stats-r

September 29, 2023

2) The heights of adult women follow an approximately normal distribution with mean 63.8inches and SD 2.9 inches

(a) What percentage of adult women are taller than 65.5 inches?

Answer : Given, mu = 63.8 and sigma = 2.9 inches. We know that with pnorm function, we can find the percentage of women who are atleast 65.5 inches. Hence, 1 - p(atleast 65.5 inches) will give us the answer.

[ ]:
```
#percentage of women taller than 63.5
1-pnorm(65.5,63.8,2.9)
```

0.278868243866999

27.88 % of women are taller than 65.5 inches

---

2) b: What is the interquartile range of adult women's heights?

[35]:
```
# IQR = q3-q1

# from the lecture, we know that IQR for a Normal random variable is 1.35 times␣
 ↪the Standard Deviation.

iqr = 1.35 * 2.9
cat("IQR from Formula from lecture", iqr, "\n")

# alternatively :

q1=qnorm(0.25,63.8,2.9)
q3 = qnorm(0.75,63.8,2.9)
cat("IQR from standard formula with qnorm",q3-q1, "\n")
```

```
IQR from Formula from lecture 3.915
IQR from standard formula with qnorm 3.912041
```

---

2 (c) Complete the following, rounding to the nearest 0.1 of an inch: "The shortest 2.5% of women are shorter than —, while the tallest 2.5% of women are taller than —."

[23]:
```
#Using the qnorm function, we can calculate the height that is corresponding to␣
 ↪women shorter than or taller than a particular perventage.
shorter_than_2.5Perc = qnorm(0.025,63.8,2.9)
shorter_than_2.5Perc
```

58.1161044448338

[22]:
```
taller_than_2.5Perc = qnorm(0.975,63.8,2.9)
taller_than_2.5Perc
```

69.4838955551662

Answer: "The shortest 2.5% of women are shorter than 58.116, while the tallest 2.5% of women are taller than 69.5."

---

3) (From the Fall 2014 takehome.) Let X be a standard normal random variable. Let Y = X2.

3(a) Find P(Y > 1).

[41]:
```
# to find p(y>1), let's generate a standard normal distribution from 1000000␣
 ↪random samples

n_samples = 1000000
x = rnorm(n_samples)

#lets calculate y = x^2
y = x^2

# add all the values where the value of x is >1

x_great_1 = sum(y>1)

#divide it by the number of samples generated

prob = x_great_1/n_samples
cat(prob)
```

```
0.316506
```

3 a) Answer: Therfore P(Y>1) = 0.3165

---

3 (b) Find the 0.9-quantile of Y:

Answer : from the lecture, we know that Y=X^2 is a chisquared distribution. By using thefunction "qchisq" function, we can calculate the 90th percentile.

[42]:
```
qchisq(0.9,1)
```

2.70554345409542

3b Answer : Therefore 0.9 percentile of y is 2.7

Question 5 :

The file IUSalaries2023.csv contains the salaries of IU Bloomington academic faculty for the 2023–24 academic year (this is publicly available information.) Read this data into R, e.g. by using the function read.csv(). Dr. Luen's salary is $81,902. What is the percentile rank of his salary among IUB academic faculty? Remember to include your code.

[25]:
```
#read the file
data <- read.csv("./IUSalaries2023.csv")
```

[34]:
```
salary  = data$Salary
cat("Mean of the salaries in IUB faculty is", mean(salary), "\n")
cat("Standard deviation of the salaries in IUB faculty is", sd(salary), "\n")
```

```
Mean of the salaries in IUB faculty is 112187
Standard deviation of the salaries in IUB faculty is 63172.93
```

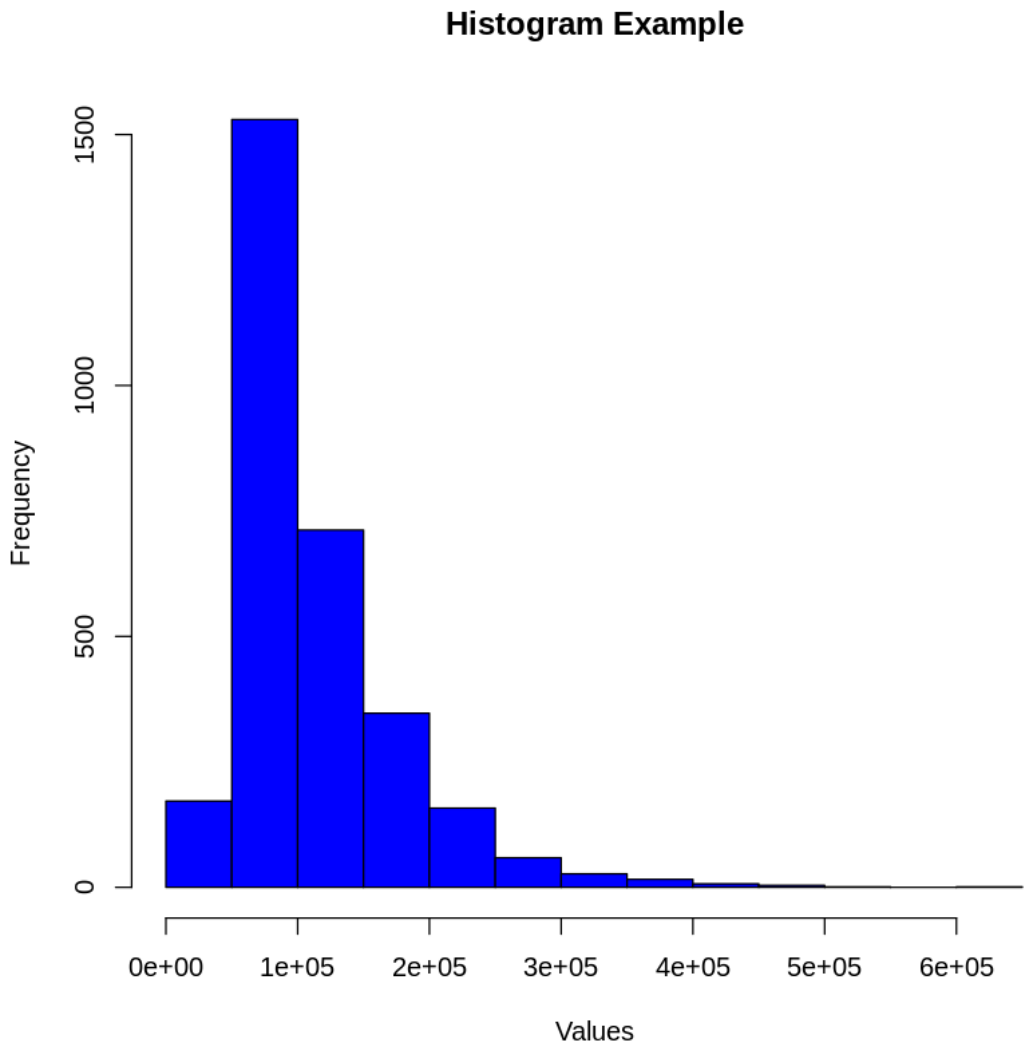
[36]:
```
# Calculate the percentile rank
percentile_rank <- (sum(data$Salary <=  81902) / length(data$Salary)) * 100 ␣
 ↪#calculate the number of salaries that are less than or equal to Professors␣
 ↪salary and divide by the total number of faculty

# Print the percentile rank
cat("Percentile Rank of Professor Brad Luen is", percentile_rank, "%\n")
```

```
Percentile Rank of Professor Brad Luen is 38.0356 %
```

5 Answer : Percentile Rank of Professor Brad Luen is 38.0356 %

[12]:
```
hist(salary,
     main = "Histogram Example",  # Main title for the plot
     xlab = "Values",             # Label for the x-axis
     ylab = "Frequency",          # Label for the y-axis
     col = "blue",                # Color of the bars
     border = "black",            # Border color of the bars
     breaks = 10)
```


Histogram Example
Frequency
0
500
1000
1500
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
6e+05
Values

[ ]: