

Week3-Basic Tests

Matthew Onimus

9/22/2020

Contents

Read in Data	1
Question 1	1
Question/Answer 1a	1
Question/Answer 1b	2
Question/Answer 1c	2
Question/Answer 1d	2
Question/Answer 1e	2
Question 2	2
Question/Answer 2a	2
Question/Answer 2b	2
Question/Answer 2c	3
Question/Answer 2d	3
Question/Answer 2e	3
Question 3	3
Question/Answer 3a	3
Question/Answer 3b	3
Question/Answer 3c	3
Question/Answer 3d	3
Question/Answer 3e	4
Question 4	4
Question/Answer 4a	4
Question/Answer 4b	4
Question/Answer 4c	4
Question/Answer 4d	4
Question?Ansewer 4e	4

Read in Data

Question 1

The HIITMAX study was a randomized controlled trial (RCT) run by David comparing the effects of high-versus moderate-intensity on the body composition and physical fitness of 28 overweight young men in the Birmingham area. The measured variables in the dataset have two versions reflecting the longitudinal design: a baseline value (BL) and a final follow-up value (F). Because both interventions involved putting previously sedentary subjects onto exercise regimens, David is interested if there was an overall change in respiratory fitness (measured by V O_{2,max}) from baseline to follow-up. Test the investigator's hypothesis using the data.

Question/Answer 1a

What test is appropriate for this data and what are the null and alternative hypotheses?

The appropriate test for this data is a paired t-test. The null hypotheses is that there is no change in VO_2 max from baseline. The alternative hypotheses is there is a change from baseline for the VO_2 max.

Question/Answer 1b

What are the sample means of V O_2 ,max pre- and post-intervention?

The mean pre-intervention is 35.59.

The mean post-intervention is 37.54.

Question/Answer 1c

What is the sample mean of the change in V O_2 ,max from baseline to follow-up?

The sample mean change from baseline to follow is 1.95.

Question/Answer 1d

What is the two-tailed p-value associated with that test statistic?

The two-sided p-value for the test statistic is 0.0163798.

Question/Answer 1e

Write a sentence summarizing this result.

From the sample of 23, there is evidence of an overall positive change in respiratory fitness from baseline based on a mean difference of 1.95 and a p -value of 0.0163798 at a significance level of $\alpha = 0.05$.

Question 2

The primary purpose of the study was to test for whether the different exercise plans had a different effect on the subjects. That is, is there a 'difference of differences' for VO_2 ,max between the HIIT and MIT groups? Test this hypothesis using the data. (Hint: The within-subject change scores for VO_2 ,max were not included in the dataset, but may be calculated using the data step.)

Question/Answer 2a

What test is appropriate for this data and what are the null and alternative hypotheses?

The appropriate test for this data is a two sample t-test. The null hypotheses is that there is no change in VO_2 difference between the treatment groups. The alternative hypotheses is there is a change VO_2 difference between the treatment groups.

Question/Answer 2b

What is the average change in V O_2 ,max for the MIT group? For HIIT?

The average change in VO_2 max for the MIT group is 3.4.

The average change in VO_2 max for the HIIT group is 0.84.

Question/Answer 2c

What are the standard deviations for the changes within each group?

The sd of change in VO_2 max for the MIT group is 3.36.

The sd of change in VO_2 max for the HIIT group is 3.5.

Question/Answer 2d

What is the two-tailed p-value associated with that test statistic?

The two-tailed p-value associated with the test statistic is 0.090288.

Question/Answer 2e

Write a sentence or two summarizing this result.

From the sample of 23 individuals, there is evidence that the average VO_2 difference between the MIT treatment group (3.4, sd = 3.36) and the HIIT treatment group (0.84, sd = 3.5) are not different. More specifically, there is an observed average difference of 2.56 in subjects who underwent the MIT treatment compared with HIIT subjects ($t_{19.8893203} = -1.78$, $p\text{-value} = 0.090288$), at a significance level of $\alpha = 0.05$.

Question 3

As these young men were overweight or obese at baseline, David is curious if they may also have issues with high blood pressure. Specifically, he is interested if their systolic blood pressure at baseline differs from the national average SBP of 120 mmHg. Test this hypothesis using the data.

Question/Answer 3a

What test is appropriate for this data and what are the null and alternative hypotheses?

The appropriate test for this data is a one-sample t-test. The null hypotheses is that there is no change in systolic blood pressure from the national average of 120 mmHg. The alternative hypotheses is there is a difference from the national average of 120 mmHg.

Question/Answer 3b

What is the sample mean?

The sample mean for sbp_bl is 129.2 mmHg.

Question/Answer 3c

What is the standard error of the sample mean?

The standard error of the mean for sbp_bl is 2.33.

Question/Answer 3d

What is the two-tailed p-value associated with that test statistic?

The two-tailed p-value associated with the test statistic is 6.908283×10^{-4} .

Question/Answer 3e

Write a sentence summarizing this result.

From the sample of 23 individuals, there is evidence that the average difference between the samples SBP (129.2 mmHg, SE = 2.33) and the national average of 120 mmHg are different. More specifically, there is an observed average difference of 9.2 mmHg in subjects from this study compared to the national average ($t_{22} = 3.94$, $p\text{-value} = 6.908283 \times 10^{-4}$), at a significance level of $\alpha = 0.05$.

Question 4

In the BasicTests Smoking.xlsx file are the results from a case-control study of 1418 individuals with or without lung cancer, separated by self-reported current smoking status.

Question/Answer 4a

What would our outcome variable be? Our exposure variable? What is the requirement to use the chi-square test? Does this sample meet that requirement?

The outcome variable is LungCancer. The exposure variable is Smoking. The requirement for a chi-square test is a relationship between two categorical variables, in this case, LungCancer and Smoking. Yes, this sample meets the requirement.

Question/Answer 4b

What is the odds ratio for smokers versus nonsmokers? (In that order!) Interpret the odds ratio.

The odds ratio for smokers versus nonsmokers is 2.83. Which means a smoker is 2.83 times more likely to develop lung cancer than a non-smoker.

NOTE: this calculated odds ratio is different than the SAS calculated ratio of 2.97. I did not dig into either functions code so I cannot give a reason why at the moment.

Question/Answer 4c

What is the test statistic for a Pearson chi-square test of independence? What is the p-value associated with this test statistic?

The χ^2 is 19.13. The associated p-value is 4.9975×10^{-4} .

Question/Answer 4d

What is the 95% confidence interval for the odds ratio in part d? (Again, the odds for smokers to the odds for nonsmokers.)

The 95% confidence interval is (1.7683, 4.852))

Question?Answer 4e

Write a sentence or two interpreting the results of your analysis, mentioning both the test and interval estimate.

In these data, we observed a significant association between lung cancer and smoking compared to non-smokers ($\chi^2 = 19.13$, $p\text{-value} = 4.9975 \times 10^{-4}$). The estimated increase in risk of developing lung cancer is 196% (95% CI: 135%, 284%) in these results.