­­­

**Revised Question 1**. The Federal Reserve has called remote deposit capture “the most important development the U.S. banking industry has seen in years.” This service allows users to scan checks and to transmit the scanned images to a bank for posting. In its annual survey of community banks, the American Bankers Association asked banks whether or not they offered this service. Hera are the results classified by the asset size (in millions of dollars) of the bank:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Offer | RDC |
|  |  | Yes | No |
| Asset | Under $100 | 63 | 309 |
| Size | $101 - $200 | 59 | 132 |
|  | $201 or more | 112 | 85 |

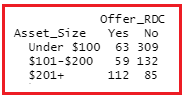
Table: Bank data

TableLab4Q1 <- as.table(rbind(c(63,309),c(59,132),c(112,85)))

dimnames(TableLab4Q1) <- list(Asset\_Size=c("Under $100", "$101-$200","$201+"),

Offer\_RDC=c("Yes","No"))

TableLab4Q1

­­

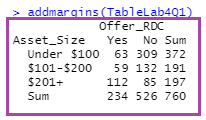
a. Set up the hypotheses to test whether there is an association between Asset size and Offer RDC.

There is no association between Asset Size and Offer RDC

 There is an association between Asset Size and Offer RDC

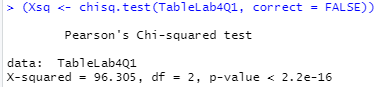
b. Create a table of expected cell counts.

* Show how you would do at least two of the expected cell counts using only a calculator for calculations. You can use R to calculate the remaining expected cell counts.
* Next, enter the observed data (from Table 4) into RStudio.
* Calculate the value of the  statistic, the df, and the *p*-value. (Do not apply Yates correction!)



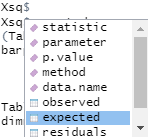
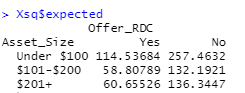
|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Offer | RDC |
|  |  | Yes | No |
| Asset | Under $100 | (372)(234)/760=114.54 | (372)(526)/760=257.46 |
| Size | $101 - $200 | (191)(234)/760 = 58.81 |  |
|  | $201 or more |  |  |

Expected count = (row total)(column total)/(grand total)



(In this case, you will get the same result if you use default: correct = TRUE.)

Expected counts:

Conclude that there is a relationship between Asset Size and Offer RDC. **This holds for the population and not simply for the sample. (Chi-square test is a technique for Inference: Using information from a sample to learn something about a population.)**

Next, we go back to our sample to explore the nature of the relationship. That’s part (d).

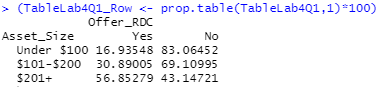
c. If the *p*-value is below 0.05, **reject the null hypothesis** in favor of the alternative. (If you reject the null hypothese, you can say that the results are **significant** – concluding that there is a relationship between Asset size and Offer RDC.).

See above.

d. To examine the nature of any association between the two variables, Asset size and Offer RDC, calculate either row or column percentages, whichever is more appropriate to the situation under study. (Which variable are you using for the explanatory variable?) Justify your choice of type of percentage. What do your percentages reveal about this situation?

We learn the most from the conditional distributions of the response variable for each level of the explanatory variable. Here the explanatory variable is Assess Size and the response variable is Offer RDC.

So, we want the distributions of Offer RDC (Yes, No) for each level of Asset Size. We will have three distributions. We will get them from row percentages.



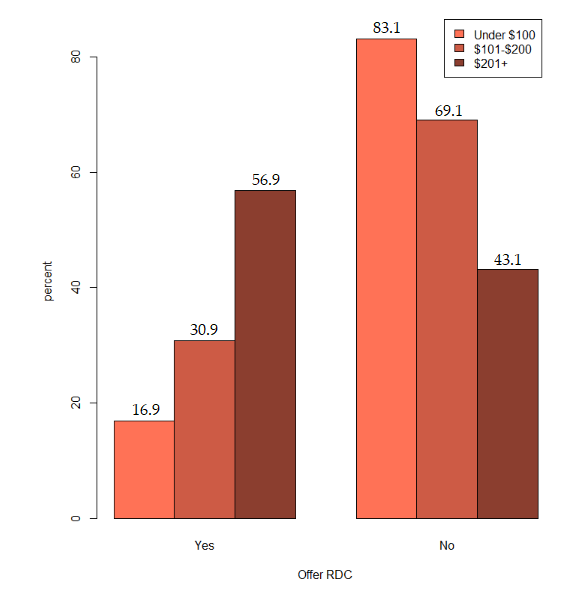
Notice there is a distribution for each row. The percents in each row will sum to 100%.

Next, we visualize the percentages in a bar chart.

barplot(TableLab4Q1\_Row, beside = TRUE,

xlab = "Offer RDC", ylab = "percent",

legend = rownames(TableLab4Q1\_Row), col = c("coral1","coral3","coral4"))



As the Asset Size increases, the percentage offered RDC increases. For “under 100” it is 16.9%; for “$100 - $200” it is 30.9% and for “$201 or more” it is 56.9%. On the other hand, as Asset Size increases the percentage not offered RDC decreases from 83.1% to 69.1% to 43.1%.

2. What if you have raw data as in Monitoring the Future data? The only thing that differs is entry of the frequency table for Sex and Religious importance.

One Variable: Sex

T\_Sex <- table(Lab5Data$Sex)

(T\_Sex\_Perc <- prop.table(T\_Sex)\*100)



Both variables

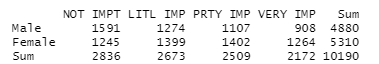
T\_Sex\_Rel <- table(Lab5Data$Sex,Lab5Data$ReligImp)

T\_Sex\_Rel



1. a. What percentage of males responded that religion was very important?

addmargins(T\_Sex\_Rel)



(908/4880)×100 = 18.61%

b. Compare that to the percentage of females.

(1264/5310) ×100 = 23.80%, which is higher than for males.

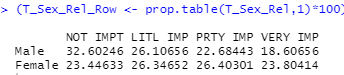
2. What percentage of participants for whom religion was very important were males?

(908/2172) ×100 = 41.80%

Why is this different from 1(a)?

3. We want the conditional distributions of religious importance for each level of sex.

(2 distributions, row percents)



4. Next, we want the conditional distributions of sex for each level of religious importance.

(4 distributions, column percents)

