Problem 9: Bon Voyage

Rakesh Vishwabrahmana

Question 1

Solution:

We want to test that the frequency at which different price points were offered to travellers was consistent among all four destinations. Therefore, the hypotheses are: H_0 = There is no strong association between price points and four destinations. i.e., add-on price was consistent. vs H_1 = There is an association.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse
1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
travel <- read.csv("travel.csv")
travel <- travel %>%
  mutate(price = as.factor(price))
tab1 <- with(travel,table(destination,price))
tab1
```

```
##             price
## destination   5  10  20
##           A  19  27  20
##           B  99 108 115
##           C  32  49  38
##           D  46  36  41
```

```
test1 <- chisq.test(tab1)
test1
```

```
##
##  Pearson's Chi-squared test
##
## data:  tab1
## X-squared = 6.3946, df = 6, p-value = 0.3805
```

The p-value of the test is 0.3804635 which is not significant at 5% level of significance. Therefore we conclude that the frequency at which different price points were offered to travelers was consistent among all four destinations.

Question 2

Solution:

We want to test that that the price offered to travelers is related to the frequency at which they accept the add-on. Therefore, the hypotheses are: H_0 = There is no strong association between offered price and acceptation. vs H_1 = There is an association.

```
tab2 <- with(travel,table(price, purchased))
tab2

##      purchased
## price   N   Y
##    5  161  35
##    10 192  28
##    20 209   5

test2 <- chisq.test(tab2)
test2

##
##  Pearson's Chi-squared test
##
## data:  tab2
## X-squared = 26.907, df = 2, p-value = 1.436e-06
```

The p-value of the test is 1.4361919^{-6} which is significant at 5% level of significance. Therefore we conclude that the price offered to travelers is related to the frequency at which they accept the add-on.

Question 3

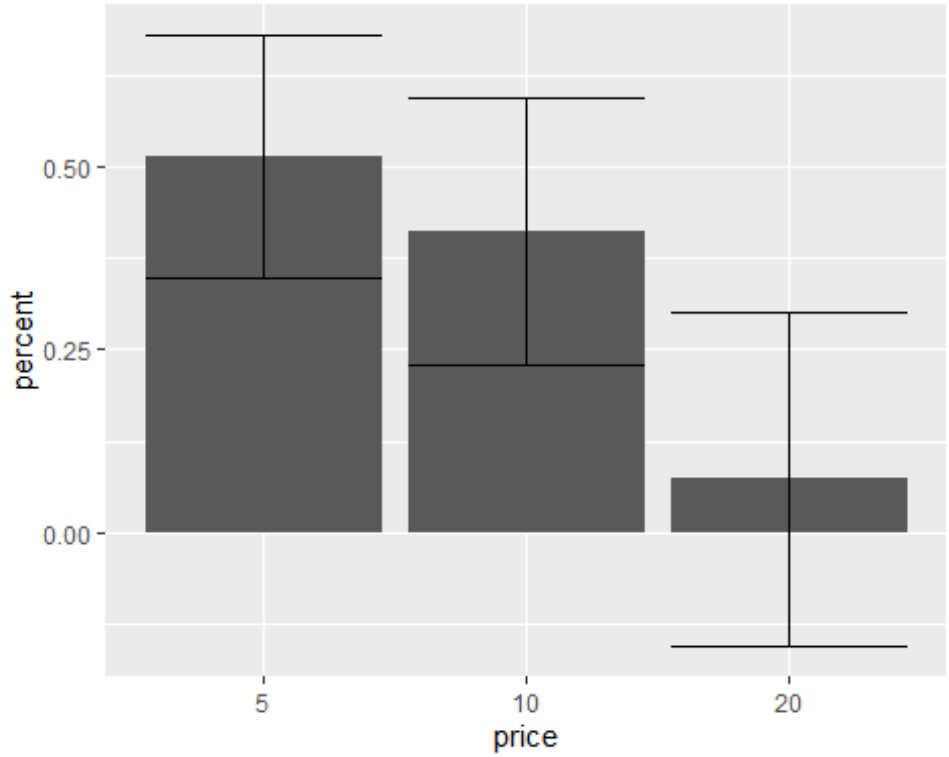
Solution:

In the given chunk below we will count the frequency as well as the proportion of travelers accepting the offer by price-point for all destinations and depict the result using a barplot with error bar.

```
dat <- travel %>%
  filter(purchased == "Y") %>%
  count(price) %>%
  mutate(percent = n / sum(n),
         error = sqrt((percent * (1-percent))/n)) %>%
  mutate(up_ci = percent+1.96*error, lw_ci = percent-1.96*error)

ggplot(dat, aes(price, percent)) +
    geom_col(position = "dodge") +
```

```
    geom_errorbar(aes(ymin = lw_ci, ymax = up_ci),
                  position = position_dodge(0.9))
```



Question 4

Solution:

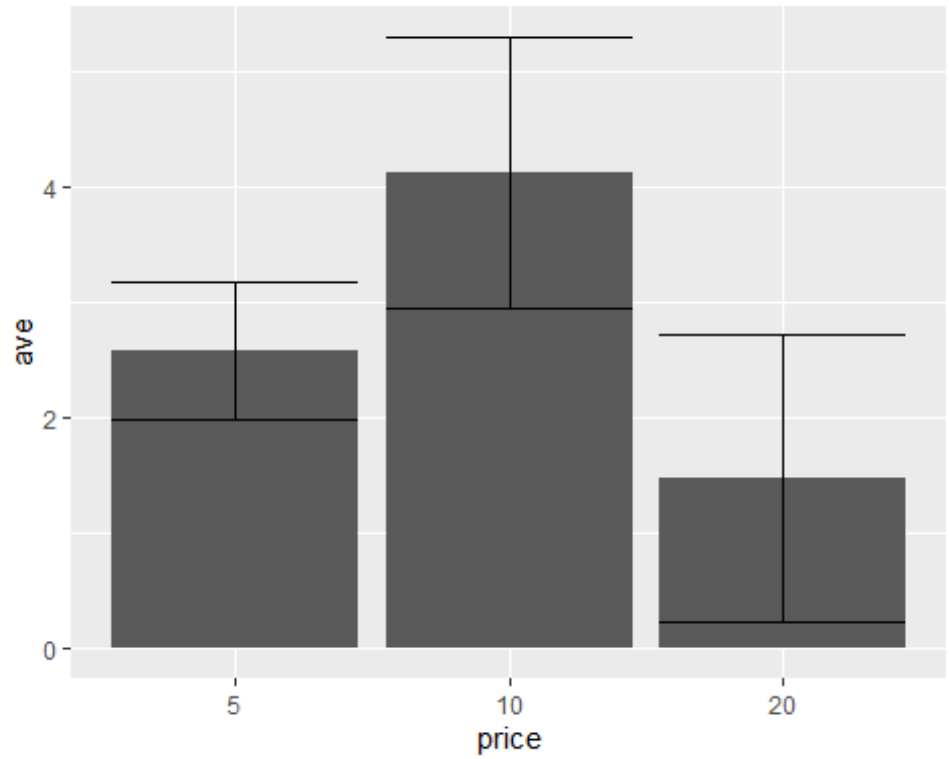
In the given chunk below we will calculate the average add-on revenue generated per traveler by price-point for all destinations and depict the result using a barplot with error bar.

```
dat2 <- travel %>%
  filter(purchased == "Y") %>%
  mutate(price_5 = ifelse(price ==5,5,0),
         price_10 = ifelse(price ==10,10,0),
         price_20 = ifelse(price ==20,20,0)) %>%
  select(price_5, price_10,price_20)

dat2 <- data.frame(price=as.factor(unique(travel$price)),
                   ave = apply(dat2,2,mean),
                   SD = apply(dat2,2,sd),
                   n = dim(dat2)[1]) %>%
  mutate(up_ci = ave + 1.96 * SD/sqrt(n), lw_ci = ave-1.96* SD/sqrt(n))

ggplot(dat2, aes(price, ave)) +
    geom_col(position = "dodge") +
```

```
    geom_errorbar(aes(ymin = lw_ci, ymax = up_ci),
                  position = position_dodge(0.9))
```



Question 5

Solution:

From the above analysis, it is observed that the maximum average add-on revenue found at price point **10**. Based on the sample data we have it is imitate that the add-on price of **10** is the optimum choice. However, we also observed that the upper error bar of add-on price of **5** crosses the lower error bar of add-on price of **10**. Therefore, there is a chance to make add-on revenue at price of **5**.

Question 6

Solution:

We can check the mean and variance of the offered price to check whether there is an over dispersion present or not.

```
travel <- travel %>%
  mutate(price = as.numeric(as.character(price)))
mean(travel$price)

## [1] 11.84127

var(travel$price)
```

```
## [1] 38.41674
```

From the output of the above chunk, we observed that the variance much higher than the mean. Now we fit a generalized linear model with family poisson to investigate whether there is evidence to suggest that the optimum price point may be different for travelers going to different destinations.

```
mod <- glm(price~destination + purchased, family=poisson(link="log"),data =
travel)
summary(mod)

##
## Call:
## glm(formula = price ~ destination + purchased, family = poisson(link =
"log"),
##     data = travel)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -2.494  -2.201  -0.715   1.970   3.667
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.45547    0.03616  67.911   <2e-16 ***
## destinationB  0.06500    0.03963   1.640   0.1010
## destinationC  0.09525    0.04537   2.100   0.0358 *
## destinationD  0.01508    0.04496   0.335   0.7373
## purchasedY   -0.42395    0.04469  -9.486   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 2053.2  on 629  degrees of freedom
## Residual deviance: 1949.4  on 625  degrees of freedom
## AIC: 4592.4
##
## Number of Fisher Scoring iterations: 4
```

This coefficient **purchesed** is highly significant (p < 2e-16). The negative coefficient for purchased indicates that as purchased increase, the mean offered price is smaller. i.e. either **10** or **5**.