lead_pipes_analysis_markdown

Ryan Heslin

10/3/2020

1

1a. Alt: People who work >5 hours a day on computers have worse eyesight than those who do not Null: Computer workers do not have worse eyesight than non-computer workers 1b. Alt. Incubated babies have higher rates of claustrophobia than nonincubated babies Null: Incubated babies have about the same rates of claustrphobia as nonincubated. 1c. Offices with plants have fewer sick days than those without plants. Null: Offices without plants have about the same sick days as those with plants

2

Type I: Reject null hypothesis despite alternative hypothesis being false Type II: Fail to reject false null hypothesis

Here I write code to manually compute the statistic. The result is the same as R's chi-squared function.

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.1     v dplyr   1.0.0
## v tidyr   1.1.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------------------------------------------------ tidy
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
mortgage = tibble(race = c("White", "Black"), Approved = c(720, 30), `Not Approved` = c(130, 10), total
mortgage
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 2 x 4
##   race  Approved `Not Approved` total
##   <chr>    <dbl>          <dbl> <dbl>
## 1 White      720            130   850
## 2 Black       30             10    40
```

```
expected = tibble(race = c("White", "Black"), Approved = (sum(mortgage$Approved) * mortgage$total) /sum

expected
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?

## # A tibble: 2 x 3
##   race  Approved `Not Approved`
##   <chr>    <dbl>          <dbl>
## 1 White    716.           134.
## 2 Black     33.7            6.29
```

```
chisq <- mortgage %>% select(c(Approved, `Not Approved`)) %>%
  imodify(~{((.x - expected[[.y]])^2) /expected[[.y]]}) %>% sum()

mortgage %>% select(-c(race, total)) %>% chisq.test(correct = FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  .
## X-squared = 2.7149, df = 1, p-value = 0.09942
```

5

To conduct a dependent samples t-test, we divide the mean of differences by the sample standard deviation c. The t-statistic for p = .05 for n = 25 is about 1.71. Since this is a 1-sided test, that is the cutoff value, and the result is not signnifcant. However, the sample size is less than 30, and we don't know if the population distriubtion is normal, so even using the t distribution here is questionable. The study should have used a much larger sample size.

In any case, even if the result is signficant, we have no way of knowing if a 10-point improvement is substantial.

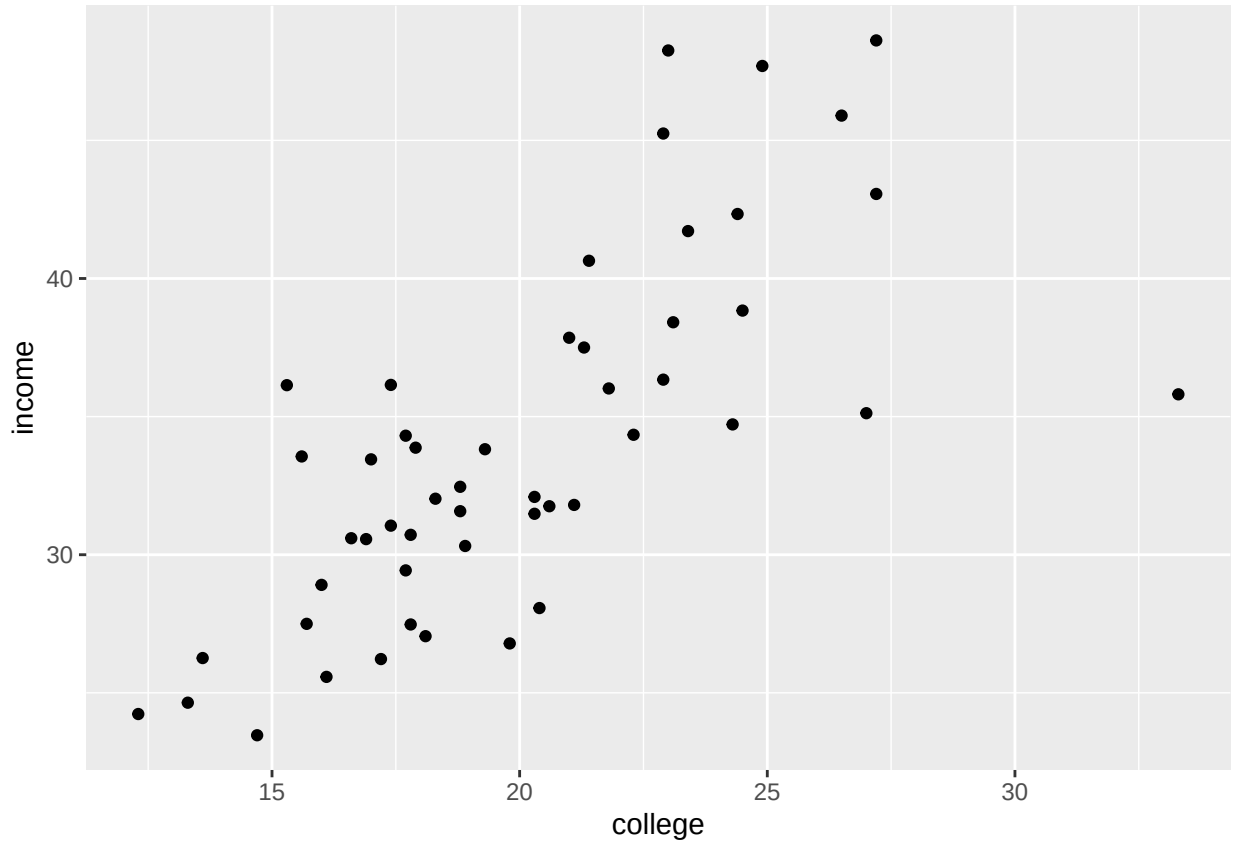
```
s <- 4 * sqrt(25)
test_stat <- (10 -0)/4
t <- qt(.95, 24, lower.tail = T)
confint <- c(10 - t*4, 10 + t*4)
```

##6

A rival hypothesis might include populationd density as a predictor, since cities are more expensive than less dense areas. Sure enough, it improves the model R squared.

```
library(foreign)
states <- foreign::read.dta("C:/Users/heslinr1/Documents/American University/Classes/Fall 2020/PUAD 601/
```

```
states %>% ggplot(aes(college, income)) +
  geom_point()
```




```
cor(states$college, states$income)
```

```
## [1] 0.723354
```

```
mod <- lm(data = states, income ~ college)
mod %>% confint()
```

```
##                 2.5 %    97.5 %
## (Intercept) 5.3781635 17.873857
## college     0.8096879  1.420962
```

```
mod %>% broom::tidy()
```

```
## Warning: '...' is not empty.
##
## We detected these problematic arguments:
## * 'needs_dots'
```

```
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic       p.value
##   <chr>          <dbl>     <dbl>     <dbl>         <dbl>
## 1 (Intercept)    11.6      3.11       3.74 0.000483
## 2 college         1.12     0.152      7.33 0.00000000202
```

```
mod %>% broom::glance()
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 1 x 12
##   r.squared adj.r.squared sigma statistic p.value    df logLik   AIC   BIC
##       <dbl>         <dbl> <dbl>     <dbl>   <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1     0.523         0.514  4.48      53.8 2.02e-9     1  -148.  302.  307.
## # ... with 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```

```
lm(data = states, income ~ college + density) %>% broom::glance()
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 1 x 12
##   r.squared adj.r.squared sigma statistic  p.value    df logLik   AIC   BIC
##       <dbl>         <dbl> <dbl>     <dbl>    <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1     0.695         0.682  3.66      53.5 7.62e-13     2  -134.  276.  284.
## # ... with 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```