

PSET5 - Boseong Yun

Boseong Yun

12/03/2020

In Lopez-Osorio et al (2017), posted on Canvas, the authors undertake an exploratory study. They ask what sort of factors relating to incidents of reported domestic violence (DV), and the parties involved, can predict DV recidivism, where DV recidivism is defined as one or more repeat calls to the police within six months. Although they test 66 candidate predictors (by my count), they take no account of multiple testing. Here you re-analyze their results, employing methods to control both FWER and FDR.

1. Extract data from Tables 1 through 4 of the article, constructing a dataset or dataframe that can be analyzed with a statistical package.

```
# Data Preparation -----

tab1a <- read_xlsx("table1.xlsx", skip = 1)
tab2a <- read_xlsx("table2.xlsx", skip = 1)
tab3a <- read_xlsx("table3.xlsx", skip = 1)
tab4a <- read_xlsx("table4.xlsx", skip = 1)

# Creating a function that cleans the data -----

clean <- function(data) {
  data %>%
    set_names(., nm = tolower(names(data))) %>%
    separate(coefficient, into = c("coef", "stdr"), sep = 4) %>%
    mutate(stdr = str_replace_all(stdr, "\\*|\\[|\\]", "")) %>%
    separate(stdr, into = c("l.conf", "h.conf"), sep = "-") %>%
    mutate_at(.vars = c("l.conf", "h.conf"), as.numeric) %>%
    mutate(
      chi_95 = qchisq(p = 0.95, df = 1),
      chi_90 = qchisq(p = 0.90, df = 1)
    ) %>%
    rename(valid_prop = '% valid')
}

# Data Cleaned -----

# Cleaning each table
tab1 <- clean(tab1a)
tab2 <- clean(tab2a)
tab3 <- clean(tab3a)
tab4 <- clean(tab4a)

# Combining the tables
combined <- tab1 %>%
  bind_rows(tab2) %>%
```

```
bind_rows(tab3) %>%
bind_rows(tab4)
```

2. The authors use chi-square statistics to test for association between their candidate predictors and DV recidivism. On the basis of the authors' analysis, how many significant predictors are there at the 10 percent level?

Answer: There are 35 predictions that are significant at the 10 percent level.

```
# The general rule for calculating degrees of freedom a chi-square test is (r-1)(c-1).
# Since the predictors are indicator variables, we can think of 3 by 2 contingency tables.
# Therefore, the right df is (2-1)(2-1) = 1
combined %>%
  filter(chi_squared > chi_90) %>%
  select(indicators, chi_squared) %>%
  kable(
    caption = "Significanct Predictors",
    col.names = c("Indicators", "$ ^2$"),
  )
```

Table 1: Significanct Predictors

| Indicators | χ^2 |
|---------------------------------------------------------------------------------------------------------------|----------|
| I-2. Psychological Violence | 3.370 |
| Serious psychological violence | 8.810 |
| I-6. Use of weapons/objects against partner | 9.510 |
| Use of a cold weapon | 13.481 |
| Serious threats from the aggressor | 21.610 |
| Suicide threats from the aggressor | 11.350 |
| Death threats from the aggressor | 2.750 |
| I-10. An escalation in aggression or threats in the last 6 months | 9.700 |
| I-11. They have shown exaggerated jealousy or have suspected their partner of infidelity in the last 6 months | 21.940 |
| I-12. They have shown controlling behaviors in the last 6 months | 38.360 |
| Physically controlling behaviors | 33.280 |
| Psychologically controlling behaviors | 33.500 |
| Controlling behaviors towards academic/labor aspects | 9.810 |
| Economically controlling behaviors | 8.270 |
| Cybernetically controlling behaviors | 23.860 |
| I-13. In the last 6 months they have shown harassing behaviors | 11.900 |
| I-14. In the last year they have done material damage | 7.540 |
| I-15. In the last year they have disrespected an authority figure | 12.000 |
| I-16. In the last year they have committed an aggression against third parties and/or animals. | 4.180 |
| I-17. In the last year they have made threats and slights towards third parties. | 5.050 |
| I-18. Presence of problems in their life in the last 6 months. | 13.170 |
| Economic or work-related problems | 9.310 |
| Problems with the justice system | 10.900 |
| I-19. Presence of a criminal record | 31.570 |
| I-20. Presence of past breakings of sentence conditions | 13.730 |
| I-21. Presence of physical/sexual aggression records | 18.360 |
| I-22. Presence of gender violence records against other partners | 10.290 |
| I-24. Presence of suicidal ideas or attempts | 17.550 |
| I-25. Presence of any type of addiction or substance abuse (alcohol and drugs) | 12.180 |
| I-26. Presence of a gender or domestic violence history within their family | 11.620 |
| I-30. Victim is a foreigner | 6.980 |

| Indicators | χ^2 |
|------------------------------------------------------------------------------------------------------------------|----------|
| I-39. The victim has suffered gender violence by the hands of other aggressors in the past | 6.290 |
| I-40. The victim has expressed their desire to end the relationship to the aggressor at least 6 months ago | 16.420 |
| I-42. The victim fears for the physical integrity of the minors or family members under their care | 2.980 |
| I-43. The victim thinks that the aggressor is capable of attacking them in a violent way or even of killing them | 9.180 |

3. Obtain p-values associated with the chi-square statistics and plot their distribution. Under the null hypothesis, one can show that p-values are uniformly distributed. Given that, would you say the predictors are basically all noise, or does the histogram suggest there is some predictive signal among them?

Answer: We expect the distribution of p-value to be uniform if the null hypothesis is true and the relevant assumptions are met. In this case, however, the distribution is not uniform and there is a high frequency with the low p-values. This suggests that there is some predictive signal among them.

<https://stats.stackexchange.com/questions/10613/why-are-p-values-uniformly-distributed-under-the-null>

```
# Calculating P-Values
p_val_df <- combined %>%
  mutate(p_value = pchisq(chi_squared, 1, lower.tail = FALSE)) %>%
  dplyr::select(indicators, chi_squared, p_value)

p_val_df %>%
  round_df(digits = 5) %>%
  kable(
    caption = "The P-Values Associated with the  $\chi^2$  Statistics",
    col.names = c("Indicators", " $\chi^2$ ", "P-Value")
  )
```

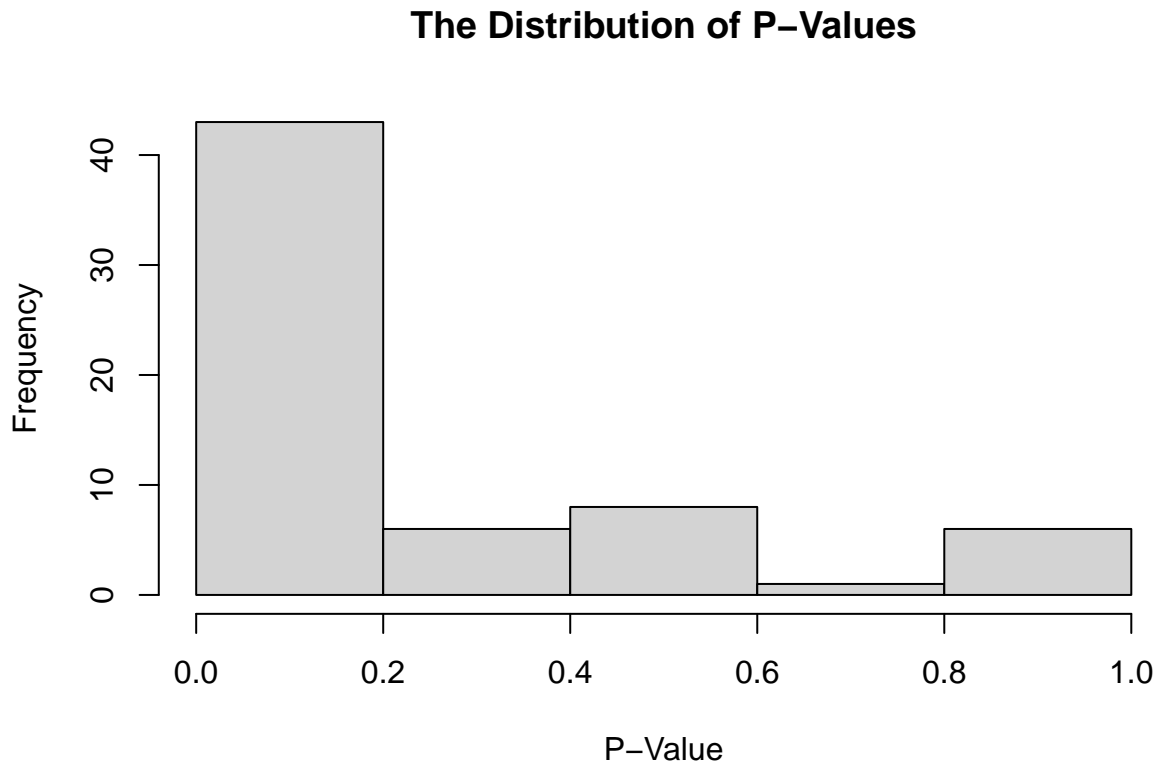
Table 2: The P-Values Associated with the χ^2 Statistics

| Indicators | χ^2 | P-Value |
|--------------------------------------------------------------------|----------|---------|
| I-1. Presence of any type of violence carried out by the aggressor | 0.810 | 0.36812 |
| I-2. Psychological Violence | 3.370 | 0.06639 |
| Serious psychological violence | 8.810 | 0.00300 |
| I-3. Physical violence | 0.000 | 1.00000 |
| Physical violence with injuries | 0.910 | 0.34011 |
| Serious physical violence | 2.600 | 0.10686 |
| I-4. Sexual violence | 0.520 | 0.47084 |
| Sexual violence with injuries | 2.150 | 0.14257 |
| Serious sexual violence | 0.470 | 0.49299 |
| I-5. Victim's defensive reaction to physical aggression | 1.234 | 0.26663 |
| I-6. Use of weapons/objects against partner | 9.510 | 0.00204 |
| Use of a cold weapon | 13.481 | 0.00024 |
| Use of a firearm | 0.370 | 0.54300 |
| Use of objects | 1.743 | 0.18676 |
| I-7. The aggressor has access to firearms | 0.430 | 0.51199 |
| I-8. The aggressor is an expert in combat techniques | 1.710 | 0.19099 |
| I-9. Presence of threats/plans aimed to cause the victim harm | 0.530 | 0.46661 |
| Serious threats from the aggressor | 21.610 | 0.00000 |

| Indicators | χ^2 | P-Value |
|---------------------------------------------------------------------------------------------------------------|----------|---------|
| Suicide threats from the aggressor | 11.350 | 0.00075 |
| Economic threats from the aggressor | 0.100 | 0.75183 |
| Death threats from the aggressor | 2.750 | 0.09725 |
| Threats to social reputation | 0.970 | 0.32468 |
| Threats to children's integrity/custody | 0.060 | 0.80650 |
| I-10. An escalation in aggression or threats in the last 6 months | 9.700 | 0.00184 |
| I-11. They have shown exaggerated jealousy or have suspected their partner of infidelity in the last 6 months | 21.940 | 0.00000 |
| I-12. They have shown controlling behaviors in the last 6 months | 38.360 | 0.00000 |
| Physically controlling behaviors | 33.280 | 0.00000 |
| Psychologically controlling behaviors | 33.500 | 0.00000 |
| Controlling behaviors towards academic/labor aspects | 9.810 | 0.00174 |
| Economically controlling behaviors | 8.270 | 0.00403 |
| Cybernetically controlling behaviors | 23.860 | 0.00000 |
| I-13. In the last 6 months they have shown harassing behaviors | 11.900 | 0.00056 |
| I-14. In the last year they have done material damage | 7.540 | 0.00603 |
| I-15. In the last year they have disrespected an authority figure | 12.000 | 0.00053 |
| I-16. In the last year they have committed an aggression against third parties and/or animals. | 4.180 | 0.04090 |
| I-17. In the last year they have made threats and slights towards third parties. | 5.050 | 0.02463 |
| I-18. Presence of problems in their life in the last 6 months. | 13.170 | 0.00028 |
| Economic or work-related problems | 9.310 | 0.00228 |
| Problems with the justice system | 10.900 | 0.00096 |
| I-19. Presence of a criminal record | 31.570 | 0.00000 |
| I-20. Presence of past breakings of sentence conditions | 13.730 | 0.00021 |
| I-21. Presence of physical/sexual aggression records | 18.360 | 0.00002 |
| I-22. Presence of gender violence records against other partners | 10.290 | 0.00134 |
| I-23. Presence of a mental/psychiatric disorder | 1.810 | 0.17851 |
| I-24. Presence of suicidal ideas or attempts | 17.550 | 0.00003 |
| I-25. Presence of any type of addiction or substance abuse (alcohol and drugs) | 12.180 | 0.00048 |
| I-26. Presence of a gender or domestic violence history within their family | 11.620 | 0.00065 |
| I-27. Presence of any type of disability | 0.970 | 0.32468 |
| I-28. Victim is in gestation period | 0.050 | 0.82306 |
| I-29. Victim suffers from a serious illness | 0.000 | 1.00000 |
| I-30. Victim is a foreigner | 6.980 | 0.00824 |
| I-31. Victim has no social/family support | 0.010 | 0.92034 |
| I-32. Victim has a mental/psychiatric disorder | 0.530 | 0.46661 |
| I-33. Victim has suicidal ideas/attempts | 1.810 | 0.17851 |
| I-34. Victim has any kind of addiction or engages in substance abuse (alcohol and drugs) | 1.020 | 0.31252 |
| I-35. There is a history of gender or domestic violence within their family | 2.390 | 0.12211 |
| I-36. Victim is economically dependent on the aggressor | 1.690 | 0.19360 |
| I-37. The victim has minors or family members under their care | 0.400 | 0.52709 |
| I-38. The victim has withdrawn complaints in the past | 0.430 | 0.51199 |
| I-39. The victim has suffered gender violence by the hands of other aggressors in the past | 6.290 | 0.01214 |
| I-40. The victim has expressed their desire to end the relationship to the aggressor at least 6 months ago | 16.420 | 0.00005 |
| I-41. There have been reports of mutual violence | 0.010 | 0.92034 |
| I-42. The victim fears for the physical integrity of the minors or family members under their care | 2.980 | 0.08430 |

| Indicators | χ^2 | P-Value |
|------------------------------------------------------------------------------------------------------------------|----------|---------|
| I-43. The victim thinks that the aggressor is capable of attacking them in a violent way or even of killing them | 9.180 | 0.00245 |

```
# Creating a histogram
hist(p_val_df$p_value,
     main = "The Distribution of P-Values" ,
     xlab = "P-Value")
```



4. Construct Bonferroni p-values. Controlling the FWER at 10 percent, which predictors are significant? Which predictors lose their significance when you account for multiple testing using the Bonferroni correction?

```
# Creating a p-value df
p_df <- combined %>%
  mutate(
    p_value = pchisq(chi_squared, 1, lower.tail = FALSE),
    p_bf = p.adjust(p_value, "bonferroni"),
    p_bh = p.adjust(p_value, "BH")
  )

# Creating the FWER at 10 percent

# Constructing Bonferroni p-values: predictors still significant
p_df %>%
  mutate(p_bf = p.adjust(p_value, "bonferroni")) %>%
  filter(p_bf < 0.1) %>%
  dplyr::select(indicators, p_bf) %>%
```

```
round_df(digits = 5) %>%
kable(
  caption = "Siginitificant Predictors with Bonferroni Adjusted P-Value (FWER at 10%)",
  col.names = c("Indicators", "Adj. P-Value")
)
```

Table 3: Siginitificant Predictors with Bonferroni Adjusted P-Value (FWER at 10%)

| Indicators | Adj. P-Value |
|---------------------------------------------------------------------------------------------------------------|-----------------|
| Use of a cold weapon | 0.01542 |
| Serious threats from the aggressor | 0.00021 |
| Suicide threats from the aggressor | 0.04829 |
| I-11. They have shown exaggerated jealousy or have suspected their partner of infidelity in the last 6 months | 0.00018 |
| I-12. They have shown controlling behaviors in the last 6 months | 0.00000 |
| Physically controlling behaviors | 0.00000 |
| Psychologically controlling behaviors | 0.00000 |
| Cybernetically controlling behaviors | 0.00007 |
| I-13. In the last 6 months they have shown harassing behaviors | 0.03593 |
| I-15. In the last year they have disrespected an authority figure | 0.03405 |
| I-18. Presence of problems in their life in the last 6 months. | 0.01821 |
| Problems with the justice system | 0.06155 |
| I-19. Presence of a criminal record | 0.00000 |
| I-20. Presence of past breakings of sentence conditions | 0.01351 |
| I-21. Presence of physical/sexual aggression records | 0.00117 |
| I-22. Presence of gender violence records against other partners | 0.08560 |
| I-24. Presence of suicidal ideas or attempts | 0.00179 |
| I-25. Presence of any type of addiction or substance abuse (alcohol and drugs) | 0.03091 |
| I-26. Presence of a gender or domestic violence history within their family | 0.04176 |
| I-40. The victim has expressed their desire to end the relationship to the aggressor at least 6 months ago | 0.00325 |

```
# Constructing Bonferroni p-values: predictors losing significance
p_df %>%
  mutate(p_bf = p.adjust(p_value, "bonferroni")) %>%
  filter(chi_squared > chi_90) %>%
  filter(p_bf > 0.1) %>%
  dplyr::select(indicators, p_bf) %>%
  round_df(digits = 5) %>%
  kable(
    caption = "Predictors that lose significance with Bonferroni Adjusted P-Value (FWER at 10%)",
    col.names = c("Indicators", "Adj. P-Value")
  )
```

Table 4: Predictors that lose significance with Bonferroni Adjusted
P-Value (FWER at 10%)

| Indicators | Adj. P-Value |
|------------------------------------------------------------------------------------------------------------------|-----------------|
| I-2. Psychological Violence | 1.00000 |
| Serious psychological violence | 0.19173 |
| I-6. Use of weapons/objects against partner | 0.13079 |
| Death threats from the aggressor | 1.00000 |
| I-10. An escalation in aggression or threats in the last 6 months | 0.11793 |
| Controlling behaviors towards academic/labor aspects | 0.11108 |
| Economically controlling behaviors | 0.25796 |
| I-14. In the last year they have done material damage | 0.38620 |
| I-16. In the last year they have committed an aggression against third parties and/or animals. | 1.00000 |
| I-17. In the last year they have made threats and slights towards third parties. | 1.00000 |
| Economic or work-related problems | 0.14586 |
| I-30. Victim is a foreigner | 0.52752 |
| I-39. The victim has suffered gender violence by the hands of other aggressors in the past | 0.77709 |
| I-42. The victim fears for the physical integrity of the minors or family members under their care | 1.00000 |
| I-43. The victim thinks that the aggressor is capable of attacking them in a violent way or even of killing them | 0.15659 |

5. Now test for significance using the Benjamini-Hochberg approach, controlling the FDR at 10 percent. Produce a useful visualization that illustrates how the procedure works. According to this criterion, how many of the features are significant? In expectation, how many of these are false discoveries?

Answer: The test for significance using the Benjamin-Hochberg approach with the False Discovery Rate at 10 percent shows that we reject 32 tests. The visualization also shows that there are 32 significant predictors below the line. In expectation, there are $0.1 * (32) = 3.2$ false discoveries.

```
# Constructing Benjamini-Hochberg p-values: predictors still significant: 25
p_df %>%
  arrange(p_value) %>%
  mutate(p_bh = p.adjust(p_value, "BH")) %>%
  filter(p_bh < 0.1) %>%
  select(indicators) %>%
  kable(
    caption = "Significant Predictors with Benjamin-Hochberg Approach (FDR at 10%)",
    col.names = c("Indicators")
  )
```

Table 5: Significant Predictors with Benjamin-Hochberg Approach (FDR at 10%)

| Indicators |
|---------------------------------------------------------------------------------------------------------------|
| I-12. They have shown controlling behaviors in the last 6 months |
| Psychologically controlling behaviors |
| Physically controlling behaviors |
| I-19. Presence of a criminal record |
| Cybernetically controlling behaviors |
| I-11. They have shown exaggerated jealousy or have suspected their partner of infidelity in the last 6 months |

Indicators

Serious threats from the aggressor

I-21. Presence of physical/sexual aggression records

I-24. Presence of suicidal ideas or attempts

I-40. The victim has expressed their desire to end the relationship to the aggressor at least 6 months ago

I-20. Presence of past breakings of sentence conditions

Use of a cold weapon

I-18. Presence of problems in their life in the last 6 months.

I-25. Presence of any type of addiction or substance abuse (alcohol and drugs)

I-15. In the last year they have disrespected an authority figure

I-13. In the last 6 months they have shown harassing behaviors

I-26. Presence of a gender or domestic violence history within their family

Suicide threats from the aggressor

Problems with the justice system

I-22. Presence of gender violence records against other partners

Controlling behaviors towards academic/labor aspects

I-10. An escalation in aggression or threats in the last 6 months

I-6. Use of weapons/objects against partner

Economic or work-related problems

I-43. The victim thinks that the aggressor is capable of attacking them in a violent way or even of killing them

Serious psychological violence

Economically controlling behaviors

I-14. In the last year they have done material damage

I-30. Victim is a foreigner

I-39. The victim has suffered gender violence by the hands of other aggressors in the past

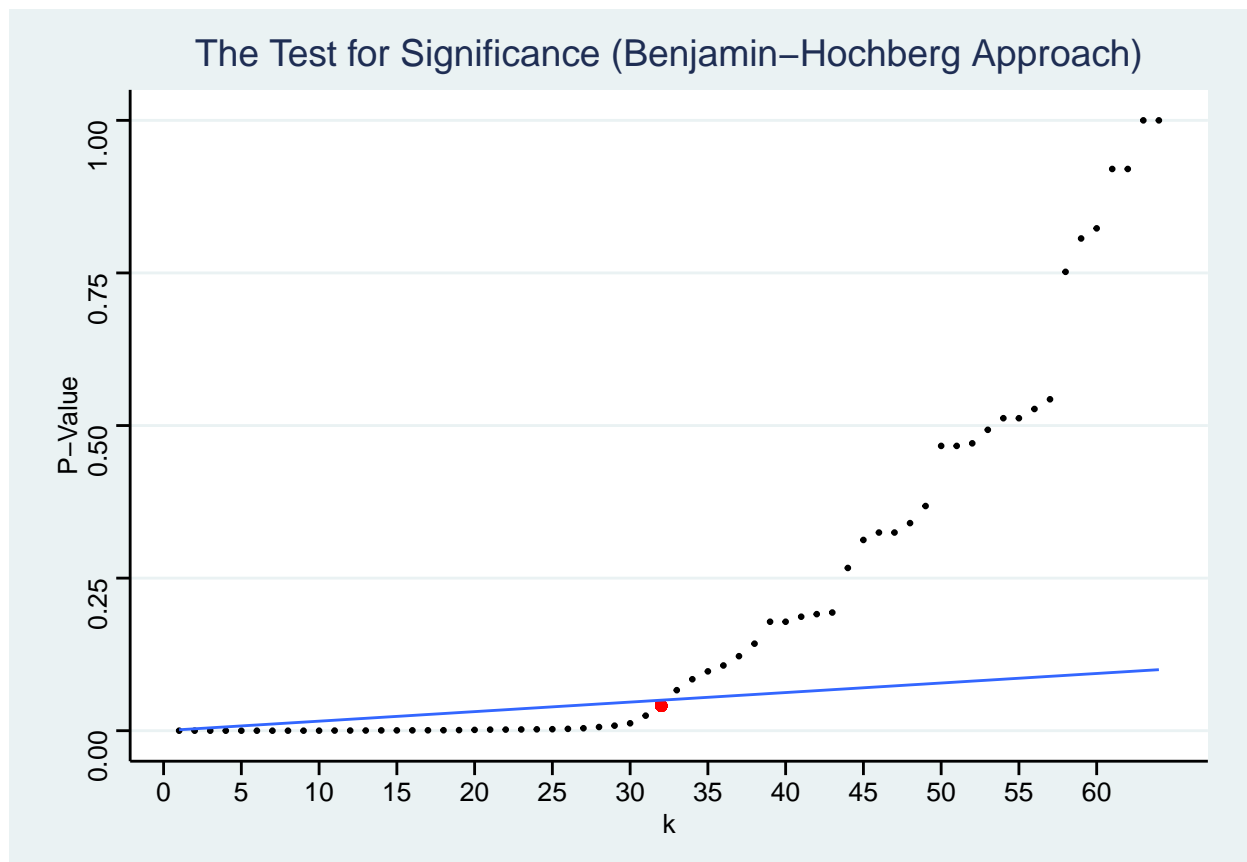
I-17. In the last year they have made threats and slights towards third parties.

I-16. In the last year they have committed an aggression against third parties and/or animals.

```
# Constructing Benjamini-Hochberg p-values manually: predictors still significant: 32
# Also refer to: https://brainder.org/2011/09/05/fdr-corrected-fdr-adjusted-p-values/
```

```
# Visualization that illustrates how the procedure works
```

```
p_df %>%
  arrange(p_value) %>%
  mutate(
    k = order(p_value),
    fdr_c = (k/nrow()) * 0.1 # false discovery rate set to 10 percent
  ) %>%
  ggplot(aes(x = k, y = p_value)) +
  geom_point(size = 0.5) +
  geom_point(aes(x = k[32], y = p_value[32]), color = "red") +
  geom_smooth(aes(x = k, y = fdr_c), size = 0.5) +
  scale_x_continuous(breaks = seq(0, 60, by = 5)) +
  theme_stata() +
  labs(
    title = "The Test for Significance (Benjamin-Hochberg Approach)",
    y = "P-Value"
  )
```

6. Compute the probability that each test represents a false discovery. Despite what I said in lecture, this does not require simulation. It merely requires solving one equation in one unknown for each test.

Notes: I have computed the q-value by hand. There many available packages that compute the q-value such as qvalue and fdr.tool but they produce slighgtly different results due to their respective assumptions about the number and characteristics of tests. I have thus calculated q-value by relying on the lecture slides and helps from Piazza post and emails

```
p_df %>%
  arrange(p_value) %>%
  mutate(
    rank = order(p_value),
    q = p_value * (nrow())/rank
  ) %>%
  select(indicators, p_value, q) %>%
  kable(caption = "The Probability That Each Test Represents a False Discovery")
```

Table 6: The Probability That Each Test Represents a False Discovery

| indicators | p_value | q |
|------------------------------------------------------------------|-----------|-----------|
| I-12. They have shown controlling behaviors in the last 6 months | 0.0000000 | 0.0000000 |
| Psychologically controlling behaviors | 0.0000000 | 0.0000002 |
| Physically controlling behaviors | 0.0000000 | 0.0000002 |
| I-19. Presence of a criminal record | 0.0000000 | 0.0000003 |
| Cybernetically controlling behaviors | 0.0000010 | 0.0000133 |

| indicators | p_value | q |
|------------------------------------------------------------------------------------------------------------------|-----------|-----------|
| I-11. They have shown exaggerated jealousy or have suspected their partner of infidelity in the last 6 months | 0.0000028 | 0.0000300 |
| Serious threats from the aggressor | 0.0000033 | 0.0000305 |
| I-21. Presence of physical/sexual aggression records | 0.0000183 | 0.0001463 |
| I-24. Presence of suicidal ideas or attempts | 0.0000280 | 0.0001990 |
| I-40. The victim has expressed their desire to end the relationship to the aggressor at least 6 months ago | 0.0000507 | 0.0003248 |
| I-20. Presence of past breakings of sentence conditions | 0.0002111 | 0.0012280 |
| Use of a cold weapon | 0.0002410 | 0.0012853 |
| I-18. Presence of problems in their life in the last 6 months. | 0.0002845 | 0.0014005 |
| I-25. Presence of any type of addiction or substance abuse (alcohol and drugs) | 0.0004830 | 0.0022082 |
| I-15. In the last year they have disrespected an authority figure | 0.0005320 | 0.0022699 |
| I-13. In the last 6 months they have shown harassing behaviors | 0.0005613 | 0.0022454 |
| I-26. Presence of a gender or domestic violence history within their family | 0.0006525 | 0.0024563 |
| Suicide threats from the aggressor | 0.0007545 | 0.0026826 |
| Problems with the justice system | 0.0009616 | 0.0032392 |
| I-22. Presence of gender violence records against other partners | 0.0013375 | 0.0042801 |
| Controlling behaviors towards academic/labor aspects | 0.0017357 | 0.0052896 |
| I-10. An escalation in aggression or threats in the last 6 months | 0.0018427 | 0.0053605 |
| I-6. Use of weapons/objects against partner | 0.0020436 | 0.0056864 |
| Economic or work-related problems | 0.0022791 | 0.0060775 |
| I-43. The victim thinks that the aggressor is capable of attacking them in a violent way or even of killing them | 0.0024467 | 0.0062637 |
| Serious psychological violence | 0.0029958 | 0.0073744 |
| Economically controlling behaviors | 0.0040306 | 0.0095539 |
| I-14. In the last year they have done material damage | 0.0060344 | 0.0137929 |
| I-30. Victim is a foreigner | 0.0082426 | 0.0181905 |
| I-39. The victim has suffered gender violence by the hands of other aggressors in the past | 0.0121421 | 0.0259032 |
| I-17. In the last year they have made threats and slights towards third parties. | 0.0246259 | 0.0508406 |
| I-16. In the last year they have committed an aggression against third parties and/or animals. | 0.0409037 | 0.0818074 |
| I-2. Psychological Violence | 0.0663938 | 0.1287636 |
| I-42. The victim fears for the physical integrity of the minors or family members under their care | 0.0842993 | 0.1586810 |
| Death threats from the aggressor | 0.0972544 | 0.1778367 |
| Serious physical violence | 0.1068637 | 0.1899799 |
| I-35. There is a history of gender or domestic violence within their family | 0.1221136 | 0.2112236 |
| Sexual violence with injuries | 0.1425699 | 0.2401177 |
| I-23. Presence of a mental/psychiatric disorder | 0.1785082 | 0.2929366 |
| I-33. Victim has suicidal ideas/attempts | 0.1785082 | 0.2856132 |
| Use of objects | 0.1867592 | 0.2915265 |
| I-8. The aggressor is an expert in combat techniques | 0.1909854 | 0.2910253 |
| I-36. Victim is economically dependent on the aggressor | 0.1936010 | 0.2881503 |
| I-5. Victim's defensive reaction to physical aggression | 0.2666305 | 0.3878262 |
| I-34. Victim has any kind of addiction or engages in substance abuse (alcohol and drugs) | 0.3125190 | 0.4444715 |
| Threats to social reputation | 0.3246802 | 0.4517289 |
| I-27. Presence of any type of disability | 0.3246802 | 0.4421177 |
| Physical violence with injuries | 0.3401144 | 0.4534859 |
| I-1. Presence of any type of violence carried out by the aggressor | 0.3681203 | 0.4808101 |
| I-9. Presence of threats/plans aimed to cause the victim harm | 0.4666069 | 0.5972568 |

| indicators | p_value | q |
|----------------------------------------------------------------|-----------|-----------|
| I-32. Victim has a mental/psychiatric disorder | 0.4666069 | 0.5855459 |
| I-4. Sexual violence | 0.4708417 | 0.5794975 |
| Serious sexual violence | 0.4929872 | 0.5953053 |
| I-7. The aggressor has access to firearms | 0.5119889 | 0.6068017 |
| I-38. The victim has withdrawn complaints in the past | 0.5119889 | 0.5957690 |
| I-37. The victim has minors or family members under their care | 0.5270893 | 0.6023877 |
| Use of a firearm | 0.5430043 | 0.6096890 |
| Economic threats from the aggressor | 0.7518296 | 0.8296051 |
| Threats to children's integrity/custody | 0.8064959 | 0.8748431 |
| I-28. Victim is in gestation period | 0.8230633 | 0.8779342 |
| I-31. Victim has no social/family support | 0.9203443 | 0.9656072 |
| I-41. There have been reports of mutual violence | 0.9203443 | 0.9500329 |
| I-3. Physical violence | 1.0000000 | 1.0158730 |
| I-29. Victim suffers from a serious illness | 1.0000000 | 1.0000000 |

7. Produce a concise listing of the predictors and the various test results indicating which of them are significant according to the various testing procedures. Also show which are not significant according to any test. Do the most- or least-significant predictors follow any pattern? What about those in between, i.e., those which are significant according to some procedures but not others?

Answer: Yes, most- or least- significant predictors follow a pattern. Specifically, the predictors that are most significant in one test also tend to be significant in other tests. Accordingly, the least significant predictors in one test tend to be insignificant in other testing procedures. VPER survey has high number of most significant predictors and the VPR survey has high number of least significant predictors. Those predictors that are in between mainly come from the VPR survey.

```
p_df %>%
  mutate_at(vars(p_value:p_bh), ~ifelse(. < 0.1, 1, 0)) %>%
  select(indicators, p_value:p_bh) %>%
  arrange(p_value) %>%
  kable(
    caption = "Significance Under Various Testing Procedures",
    col.names = c("Indicators", "raw", "bf (FWER)", "bh (FDR)")
  )
```

Table 7: Significance Under Various Testing Procedures

| Indicators | raw | bf (FWER) | bh (FDR) |
|--------------------------------------------------------------------|-----|-----------|----------|
| I-1. Presence of any type of violence carried out by the aggressor | 0 | 0 | 0 |
| I-3. Physical violence | 0 | 0 | 0 |
| Physical violence with injuries | 0 | 0 | 0 |
| Serious physical violence | 0 | 0 | 0 |
| I-4. Sexual violence | 0 | 0 | 0 |
| Sexual violence with injuries | 0 | 0 | 0 |
| Serious sexual violence | 0 | 0 | 0 |
| I-5. Victim's defensive reaction to physical aggression | 0 | 0 | 0 |
| Use of a firearm | 0 | 0 | 0 |
| Use of objects | 0 | 0 | 0 |
| I-7. The aggressor has access to firearms | 0 | 0 | 0 |
| I-8. The aggressor is an expert in combat techniques | 0 | 0 | 0 |
| I-9. Presence of threats/plans aimed to cause the victim harm | 0 | 0 | 0 |
| Economic threats from the aggressor | 0 | 0 | 0 |

| Indicators | raw | bf (FWER) | bh (FDR) |
|---------------------------------------------------------------------------------------------------------------|-----|--------------|-------------|
| Threats to social reputation | 0 | 0 | 0 |
| Threats to children's integrity/custody | 0 | 0 | 0 |
| I-23. Presence of a mental/psychiatric disorder | 0 | 0 | 0 |
| I-27. Presence of any type of disability | 0 | 0 | 0 |
| I-28. Victim is in gestation period | 0 | 0 | 0 |
| I-29. Victim suffers from a serious illness | 0 | 0 | 0 |
| I-31. Victim has no social/family support | 0 | 0 | 0 |
| I-32. Victim has a mental/psychiatric disorder | 0 | 0 | 0 |
| I-33. Victim has suicidal ideas/attempts | 0 | 0 | 0 |
| I-34. Victim has any kind of addiction or engages in substance abuse (alcohol and drugs) | 0 | 0 | 0 |
| I-35. There is a history of gender or domestic violence within their family | 0 | 0 | 0 |
| I-36. Victim is economically dependent on the aggressor | 0 | 0 | 0 |
| I-37. The victim has minors or family members under their care | 0 | 0 | 0 |
| I-38. The victim has withdrawn complaints in the past | 0 | 0 | 0 |
| I-41. There have been reports of mutual violence | 0 | 0 | 0 |
| I-2. Psychological Violence | 1 | 0 | 0 |
| Serious psychological violence | 1 | 0 | 1 |
| I-6. Use of weapons/objects against partner | 1 | 0 | 1 |
| Use of a cold weapon | 1 | 1 | 1 |
| Serious threats from the aggressor | 1 | 1 | 1 |
| Suicide threats from the aggressor | 1 | 1 | 1 |
| Death threats from the aggressor | 1 | 0 | 0 |
| I-10. An escalation in aggression or threats in the last 6 months | 1 | 0 | 1 |
| I-11. They have shown exaggerated jealousy or have suspected their partner of infidelity in the last 6 months | 1 | 1 | 1 |
| I-12. They have shown controlling behaviors in the last 6 months | 1 | 1 | 1 |
| Physically controlling behaviors | 1 | 1 | 1 |
| Psychologically controlling behaviors | 1 | 1 | 1 |
| Controlling behaviors towards academic/labor aspects | 1 | 0 | 1 |
| Economically controlling behaviors | 1 | 0 | 1 |
| Cybernetically controlling behaviors | 1 | 1 | 1 |
| I-13. In the last 6 months they have shown harassing behaviors | 1 | 1 | 1 |
| I-14. In the last year they have done material damage | 1 | 0 | 1 |
| I-15. In the last year they have disrespected an authority figure | 1 | 1 | 1 |
| I-16. In the last year they have committed an aggression against third parties and/or animals. | 1 | 0 | 1 |
| I-17. In the last year they have made threats and slights towards third parties. | 1 | 0 | 1 |
| I-18. Presence of problems in their life in the last 6 months. | 1 | 1 | 1 |
| Economic or work-related problems | 1 | 0 | 1 |
| Problems with the justice system | 1 | 1 | 1 |
| I-19. Presence of a criminal record | 1 | 1 | 1 |
| I-20. Presence of past breakings of sentence conditions | 1 | 1 | 1 |
| I-21. Presence of physical/sexual aggression records | 1 | 1 | 1 |
| I-22. Presence of gender violence records against other partners | 1 | 1 | 1 |
| I-24. Presence of suicidal ideas or attempts | 1 | 1 | 1 |
| I-25. Presence of any type of addiction or substance abuse (alcohol and drugs) | 1 | 1 | 1 |
| I-26. Presence of a gender or domestic violence history within their family | 1 | 1 | 1 |
| I-30. Victim is a foreigner | 1 | 0 | 1 |
| I-39. The victim has suffered gender violence by the hands of other aggressors in the past | 1 | 0 | 1 |

| Indicators | raw | bf (FWER) | bh (FDR) |
|------------------------------------------------------------------------------------------------------------------|-----|--------------|-------------|
| I-40. The victim has expressed their desire to end the relationship to the aggressor at least 6 months ago | 1 | 1 | 1 |
| I-42. The victim fears for the physical integrity of the minors or family members under their care | 1 | 0 | 0 |
| I-43. The victim thinks that the aggressor is capable of attacking them in a violent way or even of killing them | 1 | 0 | 1 |

```
# Sum of Individual Test Significance
p_df %>%
  mutate_at(vars(p_value:p_bh), ~ifelse(. < 0.1, 1, 0)) %>%
  select(p_value:p_bh) %>%
  colSums()
```

```
## p_value    p_bf    p_bh
##      35      20      32
```

8. Considering the objectives of the research study, which do you think is the more appropriate approach to multiple testing, FWER control or FDR control? Explain.

Answer: I believe that FDR control is more appropriate for the study. In this reserach study, FWER control is too conservative. Although FDR allows TYPE 1 error, it increases power. For instance, we lose a number of predictive indicators when using the FWER control Therefore, I think FDR is more appropriate for the study.

9. It is a complete coincidence that the authors' unadjusted testing procedure and the Benjamini-Hochberg procedure produce similar numbers of significant predictors. Can you explain why this is the case? Hint: think about the graph you produced in question 5 and the listing you produced in question 7.

Answer: The authors' uadjusted testing procedure and the Benjamin-Hochberg procedures happen to produce similar number of significant predictors because the threshold the authors use do not differ significantly from the adjusted threshold we obtained from the BH procedure. Because the difference is very small, the authors' unadjusted testing procedure and the BH procedure produce similar number of significant predictors.