Econometircs Lab02

## Working Directory

setwd("C:/Users/22700/Desktop/PhD Econometrics/26.10.2020")

## Packages

library(ggplot2)  
library(stargazer)

##   
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

library(data.table)

## Data Analysis

load("lecture02/Lab02/ceosal2(1).RData")  
dt.ceo.salaries <- data.table(data)  
rm(data)

## Descriptive Statistics

## How many CEOS are in the sample?

nrow(dt.ceo.salaries)

## [1] 177

## How many CEOS have a graduate degree?

dt.ceo.salaries[, sum(grad)]

## [1] 94

## Alternatively, we can count the number of rows in which the variable grad takes the value of 1.

nrow(dt.ceo.salaries[grad==1,])

## [1] 94

## What is the percentage of CEOs with graduate degrees?

dt.ceo.salaries[, sum(grad)]/nrow(dt.ceo.salaries)

## [1] 0.5310734

## Or we can simply calculate the mean of grad:

dt.ceo.salaries[, mean(grad)]

## [1] 0.5310734

## What is the average CEO salary?

dt.ceo.salaries[, mean(salary)]

## [1] 865.8644

## Alternatively

mean(dt.ceo.salaries[, salary])

## [1] 865.8644

## What is the mean CEO salary for those with a graduate degree?

dt.ceo.salaries[grad==1, mean(salary)]

## [1] 864.2128

## What is the mean CEO salary for those without a graduate degree?

dt.ceo.salaries[grad==0, mean(salary)]

## [1] 867.7349

## How many CEOs are have/don’t have a college degree?

dt.ceo.salaries[ , list(n\_ceo=.N), by = college]

## college n\_ceo  
## 1: 1 172  
## 2: 0 5

## Can we say that the mean salary is statistically different from 800?

t.test(dt.ceo.salaries[, salary], mu = 800)

##   
## One Sample t-test  
##   
## data: dt.ceo.salaries[, salary]  
## t = 1.4913, df = 176, p-value = 0.1377  
## alternative hypothesis: true mean is not equal to 800  
## 95 percent confidence interval:  
## 778.7015 953.0274  
## sample estimates:  
## mean of x   
## 865.8644

## Is the average salary different for CEOs with a graduate degree and those without?

t.test(dt.ceo.salaries[, salary] ~ dt.ceo.salaries[, grad])

##   
## Welch Two Sample t-test  
##   
## data: dt.ceo.salaries[, salary] by dt.ceo.salaries[, grad]  
## t = 0.038973, df = 149.94, p-value = 0.969  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -175.0489 182.0932  
## sample estimates:  
## mean in group 0 mean in group 1   
## 867.7349 864.2128

## Alternatively

dt.ceo.salaries[ , t.test (salary ~ grad)]

##   
## Welch Two Sample t-test  
##   
## data: salary by grad  
## t = 0.038973, df = 149.94, p-value = 0.969  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -175.0489 182.0932  
## sample estimates:  
## mean in group 0 mean in group 1   
## 867.7349 864.2128

## Alternatively:

t.test(dt.ceo.salaries[grad==0, salary] , dt.ceo.salaries[grad==1, salary])

##   
## Welch Two Sample t-test  
##   
## data: dt.ceo.salaries[grad == 0, salary] and dt.ceo.salaries[grad == 1, salary]  
## t = 0.038973, df = 149.94, p-value = 0.969  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -175.0489 182.0932  
## sample estimates:  
## mean of x mean of y   
## 867.7349 864.2128

## Creating a table with descriptive statistics:

dt.ceo.salaries[, list( mean\_salary = mean(salary)  
, sd\_salary = sd(salary)  
, min\_salary = min(salary)  
, max\_salary = max(salary)  
, median\_salary = median(salary))]

## mean\_salary sd\_salary min\_salary max\_salary median\_salary  
## 1: 865.8644 587.5893 100 5299 707

## You can also compute the summary statistics for different groups:

dt.ceo.salaries[, list( mean\_salary = mean(salary)  
, sd\_salary = sd(salary)  
, min\_salary = min(salary)  
, max\_salary = max(salary)), by = list(grad, college)]

## grad college mean\_salary sd\_salary min\_salary max\_salary  
## 1: 1 1 864.2128 501.3924 100 2265  
## 2: 0 1 853.0897 679.0268 174 5299  
## 3: 0 0 1096.2000 633.4569 300 1738

## Alternatively, you can use the stargazer function to get summary statistics for all variables in your data set:

stargazer(dt.ceo.salaries, type = "text")

##   
## ===================================================================  
## Statistic N Mean St. Dev. Min Pctl(25) Pctl(75) Max   
## -------------------------------------------------------------------  
## salary 177 865.864 587.589 100 471 1,119 5,299   
## age 177 56.429 8.422 33 52 62 86   
## college 177 0.972 0.166 0 1 1 1   
## grad 177 0.531 0.500 0 0 1 1   
## comten 177 22.503 12.295 2 12 33 58   
## ceoten 177 7.955 7.151 0 3 11 37   
## sales 177 3,529.463 6,088.654 29 561 3,500 51,300  
## profits 177 207.831 404.454 -463 34 208 2,700   
## mktval 177 3,600.316 6,442.276 387 644 3,500 45,400  
## lsalary 177 6.583 0.606 4.605 6.155 7.020 8.575   
## lsales 177 7.231 1.432 3.367 6.330 8.161 10.845  
## lmktval 177 7.399 1.133 5.958 6.468 8.161 10.723  
## comtensq 177 656.684 577.123 4 144 1,089 3,364   
## ceotensq 177 114.124 212.566 0 9 121 1,369   
## profmarg 177 6.420 17.861 -203.077 4.231 10.947 47.458  
## -------------------------------------------------------------------

## Can also get summary statistics for a subset of your observations and for a specific list of variables.

stargazer(dt.ceo.salaries[grad==1, list(age, salary)], type = "text")

##   
## =========================================================  
## Statistic N Mean St. Dev. Min Pctl(25) Pctl(75) Max   
## ---------------------------------------------------------  
## age 94 55.457 8.155 38 50 61 86   
## salary 94 864.213 501.392 100 481.5 1,167.8 2,265  
## ---------------------------------------------------------

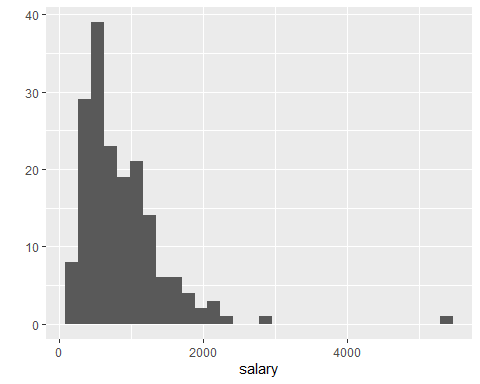
## Quick plots

## Histogram

## Salary

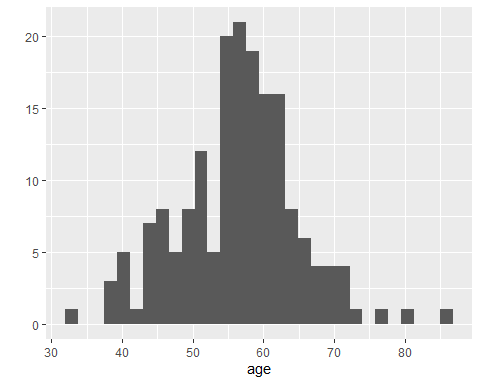
qplot( data = dt.ceo.salaries  
, x = salary  
, geom = "histogram")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

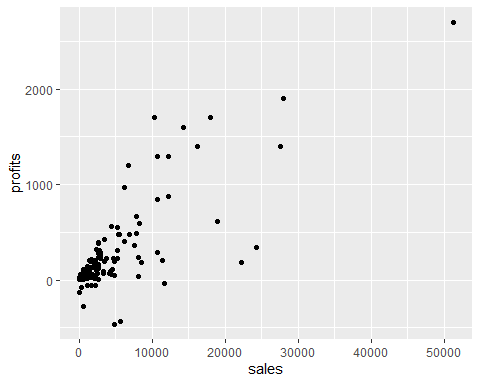
 ## Age

qplot( data = dt.ceo.salaries  
, x = age  
, geom = "histogram")

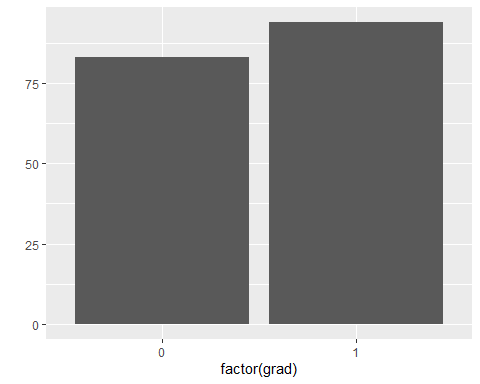
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

 ## Scatterplot

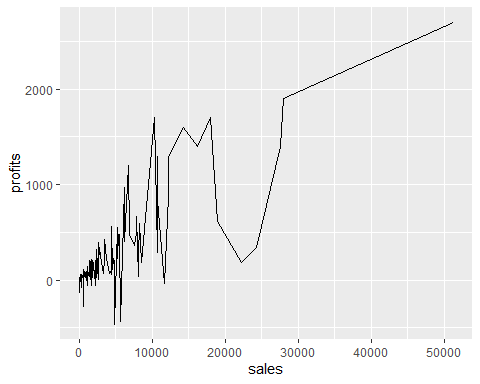
qplot( data = dt.ceo.salaries  
, x = sales  
, y = profits  
, geom = "point")

 ## Barplot ## Graduate

qplot( data = dt.ceo.salaries  
, x = factor(grad)  
, geom = "bar")

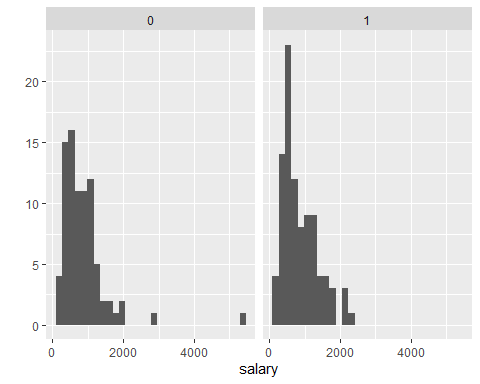
 ## Line

qplot( data = dt.ceo.salaries  
, x = sales  
, y = profits  
, geom = "line")

 ## Facet Wrap

qplot( data = dt.ceo.salaries  
, x = salary  
, geom = "histogram") + facet\_wrap(~ grad)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

 ## Customizing plots: ## Salary

qplot( data = dt.ceo.salaries  
, x = salary  
, geom = "histogram"  
, fill = factor(grad, levels = c(0,1), labels = c("Yes", "No"))) +  
theme\_bw() +  
ylim(0,50) +  
xlim(0, 4000) +  
labs( title = "MY PLOT", x = "CEO Salary", y = "Number of CEOs", fill = "Grad. Degree")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 1 rows containing non-finite values (stat\_bin).

## Warning: Removed 4 rows containing missing values (geom\_bar).

