Case 1: Employee Retention at D&Y\*

Due Feb 27

Demand for systems analysts in the consulting industry is very strong. Graduates with experience in the consulting business and those who have extensive computer knowledge are getting great offers from consulting companies. Once these people are hired, they frequently switch from one company to another as competing companies lure them away with even better offers. One consulting company, D&Y, has collected data on a sample of system analysts they hired with an undergraduate degree several years ago. The data are in the attached file.

StartSal: Employee’s starting salary at D&Y.

OnRoadPct: Percentage of time employee has spent on the road with clients.

StateU: Whether the employees graduated from the State University.

CISDegree: Whether the employee majored in computer Information Systems or a similar program.

Stayed3Yrs: Whether the employee stayed at least 3 years.

Tenure: Tenure of employee at D&Y (months) if he or she moved before 3 years.

D&Y is trying to learn everything it can about retention of these valuable employees. You can help by solving the following problems and then, based on your analysis, presenting a report to D&Y.

1. Executive Summary
2. Problems:

1. Although starting salaries are in a fairly narrow band, D&Y wonders whether they have anything to do with retention.

**a.** Find the 95% confidence interval for the mean starting salary of all employees who stay at least 3 years with D&Y.

**One-Sample T: StartSal**

Variable N Mean StDev SE Mean 95% CI

StartSal 32 37381 1806 319 (36730, 38032)

95% confident that the population mean will fall between 36,730 and 38,032.

**b.** Do the same for those who leave before 3 years.

**One-Sample T: StartSal**

Variable N Mean StDev SE Mean 95% CI

StartSal 34 38556 1887 324 (37897, 39214)

95% confident that the population mean will fall between 37,897 and 39,214.

**c.**. Management believes that, on average, those employees that left before 3 years might be getting paid less than those that stayed beyond 3 years. Although there might be a point estimate difference, is there enough evidence in this data to support this theory? Perform a hypothesis test at the 10% level of significance.

**Two-Sample T-Test and CI: StartSal, StartSal\_1**

Two-sample T for StartSal vs StartSal\_1

N Mean StDev SE Mean

StartSal 34 38556 1887 324

StartSal\_1 32 37381 1806 319

Difference = mu (StartSal) - mu (StartSal\_1)

Estimate for difference: 1175

90% upper bound for difference: 1763

T-Test of difference = 0 (vs <): T-Value = 2.58 P-Value = 0.994 DF = 63

Ho mu employees left before 3yrs >= 37381 mu employees that stayed 3yrs

Ha mu employees left before 3yrs < 37381 mu employees that stayed 3yrs

P-Value = 0.994 > .10 (alpha)We would fail to reject Ho and reject Ha.

**d.** Among all employees whose starting salary is below the median ($37,750), find a 95% confidence interval for the proportion who stay with D&Y for at least 3 years.

The analysis actually shows the opposite of the theory. Employees making less have stayed longer.

**Results for: Worksheet 6**

**Test and CI for One Proportion: Stayed3Yrs**

Event = Yes

Variable X N Sample p 95% CI

Stayed3Yrs 20 33 0.606061 (0.439350, 0.772771)

Using the normal approximation.

95% confident that the population mean will fall between 0.439350 and 0.772771.

**e.** Find a 95% confidence interval for the proportion who stay with D&Y for at least 3 years but this time with starting salaries above the median.

**Results for: Worksheet 6**

**Test and CI for One Proportion: Stayed3Yrs\_1**

Event = Yes

Variable X N Sample p 95% CI

Stayed3Yrs\_1 12 33 0.363636 (0.199510, 0.527762)

Using the normal approximation.

95% confident that the population mean will fall between 0.199510 and 0.527762.

**f.** Find a 95% confidence interval for the difference between these proportions. From this interval, can we conclude that there is a significant difference in the proportions at the 5% level of significance?

**Results for: Worksheet 6**

**Test and CI for Two Proportions: Stayed3Yrs, Stayed3Yrs\_1**

Event = Yes

Variable X N Sample p

Stayed3Yrs 20 33 0.606061

Stayed3Yrs\_1 12 33 0.363636

Difference = p (Stayed3Yrs) - p (Stayed3Yrs\_1)

Estimate for difference: 0.242424

95% CI for difference: (0.00848011, 0.476368)

Test for difference = 0 (vs not = 0): Z = 2.03 P-Value = 0.042

Fisher's exact test: P-Value = 0.084

Ho mu stay 3yr below median = mu stay 3yr above median

Ha mu stay 3yr below median not = mu stay 3yr above median

Since P-Value = 0.042 < .05 (alpha) we reject the null hypothesis. We can conclude that there is a statistical difference.

2. D&Y wonders whether the percentage of time on the road might influence who stays and who leaves.

**a.** Find the 95% confidence interval for the percentage of time employees stay on the road for those that stay at least 3 years with D&Y.

**One-Sample T: OnRoadPct\_1**

Variable N Mean StDev SE Mean 95% CI

OnRoadPct\_1 32 0.4503 0.1665 0.0294 (0.3903, 0.5104)

95% confident that the population mean will fall between 0.3903 and 0.5104.

**b.** Find the 95% confidence interval for the percentage of time employees stay on the road for those that leave before 3 years.

**One-Sample T: OnRoadPct**

Variable N Mean StDev SE Mean 95% CI

OnRoadPct 34 0.5868 0.1652 0.0283 (0.5291, 0.6444)

95% confident that the population mean will fall between 0.5291 and 0.6444.

**c.** Can we conclude that there is a statistical difference between the mean percentages of time on the road of those that stayed at least 3 years and those that did not? Test at the 5% level of significance.

**Results for: Worksheet 1**

**Two-Sample T-Test and CI: OnRoadPct, OnRoadPct\_1**

Two-sample T for OnRoadPct vs OnRoadPct\_1

N Mean StDev SE Mean

OnRoadPct 34 0.587 0.165 0.028 leave before

OnRoadPct\_1 32 0.450 0.167 0.029 stay

Difference = mu (OnRoadPct) - mu (OnRoadPct\_1)

Estimate for difference: 0.1365

95% CI for difference: (0.0548, 0.2181)

T-Test of difference = 0 (vs not =): T-Value = 3.34 P-Value = 0.001 DF = 63

Ho mu1 stay 3 years road time = mu2 did not stay 3 years road time

Ha mu1 stay 3 years road time not = mu2 did not stay 3 years road time

P-Value = 0.001 < .05 (alpha) we reject the null hypothesis.

**d.** Among all employees who travel less than the median (54%), find the 95% confidence interval for the proportion who stay with D&Y for at least 3 years.

**Test and CI for One Proportion: Stayed3Yrs**

Event = Yes

Variable X N Sample p 95% CI

Stayed3Yrs 21 33 0.636364 (0.472238, 0.800490)

Using the normal approximation.

95% confident that the population mean will fall between 0.472238 and 0.800490.

**e.** Do the same for the employees that travel more than the median who have stayed beyond 3 years.

**Test and CI for One Proportion: Stayed3Yrs\_1**

Event = Yes

Variable X N Sample p 95% CI

Stayed3Yrs\_1 11 33 0.333333 (0.172497, 0.494170)

95% confident that the population mean will fall between 0.172497 and 0.494170.

f. Then find the 95% confidence interval for the difference between proportions. From this confidence interval, can we conclude that there is a significant difference in the proportions?

**Test and CI for Two Proportions: Stayed3Yrs, Stayed3Yrs\_1**

Event = Yes

Variable X N Sample p

Stayed3Yrs 21 33 0.636364

Stayed3Yrs\_1 11 33 0.333333

Difference = p (Stayed3Yrs) - p (Stayed3Yrs\_1)

Estimate for difference: 0.303030

95% CI for difference: (0.0732353, 0.532825)

Test for difference = 0 (vs not = 0): Z = 2.58 P-Value = 0.010 < .05a

Fisher's exact test: P-Value = 0.026

Ho = 3yremployeeyes = 3yremployeeno

Ho = 3yremployeeyes not = 3yremployeeno

Reject Ho because P is less than alpha (.05). There is a significant difference in the number of employees that travel and stay with the company.

3. The state university students have traditionally been among the best of D&Y’s recruits. The managers at D&Y believe that those folks that attended the state university are less likely to move away to another company. Is there enough evidence to conclude that this is true? Test at the 5% level of significance.

4. D&Y also collected data on whether or not those employees that have CIS degrees are more likely to leave before 3 yrs.

a. Perform a hypothesis test to see if there is a statistically significant difference in those that have CIS degrees are more likely to leave before 3 years and those that don’t have CIS degrees. Test at the 5% level of significance.

b. Also, test to see if there is a statistically significant difference in salary between those that have a CIS degree and those that don’t. Test at the 5% level of significance.

5. Given the analyses above, what other analysis could you do that might be helpful with this data? What do you think your conclusions might be from that analysis?

* \* This case was adapted from a case authored by Albright, Winston and Zapp, ‘Data Analysis and Decision Making’, 2006

**Case Deliverables:**

Please submit an Executive Summary of the findings from your data analysis. Also, in the executive summary, provide D&Y any advice that you can based on the data that you analyzed. Don’t base your advice on your feelings, but purely on the suggestions and conclusions drawn from the data.

Underneath your executive summary, you should have a section for each of the six areas that you evaluate.

1. Analysis of the variation in starting salary and its impact on retention.

2. Analysis of the percentage of time of being on the road and its impact on retention.

3. Analysis of graduates from the state university and its impact on retention.

4. Analysis of those with CIS degrees and its impact on retention.

5. What other analyses do you think might be helpful? What conclusions can you draw?

For every confidence interval that you calculate, **give your interpretation of the confidence interval.** Also, for every hypothesis test that you perform, **clearly state the null and alternative hypotheses, and interpret your findings**. You can use the p-value approach or the critical value approach. Obviously, the p-value approach is the easiest given the alpha levels that I have specified.

My intent is for you do all of the analysis in MINITAB. You will have to sort the data numerous times based on different variables, so you will need to learn this function. Of course you can do this in excel and copy it over to MINITAB if you wish. Also, you will need to dissect the data (cut and paste) because MINITAB will not look at a partial column.

Most, if not all, of your analysis work will come in the STAT drop down. Go to STAT>Basic Stats and then go to 1-sample t, 2-sample t, 1-sample proportion, and 2-sample proportion. You can generate a confidence interval and perform a hypothesis test in each of these functions.

In addition to your case write-up, I would like for you to submit your MINITAB work as well. Please perform and submit all of your data analysis work in one project file. You may have multiple worksheets within the project file where you sorted and dissected the data. For example, I had 4 different worksheets within my project file.

I expect that each team member will contribute equally to the case analysis and write-up, but in the event that one team member does more than 50% of the work, please specify that on your write-up. I will make grade assignments accordingly.

Also, although I expect you to work with your team member, you are prohibited from discussing the case or your solutions with someone from another team.

Let me know if you have any questions.

Here are some helpful hints in arranging the data:

1. You will have to divide the data based on certain variables and then perform a 2-sample hypothesis test. The easiest way to do this is to first sort the data according to the variable of interest. In the first problem that variable is “Stayed3Yrs”. Sorting in Minitab is different than sorting in excel. Here are the commands: Data>Sort>Select which columns you want to sort>then select which variable you want to sort on>then select where you want to place the sorted data>OK. I usually sort the data and put the data back in the same columns of the same worksheet.
2. Once you have sorted the data, you will need to divide the data. Do this by cutting and pasting the lower half of the data to another set of columns on the same spreadsheet. If the data is continuous, I usually divide the data based on the mean. As a matter of consistency, if there are multiple values for that point, put all of the mean values with the group with the smaller numbers. Once you have done this, you can easily do the 2-sample test by comparing the data in one column to the data in another column. I think I do this in the tutorial, so if you have any questions, refer to that DVD, or feel free to give me a call.

Let me know if you have any questions.