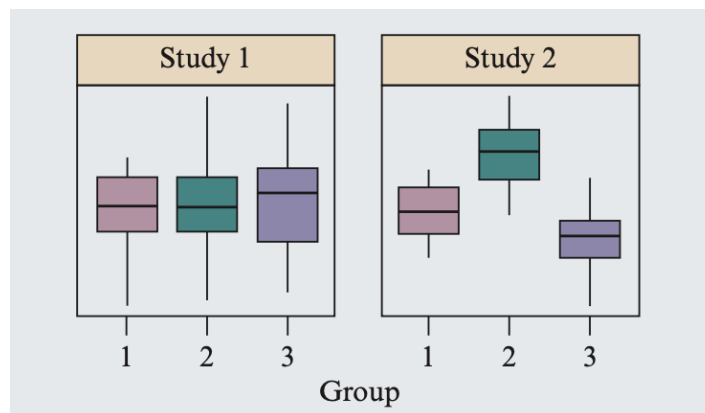


Chapters covered: Chapter 11, 14**Show your work to receive full credit.**

- **Problem 1** A bank conducts a survey in which it randomly samples 400 of its customers. The survey asks the customers which way they use the bank the most: (1) interacting with a teller at the bank, (2) using ATMs, or (3) using the bank's online banking service. It also asks their level of satisfaction with the service they most often use (on a scale of 0 to 10 with 0 = very poor and 10 = excellent). Does mean satisfaction differ according to how they most use the bank?
 - (a). Identifying notations, state the null and alternative hypotheses for conducting an ANOVA with data from the survey.
 - (b). Report the df values for this ANOVA. Above what F test statistic values give a P-value below 0.05?
 - (c). From the data, we have $SSG=0.92$, $SSE=397$. What is our observed F statistics and p-value? How do we interpret the p-value? And what can you conclude?
 - (d). What were the assumptions on which the ANOVA was based? Which assumption is the most important?
- **Problem 2** For two studies, each comparing three groups, the box plots below show results. (Each box plot is based on a random sample of size 40.)
 - (a). Judging from the box plots, which study will more likely lead to a rejection of the ANOVA null hypothesis of equal population means? Explain.
 - (b). Which study will have the larger value for the F test statistic? Why?
 - (c). The P-value for the ANOVA F test for the second study equals to 0.001. Does this necessarily imply that all three population means are different from each other?



- **Problem 3** In pro basketball games during 1980-1982, when Larry Bird of the Boston Celtics made his first free throw, 251 out of 285 times he made the second one, and when he missed his first free throw, 48 out of 53 times he made the second one.

- (a). Form a 2×2 contingency table that cross-tabulates the outcome of the first free throw (with categories made and missed) and the outcome of the second free throw (made and missed). Hint: For part a), use the following contingency table.

First free throw	Second free throw		Total
	Made	Missed	
Made	251		
Missed			53
Total			

- (b). Calculate the expected cell counts for each cell and calculate the chi-squared test statistic.
 (c). Calculate the P-value and make a conclusion based on the results.

R Problem 1

An airline recently conducted a randomized experiment to analyze whether callers would remain on hold longer, on average, if they heard (a) an advertisement about the airline and its current promotions, (b) Muzak, or (c) classical music. The company randomly selected one out of every 1000 calls in a particular week. For each call, they randomly selected one of the three recordings to play and then measured the number of minutes that the caller remained on hold before hanging up (these calls were purposely not answered). The total sample size was 15. The company kept the study small, hoping it could make conclusions without alienating too many potential customers.

For Advertisement group, observed holding times were : 5, 1, 11, 2, 8

For Muzak, observed holding times were : 0, 1, 4, 6, 3

For Classical, observed holding times were: 13, 9, 8, 15, 7.

Use the following R commands to enter the data into R manually.

```
## Wait time for each type of Recorded message
advertisement<-c(5, 1, 11, 2, 8)
muzak<-c(0, 1, 4, 6, 3)
classical<-c(13, 9, 8, 15, 7)

## Create vectors (columns)
## Column time contains all numeric values
time<-c(advertisement, muzak, classical)

## column message contains the three types of recorded message
message<-c(rep("advertisement", 5), rep("muzak", 5), rep("classical", 5))
## rep("advertisement", 5)' means repeat "advertisement", 5 times.
## Type ?rep for more details
## Create data set by putting a "frame" around two columns
dat<-data.frame(time, message)
## Use command dat to view dataframe 'dat'
```

- (a) Identify the explanatory variable and response variable.
 (b) Identify the parameters of interest using correct notation.
 (c) Using $\alpha = 0.05$, conduct an ANOVA F-test to test whether at least two means of wait time are different. Conduct all five steps of the ANOVA F-test, that is
 (i) check all necessary assumptions,

- (ii) state the null and alternative hypotheses,
 - (iii) use the `summary(aov())` command to calculate F-test statistic and
 - (iv) p -value,
 - (v) draw the conclusion in the context of the problem.
- (d) Based on your answer in part (c), do you think it is necessary to conduct Tukey's Honest Significant multiple comparison? Explain.
- (e) Regardless of your answer from part (d), use the output given below and interpret. If any two groups have statistically significant difference in their means, interpret the result.

```
> TukeyHSD(aov(dat$time~dat$message))
Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = dat$time ~ dat$message)
$`dat$message`
              diff          lwr          upr          p adj
classical-advertisement 5.0 -0.7467542 10.746754 0.0910409
muzak-advertisement    -2.6 -8.3467542  3.146754 0.4715973
muzak-classical        -7.6 -13.3467542 -1.853246 0.0107397
```

- (f) The airline wants their customers to remain on hold longer. Based on the analysis, which recorded message would you recommend? Do you think you can draw 'cause-and-effect' relationship from this data? Why or why not?

R problem 2

In this problem, we consider the problem of whether there is an association between the University of Minnesota students' genders and eye colors, based on the results of the Getting to Know You Survey conducted at the beginning of this semester. You may assume that the Getting to Know You Survey represents a random sample from the population of all University of Minnesota students.

- (a) Run the following R codes to obtain the contingency table of the distribution of gender and eye color for the sampled students.

```
NoU <- read.delim("http://stat.umn.edu/~wuxxx725/data/Getting2NoUF2016.txt",header=T)
freq.table <- table(NoU$gender, NoU$eye.color)
freq.table
```

- (b) Use `chisq.test()` to perform the chi-squared test on whether there is an association between gender and eye color for the University of Minnesota students. Please show all five steps.