Project 9 Solutions

Statistics 401: Spring 2006

1. (1 pt, Problem 15.4 on page 676) It is not appropriate to perform an ANOVA since we can not assume that the constant variance assumption is met: $\frac{\text{largest s}}{\text{smallest s}} = \frac{7.38}{2.66} \approx 2.77 > 2$.

2. Do problem 15.6 on page 677.

   (a) The numerator degrees of freedom is $DFTr = k - 1 = 2$. The denominator degrees of freedom is $DFE = N - k = 65$ since $N = 68$.

   (b) (3 pts) The ANOVA:

   i. Hypotheses:
   $\text{H}_0: \mu_{\text{sharer}} = \mu_{\text{full time}} = \mu_{\text{part time}}$
   $\text{H}_a: \mu_{\text{sharer}} \neq \mu_{\text{full time}}$ OR $\mu_{\text{sharer}} \neq \mu_{\text{part time}}$ OR $\mu_{\text{full time}} = \mu_{\text{part time}}$

   ii. Test statistic value:
   This value is given as $F = 6.62$.

   iii. Distribution of the test statistic:
   $F \sim F(2, 65)$.

   iv. p-value:
   R gives the value to be 0.0024

   v. Decision at $\alpha = .05$:
   Since the p-value=.0024 $< .05 = \alpha$, then REJECT H_0 in favor of H_a.

   vi. Conclusion:
   The evidence suggests that there is a difference in mean job satisfaction for job sharers, full-time employees, and part-time employees.

   (c) (1 pt) The test statistic is $F = \frac{MSTr}{MSE} = 6.62$. Thus, to find MSE, first find $MSTr = \frac{\text{SSTr}}{\text{DFTr}} \approx 13.43$. To find $MSTr$, $SSTr = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{\bar{x}})^2 \approx 26.85$ since the grand mean is $\bar{\bar{x}} = \frac{\sum n_i \bar{x}_i}{\sum n_i} = 5.75$. Thus, $MSE \approx 2.0283$.

   (d) (2 pts) In addition to the calculations above, we only need to compute $SSE = MSE(DFE)$ and $SST = SSTr + SSE$ to fill in the rest of the ANOVA table in Table 1.

Table 1: One-way ANOVA Table for Job Satisfaction by Employee Type

Source	DF	Sum of Squares (SS)	Mean Squares (MS)	F	p-value
Treatments	2	26.85	13.42722	6.62	0.0024
Error	65	133.87	2.0283		
Total	68	160.72			

   (e) (2 pts) Table 2 contains estimates of μ_3 and σ.

Table 2: Estimate of some ANOVA Parameters

parameter	estimate	Explanation in English words of the parmater
Estimate for μ_1	6.6	The mean level of satisfaction of the Job Sharers
Estimate of σ	$\sqrt{2.0283} = 1.4258$	The constant standard deviation for each population of workers

3. Problem 15.22 on page 685:

   (a) (1 pt) A follow-up analysis to #2: problem 15.6 on page 677 is appropriate since we rejected H_0 in the ANOVA test.

   (b) (3 pts) Three 95% Tukey's CI's were calculated using the formula: $\bar{x}_i - \bar{x}_j \pm \frac{q_{1-\alpha,k,DFE}}{\sqrt{2}}\sqrt{MSE(\frac{1}{n_i}+\frac{1}{n_j})}$. Table 8 in the text book, with $k = 3$ and $DFE = 66$ rounded down to 60, shows that the Tukey critical value is $q_{1-\alpha,k,DFE} \approx 3.4$. Table 3 shows these three CI's.

Table 3: 95% Tukey CI's for the ANOVA of Job Sharing

Comparison	Estimate	Lower	Upper
$\mu_{\text{sharer}} - \mu_{\text{full time}}$	1.23	0.2442	2.2158
$\mu_{\text{sharer}} - \mu_{\text{part time}}$	1.4	0.3661	2.4339
$\mu_{\text{full time}} - \mu_{\text{part time}}$	0.17	-0.8639	1.2039

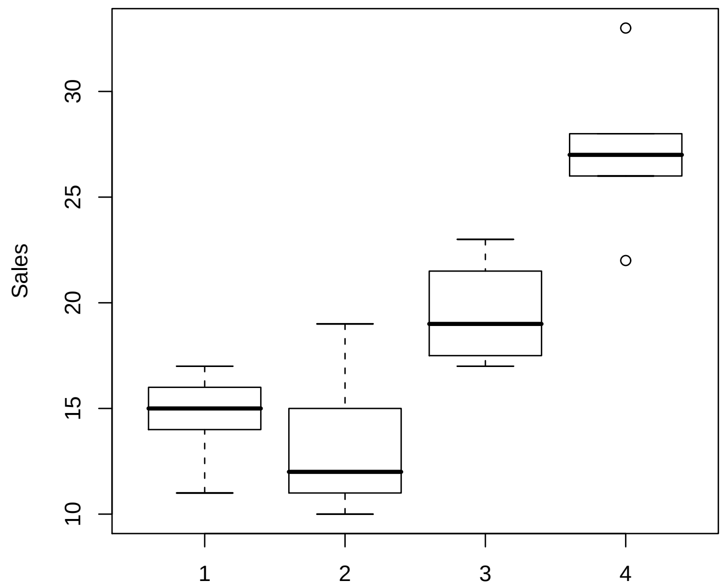
   (c) (1 pt) Among hospital employees, the mean level of job satisfaction for job sharers is significantly larger than the mean level of job satisfaction for either full-timers or part-timers.

4. The Kenton Food Company wanted to test four different package designs (Design 1, Design 2, Design 3, and Design 4) for a new breakfast cereal. Nineteen stores were selected as the experimental units. Each store was randomly assigned one of the package designs, with each package design, except for Design 3, assigned to five stores. The stores were chosen to be comparable in location and sales volume. Other relevant conditions that could affect sales, such as price, amount and location of shelf space, and special promotional efforts, were kept the same for all of the stores in the experiment.

   (a) (1 pt) The variable $x_{2,3} = 15$, which means that the monthly sales for the third store using Design #2 is $1500.

   (b) (1 pt) Side-by-side box-plots of cereal sales for each box design are given in Figure 1.

Figure 1: Cereal Sales (in hundreds of dollars) per Box Design



(c) (1 pt) Hypotheses:
$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$
$H_a: \mu_i \neq \mu_j$ for some i and j

(d) (2 pts) Fitting the ANOVA model yields the one-way ANOVA table in Table 4.

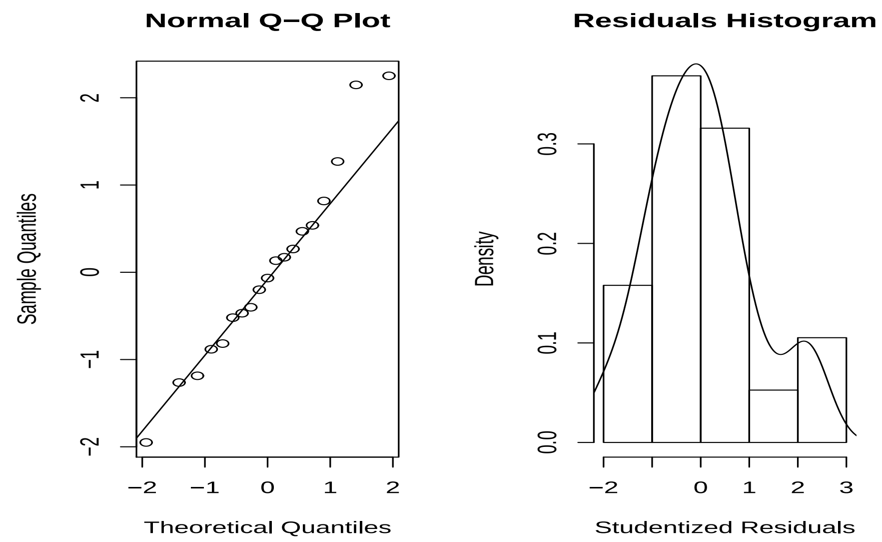
Table 4: One-way ANOVA Table for Sales by Design

Source	DF	Sum of Squares (SS)	Mean Squares (MS)	F	p-value
Treatments	3	588.22	196.07	18.591	2.585×10^{-5}
Error	15	158.20	10.55		
Total	18	746.42			

(e) (3 pts) Check the assumptions.

i. The evidence fails to suggest that the data for each group are not normal. The normal points in the normal probability plot have a linear pattern, and the smoothed histogram of the studentized residuals appears normal (see Figure 2). The correlation of the studentized residuals is .98, which is much larger than the critical correlation value of $r_{\text{critical}} = .929$.

Figure 2: Checking the normality assumption



ii. Since $\frac{\text{largest s}}{\text{smallest s}} \approx \frac{3.962323}{2.302173} \approx 1.721123 < 2$, the constant variance assumption appears to hold.

(f) (1 pt) The distribution of the test statistic assuming the H_0 is true is $F \sim F(3, 15)$.

(g) Since the p-value is tiny, then we REJECT H_0.

(h) (1 pt) Since this is a completely randomized experiment (CRD), then we can make cause-and-effect conclusions. Thus, the evidence suggests that there is an effect of package design on the mean sales of cereal.

5. (a) (1 pt) It is appropriate to conduct a follow-up test since the ANOVA null hypothesis was rejected.

(b) (3 pts) Table 5 displays the 95% Tukey confidence intervals for the pairwise differences between means

Table 5: 95% Tukey CI's

Comparison	Estimate	Lower	Upper
$\mu_2 - \mu_1$	-1.2	-7.1197584	4.719758
$\mu_3 - \mu_1$	4.9	-1.3788520	11.178852
$\mu_4 - \mu_1$	12.6	6.6802416	18.519758
$\mu_3 - \mu_2$	6.1	-0.1788520	12.378852
$\mu_4 - \mu_2$	13.8	7.8802416	19.719758
$\mu_4 - \mu_3$	7.7	1.4211480	13.978852

(c) (1 pt) Package design 4 has significantly higher mean sales than the other three package designs.

(d) (2 pts) Table 5 gives the table of parameters and estimates:

Table 5: Estimate of some ANOVA Parameters

parameter	estimate	Explanation in English words of the parmater
Estimate for μ_4	27.2	The mean sales of cereal using box Design 4
hline Estimate of σ	$\sqrt{10.55} = 3.248$	The constant standard deviation for each group of cereal boxes

Appendix

```
> # PROBLEM 2b
> pf(6.62,df1=2,df2=66,lower.tail=FALSE)
[1] 0.002397632

> # PROBLEM 2c
> n=c(24,24,20)
> xbar=c(6.6,5.37,5.2)
> xgrand=n%*%xbar/sum(n)
> SSTr=n%*%(xbar-xgrand)^2
> MSTr=SSTr/2
> MSTr
         [,1]
[1,] 13.42722
> F=6.62
> MSE=MSTr/F
> MSE
         [,1]
[1,] 2.028281


> # PROBLEM 3
> i=1;j=2;xbar[i]-xbar[j];xbar[i]-xbar[j] + c(-1,1)*qtukey(.95,nmeans=3,df=DFE)/sqrt(2)*
sqrt(MSE*(1/n[i] + 1/n[j]))
[1] 1.23
[1] 0.2442454 2.2157546
> i=1;j=3;xbar[i]-xbar[j];xbar[i]-xbar[j] + c(-1,1)*qtukey(.95,nmeans=3,df=DFE)/sqrt(2)*
sqrt(MSE*(1/n[i] + 1/n[j]))
[1] 1.4
[1] 0.3661318 2.4338682
> i=2;j=3;xbar[i]-xbar[j];xbar[i]-xbar[j] + c(-1,1)*qtukey(.95,nmeans=3,df=DFE)/sqrt(2)*
sqrt(MSE*(1/n[i] + 1/n[j]))
[1] 0.17
[1] -0.8638682  1.2038682



> # PROBLEM 4
> D=read.table("project9stores.txt",header=T)
> attach(D)
> Design
 [1] 1 1 1 1 1 2 2 2 2 2 3 3 3 3 4 4 4 4 4
> Design=factor(Design)
> Design
 [1] 1 1 1 1 1 2 2 2 2 2 3 3 3 3 4 4 4 4 4
Levels: 1 2 3 4

> # PROBLEM 4a
> boxplot(Sales ~ Design,names=levels(Design),ylab="Sales",main="Cereal Sales per Box Design")
```

```
> # PROBLEM 4c
> tapply(Sales,Design,mean)
   1    2    3    4
14.6 13.4 19.5 27.2

> sales.aov=aov(Sales ~ Design)
> summary(sales.aov)
            Df Sum Sq Mean Sq F value    Pr(>F)
Design       3 588.22  196.07  18.591 2.585e-05 ***
Residuals   15 158.20   10.55
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> # PROBLEM 4di
> par(mfrow=c(1,2))
> qqnorm(studres(sales.aov))
> qqline(studres(sales.aov))
> hist(studres(sales.aov),freq=FALSE,main="Residuals Histogram",xlab="Studentized Residuals")
> lines(density(studres(sales.aov)))
> xy=qqnorm(studres(sales.aov))
> cor(xy$x,xy$y)
[1] 0.9841437

> # PROBLEM 4dii
> tapply(Sales,Design,sd)
       1        2        3        4
2.302173 3.646917 2.645751 3.962323



> # PROBLEM 5
> TukeyHSD(sales.aov,which="Design",conf.level=.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Sales ~ Design)

$Design
    diff        lwr       upr     p adj
2-1 -1.2 -7.1197584  4.719758 0.9352978
3-1  4.9 -1.3788520 11.178852 0.1548895
4-1 12.6  6.6802416 18.519758 0.0001013
3-2  6.1 -0.1788520 12.378852 0.0582866
4-2 13.8  7.8802416 19.719758 0.0000368
4-3  7.7  1.4211480 13.978852 0.0142180
```