

Using the words we learned in this class, answer all questions succinctly; you will lose points for rambling. Where you can't otherwise do the math, provide all the R code necessary to answer the problem, indicating clearly which variable(s) contain the answer.

1. The Capital Asset Pricing Model (CAPM) is a financial model that assumes returns on a portfolio are normally distributed. Suppose a portfolio has an average annual return of 14.7% (i.e. an average gain of 14.7%) with a standard deviation of 33%. A return of 0% means the value of the portfolio doesn't change, a negative return means that the portfolio loses money, and a positive return means that the portfolio gains money.
 - (a) What percent of years does this portfolio lose money?
 - (b) What is the cutoff for the highest 15% of annual returns with this portfolio?
 - (c) What is the probability that a portfolio makes money?

```
m <- 14.7
sd <- 33

pnorm((0-m)/sd)                # A
## [1] 0.3279957

qnorm(1-.15)*sd + m           # B
## [1] 48.9023

1-pnorm((0-m)/sd)            # C
## [1] 0.6720043
```

2. In 1846, the Donner party (Donner and Reed families) left Springfield, Illinois for California in covered wagons. After reaching Fort Bridger, Wyoming, the leaders decided to find a new route to Sacramento. They became stranded in the eastern Sierra Nevada mountains at a place now called Donner Pass when the region was hit by heavy snows. By the time the survivors were rescued on April 21, 1847, 40 out of 87 had died. Below is a reduced dataframe, named `donner`, of some of the members. Answer the questions following the dataset.

```
##   ageLess30 ageGreater30 female male survived
## 1         1           0      0     1         0
## 2         1           0      1     0         1
## 3         0           1      0     1         0
## 4         1           0      0     1         0
```

```
## 5      1      0      0      1      0
## 6      1      0      0      1      1
## 7      0      1      0      1      0
## 8      1      0      1      0      1
## 9      1      0      1      0      1
## 10     0      1      1      0      1
## 11     1      0      1      0      0
## 12     0      1      1      0      0
## 13     1      0      0      1      1
## 14     0      1      0      1      0
## 15     1      0      0      1      0
## 16     0      1      1      0      0
## 17     1      0      1      0      1
## 18     1      0      0      1      1
## 19     0      1      0      1      0
## 20     1      0      0      1      1
```

- (a) Compute the lift for the following two rules: $\{age > 30, female\} \rightarrow \{survived\}$ and $\{age > 30, male\} \rightarrow \{survived\}$.
- (b) Suppose you found the following numbers. Interpret the two values, separately and together.

```
c(female = with(donner,
                round(mean(ageGreater30 & female & survived)/
                      (mean(ageGreater30 & female)*mean(survived)), 2)),
  male = with(donner,
              round(mean(ageGreater30 & male & survived)/
                    (mean(ageGreater30 & male)*mean(survived)), 2)))

## female  male
##   0.74   0.00
```

The data suggest that being older than 30 and female decreases your chances of survival. There was no older than 30 males that survived.

3. An airline charges the following baggage fees: \$25 for the first bag and \$35 total for two bags. Suppose 54% of passengers have no checked luggage, 34% have one piece of checked luggage and 12% have two pieces. We suppose a negligible portion of people check more than two bags.
- (a) What is the probability mass function for baggage fees? Hint, draw a table.
- (b) About how much revenue should the airline expect for a flight of 120 passengers?
- (c) The current exchange rate for one U.S. dollar (\$) to one Euro is .82. What is the expected revenue in Euros?

```
x <- c(0, 25, 35) # A
px <- c(54, 34, 12)/100
E <- sum(px*x) # B
120*E

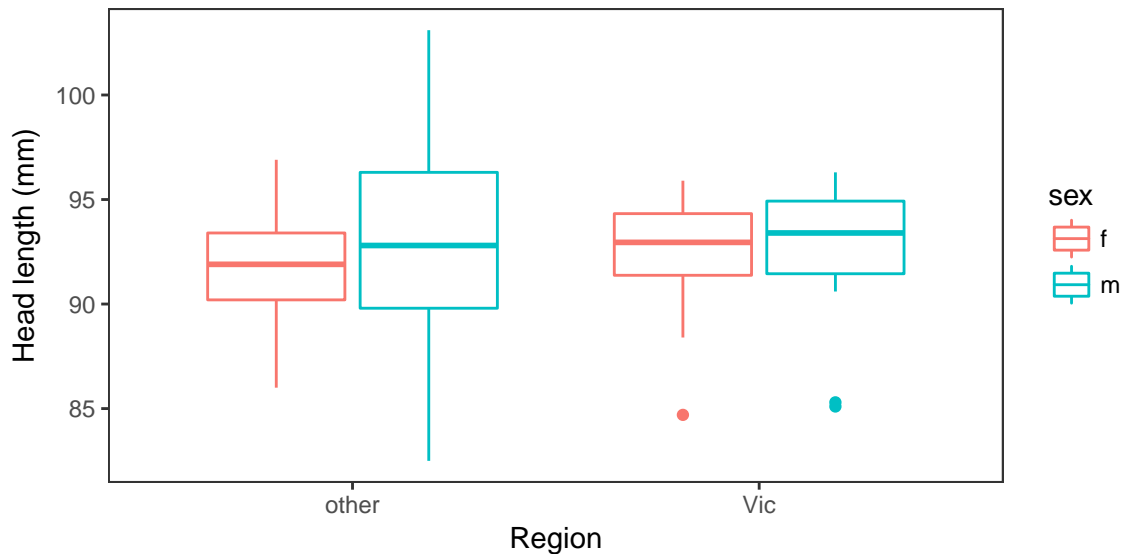
## [1] 1524

.82*E # C

## [1] 10.414
```

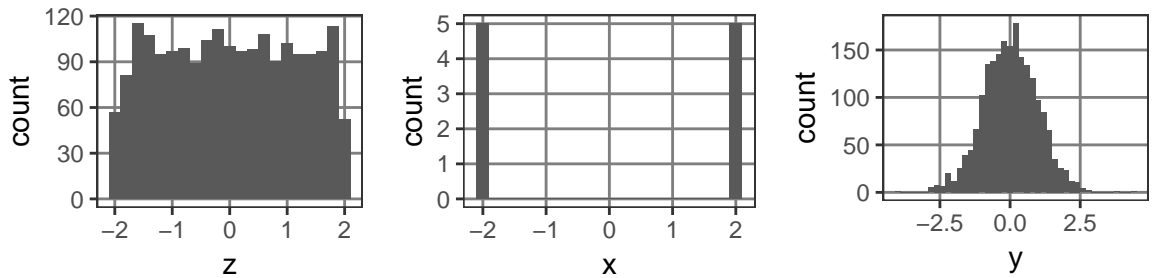
4. A random sample of 104 Australian possum were trapped in either Victoria or Other (New South Wales or Queensland).

```
possum <- read.csv("https://roualdes.us/data/possum.csv")
qplot(factor(pop), headL, data=possum, geom="boxplot", colour=sex) +
  labs(x="Region", y="Head length (mm)") +
  theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank())
```



- Identify the observations and sample size. **Possums and N=104**
- Identify the variables and their types. **Region (factor), Head length (continuous numeric), sex (factor)**
- Identify the explanatory and response variables. **response: head length; explanatory: sex and region**
- Identify the type or design of the study. **observational**

- (e) Describe at least two interesting things about the relationship between the variables. Median head length does not appear to be different amongst Australian possums. Males possums appear to vary more widely, as judged by the interquartile range, for non Victoria possums.
5. The three histograms below have the same mean, $\mu_x = \mu_y = \mu_z = 0$. Without doing any calculations, rank the variables x , y , z in order of increasing standard deviation. Explain your answer.



$\sigma_y < \sigma_z < \sigma_x$, as ordered by average squared distance from the mean.

6. Assume $x_1, \dots, x_N \sim_{iid} \text{Poisson}(\lambda)$. The simplified log-likelihood for the Poisson distribution is

$$\log(\lambda) \sum_{n=1}^N x_n - N\lambda$$

- (a) Write the simplified log-likelihood in R code.

```
## start simplified log-likelihood
ll_poisson <- function(lamdba, X) {
  N <- length(X)
  -log(lamdba)*sum(X) + N*lamdba
}
## end
```

- (b) Write R code to call `optim` to find the maximum likelihood estimator for λ . Some hints are in comments.

```
poisson_data <- rpois(1001, 5)
optim(pi, # starting point
      ll_poisson, # function
      method="L-BFGS",
      lower=1e-6, # lambda can't be 0
      X=poisson_data # pass in data
)
```

```
## $par
## [1] 5.033966
##
## $value
## [1] -3105.073
##
## $counts
## function gradient
##          7          7
##
## $convergence
## [1] 0
##
## $message
## [1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"
```