Intermediate Statistics: Final Project

DS105

Scenario 1:

Fawn is a private investigator. She has taken a job for an insurance company. The insurance company has been paying out large amounts of money for workman's comp claims, and they believe some of those claims are fraudulent. The insurance company wants Fawn and her associates to run surveillance to determine if the rate of fraud is higher than what the industry normally reports. Specifically, they want to know if the incidence of fraud among claimants is more than 16%.

They provide Fawn with a random sample of 94 claimants who have been diagnosed as unable to perform any physical labor beyond light housekeeping activities. They have asked Fawn to observe the sample, and report back to them how many of the claimants are obviously doing things that are much more strenuous than light housekeeping, such as strenuous yard work, weightlifting at the gym or other strenuous sporting activities, etc.

Fawn and her team spend about 8 weeks completing their observations, and report back to the insurance company that 28 of the 94 claimants are not nearly as "disabled" as their diagnosis suggests. The insurance company wants to test the data.

Identify the statistic to be run:

I will run a one proportion z test as I am comparing one thing to a whole and 1 variable is categorical

Test the appropriate assumptions

There was no necessary assumptions to test for this statistic

Conclusion

The rate of fraud in Fawn's investigation is 30%, that is nearly twice the expected fraud rate of 16%.

Scenario 2:

Medical researchers are trying to understand if four topical antiseptics are being used in the same ratio at three different clinics in town. They have access to medical records over the past 3 years, and have recorded each treatment where a topical antiseptic was used.

Using a tally sheet, they have determined how many times each antiseptic was used in each of the three clinics, and they want to compare the antiseptic usage across the three clinics.

Identify the statistic to be run:

I will run a independent chi square test, because the independent and dependent variable are categorical and I'm looking to compare frequencies by category.

Test the appropriate assumptions

The variables are independent and the expected frequencies are greater than 5 in each cell

Conclusion

There was not a significant difference in antiseptic usage between the clinics.

Scenario 3:

A financial institution is interested in the savings practices of different demographic groups. They have demographic data for all of their account holders, and have used those criteria to split their customers up into 4 groups.

They are going to use the results to do some targeted marketing. In order to determine savings practices, they are going to use the average savings account balance over the past 3 months for their account holders. In other words, they will have one number (average account balance) for each account.

Each demographic group has between 40 and 60 accounts they will look at.

Identify the statistic to be run:

I will run a one way ANOVA , because I am comparing means of the categories with the IV being categorical and DV being continuous.

Test the appropriate assumptions

The DV distribution was confirmed to be normal with a look at a histogram.

The bartlett test was then used to confirm homogeneity of variance

Sample size was confirmed to have at least 20 per IV

There is a assumption of independence that is assumed

Conclusion

There is a difference between the demographics and their average savings balance.

	Group	Mean
1	Group.A	23401.849
2	Group.B	18566.308
3	Group.C	9227.457
4	Group.D	15017.776

Scenario 4:

The local school board conducted a poll to gauge public sentiment about a school bond. They asked respondents if they favored or opposed a bond in the upcoming election. The respondents were asked some demographic questions, too.

With school age children and favorable - 374

With school age children and unfavorable - 129

Without school age children and favorable - 171

Without school age children and unfavorable - 74

Identify the statistic to be run:

I will run a 2 proportion z test, because I'm comparing ratio to whole and there are 2 categorical components

Test the appropriate assumptions

The DV distribution was confirmed to be normal with a look at a histogram.

The bartlett test was then used to confirm homogeneity of variance

Sample size was confirmed to have at least 20 per IV

There is a assumption of independence that is assumed

Conclusion

There is a significant difference in sentiment between voters with school age children compared to those without school age children