Assignment 5-Problem 3

Laura Ingulsrud

11/28/2018

1. Male and female salaries for starting postdoctoral and other employment positions (2015). Compare median salaries for male and female doctorate recipients in 2015. Answer these two questions: Does median salary differ significantly between male and female starting postdoc positions? Does median salary differ significantly between male and female PhD recipients in non-postdoc employment positions?

library(tidyverse)

## ── Attaching packages ────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 3.0.0 ✔ purrr 0.2.5  
## ✔ tibble 1.4.2 ✔ dplyr 0.7.6  
## ✔ tidyr 0.8.1 ✔ stringr 1.3.1  
## ✔ readr 1.1.1 ✔ forcats 0.3.0

## ── Conflicts ───────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(knitr)  
library(kableExtra)  
library(plotly)

##   
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

library(ggrepel)  
library(extrafont)

## Registering fonts with R

library(pwr)  
library(effsize)  
library(vcdExtra)

## Loading required package: vcd

## Loading required package: grid

## Loading required package: gnm

##   
## Attaching package: 'vcdExtra'

## The following object is masked from 'package:plotly':  
##   
## summarise

## The following object is masked from 'package:dplyr':  
##   
## summarise

library(grid)  
  
  
  
#load packages needed

med\_sal\_2015 <- read\_csv("Med\_Sal\_2015.csv")

## Parsed with column specification:  
## cols(  
## Field = col\_character(),  
## Status = col\_character(),  
## Male = col\_number(),  
## Female = col\_number()  
## )

#read in data (data wrangled directly in excel)  
  
med\_sal\_2015

## # A tibble: 30 x 4  
## Field Status Male Female  
## <chr> <chr> <dbl> <dbl>  
## 1 Ag\_sci\_nat\_resources Post\_Doc 42750 44000  
## 2 Ag\_sci\_nat\_resources Non\_Post\_Doc 78000 66000  
## 3 Biological and biomedical sciences Post\_Doc 42000 42000  
## 4 Biological and biomedical sciences Non\_Post\_Doc 75000 66000  
## 5 Health sciences Post\_Doc 43000 43250  
## 6 Health sciences Non\_Post\_Doc 75000 75000  
## 7 Chemistry Post\_Doc 42000 42000  
## 8 Chemistry Non\_Post\_Doc 80000 75000  
## 9 Geosciences, atmospheric, and ocean sciences Post\_Doc 50000 50000  
## 10 Geosciences, atmospheric, and ocean sciences Non\_Post\_Doc 75167 71750  
## # ... with 20 more rows

med\_sal\_15\_Employ <- med\_sal\_2015 %>%   
 select(Field, Male, Female, Status) %>%   
 filter(Status!= "Post\_Doc")  
med\_sal\_15\_Employ

## # A tibble: 15 x 4  
## Field Male Female Status   
## <chr> <dbl> <dbl> <chr>   
## 1 Ag\_sci\_nat\_resources 78000 66000 Non\_Post\_Doc  
## 2 Biological and biomedical sciences 75000 66000 Non\_Post\_Doc  
## 3 Health sciences 75000 75000 Non\_Post\_Doc  
## 4 Chemistry 80000 75000 Non\_Post\_Doc  
## 5 Geosciences, atmospheric, and ocean sciences 75167 71750 Non\_Post\_Doc  
## 6 Physics and astronomy 95000 97650 Non\_Post\_Doc  
## 7 Mathematics and computer sciences 105000 90000 Non\_Post\_Doc  
## 8 Psychology 63000 60000 Non\_Post\_Doc  
## 9 Economics 105000 95750 Non\_Post\_Doc  
## 10 Social sciences 64000 62000 Non\_Post\_Doc  
## 11 Engineering 95000 90000 Non\_Post\_Doc  
## 12 Education 71000 63000 Non\_Post\_Doc  
## 13 Humanities and arts 52000 50000 Non\_Post\_Doc  
## 14 Business management and administration 123500 120000 Non\_Post\_Doc  
## 15 Other non-S&E fields 62800 61000 Non\_Post\_Doc

#made dataset for just Non\_Post\_Doc values  
  
med\_sal\_15\_PostDoc <- med\_sal\_2015 %>%   
 select(Field, Male, Female, Status) %>%   
 filter(Status!= "Non\_Post\_Doc")  
  
#made dataset for just Post\_Doc values  
  
med\_sal\_15\_PostDoc

## # A tibble: 15 x 4  
## Field Male Female Status   
## <chr> <dbl> <dbl> <chr>   
## 1 Ag\_sci\_nat\_resources 42750 44000 Post\_Doc  
## 2 Biological and biomedical sciences 42000 42000 Post\_Doc  
## 3 Health sciences 43000 43250 Post\_Doc  
## 4 Chemistry 42000 42000 Post\_Doc  
## 5 Geosciences, atmospheric, and ocean sciences 50000 50000 Post\_Doc  
## 6 Physics and astronomy 50000 53000 Post\_Doc  
## 7 Mathematics and computer sciences 58000 55000 Post\_Doc  
## 8 Psychology 42000 42000 Post\_Doc  
## 9 Economics 65000 65000 Post\_Doc  
## 10 Social sciences 48000 49250 Post\_Doc  
## 11 Engineering 45000 45000 Post\_Doc  
## 12 Education 50000 45000 Post\_Doc  
## 13 Humanities and arts 45000 45000 Post\_Doc  
## 14 Business management and administration 60000 63500 Post\_Doc  
## 15 Other non-S&E fields 50000 44000 Post\_Doc

wsr\_Employ\_Gender <- wilcox.test(med\_sal\_15\_Employ$Male, med\_sal\_15\_Employ$Female, paired = TRUE)

## Warning in wilcox.test.default(med\_sal\_15\_Employ$Male,  
## med\_sal\_15\_Employ$Female, : cannot compute exact p-value with ties

## Warning in wilcox.test.default(med\_sal\_15\_Employ$Male,  
## med\_sal\_15\_Employ$Female, : cannot compute exact p-value with zeroes

#Run Wilcox Signed Rank for Non Post Doc dataset (significant value found)  
  
wsr\_Employ\_Gender

##   
## Wilcoxon signed rank test with continuity correction  
##   
## data: med\_sal\_15\_Employ$Male and med\_sal\_15\_Employ$Female  
## V = 101, p-value = 0.002572  
## alternative hypothesis: true location shift is not equal to 0

wsr\_Post\_Doc\_Gender <- wilcox.test(med\_sal\_15\_PostDoc$Male, med\_sal\_15\_PostDoc$Female, paired = TRUE)

## Warning in wilcox.test.default(med\_sal\_15\_PostDoc$Male,  
## med\_sal\_15\_PostDoc$Female, : cannot compute exact p-value with ties

## Warning in wilcox.test.default(med\_sal\_15\_PostDoc$Male,  
## med\_sal\_15\_PostDoc$Female, : cannot compute exact p-value with zeroes

#Run Wilcox Signed Rank for Post Doc dataset (p-value is non-significant)  
  
wsr\_Post\_Doc\_Gender

##   
## Wilcoxon signed rank test with continuity correction  
##   
## data: med\_sal\_15\_PostDoc$Male and med\_sal\_15\_PostDoc$Female  
## V = 19.5, p-value = 0.8884  
## alternative hypothesis: true location shift is not equal to 0

med\_sal\_bar <- read\_csv("Med\_Sal\_2015\_bar chart.csv")

## Parsed with column specification:  
## cols(  
## Field = col\_character(),  
## Status = col\_character(),  
## Gender = col\_character(),  
## Salary = col\_number()  
## )

med\_sal\_bar\_15\_Employ <- med\_sal\_bar %>%   
 select(Field, Gender, Salary, Status) %>%   
 filter(Status!= "Postdoc Employment")  
  
med\_sal\_bar\_15\_PostDoc <- med\_sal\_bar %>%   
 select(Field, Gender, Salary, Status) %>%   
 filter(Status!= "Non-Postdoc Employment")  
  
Med\_sal\_barchart <- ggplot(med\_sal\_bar, aes(Field, Salary)) +  
 geom\_bar(aes(fill = Gender), width = 0.5, position = position\_dodge(width = 0.6), stat = "identity") +  
 theme(legend.position = "right", legend.title = element\_blank()) +  
 scale\_x\_discrete(breaks = c("Agricultural sciences and natural resources", "Biological and biomedical sciences", "Health sciences", "Chemistry", "Geosciences, atmospheric, and ocean sciences", "Physics and astronomy", "Mathematics and computer sciences", "Psychology", "Economics", "Social sciences", "Engineering", "Education", "Humanities and arts", "Business management and administration", "Other (Non-Science and Engineering)" )) +  
 coord\_flip() +  
 facet\_wrap(~Status) +  
 scale\_fill\_manual(values = c("peachpuff2", "darkslateblue")) +  
 ylab("Salary (dollars)") +  
 xlab("Field of Employment") +  
 scale\_y\_continuous(expand = c(0,0), limits = c(0,130000), breaks = c(0, 40000, 80000, 130000)) +  
 theme(axis.text.x = element\_text(size = 7, angle = 45, hjust = 1), axis.text.y = element\_text(size = 7), axis.title.x = element\_text(size = 8, face = "bold"), axis.title.y = element\_text(size = 8, face = "bold")) +  
 theme(panel.spacing = unit(1.75, "lines"), panel.grid.major = element\_blank(), panel.grid.minor = element\_blank(), panel.background = element\_blank(), axis.line = element\_line(colour = "black"), strip.background = element\_blank(), strip.text.x = element\_text(face = "bold", size = 7.5)) +  
 guides(fill = guide\_legend(reverse = TRUE)) +  
 theme(legend.text = element\_text (size = 7))  
   
  
   
   
  
  
  
  
Med\_sal\_barchart

