Business Analytics

Prof. Phil Jones

Fall, 2016

**Exam #3**

**Instructions:**

1. Please write your name at the top of each page.
2. The exam is open book, open note, and open computer.
3. You are expected to do your work entirely within your group.

Note: Problems are worth points as indicated.

Providing short explanations may allow awarding partial credit in some cases.

The exam is due **November 9**. Please return your exam (one copy per group!) to the dropbox NLT 6PM on the 9th.

*Good Luck!*

|  |  |
| --- | --- |
| 1. 20 points 2. 10 points 3. 10 points 4. 20 points 5. 10 points 6. 20 points 7. 10 points | Jerry Jacob  Danen Sorenson  Purna Chandra Kuntla  Andrew Paterson |

1) (20 points ) **Nurse Scheduling**. As the CFO of a major metropolitan hospital, Susan is concerned about the upcoming negotiations with the healthcare workers union. As part of an earlier contract, the union negotiated for and got an agreement that nurses and other workers would be scheduled on only one shift during any given week for the 5 days they are assigned to work. That is, during a given week, the worker would be assigned, for example, only to the 8AM – 5PM shift. Even though management retained flexibility to change a worker’s assigned shift from week to week, there was no flexibility of shift assignment within the week. Susan has estimated that agreeing to this demand has raised the hospital’s salary and wage budget by approximately 5%. Management still retains the flexibility to schedule a worker’s two days off to be any two of the seven days within the week, not necessarily consecutive. This time around, however, the union is demanding that every worker be guaranteed that the two days off be consecutive days (eg. Monday and Tuesday would be fine, but Monday and Friday would not.) Susan is concerned that the hospital’s budget cannot afford to take another 5% hit, and has gathered some information about the staffing requirements of one particular shift in one of the hospital’s units. The personnel required at this unit for the specified shift are, by day:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Day of Week | Sat | Sun | Mon | Tues | Wed | Thur | Fri |
| Personnel Required | 16 | 11 | 12 | 12 | 13 | 10 | 16 |

Currently, Bob, the manager of this unit has assigned 18 workers to this shift with days off (they work the other 5 days)as follows:

Sat – Sun (2 workers)

Sun – Mon (3 workers)

Sunday and Tuesday (3 workers)

Mon – Tues (3 workers)

Wed – Thurs (5 workers)

Thurs – Fri (2 workers)

Note that everyone (with the exception of the 3 workers assigned Sunday and Tuesday as their days off) already gets consecutive days off. Bob claims that he sees no way to make everyone’s days off consecutive at the current staffing level. Further, he claims that there is no way he can meet his staffing requirements with fewer than 18 workers assigned to this shift. He thinks it might be possible to give consecutive days off with more staff, but isn’t sure how many would be needed.

a. (5 points) Does Bob’s assignment of days off to the 18 workers assigned this shift meet the staffing requirements outlined in the first table? Are there any days that are over-staffed? Can Bob currently (under the existing union contract) meet the staffing requirements with fewer than 18 workers?

Bob’s assignment does not meet staffing requirements since there will be a worker short on Sundays.

Shift will be overstaffed on Thursdays.

Cannot meet staffing requirements under existing union contract with fewer than 18 workers.

b. (10 points) Develop an optimization model to determine how many workers will be needed to satisfy the staffing requirements specified in the first table if the hospital agrees to the union’s demand for days off being consecutive. (Hint: if the days off must be consecutive, there are only 7 possibilities for assigning days off to a worker – Sat& Sun, Sun & Mon, etc.) Define the decision variables, and express the objective function and constraints in terms of the decision variables.

Decision Variables:

Number of workers taking off on consecutive days.

Define: d1,d2,d3,d4,d5,d6,d7 represent the 7 respective days

A,B,C,D,E,F,G represents the number of workers taking day off on d1,d2,d3,d4,d5,d6,d7 days respectively.

Objective Function

Minimize the total workers

MIN(A\*d1+B\*d2+C\*d3+D\*d4+E\*d5+F\*d6+G\*d7)

Constraints

Minimum staffing requirements should be met

Every worker should get a 2 consecutive weekly day offs

C\*d3+D\*d4+E\*d5+F\*d6+G\*d7>=16

A\*d1+D\*d4+E\*d5+F\*d6+G\*d7>=11

A\*d1+B\*d2 +E\*d5+F\*d6+G\*d7>=12

A\*d1+B\*d2+C\*d3+ F\*d6+G\*d7>=12

A\*d1+B\*d2+C\*d3+D\*d4+G\*d7>=13

A\*d1+B\*d2+C\*d3+D\*d4+E\*d5>=10

B\*d2+C\*d3+D\*d4+E\*d5+F\*d6>=16

c. (5 points) Solve your model in Excel using the solver. How many workers will be needed, will 18 be enough or will Bob need to add additional workers and if so, how many? Which days (if any) will be over-staffed (provided Bob adds only those workers that are needed to fill the staffing requirements) and if so, which ones and by how much? What will be the assignments of days off to the various workers?

Bob will need to add 1 additional worker, making a total of 19 workers.

Overstaffed on Thursday with 5 additional workers.

Days Off:

Friday-Saturday: 0 workers

Saturday-Sunday: 3 workers

Sunday-Monday: 5 workers

Monday-Tuesday: 2 workers

Tuesday-Wednesday: 5 workers

Wednesday-Thursday: 1 worker

Thursday-Friday: 3 workers

Excel Solver:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | sat | sun | mon | tue | wed | thu | fri |  |  |  |  |  |  |
| Workers | 0 | 3 | 5 | 2 | 5 | 1 | 3 |  |  | Total |  |  |  |
|  | 1 | 1 | 1 | 1 | 1 | 1 | 1 |  |  | 19 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Days |  |  |  |  |  |  |  |  | Require | Working | Addl. |  | Days Off |
| Sat | 0 | 0 | 1 | 1 | 1 | 1 | 1 |  | 16 | 16 | 0 |  | 3 |
| Sun | 1 | 0 | 0 | 1 | 1 | 1 | 1 |  | 11 | 11 | 0 |  | 8 |
| Mon | 1 | 1 | 0 | 0 | 1 | 1 | 1 |  | 12 | 12 | 0 |  | 7 |
| Tue | 1 | 1 | 1 | 0 | 0 | 1 | 1 |  | 12 | 12 | 0 |  | 7 |
| Wed | 1 | 1 | 1 | 1 | 0 | 0 | 1 |  | 13 | 13 | 0 |  | 6 |
| Thu | 1 | 1 | 1 | 1 | 1 | 0 | 0 |  | 10 | 15 | 5 |  | 4 |
| Fri | 0 | 1 | 1 | 1 | 1 | 1 | 0 |  | 16 | 16 | 0 |  | 3 |

2. **Nurse Scheduling revisited** (10 points).

* 1. (5 points) In part a of this problem staffing must minimize the total number of nurses required and nurses must be given two consecutive days off as in the previous problem. If it will be impossible to avoid overstaffing, management wishes to spread the overstaffing as evenly as possible throughout the week. That is, imagine a solution meeting the staffing requirements as well as the union contract requirements with the minimum number of nurses (eg. it solves your LP from the previous problem) that had (for the sake of an example) 17 nurses assigned to Sunday and exactly met the staffing requirements for other days so that there were exactly 7 units of overstaffing throughout the week. Imagine another solution that optimally solves your LP from the previous problem that had (for the sake of an example) exactly one unit of overstaffing each day so that again there were a total of 7 units of overstaffing throughout the week. The second solution would be preferred by management since it spreads the overstaffing evenly. Develop a new LP to spread overstaffing as evenly as possible while using the same number of nurses you determined in problem 1 and while still meeting the union’s contract requirement for consecutive days off. Hint: your objective can be to minimize the largest overstaffing during the week.

Excel Solver:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | sat | sun | mon | tue | wed | thu | fri | z |  |  |  |  |  |
| Workers | 0 | 3 | 3 | 4 | 1 | 5 | 3 | 2 |  | Total |  |  |  |
|  | 1 | 1 | 1 | 1 | 1 | 1 | 1 |  |  | 19 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Days |  |  |  |  |  |  |  |  | Require | Working | Addl. | Z | Days Off |
| sat | 0 | 0 | 1 | 1 | 1 | 1 | 1 |  | 16 | 16 | 0 | 2 | 3 |
| sun | 1 | 0 | 0 | 1 | 1 | 1 | 1 |  | 11 | 13 | 2 | 2 | 6 |
| mon | 1 | 1 | 0 | 0 | 1 | 1 | 1 |  | 12 | 12 | 0 | 2 | 7 |
| tue | 1 | 1 | 1 | 0 | 0 | 1 | 1 |  | 12 | 14 | 2 | 2 | 5 |
| wed | 1 | 1 | 1 | 1 | 0 | 0 | 1 |  | 13 | 13 | 0 | 2 | 6 |
| thu | 1 | 1 | 1 | 1 | 1 | 0 | 0 |  | 10 | 11 | 1 | 2 | 8 |
| fri | 0 | 1 | 1 | 1 | 1 | 1 | 0 |  | 16 | 16 | 0 | 2 | 3 |

* 1. (5 points) For part b of the problem, imagine that the union comes back next year with a demand that the all nurses not only receive consecutive days off, but that they get either Friday/Saturday off or Saturday/Sunday off or Sunday/Monday off. How many nurses will it take to meet this union demand? Note that management will still want to spread any necessary overstaffing as evenly as possible.

A total of 28 nurses will be required:

Excel Solver:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | sat | sun | mon | fri |  |  |  |  |  |  |
| Workers | 12 | 0 | 16 | 0 | 18 |  | Total |  |  |  |
|  | 1 | 1 | 1 | 1 |  |  | 28 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Days |  |  |  |  |  | Require | Working | Addl. | g | Days Off |
| Sat | 0 | 0 | 1 | 0 |  | 16 | 16 | 0 | 18 | 12 |
| Sun | 1 | 0 | 0 | 1 |  | 11 | 12 | 1 | 18 | 16 |
| Mon | 1 | 1 | 0 | 1 |  | 12 | 12 | 0 | 18 | 16 |
| Tue | 1 | 1 | 1 | 1 |  | 12 | 28 | 16 | 18 | 0 |
| Wed | 1 | 1 | 1 | 1 |  | 13 | 28 | 15 | 18 | 0 |
| Thu | 1 | 1 | 1 | 1 |  | 10 | 28 | 18 | 18 | 0 |
| Fri | 0 | 1 | 1 | 0 |  | 16 | 16 | 0 | 18 | 12 |

3. (10 points) **Pollution.** Your manufacturing process must release effluent into a neighboring river. To make sure there is no pollutant, you run the effluent through a purification process before releasing it. If the resulting effluent contains more than 0.2 ppm of pollutant, the Iowa DNR and the EPA will come in and shut you down. To make sure this doesn’t happen, you store the effluent in storage tanks and measure its pollution content before releasing it to the river. Due to inherent variation in the measuring process, you can’t rely on a single sample to provide a good measurement. Accordingly, each time you sample, you take a set of 10 samples, measure the pollution content for each sample, and perform a statistical analysis. You will release the stored effluent only if the statistical analysis indicates it is safe to do so. If the effluent does not pass the test, it must be run through the purification process again.

1. (5 points) What is the appropriate null hypothesis:

Mean of the sample >= 0.2 ppm

1. Suppose your sample mean is 0.15, and your sample standard deviation is 0.025. At the 95% level, should you release the stored effluent?

Test statistic = (sample mean – hypothesis mean) /(std dev /SQRT(n))

= -6.32456

T-test = -T.INV(0.95,9) = -1.833113

We reject hypothesis since test statistic is lower than 95% level, using T-test. Conversely, it is safe to release the stored effluent.

4. (20 points) **Cereal Manufacturing**. A manufacturer of breakfast cereals manufactures and markets a box of cereal with a nominal net weight of 500 grams. After the cereal flakes are manufactured, they are dispensed into boxes. Data listing actual weights of cereal dispensed when the machinery is set up to dispense an average nominal weight of 500 grams is given in the cereal dispensing worksheet in the accompanying excel spreadsheet.

1. (5 points) Construct a histogram of the sampled weights and compare the actual number of weights in each “bin” of the histogram to the expected number in each “bin” if the weights actually are sampled from a normally distributed population with mean and standard deviation equal to your sample mean and sample standard deviation. In constructing the histogram, use seven bins 480 – 485, 485-490, …,510-515. Conduct an hypothesis test for the null hypothesis that the weights are sampled from a normally distributed population.

Histogram

|  |  |
| --- | --- |
| **Row Labels** | **Count of Actual Weight** |
| 485-490 | 1 |
| 490-495 | 2 |
| 495-500 | 11 |
| 500-505 | 31 |
| 505-510 | 35 |
| 510-515 | 15 |
| 515-520 | 5 |
| **Grand Total** | **100** |

Null hypothesis: Mean of the weights = 500

Sample mean = 505.58

Sample Std Dev = 5.33

Samples within 1 Std Dev. = 70 (~68%)

Samples within 2 Std Dev. = 95 (~95%)

Samples within 3 Std Dev. = 100 (~99%)

The above numbers prove that the data from the sample is normally distributed

1. (2.5 points) Compute the proportion of sampled weights that actually exceed 500 grams. This is your sample proportion. Conduct an hypothesis test for the null hypothesis that the true population proportion is greater than or equal to 75%. Use a 95% confidence level.

The proportion of sampled weights that exceed 500 grams is 86%

True proportion = 75%

Sample proportion p=0.86

The sample stdev is sqrt[0.86 x (1 – 0.86)]

= sqrt[0.86 x 0.14] = 0.346987031

Stdev of Sampling Distribution = std. error

= 0.346987031/sqrt(100)

= 0.034698703

A 95% confidence interval for the sample proportion is given by:

[0.86-normsinv(.975)x0.034698703, 0.86+normsinv(.975)x0.034698703]

= [0.791991792,0.928008208]

The true population proportion is greater than 75%. Failed to reject the hypothesis.

1. (2.5 points) Construct a 95% confidence interval about your estimate of the true population proportion.

A 95% confidence interval for the true population proportion is

[0.75-normsinv(.975)x0.034698703, 0.75+normsinv(.975)x0.034698703]

= [0.681991792,0.818008208]

**Regardless of your answers to parts a through b above, assume for the remaining parts of this question that:**

* + 1. **The cereal dispensing machinery can be set to dispense any nominal weight between 50 and 1000 grams.**
    2. **Further, when the machinery is set to dispense a nominal weight equal to W, the actual weight dispensed will be a random variable that is normally distributed with a mean of W and a *coefficient of variance* = 0.01. The *coefficient of variance* is simply the standard deviation divided by the mean, so if the nominal weight were set to equal 500 grams, the standard deviation would be equal to 0.01 x 500 = 5 grams.**
  1. (5 points) Suppose that government fair trade regulations require that a box labeled as having a nominal weight of 500 grams must have a probability of at least 99% of containing 500 grams or more.
     1. If the company sets the nominal weight setting of its dispensing machinery to 510 grams, what is the probability that a randomly chosen box of cereal will actually contain 500 grams or more?

Mean=510

Std dev=0.01\*510=5.1

P(X>=500) = 1- P(X<=500)

=1-NORMDIST(500,510,5.1,1)

= 0.975047906 or 97.5%

* + 1. How high would you have to set the nominal weight to guarantee that your company meets the 99% standard? Choose the lowest nominal weight that will meet the standard. If you do so, what would be the average or expected excess weight in each box (excess = actual weight in grams – 500 grams)?

If company sets 512 grams as nominal weight it will meet the standard of 99%

1-NORMDIST(500,512,5.12,1)= 0.990454518 or 99%

Excess=512-500=12 grams.

* 1. (5 points) Suppose the manufacturer decides to fill the boxes by first dispensing two separate batches, each with a nominal weight setting of 253 grams and then adding the two separate batches into a single box.
     1. What is the probability that the box will contain at least 500 grams of cereal?

Total Mean=Box1 Mean + Box2 Mean= 253+253=506

Total Stdev= SQRT ((Box1 Stdev) ^2 + (Box1 Stdev)^2)

=SQRT ((253\*0.01)^2+(253\*0.01)^2)

= 3.577960313

P(X>=500) = 1- P(X<=500)

=1-NORMDIST(500,506, 3.577960313,1)

=0.953222216 or 95.3%

* + 1. What will be the average or expected excess weight in each box?

Expected excess weight is 6 grams.

* + 1. How large do you have to set the nominal weight (253 grams for parts i and ii) to make sure that the probability of the box containing at least 500 grams is 99% and what will the average excess in each box be? Is there any advantage to using separate batches as in part e?

To meet 99% standard the stdev in each box should be P(X>=500)=1-NORMDIST(500,255\*2,SQRT((255\*0.01)^2+(255\*0.01)^2),1)

= 0.997222619 or 99.72% (approximately)

So the nominal weight required for each box is 255 grams.

The average excess weight = 2\*255-500=10 grams.

There is an advantage in using separate batches.

The advantage is 2 grams.

5. (10 points) **Artsy Regression Analysis.**  Use the Artsy Regression data from the exam 3 data file..

In homework 8 we ran the original regression analysis in which rate is the y-variable and grade, sex, and tingrade are the explanatory variables. We used pivot table to group the residuals by grade and conducted t-tests to determine whether or not there is any systematic misprediction by grade and found that there were problems with grades 3, 5, and 8.

1. (5 points) Add an additional “male grade 6” column to your data in which each individual is given a “1” if they are both male and grade 6. Otherwise give the individual a “0”. Do the same for “male grade 7” and “male grade 8”. Rerun the regression, now using grade, male grade 6, male grade 7, male grade 8, and tingrade (but NOT sex!) as the explanatory variables.
   1. Redo the residual analysis performed in part a to see whether or not systematic mispredictions by grade still exist.

We still see systematic mispredictions by grade.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| GRADE | Data | Total | SE | Tstat | Cvalue | Reject? |
| 1 | Average of Residuals | 0.827249145 |  |  |  |  |
|  | StdDev of Residuals | 26.34229123 | 5.492747 | 0.150608 | 2.073873 | N |
|  | Count of Residuals | 23 |  |  |  |  |
| 2 | Average of Residuals | -1.265601109 |  |  |  |  |
|  | StdDev of Residuals | 17.85352959 | 2.499994 | -0.50624 | 2.008559 | N |
|  | Count of Residuals | 51 |  |  |  |  |
| 3 | Average of Residuals | 11.75806954 |  |  |  |  |
|  | StdDev of Residuals | 15.52899986 | 2.789091 | 4.215736 | 2.042272 | Y |
|  | Count of Residuals | 31 |  |  |  |  |
| 4 | Average of Residuals | -5.663343564 |  |  |  |  |
|  | StdDev of Residuals | 27.7461784 | 5.785478 | -0.97889 | 2.073873 | N |
|  | Count of Residuals | 23 |  |  |  |  |
| 5 | Average of Residuals | -11.16042361 |  |  |  |  |
|  | StdDev of Residuals | 30.11910714 | 5.091058 | -2.19216 | 2.032245 | Y |
|  | Count of Residuals | 35 |  |  |  |  |
| 6 | Average of Residuals | 1.53669609 |  |  |  |  |
|  | StdDev of Residuals | 32.36172906 | 6.472346 | 0.237425 | 2.063899 | N |
|  | Count of Residuals | 25 |  |  |  |  |
| 7 | Average of Residuals | 0.952466596 |  |  |  |  |
|  | StdDev of Residuals | 37.74601345 | 5.338092 | 0.178428 | 2.009575 | N |
|  | Count of Residuals | 50 |  |  |  |  |
| 8 | Average of Residuals | 6.436098138 |  |  |  |  |
|  | StdDev of Residuals | 42.33941802 | 9.979497 | 0.644932 | 2.109816 | N |
|  | Count of Residuals | 18 |  |  |  |  |

* 1. Also, perform the analysis necessary to answer the question*: is there any misprediction by sex*? Look at both average residual within each sex and variance within each sex.

There isn’t any mispredictions by sex.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SEX | Data | Total | SE | Tstat | Cvalue | Reject? |
| 0 | Average of Residuals | 0.031189494 |  |  |  |  |
|  | Var of Residuals | 643.4056132 | 1.939744 | 0.016079 | 1.974017 | N |
|  | Count of Residuals | 171 |  |  |  |  |
| 1 | Average of Residuals | -0.062745924 |  |  |  |  |
|  | Var of Residuals | 1340.379676 | 3.971042 | -0.0158 | 1.98861 | N |
|  | Count of Residuals | 85 |  |  |  |  |

1. (5 points) Recalling the answer to the homework 8 question, it might make sense to try adding variables that might address the problems in grades 3, 5, and 8. Try re-running the regression, but this time using grade, tingrade, male grade 3, malegrade 5, and male grade 8 as explanatory variables. Which of the two models (this one or the one considered in part a) is better? Support your conclusion with appropriate analysis and explanation.
2. ~~(5 points) Try fitting various polynomial models (as in HW 8) to the data and come up with the best model. Explain your choice.~~

6. (20 points) **Analyzing Artsy**. Use the Artsy data from the Exam 3 data file.

1. (5 points) Is there any significant difference between male and female average pay rates in grade 5?
2. (5 points) Is there any significant difference between the variance of male pay rates and the variance of female pay rates in grade 5?
3. (5 points) Analyze the null hypothesis that the variance of salaries in grade 2 is greater than or equal to 370.
4. (2.5 points) Analyze the null hypothesis that the average pay rate is independent of sex.
5. (2.5 points) Analyze the null hypothesis that the number of employees in each grade is independent of sex.

**Use the appropriate statistical tools and explain your results. You need to use the appropriate tool(s) to receive credit for this problem.**

7. (10 points) **A Dice Problem.** Use the Dice data from the Exam 3 data file.

Management at the Bar-None Casino are concerned that a recent shipment of dice they have received are a bit fishy. They wish to conduct a test of the hypothesis that the numbers generated by throws of a single die are uniformly distributed.

Develop two different methodologies for conducting a statistical test of this hypothesis and carry out the tests. Explain specifically what each test is and how it will work. Report the p-value for both tests and whether or not the tests reach identical conclusions.