris_proportion("Partner", prop = .61, direction = "equal")
```

```
The hypothesis is TRUE because the proportion of 61% is in the 95% range between
60.03% and 63.36%.
```

TRUE

[8]:
```
#'33aa2073-ed11-4bb1-bffa-3513c475ce74'
#14% of the breaches involved Error
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("action", ci.method = "bootstrap") %>%
    verisr::test_veris_proportion("Error", prop = .14, direction = "equal")
```

```
The hypothesis is TRUE because the proportion of 14% is in the 95% range between
13.28% and 15.22%.
```

TRUE

0.2 Actors

[ ]:
```
### 74a81eba-e791-4b1a-9900-8aeed08e248b
## 50% of actor motives are NA

chunk <- vz %>%
    filter(plus.dbir_year==2022, subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(actor.Internal) %>%
    verisr::getenumCI2022("actor.*.motive", na.rm=FALSE, ci.method="bayes")

chunk

verisr::test_veris_proportion(chunk, "NA", 0.5, direction="greater")
```

0.3 Assets

[27]:
```
#'ade8b50b-08a4-4e62-a2d2-c446a55aa189'
#Claim there were more OT incidents in 2022 than in 2021
## results: TRUE
chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("asset.role", by="plus.dbir_year", ci.method =␣
 ↪"bayes", na.rm = TRUE)

chunk %>%
    filter(enum=="OT")%>%
    mutate(enum=by) %>%
    verisr::test_veris_hypothesis(Ea = "2022",Eb="2021", direction = "greater")
```

```
Warning message in verisr::getenumCI2022(., "asset.role", by = "plus.dbir_year",
:
"Removing all rows with n < 5.  5 is the smallest samples size appropriate for
the DBIR. Use force = TRUE to avoid this removal.  If this leaves n othing but
columns with no sample size, those will be removed as well."
The hypothesis is TRUE because the p value of 0.20% is less than or equal to the
confidence level of 5%.
```

TRUE

[25]:
```
#'011cffc6-372f-4dc1-aa3c-5deed63e275c'
# Claim: OT in 2022 equals approximately 3% of incidents

chunk <- vz %>%
    filter(plus.dbir_year==2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("asset.role", ci.method = "bayes", na.rm = TRUE)


chunk %>%
    verisr::test_veris_proportion(Ea = "OT",prop = .03, direction = "equal")
```

```
The hypothesis is TRUE because the proportion of 3% is in the 95% range between
1.72% and 5.46%.
```

TRUE

[24]:
```
#'e8162a42-b161-471c-8a20-e87845dcbd2f'
# Claim: OT in 2021 equals apprxomiately than 1% of incidents

chunk <- vz %>%
```

```
    filter(plus.dbir_year==2021,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("asset.role", ci.method = "bayes", na.rm = TRUE)


chunk %>%
    verisr::test_veris_proportion(Ea = "OT",prop = .01, direction = "equal")
```

The hypothesis is TRUE because the proportion of 1% is in the 95% range between 0.17% and 1.32%.

TRUE

[18]:
```
#dfd45623-87d0-4198-8847-b0eaedbc31a
# Webservers makeup 56% of breached assets

chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("asset.assets.variety", ci.method = "bootstrap",␣
 ↪top=10)


chunk %>%
    verisr::test_veris_proportion( Ea="S - Web application", prop = .56,␣
 ↪direction = "equal")
```

The hypothesis is TRUE because the proportion of 56% is in the 95% range between 53.83% and 57.40%.

TRUE

[19]:
```
## '017f7d20-ac4a-4db4-8fec-cb960da12a15'
## Email servers make up approximately 28% of breached assets

chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("asset.assets.variety", ci.method = "bootstrap",␣
 ↪top=10)


chunk %>%
    verisr::test_veris_proportion( Ea="S - Mail", prop = .28, direction =␣
 ↪"equal")
```

The hypothesis is TRUE because the proportion of 28% is in the 95% range between

27.32% and 30.58%.

TRUE

0.4 Attribute

[34]:
```
## 'be157725-b728-45a6-86d6-3a80501b5512'
## Credentials and Personal are the top impacted data variety

chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("attribute.confidentiality.data.variety", ci.method =␣
 ↪"bootstrap") %>%
    slice(2:3)

second_place <- chunk %>% slice(1) %>% pull(enum) %>% as.character()
third_place <- chunk %>% slice(2) %>% pull(enum) %>% as.character()

chunk %>%
    verisr::test_veris_hypothesis(Ea=second_place, Eb=third_place, direction =␣
 ↪"greater")
```

```
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

[36]:
```
## 80ebe7d5-14ce-4101-a43c-2add0abbdbcb
## Social are approximately .2 of breaches
## Malware are greater than .3 of breaches

chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022('action')

chunk %>%
    verisr::test_veris_proportion(Ea = "Social", prop = .2, direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea="Malware", prop = .3, direction =␣
 ↪"greater")
```

The hypothesis is TRUE because the proportion of 20% is in the 95% range between 19.47% and 21.57%.

TRUE

The hypothesis is TRUE because the confidence of 100.00% at 30% is greater than the confidence level of 95%.

TRUE

[38]:
```
#'5f6bad57-6328-4e09-92db-0951cdada211'
# 38% of ransomware cases have some confidnetiality impact

chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(action.malware.variety.Ransomware) %>%
    verisr::getenumCI2022('attribute.confidentiality.data_disclosure', ci.
↪method = "bootstrap")

chunk %>%
    verisr::test_veris_proportion(Ea="Yes", prop = .38, direction = "equal")
```

The hypothesis is TRUE because the proportion of 38% is in the 95% range between 36.51% and 40.33%.

TRUE

0.5 Timeline

[42]:
```
#'47783179-f7f6-4553-bedf-58b6c3841ea5
# Discovery method is  greater 50% actor disclosed

chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("discovery_method.*.variety", ci.method = "bootstrap")

chunk %>%
    verisr::test_veris_proportion(Ea="Actor disclosure", prop = .5, direction =␣
↪"greater")
```

The hypothesis is TRUE because the confidence of 98.80% at 50% is greater than the confidence level of 95%.

TRUE

1 Patterns

1.1 Patterns: System Intrusion

[45]:
```
#'441c3a6b-6980-48b0-a7ea-34f415cf94d5'
# 14% of incidents involved remote desktop

chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(`pattern.System Intrusion`) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("action.*.vector", ci.method = "bootstrap")

chunk %>%
    verisr::test_veris_proportion(Ea="Desktop sharing software", prop = .14,␣
 ↪direction = "equal")
```

The hypothesis is TRUE because the proportion of 14% is in the 95% range between 13.16% and 15.55%.

TRUE

[47]:
```
# '6110b7b0-8759-4cf1-9ea7-b9788677ffee'
# 9% of incidents involved email as a vector

chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(`pattern.System Intrusion`) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("action.*.vector", ci.method = "bootstrap")

chunk %>%
    verisr::test_veris_proportion(Ea="Email", prop = .09, direction = "equal")
```

The hypothesis is TRUE because the proportion of 9% is in the 95% range between 8.40% and 10.29%.

TRUE

[ ]:
```
#'534f7092-242e-467f-9cb3-c7e6cb48122d'
#Increase in ransomware


chunk <- vz %>%
    filter(plus.dbir_year %in% 2021:2022,subset.2022dbir) %>%
    verisr::getenumCI2021("action.*.variety", by="timeline.incident.year", ci.
 ↪method="bootstrap") %>%
```

```
    filter(enum=="Ransomware") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_hypothesis(Ea="2022", Eb="2021", direction = "greater")
```

[7]:
```
#'67e6ec84-c390-4ea1-b61b-427f6a561c94
# CLAIM:  3rd party breaches are 1% of our overall breaches

chunk <- vz %>%
    filter(plus.dbir_year ==2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    mutate(extra.Rd_party = asset.ownership.Partner |
           attribute.confidentiality.data_victim.Partner |
           victim.secondary.amount>0 |
          action.social.vector.Partner) %>%
    mutate(extra.Not_rd_party = TRUE) %>%
    #filter(victim.secondary.amount>0) %>%
    #filter(asset.ownership.Partner) %>%
    #filter(attribute.confidentiality.data_victim.Partner)%>%
    verisr::getenumCI2022("extra")


chunk %>%
    verisr::test_veris_proportion(Ea = "Rd_party", prop = .01,␣
 ↪direction="equal" )
```

The hypothesis is TRUE because the proportion of 1% is in the 95% range between 0.90% and 1.53%.

TRUE

[5]:
```
#'8369f624-e065-45b5-ad90-229171894d27
#Supply chain represented 8% of our incidentss
#Supply chain represents 6% of our incidentss

vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    verisr::getenumCI2022("subset.Supply chain") %>%
    verisr::test_veris_proportion(Ea="Supply chain", prop = .09, direction =␣
 ↪"equal")

vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("subset.Supply chain") %>%
```

```
    verisr::test_veris_proportion(Ea="Supply chain", prop=.006, direction =␣
↪"equal")
```

The hypothesis is TRUE because the proportion of 9% is in the 95% range between 8.49% and 9.20%.

TRUE

The hypothesis is TRUE because the proportion of 1% is in the 95% range between 0.44% and 0.88%.

TRUE

[15]:
```
#'a7612913-254c-44a2-8cb6-2c07ef1e2d40'
#majority were used to push backdoor (97%)
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(`subset.Supply chain`) %>%
    verisr::getenumCI2022("action.*.variety", ci.method = "bootstrap") %>%
    verisr::test_veris_proportion(Ea="Backdoor or C2", prop = .972, direction =␣
↪"equal")
```

The hypothesis is FALSE because the proportion of 97% is not in the 95% range between 97.24% and 98.43%.

FALSE

[12]:
```
#'557f864a-0ea8-46a8-89c4-52d2ec0eff4a'
#40% of Ransomware incidents involve Desktop sSharing sSoftware + 30% invovled␣
↪email

chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(action.malware.variety.Ransomware) %>%
    verisr::getenumCI2021("action.*.vector", ci.method = "bayes", force=TRUE) ␣
↪%>%
    arrange(desc(n))

chunk %>%
    verisr::test_veris_proportion(Ea="Desktop sharing software", prop = .4,␣
↪direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea="Email", prop = .35, direction = "equal")
```

The hypothesis is TRUE because the proportion of 40% is in the 95% range between 37.23% and 43.08%.

TRUE

The hypothesis is TRUE because the proportion of 35% is in the 95% range between 32.64% and 38.22%.

TRUE

[220]:
```
#'e373ef10-49e5-44ab-8b33-42d9ed7c884a'
# 62% of syste intrusion incidents involved partner
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(`pattern.System Intrusion`) %>%
    verisr::getenumCI2021("action.*.vector", ci.method="bootstrap") %>%
    verisr::test_veris_proportion(Ea="Partner", prop = .62, direction = "equal")
```

The hypothesis is TRUE because the proportion of 62% is in the 95% range between 60.01% and 63.18%.

TRUE

1.1.1 Vulnerability

[ ]:
```
### 91ebf5f1-b446-46f5-ae1b-39bcc2644204
## Vulnerabilities are consistent between 2021 and 2022
chunk_91ebf5f1 <- vz %>%
    filter(plus.dbir_year %in% 2021:2022, subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("action.*.variety", by="plus.dbir_year", ci.
 ↪method="bayes", ci.params=TRUE) %>%
    filter(enum == "Exploit vuln")




chunk_91ebf5f1 %>%
    select(-enum) %>% rename(enum = by) %>%
    mutate(enum = as.character(enum)) %>%
    mutate(
        freq = ifelse(enum == "2021", freq * 2, freq),
        x = ifelse(enum == "2021", x * 2, x)

    ) %>%
    verisr::test_veris_consistency("2021", "2022")
```

[11]:
```
#'422d031d-5830-46e5-9430-d3f2239245f4'
# exploit vuln is up from last year
 vz %>%
    filter(plus.dbir_year %in% 2021:2022, subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("action.*.variety", by="plus.dbir_year", ci.
↪method="bayes", ci.params=TRUE) %>%
    filter(enum=="Exploit vuln") %>%
    select(-enum) %>%
    mutate(enum=by) %>%
    verisr::test_veris_hypothesis(Ea="2022", Eb="2021", direction = "greater")
```

```
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"2 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"5 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

[12]:
```
#'6c07c479-71d6-4152-8157-6221d465dd93'
# exploit vuln is 7% of breaches this year
     vz %>%
    filter(plus.dbir_year == 2022, subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("action.*.variety",  ci.method="bayes", ci.
↪params=TRUE) %>%
    verisr::test_veris_proportion(Ea="Exploit vuln", prop = .07, direction =␣
↪"equal")
```

```
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"5 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
The hypothesis is TRUE because the proportion of 7% is in the 95% range between
6.12% and 7.64%.
```

TRUE

[10]:
```
#'954abd93-987e-40e3-b7d7-fedf53580c35'
#2/3 phishing/stolen/ransomware

vz %>%
    filter(plus.dbir_year==2022, subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    mutate(action.extra.variety.Phishing_ransomware_creds = `action.malware.
 ↪variety.Ransomware` |
             action.social.variety.Phishing | `action.hacking.variety.Use of␣
 ↪stolen creds`) %>%
    verisr::getenumCI2022("action.*.variety", ci.method="bootstrap") %>%
    verisr::test_veris_proportion(Ea="Phishing_ransomware_creds", prop = .66,␣
 ↪direction = "equal")
```

The hypothesis is TRUE because the proportion of 66% is in the 95% range between 64.70% and 67.64%.

TRUE

1.2 Patterns: Social Engineering

[13]:
```
#'8e87c9c5-4778-49de-879e-ead6cdb2d500'
#82% Of breaches invovles the human vector
vz %>%
    filter(plus.dbir_year==2022, subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    mutate(action.extra.variety.Human = actor.Internal |
             action.malware.vector.Email | `action.malware.vector.Email␣
 ↪attachment` |
             `action.malware.vector.Email autoexecute` | `action.malware.vector.
 ↪Email link` |
             `action.malware.vector.Email link` | `action.malware.vector.Email␣
 ↪other` |
             `action.malware.vector.Email unknown` | `action.malware.vector.
 ↪Instant messaging` |
             `action.malware.vector.Removable media` |
             `action.malware.vector.Web application - download` |
             `action.malware.vector.Web application - drive-by` |
             `action.hacking.variety.Use of stolen creds` |
             action.Misuse | action.Social | action.Error | asset.variety.Person
    ) %>%
    verisr::getenumCI2022("action.*.variety", ci.method="bootstrap", ci.
 ↪params=TRUE) %>%
    filter(enum == "Human")  %>%
    verisr::test_veris_proportion(Ea="Human", prop = .82, direction = "equal")
```

The hypothesis is TRUE because the proportion of 82% is in the 95% range between 80.58% and 82.97%.

TRUE

[7]:
```
#'5932084b-0435-4287-ae3e-21f38e0d4529'
#.20 of breaches involved social engineering
vz %>%
    filter(plus.dbir_year==2022, subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("pattern", ci.method = "bayes") %>%
    verisr::test_veris_proportion("Social Engineering", prop = .2, direction =␣
 ↪"equal")
```

The hypothesis is TRUE because the proportion of 20% is in the 95% range between 19.30% and 21.49%.

TRUE

[15]:
```
#'88660b3b-d54a-462a-8bf0-a4bbef1db815'
# ransomware in 17% of incidents

vz %>%
    filter(plus.dbir_year==2022, subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(`pattern.Social Engineering`) %>%
    verisr::getenumCI2022('action.*.variety', top=10, ci.method="bootstrap", ci.
 ↪params=FALSE) %>%
    filter(enum == "Ransomware") %>%
    verisr::test_veris_proportion(Ea="Ransomware", prop = .15, direction =␣
 ↪"equal")
```

The hypothesis is TRUE because the proportion of 15% is in the 95% range between 14.10% and 17.16%.

TRUE

[17]:
```
#'521b52a2-b18c-4f1b-96b8-3832695baa4a'
# BEC is between 27% and 34% of social engineering

vz %>%
    filter(plus.dbir_year==2022, subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(`pattern.Social Engineering`) %>%
    verisr::getenumCI2021("action.*.variety", ci.method = "bootstrap") %>%
    verisr::test_veris_proportion("Pretexting", prop = .27, direction = "equal")
```

The hypothesis is TRUE because the proportion of 27% is in the 95% range between 26.71% and 31.98%.

TRUE

[18]:
```
#'012972ae-377b-49ca-8fee-b2a5c3a0e897'
#Only 41% of BECs involved Phishing.
vz %>%
    filter(plus.dbir_year==2022, subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(`pattern.Social Engineering`) %>%
    filter(subset.BEC) %>%
    mutate(subset.True = TRUE) %>%
    verisr::getenumCI2022("action.social.variety", ci.method="bayes") %>%
    filter(enum == "Phishing") %>%
    verisr::test_veris_proportion(Ea="Phishing", prop = .41, direction =␣
↪"equal")
```

The hypothesis is TRUE because the proportion of 41% is in the 95% range between 37.16% and 45.93%.

TRUE

[19]:
```
# '7072fc90-ac41-4fed-8bf0-1f1cb172239b'
#43% of BEC involved Use of stolen credentials against the victim organizatio
vz %>%
    filter(plus.dbir_year==2022, subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(`pattern.Social Engineering`) %>%
    filter(subset.BEC) %>%
    filter(!action.social.variety.Phishing) %>%
    verisr::getenumCI2022("action.*.variety", ci.method="bayes", top=5) %>%
    verisr::test_veris_proportion(Ea="Use of stolen creds", prop = .43,␣
↪direction = "equal")
```

The hypothesis is TRUE because the proportion of 43% is in the 95% range between 37.37% and 48.75%.

TRUE

[25]:
```
#'d7b7ffe0-a378-4640-ae40-360fb136eafa'
# Social engineering malware 74% of malware breaches are downloader
vz %>%
    filter(plus.dbir_year==2022, subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(`pattern.Social Engineering`) %>%
    filter(action.Malware) %>%
```

```
    verisr::getenumCI2022('action.malware.variety', top=10, ci.method="bayes")␣
↪%>%
    verisr::test_veris_proportion(Ea="Downloader", prop = .74, direction =␣
↪"equal")
```

The hypothesis is TRUE because the proportion of 74% is in the 95% range between 70.11% and 77.31%.

TRUE

[26]:
```
#'507a448b-6290-49a1-a3ea-def9bd84e68b'
# Social engineering malware 64% of malware breaches are ransomware
vz %>%
    filter(plus.dbir_year==2022, subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(`pattern.Social Engineering`) %>%
    filter(action.Malware) %>%
    verisr::getenumCI2022('action.malware.variety', top=10, ci.method="bayes")␣
↪%>%
    verisr::test_veris_proportion(Ea="Ransomware", prop = .64, direction =␣
↪"equal")
```

The hypothesis is TRUE because the proportion of 64% is in the 95% range between 59.59% and 67.53%.

TRUE

1.3 Patterns: Basic web application attacks

[27]:
```
#'14458d18-5c95-4559-9a27-c49f1d826cfd'
#Over 80% of the breaches invovle the use of stolen creds
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(`pattern.Basic Web Application Attacks`) %>%
    verisr::getenumCI2022("action.*.variety", top = 10,ci.method = "bootstrap",␣
↪ci.params = TRUE) %>%
    verisr::test_veris_proportion("Use of stolen creds", prop = .8, direction =␣
↪"greater")
```

The hypothesis is TRUE because the confidence of 100.00% at 80% is greater than the confidence level of 95%.

TRUE

[34]:
```
#'dcf74ff5-fa90-40f7-a3de-f2e9dcbb1b51'
# 30% increase in stolen credentials since 2017

chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(`pattern.Basic Web Application Attacks`) %>%
    verisr::getenumCI2022("action.*.variety", by="timeline.incident.year", ci.
↪method = "bayes") %>%
    filter(enum=="Use of stolen creds")  %>% filter(by %in% c(2017,2021))  %>%␣
↪pull(freq)

vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(`pattern.Basic Web Application Attacks`) %>%
    verisr::getenumCI2022("action.*.variety", by="timeline.incident.year", ci.
↪method = "bayes") %>%
    filter(enum=="Use of stolen creds")  %>% select(-enum) %>% mutate(enum=by)␣
↪%>%
    verisr::test_veris_hypothesis(Ea="2021", Eb="2017", direction ="greater" )

chunk[2] - chunk[1]
```

```
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"2 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
Warning message in verisr::getenumCI2022(., "action.*.variety", by =
"timeline.incident.year", :
"Removing all rows with n < 5.  5 is the smallest samples size appropriate for
the DBIR. Use force = TRUE to avoid this removal.  If this leaves n othing but
columns with no sample size, those will be removed as well."
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"2 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
Warning message in verisr::getenumCI2022(., "action.*.variety", by =
"timeline.incident.year", :
"Removing all rows with n < 5.  5 is the smallest samples size appropriate for
the DBIR. Use force = TRUE to avoid this removal.  If this leaves n othing but
columns with no sample size, those will be removed as well."
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
```

confidence level of 5%.

TRUE

0.36311

[35]:
```
#'98df3778-b5dc-4b10-a9f8-6eacc63d69aa'
#mail servers less than 20%
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(`pattern.Basic Web Application Attacks`) %>%
    verisr::getenumCI2022("asset.assets.variety", ci.method = "bayes", top=5,␣
 ↪ci.params = TRUE) %>%
    verisr::test_veris_proportion("S - Mail", prop = .2, direction = "less")
```

The hypothesis is TRUE because the confidence of 96.80% at 20% is greater than the confidence level of 95%.

TRUE

[37]:
```
#'3a4bed74-0a99-4c50-9e9e-19a3107ffe47'
#Of those Mail servers, 80% use of stolen creds
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(`pattern.Basic Web Application Attacks`) %>%
    filter(`asset.assets.variety.S - Mail`) %>%
    verisr::getenumCI2021("action.*.variety", ci.method = "bootstrap") %>%
    verisr::test_veris_proportion(Ea="Use of stolen creds", prop = .8,␣
 ↪direction = "equal")
```

The hypothesis is TRUE because the proportion of 80% is in the 95% range between 73.99% and 85.55%.

TRUE

[38]:
```
# '211dafb1-d2aa-4f60-bad0-068d6bf67c6c'
#30% were compromised using some form of exploit
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(`pattern.Basic Web Application Attacks`) %>%
    filter(`asset.assets.variety.S - Mail`) %>%
    verisr::getenumCI2021("action.*.variety", ci.method = "bootstrap") %>%
    verisr::test_veris_proportion(Ea="Exploit vuln", prop = .3, direction =␣
 ↪"equal")
```

The hypothesis is TRUE because the proportion of 30% is in the 95% range between 23.70% and 37.57%.

TRUE

1.4 Pattern Miscellaneous Errors

[44]:
```
#'eae93b32-8136-40fe-b763-6578453fb262'
#Misconfiguration is frequently paired with the Discovery Method of "'Security␣
↪Researcher

vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(`pattern.Miscellaneous Errors`) %>%
    verisr::getenumCI2021("discovery_method.*.variety", ci.method =␣
↪"bootstrap") %>%
    verisr::test_veris_proportion(Ea="Security researcher", prop = .4,␣
↪direction = "equal")
```

The hypothesis is TRUE because the proportion of 40% is in the 95% range between 29.03% and 64.52%.

TRUE

1.5 Patterns: Privilege Misuse

[ ]:
```

```

1.6 Patterns: Lost and Stolen assets

[47]:
```
#'5b0b5fa6-a11f-40b5-a459-9ff0a7c7e9a3'
#approximately. 90% are considered incidents

vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(`pattern.Lost and Stolen Assets`) %>%
    verisr::getenumCI2021("attribute.confidentiality.data_disclosure", ci.
↪method = "bootstrap") %>%
    verisr::test_veris_proportion(Ea="Potentially", prop = .9, direction =␣
↪"equal")
```

The hypothesis is TRUE because the proportion of 90% is in the 95% range between 88.25% and 92.20%.

TRUE

[54]:
```
#'3c9b1b86-3846-47b1-9162-966d942307f5'
#MObiles exist in aprox 1% of oru breaches
vz %>%
    filter(timeline.incident.year %in% 2017:2021, subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    #filter(`asset.assets.variety.U - Mobile phone`) %>%
    verisr::getenumCI2022("asset.assets.variety", by="timeline.incident.year")␣
↪%>%
    filter(enum == "U - Mobile phone") %>%
    select(-enum) %>% mutate(enum=by) %>%
    verisr::test_veris_proportion(Ea="2021", prop = .01, direction = "equal" )
```

```
The hypothesis is TRUE because the proportion of 1% is in the 95% range between
0.91% and 1.86%.
```

TRUE

1.7 Pattern MIscellaneous Errors

[ ]:
```

```

1.8 Pattern: Denial of Service

[ ]:
```

```

2 Industries

2.1 Industiries: Manufacturing 31_33

[133]:
```
#'fd045ed7-670b-4b10-b801-23831a0eec62'
#increasingly attacked by organized crime
vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.31_33) %>%
    verisr::getenumCI2021('actor.*.variety', by="plus.dbir_year", ci.method =␣
↪"bootstrap", force=TRUE, na.rm = TRUE) %>%
    filter(enum=="Organized crime") %>%
    select(-enum) %>%
    mutate(enum=by) %>%
    verisr::test_veris_hypothesis(Ea="2022", Eb="2018", direction = "greater")
```

```
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
```

```
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

[134]:
```
#'68785e2e-a986-4261-855d-5b8dcda40e8c'
# 2016 55% of incidents involved espionage
chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.31_33) %>%
    verisr::getenumCI2021('actor.*.motive', by="plus.dbir_year", ci.method =␣
 ↪"bootstrap", force=TRUE, na.rm = TRUE) %>%
    filter(enum=="Espionage") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_proportion(Ea="2016", prop = .55, direction = "equal")
```

```
The hypothesis is TRUE because the proportion of 55% is in the 95% range between
47.37% and 78.95%.
```

TRUE

[142]:
```
#'09c3756a-ddcc-4a19-adf4-fba460a8cab1'
#DoS over 40% in 2018
vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.31_33) %>%
    verisr::getenumCI2021('action.*.variety', by="timeline.incident.year", ci.
 ↪method = "bayes", force=TRUE, na.rm = TRUE)%>%
    filter(enum=="DoS") %>%
    select(-enum) %>%
    mutate(enum=by) %>%
    verisr::test_veris_proportion(Ea="2018", prop = .4, direction = "greater")
```

```
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"2 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
The hypothesis is TRUE because the confidence of 99.90% at 40% is greater than
the confidence level of 95%.
```

TRUE

[149]:
```
#'70eb0887-6d1d-46dd-a85a-ac7ec521d502'
#DoS approximately 70%
vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.31_33) %>%
    verisr::getenumCI2021('action.*.variety', by="plus.dbir_year", ci.method =␣
 ↪"bayes", force=TRUE, na.rm = TRUE)%>%
    filter(enum=="DoS") %>%
    select(-enum) %>%
    mutate(enum=by) %>%
    verisr::test_veris_proportion(Ea="2022", prop = .68, direction = "equal")
```

```
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"19 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"6 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
The hypothesis is TRUE because the proportion of 68% is in the 95% range between
65.75% and 69.55%.
```

TRUE

[151]:
```
#'4121a0e6-5920-40d3-bd74-fe8e519552af'
#stolen credentials (39%), Ransomware (23%) and Phishing (10%)
chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.31_33) %>%
    verisr::getenumCI2021('action.*.variety', ci.method = "bayes", force=TRUE,␣
 ↪na.rm = TRUE)

chunk %>%
    verisr::test_veris_proportion(Ea="Use of stolen creds", prop = .39,␣
 ↪direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea="Ransomware", prop = .24, direction =␣
 ↪"equal")

chunk %>%
    verisr::test_veris_proportion(Ea="Phishing", prop = .11, direction =␣
 ↪"equal")
```

```
The hypothesis is TRUE because the proportion of 39% is in the 95% range between
30.12% and 42.08%.
```

TRUE

```
The hypothesis is TRUE because the proportion of 24% is in the 95% range between
23.94% and 35.14%.
```

TRUE

```
The hypothesis is TRUE because the proportion of 11% is in the 95% range between
10.42% and 18.92%.
```

TRUE

2.2 Industries: Accommodation and Food Service (72)

[18]:
```
#'541ec7b8-6a9c-48a5-8c10-4f83439f75b1'
#no increase in payment since
vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(!!as.name(paste0("victim.industry2.72"))) %>%
    verisr::getenumCI2022("attribute.confidentiality.data.variety", by="plus.
↪dbir_year") %>%
    filter(enum=="Payment") %>%
    mutate(enum=by) %>%
    verisr::test_veris_hypothesis(Ea = "2022", Eb="2020", direction = "less")
```

```
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

[17]:
```
#'b27dc9fa-abdc-4fc0-8456-1d54f292d42
#System intrusion is less than 2017

chunk<-   vz %>%
        filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
        filter(attribute.confidentiality.data_disclosure.Yes) %>%
        filter(victim.industry2.72) %>%
        verisr::getenumCI2022("pattern", by="plus.dbir_year", ci.method =␣
↪"bootstrap")
```

```
chunk %>%
    filter(enum=="System Intrusion") %>%
    select(-enum) %>% mutate(enum=by) %>%
    verisr::test_veris_hypothesis(Ea = "2022", Eb="2017", direction = "less" )
```

```
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

[61]:
```
#'0396fe00-9910-4bdb-8b2d-8f4c8a293239'
# over 80% other in top 5 actions

vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.72) %>%
    verisr::getenumCI2021("action.*.variety", ci.method = "bootstrap",␣
 ↪force=TRUE, top=5) %>%
    verisr::test_veris_proportion(Ea="Other", prop = .8, direction = "greater")
```

```
The hypothesis is TRUE because the confidence of 95.10% at 80% is greater than
the confidence level of 95%.
```

TRUE

[63]:
```
#'4dd07a05-fedc-439e-8d52-4173c75ef3c1'
#less than 2% of breaches impacted industry 72
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2021("victim.industry2") %>%
    verisr::test_veris_proportion(Ea="72", prop = .02, direction = "less")
```

```
The hypothesis is TRUE because the confidence of 100.00% at 2% is greater than
the confidence level of 95%.
```

TRUE

2.3 Industries: Arts, Recreation and Entertainment (71)

[68]:
```
#9fb322a2-0624-4c47-bb21-a065d4da91b9
##DoS represents 20% of incidents in this sector

vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.71) %>%
    verisr::getenumCI2021("pattern", ci.method = "bootstrap") %>%
    verisr::test_veris_proportion("Denial of Service", prop = .2, direction =␣
 ↪"equal")
```

```
The hypothesis is TRUE because the proportion of 20% is in the 95% range between
17.66% and 29.30%.
```

TRUE

[69]:
```
#'fcc45d6b-9861-4a3a-a432-109cc6415754
#Personal data down from last year
vz %>%
    filter(plus.dbir_year %in% 2021:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.71) %>%
    verisr::getenumCI2021("attribute.confidentiality.data.variety", by="plus.
 ↪dbir_year", ci.method="bootstrap") %>%
    filter(enum=="Personal") %>%
    select(-enum) %>%
    mutate(enum=by) %>%
    verisr::test_veris_hypothesis(Ea="2022", Eb="2021", direction = "less")
```

```
The hypothesis is TRUE because the p value of 0.80% is less than or equal to the
confidence level of 5%.
```

TRUE

[70]:
```
#Medical data is 15% of the data
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.71) %>%
    verisr::getenumCI2021("attribute.confidentiality.data.variety", ci.
 ↪method="bootstrap") %>%
    verisr::test_veris_proportion(Ea="Medical", prop = .15, direction = "equal")
```

```
The hypothesis is TRUE because the proportion of 15% is in the 95% range between
7.69% and 23.08%.
```

TRUE

[71]:
```
#'d322a908-f410-49ca-9392-c30a67fb1a49
#MISC errors at 25% of breaches

vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.71) %>%
    verisr::getenumCI2021("pattern", ci.method = "bootstrap") %>%
    verisr::test_veris_proportion("Miscellaneous Errors", prop = .2, direction␣
↪= "equal")
```

The hypothesis is TRUE because the proportion of 20% is in the 95% range between 15.62% and 33.33%.

TRUE

[88]:
```
#0037b666-d317-4989-a177-453bbbdd9a50
#Misconfiguration is .15% of breaches

vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.71) %>%
    #filter(`pattern.Miscellaneous Errors`) %>%
    verisr::getenumCI2021("action.*.variety", ci.method = "bootstrap",␣
↪force=TRUE) %>%
    verisr::test_veris_proportion("Misconfiguration", prop = .15, direction =␣
↪"equal")
```

The hypothesis is TRUE because the proportion of 15% is in the 95% range between 3.95% and 19.74%.

TRUE

[92]:
```
#6aeb8e73-1263-497a-bcfe-921d727315a0
#In 2021 Misdelivery was more commmmon than misconfiguration (false)
#PL: Fixed
vz %>%
    filter(plus.dbir_year == 2021,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.71) %>%
    #filter(`pattern.Miscellaneous Errors`) %>%
    verisr::getenumCI2021("action.*.variety", ci.method = "bootstrap",␣
↪force=TRUE) %>%
```

```
    verisr::test_veris_hypothesis(Ea="Misdelivery", Eb="Misconfiguration",␣
↪direction = "greater")
```

The hypothesis is FALSE because the p value of 97.20% is greater than the confidence level of 5%.

FALSE

[89]:
```
#45014a93-002b-4dcd-8f99-7be04c413041
# iN 2022 misconfiguration was more common than misdelivery
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.71) %>%
    #filter(`pattern.Miscellaneous Errors`) %>%
    verisr::getenumCI2021("action.*.variety", ci.method = "bootstrap",␣
↪force=TRUE) %>%
    verisr::test_veris_hypothesis(Ea="Misconfiguration", Eb="Misdelivery",␣
↪direction = "greater")
```

The hypothesis is TRUE because the p value of 1.80% is less than or equal to the confidence level of 5%.

TRUE

[78]:
```
#e56dfbbb-870d-472b-ada4-4ada4be38180
#Misdelivery has dropped (false)
#Fixed
vz %>%
    filter(plus.dbir_year %in% 2020:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.71) %>%
    verisr::getenumCI2021("action.*.variety",by="plus.dbir_year", ci.method =␣
↪"bootstrap", force=TRUE) %>%
    filter(enum=="Misdelivery") %>%
    select(-enum) %>%
    mutate(enum=by) %>%
    verisr::test_veris_hypothesis(Ea="2022", Eb="2020", direction = "less")
```

The hypothesis is FALSE because the p value of 15.60% is greater than the confidence level of 5%.

FALSE

2.4 Industries: Educational services (61)

[29]:
```
#3eadfe17-123c-455f-84ed-4e87babe6a6b
# Ransomware increased
vz %>%
    filter(plus.dbir_year %in% 2020:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.61) %>%
    verisr::getenumCI2021("action.*.variety",by="plus.dbir_year", ci.method =␣
 ↪"bootstrap", force=TRUE) %>%
    filter(enum=="Ransomware") %>%
    select(-enum) %>%
    mutate(enum=by) %>%
    verisr::test_veris_hypothesis(Ea="2022", Eb="2021", direction = "greater")
```

```
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

[95]:
```
#aef90507-c7bd-4cbc-8b05-4561d81003fa
# Ransomware was 30% of breaches
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.61) %>%
    verisr::getenumCI2021("action.*.variety",, ci.method = "bootstrap",␣
 ↪force=TRUE) %>%
    verisr::test_veris_proportion(Ea="Ransomware", prop = .3, direction =␣
 ↪"greater")
```

```
The hypothesis is TRUE because the confidence of 98.60% at 30% is greater than
the confidence level of 95%.
```

TRUE

[96]:
```
#36ede12f-38c6-411a-9b68-fc574890bda1
#34% of errors were misdelivery

vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.61) %>%
```

```
    verisr::getenumCI2021("action.error.variety",by="plus.dbir_year", ci.method␣
↪= "bootstrap", force=TRUE) %>%
    verisr::test_veris_proportion("Misdelivery", prop = .34, direction =␣
↪"equal")
```

```
The hypothesis is TRUE because the proportion of 34% is in the 95% range between
27.27% and 59.09%.
```

TRUE

2.5 Industries: Financial (52)

[97]:
```
#'ef76eeaf-b0a3-4266-96bf-9b76788860970
#In 2016 50% of breaches in financially motived in fincnaicl involved Servers
# In 2022, 90%

chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.52)%>%
    filter(actor.external.motive.Financial) %>%
    verisr::getenumCI2021("asset.variety", by="timeline.incident.year", ci.
↪method="bayes",na.rm = TRUE) %>%
    filter(enum=="Server") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_proportion(Ea="2016", prop =.5, direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea="2021", prop =.9, direction = "equal")
```

```
Warning message in verisr::getenumCI2021(., "asset.variety", by =
"timeline.incident.year", :
"Removing all rows with n < 5.  5 is the smallest samples size appropriate for
the DBIR. Use force = TRUE to avoid this removal.  If this leaves n othing but
columns with no sample size, those will be removed as well."
The hypothesis is TRUE because the proportion of 50% is in the 95% range between
27.12% and 50.85%.
```

TRUE

```
The hypothesis is TRUE because the proportion of 90% is in the 95% range between
80.25% and 91.36%.
```

TRUE

[98]:
```
## '8bb2ff6c-8b68-4dc1-894f-f9f5bfdbbe08
## Webservers in 2016 were 12% of breaches
## Webservers in 2021 were 51% of breaches
chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.52)%>%
    #filter(actor.external.motive.Financial) %>%
    verisr::getenumCI2021("asset.assets.variety", by="timeline.incident.year",␣
 ↪ci.method="bayes",na.rm = TRUE) %>%
    filter(enum=="S - Web application") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_proportion(Ea="2016", prop = .12, direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea="2021", prop = .512, direction = "equal")
```

```
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"2 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
Warning message in verisr::getenumCI2021(., "asset.assets.variety", by =
"timeline.incident.year", :
"Removing all rows with n < 5.  5 is the smallest samples size appropriate for
the DBIR. Use force = TRUE to avoid this removal.  If this leaves n othing but
columns with no sample size, those will be removed as well."
The hypothesis is TRUE because the proportion of 12% is in the 95% range between
5.43% and 19.57%.
```

TRUE

```
The hypothesis is TRUE because the proportion of 51% is in the 95% range between
44.84% and 56.45%.
```

TRUE

[99]:
```
#2679838c-cc0d-4752-8028-a6e9f334cf1f
## Misdelivery is 16% of breaches
chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.52)%>%
```

```
    verisr::getenumCI2021("action.*.variety")

chunk %>%
    verisr::test_veris_proportion(Ea="Misdelivery", prop = .16, direction =␣
 ↪"equal")
```

```
The hypothesis is TRUE because the proportion of 16% is in the 95% range between
12.88% and 18.85%.
```

TRUE

[107]:
```
#e2221301-5c88-4cdc-80a9-253f01400dd4
# 2018 system intrusion was 14% of breaches (false fixed)
# 2021 was 30%
# 2021 was greater than 2018
chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.52)%>%
    verisr::getenumCI2021("pattern", by="timeline.incident.year", ci.
 ↪method="bayes",na.rm = TRUE) %>%
    filter(enum=="System Intrusion") %>%
    select(-enum) %>%
    mutate(enum=by)


chunk %>%
    verisr::test_veris_proportion(Ea="2016", prop = .14, direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea = "2021", prop = .3, direction = "equal")

chunk %>%
    verisr::test_veris_hypothesis(Ea="2021", Eb="2016", direction = "greater")
```

```
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"1 confidence interval failed to converge (marked by '*').
  Try changing 'tol' to a different value."
Warning message in verisr::getenumCI2021(., "pattern", by =
"timeline.incident.year", :
"Removing all rows with n < 5.  5 is the smallest samples size appropriate for
the DBIR. Use force = TRUE to avoid this removal.  If this leaves n othing but
columns with no sample size, those will be removed as well."
The hypothesis is TRUE because the proportion of 14% is in the 95% range between
7.69% and 19.66%.
```

TRUE

```
The hypothesis is TRUE because the proportion of 30% is in the 95% range between
26.30% and 34.17%.
```

TRUE

```
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

[109]:
```r
#b15108f3-46af-481f-bd99-78245ca0ebcd
## 2018 breaches organized crime == 49% (
## 2021 brecaches organized crime == 79%
chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.52)%>%
    verisr::getenumCI2021("actor.*.variety", by="timeline.incident.year", ci.
 ↪method="bayes",na.rm = TRUE) %>%
    filter(enum=="Organized crime") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_proportion(Ea="2018", prop = .49, direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea="2021", prop = .79, direction = "equal")
```

```
Warning message in verisr::getenumCI2021(., "actor.*.variety", by =
"timeline.incident.year", :
"Removing all rows with n < 5.  5 is the smallest samples size appropriate for
the DBIR. Use force = TRUE to avoid this removal.  If this leaves n othing but
columns with no sample size, those will be removed as well."
The hypothesis is TRUE because the proportion of 49% is in the 95% range between
38.46% and 59.34%.
```

TRUE

```
The hypothesis is TRUE because the proportion of 79% is in the 95% range between
72.22% and 85.19%.
```

TRUE

[110]:
```
#b1786860-eebd-4545-9cf1-0d3d646f7ba7
#Avaialbility impacted 6% of breaches in 2016 and 14% in 2021

chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.52) %>%
    verisr::getenumCI2021("attribute",by="timeline.incident.year",ci.
 ↪method="bayes",na.rm = TRUE) %>%
    filter(enum=="Availability") %>%
    select(-enum) %>%
    mutate(enum=by)
chunk %>%
    verisr::test_veris_proportion(Ea="2016", prop = .06, direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea="2021", prop = .14, direction = "equal")
```

```
Warning message in verisr::getenumCI2021(., "attribute", by =
"timeline.incident.year", :
"Removing all rows with n < 5.  5 is the smallest samples size appropriate for
the DBIR. Use force = TRUE to avoid this removal.  If this leaves n othing but
columns with no sample size, those will be removed as well."
The hypothesis is TRUE because the proportion of 6% is in the 95% range between
1.71% and 11.11%.
```

TRUE

```
The hypothesis is TRUE because the proportion of 14% is in the 95% range between
10.75% and 16.51%.
```

TRUE

[111]:
```
# 6907d83f-08a2-40d1-b47c-c2e96c376adc
#Discovery Method of Actor disclosure method of discovery was 5% (in 2016) as␣
 ↪opposed to the 58%
chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.52) %>%
    verisr::getenumCI2021("discovery_method.*.variety",by="timeline.incident.
 ↪year",ci.method="bayes",na.rm = TRUE) %>%
    filter(enum=="Actor disclosure") %>%
    select(-enum) %>%
    mutate(enum=by)
```

```
chunk %>%
    verisr::test_veris_proportion(Ea="2016", prop = .05, direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea="2021", prop = .58, direction = "equal")
```

```
Warning message in verisr::getenumCI2021(., "discovery_method.*.variety", by =
"timeline.incident.year", :
"Removing all rows with n < 5.  5 is the smallest samples size appropriate for
the DBIR. Use force = TRUE to avoid this removal.  If this leaves n othing but
columns with no sample size, those will be removed as well."
The hypothesis is TRUE because the proportion of 5% is in the 95% range between
1.39% and 11.11%.
```

TRUE

```
The hypothesis is TRUE because the proportion of 58% is in the 95% range between
44.07% and 69.53%.
```

TRUE

[113]:
```
#44b4c8a4-86ff-4136-9993-b9a14619d634
#DoS is 58% of incidents
 vz %>%
    filter(plus.dbir_year==2022,subset.2022dbir) %>%
    filter(victim.industry2.52) %>%
    verisr::getenumCI2021("pattern", ci.method = "bootstrap") %>%
    verisr::test_veris_proportion(Ea="Denial of Service", prop = .58, direction␣
 ↪= "equal")
```

```
The hypothesis is TRUE because the proportion of 58% is in the 95% range between
56.83% and 60.74%.
```

TRUE

2.6 Industries: Information (51)

[119]:
```
#2e650430-d939-4fcb-9abc-40bae6452ced
#Malware has risen
vz %>%
    filter(plus.dbir_year %in% 2021:2022,subset.2022dbir) %>%
    filter(victim.industry2.51) %>%
    verisr::getenumCI2021("action", ci.method = "bootstrap", by="plus.
 ↪dbir_year") %>%
    filter(enum=="Malware") %>%
```

```
    select(-enum) %>%
    mutate(enum=by) %>%
    verisr::test_veris_hypothesis(Ea="2022", Eb="2021", direction = "greater")
```

```
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

[14]:
```
#c0f5d90c-06f4-44ef-b9e0-9247c000d66e'
#system intrusion is greater than bwaa

vz %>%
    filter(plus.dbir_year==2022,subset.2022dbir) %>%
    filter(victim.industry2.51) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2021("pattern", ci.method = "bootstrap") %>%
    verisr::test_veris_hypothesis(Ea="System Intrusion", Eb="Basic Web␣
 ↪Application Attacks", direction = "greater")
```

```
The hypothesis is TRUE because the p value of 2.80% is less than or equal to the
confidence level of 5%.
```

TRUE

[126]:
```
#82926d84-c9c0-42ff-8b78-c3deab9945d0
#Error has decrased
vz %>%
    filter(plus.dbir_year %in% 2017:2022,subset.2022dbir) %>%
    filter(victim.industry2.51) %>%
    verisr::getenumCI2021("action", ci.method = "bootstrap", by="plus.
 ↪dbir_year", force=TRUE) %>%
    filter(enum=="Error") %>%
    select(-enum) %>%
    mutate(enum=by) %>%
    verisr::test_veris_hypothesis(Ea="2022", Eb="2018", direction = "less")
```

```
The hypothesis is TRUE because the p value of 4.20% is less than or equal to the
confidence level of 5%.
```

TRUE

[127]:
```
#6983eec1-bee8-43c3-8357-415001497746
#Use of stolen creds top
chunk <- vz %>%
    filter(plus.dbir_year==2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.51) %>%
    verisr::getenumCI2021("action.*.variety")

top_2 <- chunk %>% slice(2) %>% pull(enum) %>% as.character()

chunk %>%
    verisr::test_veris_hypothesis(Ea="Use of stolen creds", Eb=top_2, direction␣
 ↪= "greater")
```

```
The hypothesis is TRUE because the p value of 2.40% is less than or equal to the
confidence level of 5%.
```

TRUE

2.7 Industries: Professional Services (54)

[155]:
```
#da1f019d-c1c0-4afb-9b4c-355c060303eb
# DoS almost equals over half of incidents for professional services

chunk <- vz %>%
    filter(plus.dbir_year==2022,subset.2022dbir) %>%
    filter(victim.industry2.54) %>%
    verisr::getenumCI2022("pattern", ci.method = "bayes")

chunk %>%
    verisr::test_veris_proportion(Ea = "Denial of Service", prop = .5,␣
 ↪direction = "equal")
```

```
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"2 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
The hypothesis is TRUE because the proportion of 50% is in the 95% range between
47.48% and 50.76%.
```

TRUE

[156]:
```
#'38688ce5-6e71-43d0-ba72-8e1e6ed77bbb
# Internal actor is less in 2022 than in 2021 report

chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
```

```
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.54) %>%
    verisr::getenumCI2022("actor", by="plus.dbir_year", ci.method = "bayes") %>%
    filter(enum=="Internal") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_hypothesis(Ea = "2022", Eb= "2021", direction = "less")
```

```
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

[157]:
```
# ddc2575e-9d80-4275-b1a0-754789b32631
# Mulitple actors has increased

chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.54) %>%
    verisr::getenumCI2022("actor", by="plus.dbir_year", ci.method = "bayes") %>%
    filter(enum=="Multiple") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_hypothesis(Ea = "2022", Eb= "2021", direction =␣
 ↪"greater")
```

```
The hypothesis is TRUE because the p value of 3.80% is less than or equal to the
confidence level of 5%.
```

TRUE

2.8 Industries: Mining and Utilities 21__22

[153]:
```
#61114788-9de5-4cb6-bc85-917f9028835f
#more than 60% of breaches are phishing
vz %>%
    filter(plus.dbir_year==2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
```

```
    filter(victim.industry2.21 | victim.industry2.22) %>%
    verisr::getenumCI2021("action.*.variety", ci.method = "bootstrap", force =␣
↪TRUE) %>%
    verisr::test_veris_proportion(Ea="Phishing", prop = .6, direction =␣
↪"greater")
```

```
The hypothesis is TRUE because the confidence of 99.20% at 60% is greater than
the confidence level of 95%.
```

TRUE

[154]:
```
#348639da-b087-4edc-94f1-39d28a76ed0b
#Email servers most commonet

chunk <- vz %>%
    filter(plus.dbir_year==2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.21 | victim.industry2.22) %>%
    verisr::getenumCI2021("asset.assets.variety", ci.method="bootstrap", force␣
↪= TRUE)

top2 <- chunk %>% slice(2) %>% pull(enum) %>% as.character()

chunk %>%
    verisr::test_veris_hypothesis(Ea="S - Mail", Eb=top2, direction = "greater")
```

```
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

2.9 Industries: Public Admin

[158]:
```
#'45c3f2f0-70bb-4686-b828-a2ffbf3ed758'
# Loss / Misconfiguration and Misdelivery are equal breach contributors
# TRUE
chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.92) %>%
    verisr::getenumCI2022("action.*.variety", ci.method = "bayes")

chunk %>%
```

```
    verisr::test_veris_hypothesis(Ea="Misconfiguration", Eb="Misdelivery",␣
 ↪direction = "greater")

chunk %>%
    verisr::test_veris_hypothesis(Ea="Misconfiguration", Eb="Misdelivery",␣
 ↪direction = "less")

chunk %>%
    verisr::test_veris_hypothesis(Ea="Loss", Eb="Misdelivery", direction =␣
 ↪"greater")

chunk %>%
    verisr::test_veris_hypothesis(Ea="Loss", Eb="Misdelivery", direction =␣
 ↪"less")


chunk %>%
    verisr::test_veris_hypothesis(Ea="Loss", Eb="Misconfiguration", direction =␣
 ↪"greater")

chunk %>%
    verisr::test_veris_hypothesis(Ea="Loss", Eb="Misconfiguration", direction =␣
 ↪"less")
```

```
The hypothesis is FALSE because the p value of 51.90% is greater than the
confidence level of 5%.
```

FALSE

```
The hypothesis is FALSE because the p value of 59.20% is greater than the
confidence level of 5%.
```

FALSE

```
The hypothesis is FALSE because the p value of 71.80% is greater than the
confidence level of 5%.
```

FALSE

```
The hypothesis is FALSE because the p value of 37.80% is greater than the
confidence level of 5%.
```

FALSE

```
The hypothesis is FALSE because the p value of 75.80% is greater than the
confidence level of 5%.
```

FALSE

```
The hypothesis is FALSE because the p value of 35.80% is greater than the
confidence level of 5%.
```

FALSE

[159]:
```
#'79ec2c13-7abf-432d-a877-b4519cd96117'
#CLAIM: Espionage has increased for public sector
#

chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.92) %>%
    verisr::getenumCI2022("actor.*.motive", by="plus.dbir_year", ci.method =␣
 ↪"bayes", na.rm = TRUE)%>%
    filter(enum=="Espionage") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_hypothesis(Ea="2022", Eb="2021", direction = "greater")
```

```
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

[163]:
```
#2021 espionage was 4%
#5306a1bb-8389-482c-a378-6e2d4ee43855

chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.92) %>%
    verisr::getenumCI2022("actor.*.motive", by="plus.dbir_year", ci.method =␣
 ↪"bayes", na.rm = TRUE)%>%
    filter(enum=="Espionage") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
```

```
    verisr::test_veris_proportion(Ea="2021", prop = .04, direction = "equal")
```

The hypothesis is TRUE because the proportion of 4% is in the 95% range between 2.33% and 5.29%.

TRUE

[160]:
```
#Credentials accounted for 80% of data comprmised in 2021
#'63f8297f-7c60-482e-8074-a0c3c91fd9e0'

chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.92) %>%
    verisr::getenumCI2022("attribute.confidentiality.data.variety", by="plus.
 ↪dbir_year", ci.method = "bayes", na.rm = TRUE)%>%
    filter(enum=="Credentials") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_proportion(Ea="2021", prop =.8, direction = "equal" )
```

The hypothesis is TRUE because the proportion of 80% is in the 95% range between 76.61% and 82.37%.

TRUE

[161]:
```
#9a8b39e3-8c21-41d2-bec0-15aa33f4326a
#Personal was 18% last year
chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.92) %>%
    verisr::getenumCI2022("attribute.confidentiality.data.variety", by="plus.
 ↪dbir_year", ci.method = "bayes", na.rm = TRUE)%>%
    filter(enum=="Personal") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_proportion(Ea="2021", prop =.18, direction = "equal" )
```

The hypothesis is TRUE because the proportion of 18% is in the 95% range between 15.98% and 21.15%.

TRUE

[ ]:
```
#077a613f-b5e7-4677-974b-a45abb2bc848
#INCREASE IN INTERNAL ACTORS
chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.92) %>%
    verisr::getenumCI2022("actor",by="plus.dbir_year", ci.method = "bayes") %>%
    filter(enum =="Internal") %>%
    select(-enum) %>%
    mutate(enum = by)



chunk %>%
    verisr::test_veris_hypothesis(Ea="2022", Eb="2021", direction = "greater")
```

[162]:
```
#214e87f2-5ee9-4c6b-b695-7edba3ec9ca0
#Just three years ago, the top motive was Espionage, with 66% of breaches. ␣
 ↪Five years ago, it was 64%,
#

chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.92) %>%
    verisr::getenumCI2022("actor.*.motive",by="plus.dbir_year", ci.method =␣
 ↪"bayes",na.rm = TRUE) %>%
    filter(enum =="Espionage") %>%
    select(-enum) %>%
    mutate(enum = by)



chunk %>%
    verisr::test_veris_proportion(Ea = "2019", prop = .66, direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea = "2017", prop = .64,direction = "equal")
```

```
The hypothesis is TRUE because the proportion of 66% is in the 95% range between
57.47% and 69.23%.
```

TRUE

```
The hypothesis is TRUE because the proportion of 64% is in the 95% range between
56.07% and 69.94%.
```

TRUE

2.10 Industries: Arts and entertainment (71)

[ ]:
```
## c1f4710c-0e44-4d63-899e-803d0a38531e
## DoS are more than 20% of incidents
chunk <- vz %>%
    filter(plus.dbir_year==2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.71) %>%
    verisr::getenumCI2022("action.*.variety")

chunk %>% verisr::test_veris_proportion(Ea="DoS", prop = .2, direction =␣
↪"greater")
```

[ ]:
```
#c1f4710c-0e44-4d63-899e-803d0a38531e
#Less medical data was compromised in 2022 than in 2021 {false changed: fixed}

chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.71) %>%
    verisr::getenumCI2021("attribute.confidentiality.data.variety", by="plus.
↪dbir_year", ci.method = "bayes") %>%
    filter(enum=="Medical") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk

chunk %>%
    verisr::test_veris_hypothesis(Ea="2022", Eb="2021", direction = "less")
```

2.11 Industries: Healthcare (62)

[114]:
```
# 0f856ba2-a46c-4198-a002-d0d7bcc305d5
#  While the make-up of the insider breach has moved from being largely␣
↪malicious Misuse incidents to the more benign (
chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.62)%>%
    filter(actor.Internal) %>%
    verisr::getenumCI2021("pattern", by="plus.dbir_year", na.rm = TRUE) %>%
    filter(enum=="Privilege Misuse" | enum=="Miscellaneous Errors")
```

```
chunk %>%
    filter(enum=="Privilege Misuse") %>%
    select(-enum) %>%
    mutate(enum=by) %>%
    verisr::test_veris_proportion(Ea = "2015", prop = .5, direction = "greater")

chunk %>%
    filter(enum=="Miscellaneous Errors") %>%
    select(-enum) %>%
    mutate(enum=by) %>%
    verisr::test_veris_proportion(Ea="2021", prop = .5, direction = "greater")
```

```
The hypothesis is TRUE because the confidence of 97.90% at 50% is greater than
the confidence level of 95%.
```

TRUE

```
The hypothesis is TRUE because the confidence of 100.00% at 50% is greater than
the confidence level of 95%.
```

TRUE

[115]:
```
#'bcedd7f3-e985-4069-8311-ea3c1c8a7768'
#With the rise of the Basic Web Application Attacks pattern in this vertical
chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.62)%>%
    #filter(actor.Internal) %>%
    verisr::getenumCI2021("pattern", by="plus.dbir_year", na.rm = TRUE) %>%
    filter(enum=="Basic Web Application Attacks") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_hypothesis(Ea="2021", Eb="2018", direction = "greater")
```

```
The hypothesis is TRUE because the p value of 1.40% is less than or equal to the
confidence level of 5%.
```

TRUE

[116]:
```
##d42e3eff-c1e5-498e-8bd7-9f8a88192605
##With the increase in ransomware, comes the Discovery Method of Actor␣
 ↪Disclosure
chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
```

```
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.62)%>%
    #filter(actor.Internal) %>%
    verisr::getenumCI2021("discovery_method.*.variety", by="plus.dbir_year", na.
↪rm = TRUE) %>%
    filter(enum=="Actor disclosure") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_hypothesis(Ea="2021", Eb="2018", direction = "greater")
```

```
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

[10]:
```
#d8a153af-a4de-4422-8b07-8eda021145ff
#For the third year in a row, Personal data is compromised more often than␣
↪Medical
#Fixed
chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.62)%>%
    #filter(actor.Internal) %>%
    verisr::getenumCI2021("attribute.confidentiality.data.variety", by="plus.
↪dbir_year", na.rm = TRUE, ci.method = "bayes") %>%
    filter(enum=="Medical" | enum=="Personal")

chunk %>%
    filter(by=="2020") %>%
    verisr::test_veris_hypothesis(Ea="Personal", Eb="Medical", direction =␣
↪"greater")

chunk %>%
    filter(by=="2021") %>%
    verisr::test_veris_hypothesis(Ea="Personal", Eb="Medical", direction =␣
↪"greater")
```

```
The hypothesis is TRUE because the p value of 0.10% is less than or equal to the
confidence level of 5%.
```

TRUE

The hypothesis is TRUE because the p value of 0.20% is less than or equal to the confidence level of 5%.

TRUE

2.12 Industries: Retail (44-45)

[165]:
```
#b23256c0-9c4e-4472-8c4f-a8741696ad18
#Social attacks, roughly split between Pphishing (53%) and Ppretexting (47%)
chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.44_45)%>%
    verisr::getenumCI2021("action.social.variety", ci.method = "bootstrap",␣
 ↪force = TRUE)

chunk %>%
    verisr::test_veris_proportion(Ea="Phishing", prop = .53, direction =␣
 ↪"equal")

chunk %>%
    verisr::test_veris_proportion(Ea="Pretexting", prop = .47, direction =␣
 ↪"equal")
```

The hypothesis is TRUE because the proportion of 53% is in the 95% range between 40.32% and 66.13%.

TRUE

The hypothesis is TRUE because the proportion of 47% is in the 95% range between 33.87% and 58.06%.

TRUE

[169]:
```
#5125ee18-2647-4c7c-bb55-b60ccfcb0f19
# Social  7% in 2016, 13% in 2018, 29% this year.

chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.44_45)%>%
    verisr::getenumCI2021("action", by="timeline.incident.year",ci.method =␣
 ↪"bayes") %>%
    filter(enum=="Social") %>%
    select(-enum) %>%
```

```
    mutate(enum=by)

chunk %>%
    verisr::test_veris_proportion(Ea="2016", prop = .07, direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea="2018", prop = .13, direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea="2021", prop=.29, direction = "equal")
```

```
Warning message in verisr::getenumCI2021(., "action", by =
"timeline.incident.year", :
"Removing all rows with n < 5.  5 is the smallest samples size appropriate for
the DBIR. Use force = TRUE to avoid this removal.  If this leaves n othing but
columns with no sample size, those will be removed as well."
The hypothesis is TRUE because the proportion of 7% is in the 95% range between
2.50% and 13.75%.
```

TRUE

```
The hypothesis is TRUE because the proportion of 13% is in the 95% range between
7.50% and 20.00%.
```

TRUE

```
The hypothesis is TRUE because the proportion of 29% is in the 95% range between
22.56% and 35.38%.
```

TRUE

[175]:
```
#'8625dd17-46ed-4d0a-9b82-0696d3a1b348
#Ccredentials are later utilized to hack into servers and, load ransomware (47%)

vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.44_45)%>%
    filter(attribute.confidentiality.data.variety.Credentials) %>%
    verisr::getenumCI2021("action.*.variety",ci.method = "bayes") %>%
    verisr::test_veris_proportion("Ransomware", prop = .47, direction = "equal")
```

```
The hypothesis is TRUE because the proportion of 47% is in the 95% range between
38.54% and 58.33%.
```

TRUE

[181]:
```
#'0f7982b4-574a-49f1-833e-299bd7bde789
# 50% of breaches were detected by fraud or law enforcement
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.industry2.44_45)%>%
    mutate(discovery_method.external.variety.Fraud_Law = `discovery_method.
↪external.variety.Fraud detection` |`discovery_method.external.variety.Law␣
↪enforcement`) %>%
    verisr::getenumCI2021("discovery_method.*.variety", ci.method =␣
↪"bootstrap")  %>%
    verisr::test_veris_proportion(Ea="Fraud_Law", prop = .5, direction =␣
↪"equal")
```

```
The hypothesis is TRUE because the proportion of 50% is in the 95% range between
36.99% and 60.27%.
```

TRUE

3 Regions

3.1 Regions: EMEA

[201]:
```
#20d08a92-6a41-4a8b-b6e8-eab208807311
#Increase in social engineering
chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "EMEA") %>%
    verisr::getenumCI2022("pattern", by="plus.dbir_year",ci.method="bayes") %>%
    filter(enum=="Social Engineering") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_hypothesis(Ea="2022", Eb="2021", direction = "greater")
```

```
Warning message in verisr::getenumCI2022(., "pattern", by = "plus.dbir_year",
ci.method = "bayes"):
"Removing all rows with n < 5.  5 is the smallest samples size appropriate for
the DBIR. Use force = TRUE to avoid this removal.  If this leaves n othing but
columns with no sample size, those will be removed as well."
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
```

confidence level of 5%.

TRUE

[204]:
```
#'30825286-b84e-43d0-96c8-31505a85f63f'
#SE pattern accoutns of approximately 60% of breaches
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "EMEA") %>%
    verisr::getenumCI2021("pattern", ci.method = "bootstrap") %>%
    verisr::test_veris_proportion(Ea="Social Engineering", prop = .6, direction␣
 ↪= 'equal')
```

The hypothesis is TRUE because the proportion of 60% is in the 95% range between 53.42% and 64.50%.

TRUE

[ ]:
```
#In 2019 BWAA represented 50% of the breaches in this region
#'eeb42c1e-7efe-43f1-8ec8-069a19cdc380
chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "EMEA") %>%
    verisr::getenumCI2022("pattern", by="plus.dbir_year",ci.method="bayes") %>%
    filter(enum=="Basic Web Application Attacks") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_proportion(Ea="2021", prop = .5, direction = "equal")
```

[205]:
```
#In 2019 BWAA has decerased
#ae36c617-a42d-40bd-aa9c-bcc4faa83213

chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "EMEA") %>%
    verisr::getenumCI2022("pattern", by="plus.dbir_year",ci.method="bayes") %>%
    filter(enum=="Basic Web Application Attacks") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_hypothesis(Ea="2022", Eb="2021", direction = "less")
```

```
Warning message in verisr::getenumCI2022(., "pattern", by = "plus.dbir_year",
ci.method = "bayes"):
"Removing all rows with n < 5.  5 is the smallest samples size appropriate for
the DBIR. Use force = TRUE to avoid this removal.  If this leaves n othing but
columns with no sample size, those will be removed as well."
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

[209]:
```
#5d93125d-9395-4830-8b39-263cca241548
#2022 BWA is 15%

chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "EMEA") %>%
    verisr::getenumCI2022("pattern", by="plus.dbir_year",ci.method="bayes") %>%
    filter(enum=="Basic Web Application Attacks") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_proportion(Ea="2022", prop = .15, direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea="2021", prop = .5, direction = "equal")
```

```
Warning message in verisr::getenumCI2022(., "pattern", by = "plus.dbir_year",
ci.method = "bayes"):
"Removing all rows with n < 5.  5 is the smallest samples size appropriate for
the DBIR. Use force = TRUE to avoid this removal.  If this leaves n othing but
columns with no sample size, those will be removed as well."
The hypothesis is TRUE because the proportion of 15% is in the 95% range between
11.73% and 19.55%.
```

TRUE

```
The hypothesis is TRUE because the proportion of 50% is in the 95% range between
49.01% and 60.26%.
```

TRUE

[210]:
```
#System intrusion is 30%
#8a0cec78-0524-463d-b7bf-a6e517a08d96
chunk <- vz %>%
    filter(plus.dbir_year %in% 2015:2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "EMEA") %>%
    verisr::getenumCI2022("pattern", by="plus.dbir_year",ci.method="bayes") %>%
    filter(enum=="System Intrusion") %>%
    select(-enum) %>%
    mutate(enum=by)

chunk %>%
    verisr::test_veris_proportion(Ea="2022", prop = .3, direction = "equal")
```

```
Warning message in verisr::getenumCI2022(., "pattern", by = "plus.dbir_year",
ci.method = "bayes"):
"Removing all rows with n < 5.  5 is the smallest samples size appropriate for
the DBIR. Use force = TRUE to avoid this removal.  If this leaves n othing but
columns with no sample size, those will be removed as well."
The hypothesis is TRUE because the proportion of 30% is in the 95% range between
26.06% and 36.48%.
```

TRUE

[211]:
```
# Finance ppl are the preferred target for phishing

# 'b9dfb082-e274-4212-a6a3-7af4777f26c0

chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "EMEA") %>%
    filter(action.social.variety.Phishing) %>%
    verisr::getenumCI2022("asset.assets.variety",ci.method="bayes")

chunk %>%
    verisr::test_veris_proportion(Ea="P - Finance", prop = .5, direction =␣
 ↪"greater")
```

```
The hypothesis is TRUE because the confidence of 100.00% at 50% is greater than
the confidence level of 95%.
```

TRUE

3.2 Regions: NA

[212]:
```
#Most common malware is ransomware
#'8276a245-14a8-46bd-a7f9-eea8606a959
chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "NA") %>%
    verisr::getenumCI2021("action.malware.variety", ci.method = "bootstrap")

top2 <- chunk %>% slice(2) %>% pull(enum) %>% as.character()

chunk %>%
    verisr::test_veris_hypothesis(Ea="Ransomware", Eb=top2, direction =␣
 ↪"greater")
```

```
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

[214]:
```
#Most common social is Phishing
#66e3362e-ea7b-479c-90cc-4f36cf512d67
chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "NA") %>%
    verisr::getenumCI2021("action.social.variety", ci.method = "bootstrap")

top2 <- chunk %>% slice(2) %>% pull(enum) %>% as.character()

chunk %>%
    verisr::test_veris_hypothesis(Ea="Phishing", Eb=top2, direction = "greater")
```

```
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

[213]:
```
#Credentials most common data taken
#50546761-7dbe-48c0-8798-2511c55ace66
chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "NA") %>%
    verisr::getenumCI2021("attribute.confidentiality.data.variety", ci.method =␣
 ↪"bootstrap")

top2 <- chunk %>% slice(2) %>% pull(enum) %>% as.character()

chunk %>%
    verisr::test_veris_hypothesis(Ea="Credentials", Eb=top2, direction =␣
 ↪"greater")
```

```
Warning message:
"Please be cautious in reporting a p-value of 0. This result is an approximation
based on the number of `reps` chosen in the `generate()` step. See
`?get_p_value()` for more information."
The hypothesis is TRUE because the p value of 0.00% is less than or equal to the
confidence level of 5%.
```

TRUE

3.3 Regions: APAC

[182]:
```
#well-known trifecta of Hacking (58%), Social (48%), and Malware (36%)
#2b62e7a2-5c30-46eb-868a-5ad66b889554

chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "APAC") %>%
    verisr::getenumCI2021("action")

chunk %>%
    verisr::test_veris_proportion(Ea="Hacking", prop=.58, direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea="Social", prop = .48, direction = "equal")

chunk %>%
    verisr::test_veris_proportion(Ea="Malware", prop = .36, direction = "equal")
```

```
The hypothesis is TRUE because the proportion of 58% is in the 95% range between
52.65% and 64.66%.
```

TRUE

The hypothesis is TRUE because the proportion of 48% is in the 95% range between 42.05% and 53.71%.

TRUE

The hypothesis is TRUE because the proportion of 36% is in the 95% range between 29.33% and 40.28%.

TRUE

[188]:
```
#The majority of attacks were perpetrated with Financial (81%) motives.
#39a4c715-2afe-417d-a545-0c5e21e3f0a5
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "APAC") %>%
    verisr::getenumCI2021("actor.*.motive", ci.method = "bayes", na.rm = TRUE)␣
↪%>%
    verisr::test_veris_proportion(Ea="Financial", prop = .81, direction =␣
↪"equal")
```

The hypothesis is TRUE because the proportion of 81% is in the 95% range between 76.31% and 85.55%.

TRUE

[189]:
```
#motive of Espionage (19%)
#442a1af0-1ba1-44b5-b9e5-1cd791a38c61
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "APAC") %>%
    verisr::getenumCI2021("actor.*.motive", ci.method = "bayes", na.rm = TRUE)␣
↪%>%
    verisr::test_veris_proportion(Ea="Espionage", prop = .19, direction =␣
↪"equal")
```

The hypothesis is TRUE because the proportion of 19% is in the 95% range between 14.06% and 23.69%.

TRUE

[191]:
```
#state-affiliated (19%) and nation-state (1%) actors
#c649c7be-dfc7-4b72-b3ee-03a903287ddb
chunk <- vz %>%
```

```
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "APAC") %>%
    verisr::getenumCI2021("actor.*.variety", ci.method = "bayes", na.rm = TRUE)


chunk %>%
    verisr::test_veris_proportion(Ea="Nation-state", prop = .01, direction =␣
↪"equal")

chunk %>%
    verisr::test_veris_proportion(Ea="State-affiliated", prop = .19, direction␣
↪= "equal")
```

The hypothesis is TRUE because the proportion of 1% is in the 95% range between 0.00% and 2.77%.

TRUE

The hypothesis is TRUE because the proportion of 19% is in the 95% range between 14.23% and 24.11%.

TRUE

[193]:
```
#Hacking action was 'Use of stolen credentials' (83%) being mostly used to␣
↪compromise a web application (60%).
#45521194-7b9b-4da5-bfc0-ef8a528a74e3
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "APAC") %>%
    verisr::getenumCI2021("action.hacking.variety", ci.method = "bayes", na.rm␣
↪= TRUE) %>%
    verisr::test_veris_proportion(Ea="Use of stolen creds", prop = .83,␣
↪direction = "equal")

vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "APAC") %>%
    verisr::getenumCI2021("action.hacking.vector", ci.method = "bayes", na.rm =␣
↪TRUE) %>%
    verisr::test_veris_proportion(Ea="Web application", prop = .60, direction =␣
↪"equal")
```

The hypothesis is TRUE because the proportion of 83% is in the 95% range between 76.67% and 90.00%.

TRUE

The hypothesis is TRUE because the proportion of 60% is in the 95% range between 52.24% and 67.74%.

TRUE

[195]:
```
# social in apac  and were comprised almost exclusively of Phishing (99%).
#f4ba2bff-954e-4887-bac8-67a4058012ff
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "APAC") %>%
    verisr::getenumCI2021("action.social.variety", ci.method = "bayes", na.rm =␣
↪TRUE) %>%
    verisr::test_veris_proportion(Ea="Phishing", prop = .99, direction =␣
↪"equal")
```

The hypothesis is TRUE because the proportion of 99% is in the 95% range between 97.79% and 100.00%.

TRUE

[196]:
```
# 10% ransomware
#'827d65dd-b993-41e9-80a3-0ccdc052abeb
chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "APAC") %>%
    verisr::getenumCI2021("action.*.variety", ci.method = "bayes", na.rm = TRUE)


chunk %>%
    verisr::test_veris_proportion(Ea="Ransomware", prop = .1, direction =␣
↪"equal")
```

The hypothesis is TRUE because the proportion of 10% is in the 95% range between 9.93% and 17.65%.

TRUE

[197]:
```
# 25% ranssomware for overall
#'8620306e-6b69-42c2-b697-90bb67de4b54
chunk <- vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
```

```
    filter(victim.Super_region != "APAC") %>%
    verisr::getenumCI2021("action.*.variety", ci.method = "bayes", na.rm = TRUE)


chunk %>%
    verisr::test_veris_proportion(Ea="Ransomware", prop = .25, direction =␣
↪"equal")
```

```
Warning message in binom::binom.bayes(as.integer(enum_subchunk[enum_subchunk$x
!= :
"4 confidence intervals failed to converge (marked by '*').
  Try changing 'tol' to a different value."
The hypothesis is TRUE because the proportion of 25% is in the 95% range between
23.90% and 26.89%.
```

TRUE

[200]:
```
# INtegrity 75%
#'72d0b752-ca74-4617-804f-dce9cb861287

 vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "APAC") %>%
    verisr::getenumCI2021("attribute", ci.method = "bayes") %>%
    verisr::test_veris_proportion(Ea="Integrity", prop = .75, direction =␣
↪"equal")
```

```
The hypothesis is TRUE because the proportion of 75% is in the 95% range between
65.72% and 75.62%.
```

TRUE

3.4 Regions: LATAM

[215]:
```
#95e88cd3-9a43-4aac-b2cc-08c1c93c9267
#Ransomware 37%
 vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "LAC") %>%
    verisr::getenumCI2021("action.*.variety", ci.method="bootstrap") %>%
    verisr::test_veris_proportion(Ea="Ransomware", prop=.37, direction =␣
↪"equal")
```

```
The hypothesis is TRUE because the proportion of 37% is in the 95% range between
26.97% and 47.19%.
```

TRUE

[217]:
```
#e881c70b-cf7b-455d-9fad-81c1e700d645
#25% Dos
 vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    #filter(attribute.confidentiality.data_disclosure.Yes) %>%
    filter(victim.Super_region == "LAC") %>%
    verisr::getenumCI2021("action.*.variety", ci.method="bootstrap") %>%
    verisr::test_veris_proportion(Ea="Ransomware", prop=.27, direction =␣
 ↪"equal")
```

```
The hypothesis is TRUE because the proportion of 27% is in the 95% range between
26.97% and 46.07%.
```

TRUE

3.5 MISC

[33]:
```
#51a935b6-eb9b-4eef-8279-ac98f0ceb9e1
# ransomware is 25% of breachers
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("action.*.variety", ci.method="bootstrap") %>%
    filter(enum=="Ransomware") %>%
    verisr::test_veris_proportion(Ea="Ransomware",prop = .25, direction =␣
 ↪"equal" )
```

```
The hypothesis is TRUE because the proportion of 25% is in the 95% range between
23.25% and 25.86%.
```

TRUE

[57]:
```
#Phishing is 19% of breaches
#'c361b4f2-ab36-4cd0-ab02-25fa01117ac9
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    verisr::getenumCI2022("action.*.variety", ci.method="bootstrap") %>%
    verisr::test_veris_proportion("Phishing", prop = .19, direction = 'equal')
```

```
The hypothesis is TRUE because the proportion of 19% is in the 95% range between
17.65% and 20.13%.
```

TRUE

[12]:
```
#which continues today with over 50% breaches leveraging remote access or web␣
↪applications.
#'b1d012b6-dad7-451f-9c94-d289108d20c3'
vz %>%
    filter(plus.dbir_year == 2022,subset.2022dbir) %>%
    filter(attribute.confidentiality.data_disclosure.Yes) %>%
    mutate(action.extra.vector.Web_desktop_vpn = `action.hacking.vector.Web␣
↪application` |
             `action.hacking.vector.Desktop sharing software` | `action.hacking.
↪vector.VPN` |
             `action.hacking.vector.Desktop sharing`) %>%
    verisr::getenumCI2021("action.*.vector", ci.method = "bootstrap") %>%
    verisr::test_veris_proportion(Ea="Web_desktop_vpn", prop = .5, direction =␣
↪'greater')
```

```
The hypothesis is TRUE because the confidence of 100.00% at 50% is greater than
the confidence level of 95%.
```

TRUE