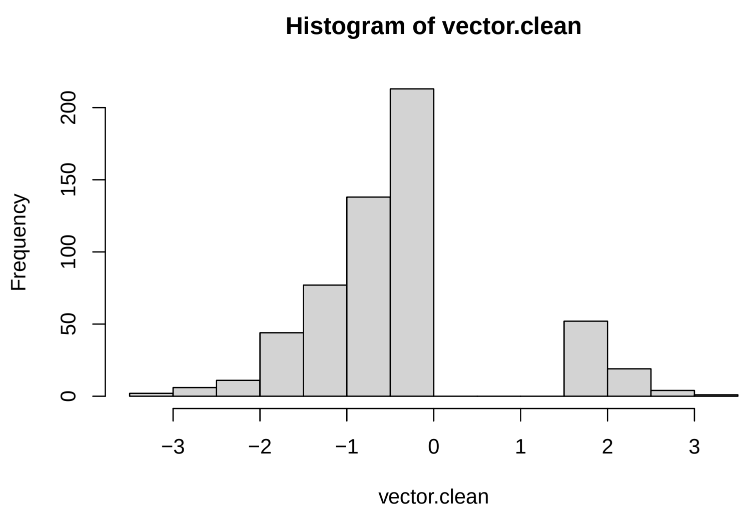
SC306 Homework 3: R Intro Part 2

Othar Zaldastani II

February 27, 2026

Question 1

```
vector.raw = rnorm(1000,mean = 0, sd = 1)
vector.clean = vector.raw[!(vector.raw >= 0 & vector.raw <= 1.5)]
hist(vector.clean)
```



Question 2

```
vector.1_100 = abs(abs(-99:99)-100)
vector.1_100
```

```
##   [1]   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
##  [19]  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36
##  [37]  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54
##  [55]  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72
##  [73]  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90
##  [91]  91  92  93  94  95  96  97  98  99 100  99  98  97  96  95  94  93  92
## [109]  91  90  89  88  87  86  85  84  83  82  81  80  79  78  77  76  75  74
## [127]  73  72  71  70  69  68  67  66  65  64  63  62  61  60  59  58  57  56
## [145]  55  54  53  52  51  50  49  48  47  46  45  44  43  42  41  40  39  38
## [163]  37  36  35  34  33  32  31  30  29  28  27  26  25  24  23  22  21  20
## [181]  19  18  17  16  15  14  13  12  11  10   9   8   7   6   5   4   3   2
## [199]   1
```

Question 3

```
library(knitr)

us_sorted <- USArrests[order(USArrests$Murder, decreasing = TRUE), ]

us_sorted
```

```
##                Murder Assault UrbanPop Rape
## Georgia          17.4     211       60 25.8
## Mississippi      16.1     259       44 17.1
## Florida          15.4     335       80 31.9
## Louisiana        15.4     249       66 22.2
## South Carolina   14.4     279       48 22.5
## Alabama          13.2     236       58 21.2
## Tennessee        13.2     188       59 26.9
## North Carolina   13.0     337       45 16.1
## Texas            12.7     201       80 25.5
## Nevada           12.2     252       81 46.0
## Michigan         12.1     255       74 35.1
## New Mexico       11.4     285       70 32.1
## Maryland         11.3     300       67 27.8
## New York         11.1     254       86 26.1
## Illinois         10.4     249       83 24.0
## Alaska           10.0     263       48 44.5
## Kentucky          9.7     109       52 16.3
## California        9.0     276       91 40.6
## Missouri          9.0     178       70 28.2
## Arkansas          8.8     190       50 19.5
## Virginia          8.5     156       63 20.7
## Arizona           8.1     294       80 31.0
## Colorado          7.9     204       78 38.7
## New Jersey        7.4     159       89 18.8
## Ohio              7.3     120       75 21.4
## Indiana           7.2     113       65 21.0
## Wyoming           6.8     161       60 15.6
## Oklahoma          6.6     151       68 20.0
## Pennsylvania      6.3     106       72 14.9
## Kansas            6.0     115       66 18.0
## Montana           6.0     109       53 16.4
## Delaware          5.9     238       72 15.8
## West Virginia     5.7      81       39  9.3
## Hawaii            5.3      46       83 20.2
## Oregon            4.9     159       67 29.3
## Massachusetts     4.4     149       85 16.3
## Nebraska          4.3     102       62 16.5
## Washington        4.0     145       73 26.2
## South Dakota      3.8      86       45 12.8
## Rhode Island      3.4     174       87  8.3
## Connecticut       3.3     110       77 11.1
## Utah              3.2     120       80 22.9
## Minnesota         2.7      72       66 14.9
## Idaho             2.6     120       54 14.2
## Wisconsin         2.6      53       66 10.8
## Iowa              2.2      56       57 11.3
## Vermont           2.2      48       32 11.2
## Maine             2.1      83       51  7.8
## New Hampshire     2.1      57       56  9.5
## North Dakota      0.8      45       44  7.3
```

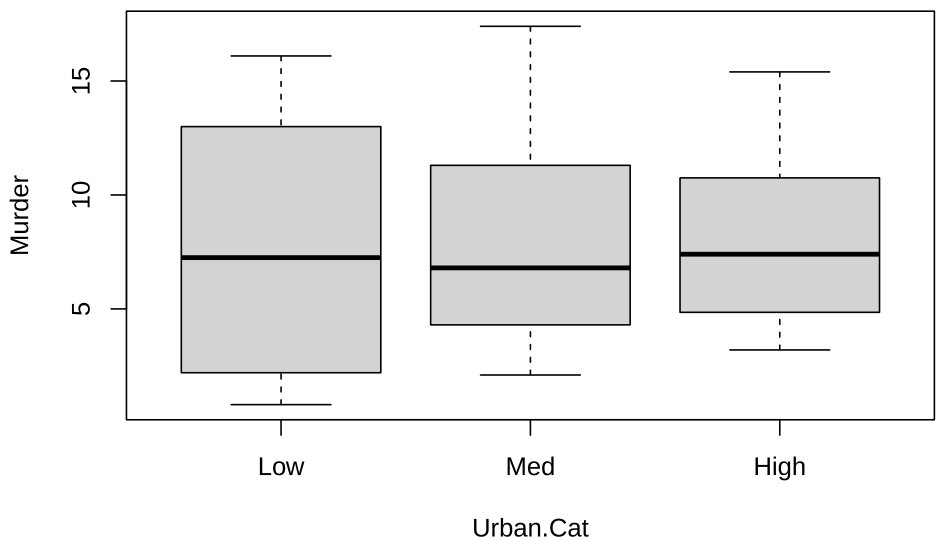
Question 4

```r
USArrests$Urban.Cat = cut(USArrests$UrbanPop,3, labels = c("Low", "Med", "High"))
USArrests
```

```
##                Murder Assault UrbanPop Rape Urban.Cat
## Alabama          13.2     236       58 21.2       Med
## Alaska           10.0     263       48 44.5       Low
## Arizona           8.1     294       80 31.0      High
## Arkansas          8.8     190       50 19.5       Low
## California        9.0     276       91 40.6      High
## Colorado          7.9     204       78 38.7      High
## Connecticut       3.3     110       77 11.1      High
## Delaware          5.9     238       72 15.8      High
## Florida          15.4     335       80 31.9      High
## Georgia          17.4     211       60 25.8       Med
## Hawaii            5.3      46       83 20.2      High
## Idaho             2.6     120       54 14.2       Med
## Illinois         10.4     249       83 24.0      High
## Indiana           7.2     113       65 21.0       Med
## Iowa              2.2      56       57 11.3       Med
## Kansas            6.0     115       66 18.0       Med
## Kentucky          9.7     109       52 16.3       Med
## Louisiana        15.4     249       66 22.2       Med
## Maine             2.1      83       51  7.8       Low
## Maryland         11.3     300       67 27.8       Med
## Massachusetts     4.4     149       85 16.3      High
## Michigan         12.1     255       74 35.1      High
## Minnesota         2.7      72       66 14.9       Med
## Mississippi      16.1     259       44 17.1       Low
## Missouri          9.0     178       70 28.2       Med
## Montana           6.0     109       53 16.4       Med
## Nebraska          4.3     102       62 16.5       Med
## Nevada           12.2     252       81 46.0      High
## New Hampshire     2.1      57       56  9.5       Med
## New Jersey        7.4     159       89 18.8      High
## New Mexico       11.4     285       70 32.1       Med
## New York         11.1     254       86 26.1      High
## North Carolina   13.0     337       45 16.1       Low
## North Dakota      0.8      45       44  7.3       Low
## Ohio              7.3     120       75 21.4      High
## Oklahoma          6.6     151       68 20.0       Med
## Oregon            4.9     159       67 29.3       Med
## Pennsylvania      6.3     106       72 14.9      High
## Rhode Island      3.4     174       87  8.3      High
## South Carolina   14.4     279       48 22.5       Low
## South Dakota      3.8      86       45 12.8       Low
## Tennessee        13.2     188       59 26.9       Med
## Texas            12.7     201       80 25.5      High
## Utah              3.2     120       80 22.9      High
## Vermont           2.2      48       32 11.2       Low
## Virginia          8.5     156       63 20.7       Med
## Washington        4.0     145       73 26.2      High
## West Virginia     5.7      81       39  9.3       Low
## Wisconsin         2.6      53       66 10.8       Med
## Wyoming           6.8     161       60 15.6       Med
```

Question 5

```r
boxplot(Murder ~ Urban.Cat, data = USArrests)
```



Percent urban, as defined by the three categories in `Urban.Cat` does not appear to have a meaningful association with the murder rate in any given city.

```r
cor(USArrests$UrbanPop, USArrests$Murder)
```

```
## [1] 0.06957262
```

```r
summary(lm(Murder ~ UrbanPop, data = USArrests))
```

```
##
## Call:
## lm(formula = Murder ~ UrbanPop, data = USArrests)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.537 -3.736 -0.779  3.332  9.728
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.41594    2.90669   2.207   0.0321 *
## UrbanPop     0.02093    0.04333   0.483   0.6312
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.39 on 48 degrees of freedom
## Multiple R-squared:  0.00484,    Adjusted R-squared:  -0.01589
## F-statistic: 0.2335 on 1 and 48 DF,  p-value: 0.6312
```

These data show the “urbanness” of a city is effectively unrelated to its murder rate.

Question 6

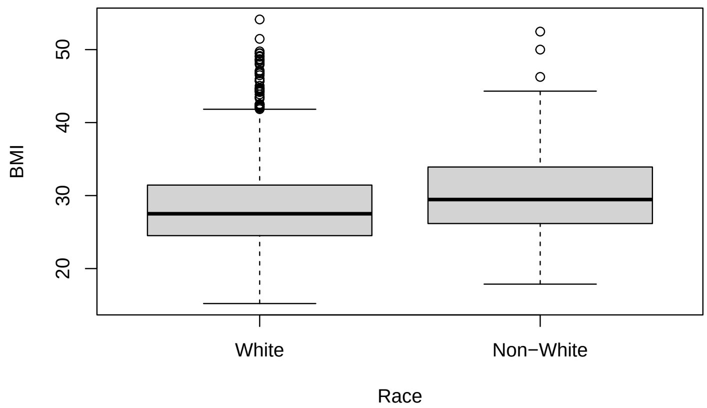
```r
hersdata <- read.csv("C:/Users/Othar Zaldastani/Downloads/hersdata.csv")
head(hersdata)
```

```
##                HT age           raceth nonwhite smoking drinkany exercise
## 1         placebo  70 African American      yes      no       no       no
## 2         placebo  62 African American      yes      no       no       no
## 3 hormone therapy  69            White       no      no       no       no
## 4         placebo  64            White       no     yes      yes       no
## 5         placebo  65            White       no      no       no       no
## 6 hormone therapy  68 African American      yes      no      yes       no
##                physact globrat poorfair medcond htnmeds statins diabetes
## 1     much more active    good       no       0     yes     yes       no
## 2     much less active    good       no       1     yes      no       no
## 3      about as active    good       no       0     yes      no      yes
## 4     much less active    good       no       1     yes      no       no
## 5 somewhat less active    good       no       0      no      no       no
## 6      about as active    good       no       0      no      no       no
##   dmpills insulin weight   BMI waist   WHR glucose weight1  BMI1 waist1  WHR1
## 1      no      no   73.8 23.69  96.0 0.932      84    73.6 23.63   93.0 0.912
## 2      no      no   70.9 28.62  93.0 0.964     111    73.4 28.89   95.0 0.964
## 3      no      no  102.0 42.51 110.2 0.782     114    96.1 40.73  103.0 0.774
## 4      no      no   64.4 24.39  87.0 0.877      94    58.6 22.52   77.0 0.802
## 5      no      no   57.9 21.90  77.0 0.794     101    58.9 22.28   76.5 0.757
## 6      no      no   60.9 29.05  96.0 1.000     116    57.7 27.52   86.0 0.910
##   glucose1 tchol   LDL HDL  TG tchol1  LDL1 HDL1 TG1 SBP DBP age10
## 1       94   189 122.4  52  73    201 137.6   48  77 138  78   7.0
## 2       78   307 241.6  44 107    216 150.6   48  87 118  70   6.2
## 3       98   254 166.2  57 154    254 156.0   66 160 134  78   6.9
## 4       93   204 116.2  56 159    207 122.6   57 137 152  72   6.4
## 5       92   214 150.6  42 107    235 172.2   35 139 175  95   6.5
## 6      115   212 137.8  52 111    202 126.6   53 112 174  98   6.8
```

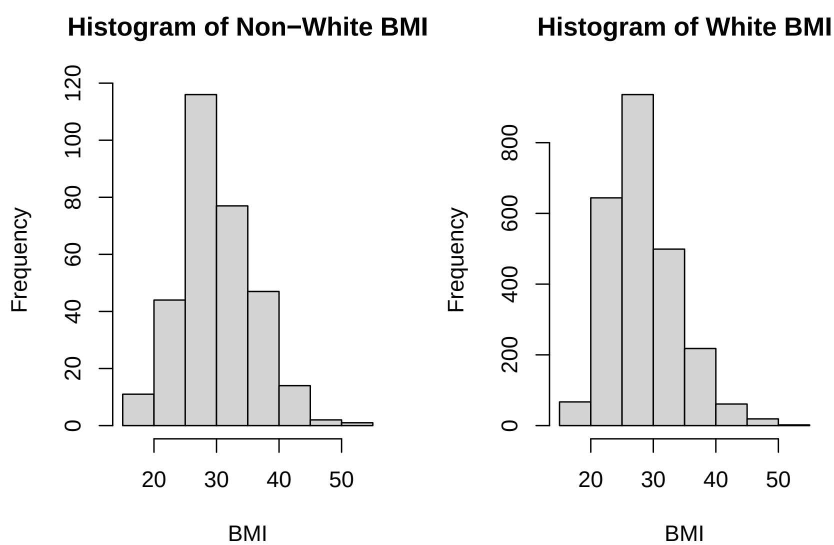
```r
names(hersdata)
```

```
##  [1] "HT"       "age"      "raceth"   "nonwhite" "smoking"  "drinkany"
##  [7] "exercise" "physact"  "globrat"  "poorfair" "medcond"  "htnmeds"
## [13] "statins"  "diabetes" "dmpills"  "insulin"  "weight"   "BMI"
## [19] "waist"    "WHR"      "glucose"  "weight1"  "BMI1"     "waist1"
## [25] "WHR1"     "glucose1" "tchol"    "LDL"      "HDL"      "TG"
## [31] "tchol1"   "LDL1"     "HDL1"     "TG1"      "SBP"      "DBP"
## [37] "age10"
```

```r
boxplot(BMI ~ nonwhite, data = hersdata, xlab = "Race", names = c("White", "Non-White"))
```




```r
par(mfrow = c(1,2))
hist(hersdata$BMI[hersdata$nonwhite == "yes"], main = "Histogram of Non-White BMI", xlab = "BMI")
hist(hersdata$BMI[hersdata$nonwhite == "no"], main = "Histogram of White BMI", xlab = "BMI")
```




```r
hersdata$LnBMI = log(hersdata$BMI)
t.test(LnBMI ~ nonwhite, hersdata)
```

```
##
##  Welch Two Sample t-test
##
## data:  LnBMI by nonwhite
## t = -5.0211, df = 386.75, p-value = 7.857e-07
## alternative hypothesis: true difference in means between group no and group yes is not equal to 0
## 95 percent confidence interval:
```

```
##  -0.08135354 -0.03556994
## sample estimates:
##  mean in group no mean in group yes
##          3.328232          3.386694
```

I ran a t-test with a null hypothesis of no BMI difference between `white` and `nonwhite` women and an alternative hypothesis of a difference in average BMI existing between the two groups. The test gave us a t-score of -5.02, for a p-value of «0.01 in favor of rejecting the null hypothesis. The data show a clear association between BMI and `white`/`nonwhite` status.

Question 7

```r
# Confidence interval function
ci.prop.conf = function(x,n,c) {
  p_hat = x/n
  se = sqrt(p_hat * (1-p_hat) / n)

  alpha = 1 - c
  z = qnorm(1-alpha/2)

  lower = p_hat - z * se
  upper = p_hat + z * se
  return(list(
    proportion = p_hat,
    confidence.interval = c,
    lower = lower,
    upper = upper
  ))}
ci.prop.conf(46, 200, .95)
```

```
## $proportion
## [1] 0.23
##
## $confidence.interval
## [1] 0.95
##
## $lower
## [1] 0.1716767
##
## $upper
## [1] 0.2883233
```

$\hat{p} = 0.23$, and the 95% confidence interval is $[0.192 : 0.268]$.

Question 8

Part A Completed Table

Occupation	# With Latex Allergy	Total # in Group	Frequency per 1,000 people
Health Care Workers	13	58	224.1
Non-Health Care Workers	25	1093	22.9
Painters	2	32	62.5
Hairdressers	3	59	50.8
Food Handlers	7	41	170.7
Cleaners	3	78	38.5

Part B The name of the measurement of disease calculated above is prevalence.
Here it is represented as $\frac{\text{Number with Latex Allergy}}{\text{Total in Group}} \times 1000$.

Part C To determine cumulative incidence, the previous positive results are excluded from the at-risk population. Baseline positives number 164 and the at-risk population is then $2030 - 164 = 1866$. The total number of new cases is 47, making the cumulative incidence $\frac{47}{1866} = 0.0252$.

Part D Person time can be calculated by taking the sum of the products of participants and the number of monthly visits they completed. $(1532 \cdot 24) + (17 \cdot 22) + (19 \cdot 19) + (41 \cdot 16) + (232 \cdot 12) + (25 \cdot)7 = 41,118$ person months, or 3426.5 person-

Part E Incidence density is the number of new cases divided by the person-time.

With 47 new cases and 41,118 person-months, the incidence density is equal to $\frac{47}{41118} \times 1000 \approx 1.14$ cases per 1000 person-months.

Part F This is an example of a proportion because it is calculated by dividing the number of cases by the number of workers. It could also be described as point prevalence because it describes the prevalence of workers with the Latex allergy at a specific point in time (the first week in June). No person-time means it's not incidence density, no follow-up eliminates incidence proportion, and it is not period prevalence because it is a measure at one point in time.