**Department of Mathematics**
**The University of Toledo**

Master of Science Degree
Comprehensive Examination
**Applied Statistics**

April 20, 2019

*Instructions*

*Do all problems.*

Show all of your computations.
Prove all of your assertions or quote appropriate theorems.
Books, notes, and calculators may be used.
This is a three hour test.

1. (30 points) A company uses six filling machines of the same make and model to place detergent into cartons that show a label weight of 32 ounces. The production manager has complained that the six machines do not place the same amount of fill into the cartons. A consultant requested that 20 filled cartons be selected randomly from each of the six machines and the content of each carton carefully weighed. The observations (stated for convenience as deviations from 32.00 ounces) follow. Assume that ANOVA cell means model $Y_{ij} = \mu_i + \varepsilon_{ij}$ is applicable. The SAS output is given in the attached paper. Please answer following questions:
(a) (5 points) Obtain the fitted values.
(b). (5 points) Obtain the analysis of variance table.
(c). (5 points) Test whether or not the mean fill differs among the six machine. Let $\alpha = 0.05$, state the alternatives, decision rule and conclusion. Does your conclusion support the production manager's complaint?
(d). (5 points) Use the Tukey multiple comparison method to determine which pairs of machines differ significantly. Summarize the results. $\alpha = 0.05$.
(e). (10 points) Suppose you want to compare the average of the first two machines with the average of the last four. Write out the Linear contrast and report the estimated value of this contrast with its standard error; state the null and alternative hypotheses, the test statistic with degrees of freedom, the p-value and your conclusion.

2. (20 points) A model fit predicting preference for U.S. President (Democrat, Republican, Independent) using $x$ = annual income (in $\$10,000$) is $\log(\hat{\pi}_D/\hat{\pi}_I) = 3.3 - 0.2x$ and $\log(\hat{\pi}_R/\hat{\pi}_I) = 1.0 + 0.3x$. Here $\hat{\pi}_D, \hat{\pi}_R, \hat{\pi}_I$ are the predicted probilities of preference for Democrat, Republican and Independent, respectively.
(a). (10 points) Find the prediction equation for $\log(\hat{\pi}_R/\hat{\pi}_D)$ and interpret the slope. For what range of $x$ is $\hat{\pi}_R > \hat{\pi}_D$?
(b). (10 points) Find the prediction equation for $\hat{\pi}_I$. Hint: $\hat{\pi}_D + \hat{\pi}_R + \hat{\pi}_I = 1$.

**3.** [25 points] Let $k$ be a positive integer. For the $3 \times 4$ contingency table

$$\begin{array}{cccc} k & k & k & k \\ k & k & 0 & k \\ k & k & k & k \end{array}$$

find $k$ such that the hypothesis about independence of two classifications will be rejected at the 0.05 significance level. Use Pearson chi-squared statistic $X^2$.

**4.** [25 points] Suppose you conduct an experiment and inject a drug into three mice. Their times for running a maze are 8, 10, and 15 seconds; the times for two control mice are 5 and 9 seconds.

(a) Compute the difference in mean times between the treatment group and the control group.

(b) Write out all possible permutations of these times to the two groups and calculate the difference in means.

(c) What proportion of the differences are as large or larger than the observed difference in mean times?

(d) For each permutation, calculate the mean of the treatment group only. What proportion of these means are as large or larger than the observed mean of the treatment group?

SAS Output for problem 1

```
data detergent,
input fill machine carton;
datalines;
  -0.14      1      1
   0.20      1      2
   ......
   0.05      6     20
;
run;

proc glm data=detergent;
class machine;
model fill=machine;
means machine / t clm;
means machine / tukey;
contrast '1&2 v 3&4&5&6' machine .5 .5 -.25 -.25 -.25 -.25;
estimate '1&2 v 3&4&5&6' machine .5 .5 -.25 -.25 -.25 -.25;
run;
quit;
```

The GLM Procedure

Class Level Information

| Class | Levels | Values |
|---|---|---|
| machine | 6 | 1 2 3 4 5 6 |

Number of Observations Read 120
Number of Observations Used 120

The GLM Procedure

Dependent Variable: fill

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 5 | 2.28934667 | 0.45786933 | 14.78 | <.0001 |
| Error | 114 | 3.53060000 | 0.03097018 | | |
| Corrected Total | 119 | 5.81994667 | | | |

| R-Square | Coeff Var | Root MSE | fill Mean |
|---|---|---|---|
| 0.393362 | 77.29873 | 0.175983 | 0.227667 |

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| machine | 5 | 2.28934667 | 0.45786933 | 14 78 | <.0001 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| machine | 5 | 2.28934667 | 0.45786933 | 14.78 | <.0001 |

The GLM Procedure

t Confidence Intervals for fill

| | |
|---|---|
| Alpha | 0.05 |
| Error Degrees of Freedom | 114 |
| Error Mean Square | 0.03097 |
| Critical Value of t | 1.98099 |
| Half Width of Confidence Interval | 0.077954 |

| machine | N | Mean | 95% Confidence Limits | |
|---|---|---|---|---|
| 3 | 20 | 0.46000 | 0.38205 | 0.53795 |
| 4 | 20 | 0.36550 | 0.28755 | 0.44345 |
| 2 | 20 | 0.19050 | 0.11255 | 0.26845 |
| 6 | 20 | 0.15150 | 0.07355 | 0.22945 |
| 5 | 20 | 0.12500 | 0.04705 | 0.20295 |
| 1 | 20 | 0.07350 | -0.00445 | 0.15145 |

The GLM Procedure

Tukey's Studentized Range (HSD) Test for fill

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

| | |
|---|---|
| Alpha | 0.05 |
| Error Degrees of Freedom | 114 |
| Error Mean Square | 0.03097 |
| Critical Value of Studentized Range | 4.09949 |
| Minimum Significant Difference | 0.1613 |

Means with the same letter are not significantly different.

| Tukey Grouping | Mean | N | machine |
|---|---|---|---|
| A | 0.46000 | 20 | 3 |
| A | | | |
| A | 0.36550 | 20 | 4 |
| | | | |
| B | 0.19050 | 20 | 2 |
| B | | | |
| B | 0.15150 | 20 | 6 |
| B | | | |
| B | 0.12500 | 20 | 5 |

| | | | |
|---|---|---|---|
| | B | | |
| B | 0.07350 | 20 | 1 |

The GLM Procedure

Dependent Variable: fill

| Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| 1&2 v 3&4&5&6 | 1 | 0.54912667 | 0.54912667 | 17.73 | <.0001 |

| Parameter | Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|
| 1&2 v 3&4&5&6 | -0.14350000 | 0.03407905 | -4.21 | <.0001 |