MATH 2310 - Probability

Jakob Balkovec

2024-05-03

Activity One

For this lab, we will be looking at data on incomes for Washington state from the 2000 census, available in the file census.txt You will use this data to answer a number of questions regarding probabilities.

For each problem, unless otherwise specified, suppose we randomly select one individual between the ages of 21 and 64 living in Washington state in 2000.

```
# Define commonly used variables/functions

total <-        3391405
total_male <-   1678835
total_female <- 1712570

pretty_print <- function(question_num, prob) {
  answer <- paste(round(prob * 100, 2), "%", sep = "")
  cat("Answer to [Q", question_num, "]: ", answer, "\n", sep = "")
}
```

[Q1] - What is the probability that that individual is older than 44?

```
# Problem 1
### __notes__
# Calculate P(A)
###
older_than_44 <- 835590 + 490185

prob_of_older_than_44 <- older_than_44/total

pretty_print(question_num = 1, prob = prob_of_older_than_44)
```

```
## Answer to [Q1]: 39.09%
```

[Q2] - What is the probability that the individual has completed at least a Bachelor's degree (that is, a Bachelor's degree or an advanced degree)?

```
# Problem 2
### __notes__
# Calculate P(A)
###
bachelor_or_advanced <- 652845 + 307405

prob_of_bachelors_or_advanced_degree <- bachelor_or_advanced/total

pretty_print(question_num = 2, prob = prob_of_bachelors_or_advanced_degree)
```

```
## Answer to [Q2]: 28.31%
```

[Q3] - What is the probability that the individual has completed an advanced degree and worked full-time year-round in 1999?

```
# Problem 3
### __notes__
# Events are independent, since an individual doesn't need an advanced degree to work full-time year-round
#
# P(A) - Individual has completed an advanced degree
# P(B) - Individual worked full-time year-round in 1999
#
# P(A && B) = P(A) x P(B)
###

completed_advanced_degree <- 307405
P_of_A <- completed_advanced_degree/total

worked_full_time <- 1735245
P_of_B <- worked_full_time/total

P_of_A_and_B <- P_of_A * P_of_B
pretty_print(question_num = 3, prob = P_of_A_and_B)
```

```
## Answer to [Q3]: 4.64%
```

[Q4] - What is the probability that the individual has completed an advanced degree or worked full-time year-round in 1999?

```
# Problem 4
### __notes__
#! Assuming events are independent, since an individual doesn't need an advanced degree to work full-time year-ro
und
#
# P(A) - Individual has completed an advanced degree
# P(B) - Individual worked full-time year-round in 1999
#
# P(A || B) = P(A) + P(B)
### DIFF

completed_advanced_degree <- 307405
P_of_A <- completed_advanced_degree/total

worked_full_time <- 1735245
P_of_B <- worked_full_time/total

P_of_A_or_B <- P_of_A + P_of_B
pretty_print(question_num = 4, prob = P_of_A_or_B)
```

```
## Answer to [Q4]: 60.23%
```

[Q5] - What is the probability that the individual has completed at least high school and worked full-time year-round in 1999?

```
# Problem 5
### __notes__
#! Assuming events are independent, since an individual doesn't need to graduate high school to work full-time ye
ar-round
#
# P(A) - Individual has completed at least high school
# P(B) - Individual worked full-time year-round in 1999
#
# P(A && B) = P(A) * P(B)
###

completed_highschool <- total - 278530
P_of_A <- completed_highschool/total

worked_full_time <- 1735245
P_of_B <- worked_full_time/total

P_of_A_and_B <- P_of_A * P_of_B
pretty_print(question_num = 5, prob = P_of_A_and_B)
```

```
## Answer to [Q5]: 46.96%
```

[Q6] - Suppose that we know that an individual did not work full-time year-round in 1999. What is the probability that they did not graduate high school?

```
# Problem 6
### __notes__
#
# P(A|B) = P(A && B)/P(B)
#
# Event A: Individual did not graduate high school
# Event B: Individual did not work full-time year-round in 1999
###

not_a_highschool_grad <- 278530 - 81375
P_of_A <- not_a_highschool_grad/total

did_not_work_full_time <- total - 1735245
P_of_B <- did_not_work_full_time/total

P_of_A_and_B <- P_of_A * P_of_B
P_of_A_given_B <- P_of_A_and_B/P_of_B
pretty_print(question_num = 6, prob = P_of_A_given_B)
```

```
## Answer to [Q6]: 5.81%
```

[Q7] - Does it appear that biological sex is independent of completing high school? Explain.

```
# Problem 7
### __notes__
# Use the definition of independent events
###

not_a_high_school_grad_male <- 144875
not_a_high_school_grad_female <- 133660

high_school_grad_male <- total_male - not_a_high_school_grad_male
high_school_grad_female <- total_female - not_a_high_school_grad_female

# Proportions
prop_high_school_male <- high_school_grad_male / total_male
prop_high_school_female <- high_school_grad_female / total_female

pct_male_str <- paste(round(prop_high_school_male * 100, 2), "%", sep = "")
pct_female_str <- paste(round(prop_high_school_female * 100, 2), "%", sep = "")

P_of_A_and_B <- prop_high_school_female * prop_high_school_male

cat("[ Male: ", pct_male_str, "] | [ Female: ", pct_female_str, "]")
```

```
## [ Male:  91.37% ] | [ Female:  92.2% ]
```

[Answer]:

- Approximately **91.37%** of males and **92.20%** of females have finished high school. This indicates a slightly higher completion rate among females. On the flip side, these proportions are not significantly different from the overall completion rates of males and females, it can be inferred that completing high school is not significantly influenced by biological sex. Therefore, these events can be considered independent

[Q8] - What percentage of people aged 25 to 44 years are female?

```
# Problem 8
### __notes__
###

female_25_to_44 <- 406110 + 488830
total_25_to_44 <- 809400 + 966115

percentage_female <- (female_25_to_44 / total_25_to_44)

pretty_print(question_num = 8, prob = percentage_female)
```

```
## Answer to [Q8]: 50.4%
```