

## 1. Matrix Computations

(a)

$$A + B = \begin{pmatrix} 2 & -3 & 0 \\ 1 & 4 & 2 \end{pmatrix} + \begin{pmatrix} 6 & 1 & -3 \\ -1 & 0 & 4 \end{pmatrix} = \begin{pmatrix} 8 & -2 & -3 \\ 0 & 4 & 6 \end{pmatrix}$$

(b)

$$AC = \begin{pmatrix} 2 & -3 & 0 \\ 1 & 4 & 2 \end{pmatrix} \begin{pmatrix} 3 & 4 \\ -2 & 0 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 6+6+0 & 8+0+0 \\ 3-8+2 & 4+0+4 \end{pmatrix} = \begin{pmatrix} 12 & 8 \\ -3 & 8 \end{pmatrix}$$

(c)

$$AB' = \begin{pmatrix} 2 & -3 & 0 \\ 1 & 4 & 2 \end{pmatrix} \begin{pmatrix} 6 & -1 \\ 1 & 0 \\ -3 & 4 \end{pmatrix} = \begin{pmatrix} 9 & -2 \\ 4 & 7 \end{pmatrix}$$

(d) Eigenvalues/vectors of  $D$ . We compute,

$$\det(D - \lambda I) = \begin{vmatrix} 2 - \lambda & -4 \\ -1 & -1 - \lambda \end{vmatrix} = \lambda^2 - \lambda - 6$$

Setting equal to zero and solving for  $\lambda$ , we arrive at the values,

$$\lambda_1 = 3, \lambda_2 = -2$$

Now,

$$\begin{aligned} D - \lambda_1 I &= \begin{pmatrix} -1 & -4 \\ -1 & -4 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 4 \\ 0 & 0 \end{pmatrix} \end{aligned}$$

Thus, we arrive at the equation,

$$x_1 + 4x_2 = 0 \Rightarrow x_1 = -4x_2$$

Fixing  $x_2 = 1$ , we arrive at the eigenvector,

$$\varepsilon_1 = \begin{pmatrix} -4 \\ 1 \end{pmatrix}$$

Similarly,

$$\begin{aligned} D - \lambda_2 I &= \begin{pmatrix} 4 & -4 \\ -1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \end{aligned}$$

Thus, our second eigenvector is,

$$\varepsilon_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

(e) By construction, we are seeking a matrix of dimension  $1 \times 3$ . As such, we now consider the method by which we compute the mean of a set of numbers, by adding together and then dividing. Thus, we consider the matrix,

$$Y = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$$

Now,

$$YC = \begin{pmatrix} 2 & 6 \end{pmatrix}$$

As desired, so  $z = 1$ 

## 2. Given,

y	0	1	2	3
p(y)	1/8	1/4	3/8	1/4

We begin by computing the mean of  $Y$  as,

$$\mu_Y = \sum (yp(y)) = 1/4 + 3/4 + 3/4 = 7/4$$

The variance is computed similarly as,

$$\begin{aligned} Var(Y) &= \sum (y^2 p(y)) - \mu_Y^2 \\ &= 1 * 1/4 + 4 * 3/8 + 9 * 1/4 - 49/16 = 15/16 = .9375 \end{aligned}$$

The standard deviation of  $Y$  is then,

$$\sigma_Y = \sqrt{Var(Y)} = \sqrt{15}/4 \approx .9682$$

3. We may compute the covariance between  $Y_1$  and  $Y_2$  using the formula,

$$Cov(Y_1, Y_2) = E(Y_1 Y_2) - \mu_{Y_1} \mu_{Y_2}$$

Here,  $\mu_{Y_i} = 1$  as this is the average number of jobs for both firms. The Expected value is then computed as,

$$\sum Y_1 Y_2 p(Y_1 Y_2)$$

Given the locations of zeros in this matrix, the only nonzero entry to this sum is the case,

$$Y_1 = Y_2 = 1 \Rightarrow Y_1 Y_2 P(Y_1 Y_2) = \frac{2}{9}$$

The covariance is then,

$$Cov(Y_1, Y_2) = \frac{2}{9} - 1 = \frac{-7}{9}$$

This value does make sense in context, as the variables are inversely correlated. If more jobs are awarded to either company, there are less available jobs for the others.

4. Reading data

- (a) Due to the relative sizes of the two groupings, 21 and 23, the best choice of analysis for this data is a Small Sample t-test for the difference between two means.
- (b) For this proposed test, we are assuming that the data comes from a continuous spectrum and follows an approximately normal distribution. By using a small sample test, we are correcting for possible violations of the CLT. Based on the experiment described, we may assume that these conditions are met.
- (c) To begin, we isolate the two groups into the treat/control categories, and compute the basic statistics. These results follow,

Statistic	Treat	Control
$\mu$	51.4762	41.5
$Var$	121.1619	308.0714
$s$	11.00736	17.55196
$n$	21	23

Because the samples are small, we must compute the degrees of freedom,  $\nu$  using the  $s$  and  $n$  values. Based on the data, we conclude that,

$$\nu = 37$$

Thus, we will use a  $t$ -test with 37 degrees of freedom. Now, let us set up our hypotheses. Our Null hypothesis,  $H_0$  is that there is no difference in the means, or that the test is ineffective i.e.

$$H_0 : \mu_T - \mu_C \leq 0$$

Our alternative is then the success of the treatment,

$$H_A : \mu_T - \mu_C > 0$$

We compute the  $t$  statistic as,

$$t = \frac{\mu_T - \mu_C - 0}{\sqrt{s_T^2/n_t + s_C^2/n_c}} = 2.27888$$

Using  $R$ , we compute the corresponding  $p$  value to be,

$$p = 0.01426957$$

Because  $p < 0.05$ , we shall reject the null hypothesis, that there is no improvement from the activity, in favor of the alternative, that the activity improves the DRP scores.

(d) For the instructor:

After analyzing the data that you have provided, we have performed a statistical test to determine whether the scores for the class with the added activity showed significant improvement over the control. Using the scores provided, we have concluded that the activities appear to have a positive effect, for the tested students at least.

#### 5. Eye Color Data:

- (a) Due to the nature of this experiment, the best approach for analysis is a one-way ANOVA. Because we have four different conditions that are being tested, the ANOVA method provides a way to compare the effects.
- (b) The assumptions for ANOVA are,
  - i. The populations are normal
  - ii. The populations have the same variance

Because we are considering data that was sampled from a large population, with each member of the population being subjected to at least one of the treatments, we may assume that these conditions are met.

- (c) To begin the analysis, we compute the weighted averages and overall mean as follows,

Eye Condition	Associated Mean
Blue	3.19403
Brown	3.724324
Green	3.85974
Down	3.107317
Overall	3.497297

Next, we compute the SSTr and SSE for the data,

$$SSTr = 24.41966 \quad SSE = 613.1387$$

From this, we can compute

$$MSTr = 8.139886667 \quad MSE = 2.774383258$$

Then,

$$F = \frac{8.139886667}{2.774383258} = 2.93394456$$

We then compute the associated p-value as,

$$1 - P(f \leq F) = 0.03$$

This distribution is,

$$F_{3,221}$$

Because  $p < 0.05$ , we shall reject the null hypothesis, that there is no difference in the Eye color effects, in favor of the alternative, that the color/position does have different effects on the results.

- (d) We now construct the CI for these results as follows, noting that our critical  $t$ -value is 1.9709.
  - i. Blue:

$$[3.19403 \pm t^* \sqrt{\frac{MSE}{J_{Blue}}}] = [2.792998, 3.595062]$$

- ii. Brown:

$$[3.724324 \pm t^* \sqrt{\frac{MSE}{J_{Brown}}}] = [3.18467, 4.263978]$$

- iii. Down:

$$[3.107317 \pm t^* \sqrt{\frac{MSE}{J_{Down}}}] = [2.594663, 3.619071]$$

iv. Green:

$$[3.85974 \pm t^* \sqrt{\frac{MSE}{J_{Green}}}] = [3.485655, 4.233826]$$

Interpreting these intervals, we see that both Blue Eyes and Downward Facing Eyes have very similar means, and confidence interval widths. These means are much lower comparatively than both the Brown and Green Eye tests. Based on the ANOVA test that we performed, we consider that the Brown and Green tests may have a higher attitude score than the Blue and Down tests.

- (e) To compute the Bonferroni CI, we adjust the width of the interval to be from 95% to 98.5%. This enables us to increase the accuracy, at the cost of a wider interval. These intervals are considered more conservative than the ones we computed before.

\*Note: I performed the computations for these tests by hand for this assignment to recall the process. I will include the relevant R code for future assignments, like you said. I just had not done these computations manually for several years, and I wanted to recall how to do them.