Practical for Session 4 Hypothesis Testing

01/11/2022

1 Inference about a proportion

Out of 80 women in a random sample of women in Manchester, 13 were asthmatic; this could be used to calculate a 95% confidence interval for the proportion of women in Manchester with asthma. This confidence interval could be compared to the suggested prevalence of 20% in Northern England. An alternative approach would be to test the hypothesis that the true proportion, π, is 0.20.

1.1   What is the expected proportion of women with asthma under the null hypothesis ?

......

1.2   What is the observed proportion of women with asthma ?   ......

1.3   What is the standard error of the expected proportion (remember from last week that the standard error of a proportion p is given by

$$\sqrt{\frac{p(1-p)}{n}}$$

......

1.4   The appropriate test statistic, T, is given by the formula:

$$\frac{\text{observed proportion - expected proportion}}{\text{standard error of proportion}}$$

      Calculate T.   ......

1.5 T should be compared to a t-distribution with how many degrees of freedom ?

......

1.6   From tables for the appropriate t-distribution, the corresponding p-value is 0.4. Is is reasonable to suppose that these women are a random sample from a population

      in which the prevalence of asthma is 20% ?   ......

2 More inference about a proportion

In the sample heights and weights we have looked at, there were 412 individuals of whom 234 were women. We wish to test that there are equal numbers of men and women in our population.

2.1   What is the null hypothesis proportion of women ?   ......

2.2   What is the observed proportion of women ?   . . . . . .

2.3   What is the null hypothesis standard error for the proportion of women ? . . . . . .

2.4   What is an appropriate statistic for testing the null hypothesis ?   . . . . . .

3   Inference about a mean

Load `htwt.dta` into stata with the commands (each command needs to be entered on a separate line).

```
 global datadir http://personalpages.manchester.ac.uk/staff/mark.lunt/stats
use $datadir/2_summarizing_data/data/htwt.dta
```

 We wish to test whether the mean height is the same in men and women.

3.1   What is the null hypothesis difference in height between men and women ?

        . . . .

3.2   Use the command `ttest nurseht, by(sex)` to test whether the mean height differs between men and women.

3.3   What is the mean height in men ?   . . . . . .

3.4   What is the mean height in women ?   . . . . . .

3.5   What is the mean difference in height between men and women, with its 95% confidence interval ?

        . . . .

3.6   Which of the three hypothesis tests is the appropriate one in this instance ?

        . . . .

4

3.7   What is the p-value from the t-test ?   . . . . . .

3.8   What would you conclude ?

....

4   Two-sample t-test

Compare BMI (based on the measured values, i.e. `bmi`) between men and women in `htwt.dta`, using the command `ttest bmi, by(sex)`.

4.1   Is there a difference in BMI between men and women ?   . . . . . .

4.2   What is the mean difference in BMI between men and women and its 95% confidence interval.

....

4.3   Is there a difference in the standard deviation of BMI between men and women ? (This can be tested with the command `sdtest bmi, by(sex)`

....

4.4   If there is, repeat the t-test you performed above, using the `unequal` option. Are your conclusions any different ?

....

5   One sample t-test

Load the `bpwide` dataset into stata with the command `sysuse bpwide`. This consists of fiction blood pressure data, taken before and after an intervention. We wish to determine whether the intervention had affected the blood pressure.

5.1   Use the `summarize` command to calculate the mean blood pressure before and after the intervention. Has the blood pressure increased or decreased ?

....

5.2   Generate a variable containing the change in blood pressure using the command `gen bp_diff = bp_after - bp_before`

5.3   Use the command `ttest bp_diff = 0` to test whether the change in blood pressure is statistically significant. Is it ?

....

5.4   Give a 95% confidence interval for the change in blood pressure.

....

6 Power Calculations

The following questions can all be answered using the `sampsi` command.

6.1   How many subjects would need to be recruited to have 90% power to detect a difference between unexposed and exposed subjects if the prevalence of the condition is 25% in the unexposed and 40% in the exposed, assuming equal numbers of exposed and unexposed subjects ?

....

6.2   If the exposure was rare, so it was decided to recruit twice as many unexposed subjects as exposed subjects, how many subjects would need to be recruited ?

....

6.3   Suppose it were only possible to recruit 100 subjects in each group. What power would the study then have ?   ......

6.4   Suppose that we expect a variable to have a mean of 15 and an SD of 5 in group 1, and a mean of 17 and an SD of 6 in group 2. How large would two equal sized groups need to be to have 90% power to detect a difference between the groups ?

....

6.5   If we wanted 95% power, how large would the groups have to be ?

....

6.6   Suppose we could only recruit 100 subjects in group 1. How many subjects would
         we have to recruit from group 2 to have 90% power ?

....

Hint: the last question can only be answered by trying different numbers for the size
of group 2 and seeing what power is achieved. Sensible choice of numbers will give a result
fairly quickly. The `PageUp` key is your friend.