Matt Isaac

Biostatistics Methods

Homework 1

Due: 5 Feb 2018

1. –
   1. A two-sample independent t-test would be appropriate to test for a significant difference between the mean spirometric increase of the two groups. Because s2A = 400 and s2S = 225, and 400/225 = 1.78, we will perform this test with the general (not pooled variance) procedure.

Let µA and µS denote the mean spirometric increase of the aminophylline group and salbutamol groups respectively.

We will set our hypotheses to be:

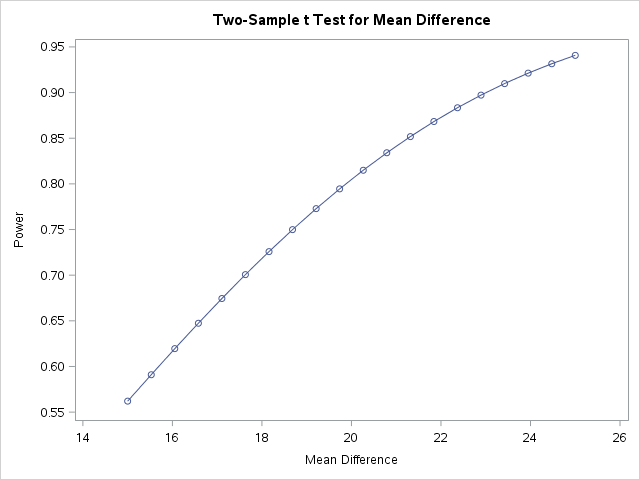
H­0: µA = µS

HA: µA ≠ µS

­­For a t-test, our test statistic is , with 8 degrees of freedom.

From SAS (see appendix for code), we see that our p-value is 0.0008. This is less than our threshold value of 0.05, so we reject the null hypothesis. We have found strong evidence that the aminophylline treatment does affect spirometric values differently than the salbutamol treatment. Because the sample mean for the salbutamol group was higher, it appears that the salbutamol treatment is associated with a greater mean spirometric increase in patients.

* 1. Recall that the p-value is the probability that we would observe the results we did (or something more extreme) by chance if the null hypothesis is true. Thus, the probability that we would have observed the spirometric values that we did if there was no difference between aminophylline and salbutamol treatments is 0.0008. In other words, there is nearly no chance that we would have seen the large difference between the two treatment groups that we did if the treatments did not affect spinometric values differently than each other.
  2. Assuming a common standard deviation of 15, the power would be 0.805 (see appendix for SAS code).
  3. In this context, the power of 0.805 means that if the true value of the difference between the mean spirometric increases of the two groups is really 20, then we have 80.5 percent chance of observing data that will cause us to reject the hypothesis that there is no difference between the two treatments.



As the true mean difference increases (along the x-axis), out test has more power (more likelihood of finding a difference in treatments if there really is one).

1. –

P-values (13)

0.350697

0.88816

0.005715

0.000684

0.010598

0.850057

0.651455

0.000005

0.075073

0.006792

0.9034

0.949922

0.0008 (from #1)

* 1. We will want to adjust the p-values above to account for multiple hypothesis tests. The more independent test of significance that are carried out, the higher likelihood we have of having Type I Errors in our results. In other words, the more likely we are to “find” something significant when there is not actually anything there.

| **p-Values** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Test** | **Raw** | **Bonferroni** | **Stepdown Bonferroni** | **Sidak** | **Hochberg** | **False Discovery Rate** |
| 1 | 0.3507 | 1.0000 | 1.0000 | 0.9964 | 0.9499 | 0.5699 |
| 2 | 0.8882 | 1.0000 | 1.0000 | 1.0000 | 0.9499 | 0.9499 |
| 3 | 0.0057 | 0.0743 | 0.0572 | 0.0718 | 0.0572 | 0.0177 |
| 4 | 0.0007 | 0.0089 | 0.0082 | 0.0089 | 0.0082 | 0.0035 |
| 5 | 0.0106 | 0.1378 | 0.0848 | 0.1293 | 0.0848 | 0.0230 |
| 6 | 0.8501 | 1.0000 | 1.0000 | 1.0000 | 0.9499 | 0.9499 |
| 7 | 0.6515 | 1.0000 | 1.0000 | 1.0000 | 0.9499 | 0.9410 |
| 8 | <.0001 | <.0001 | <.0001 | <.0001 | <.0001 | <.0001 |
| 9 | 0.0751 | 0.9759 | 0.5255 | 0.6374 | 0.5255 | 0.1394 |
| 10 | 0.0068 | 0.0883 | 0.0611 | 0.0848 | 0.0611 | 0.0177 |
| 11 | 0.9034 | 1.0000 | 1.0000 | 1.0000 | 0.9499 | 0.9499 |
| 12 | 0.9499 | 1.0000 | 1.0000 | 1.0000 | 0.9499 | 0.9499 |
| 13 | 0.0008 | 0.0104 | 0.0088 | 0.0104 | 0.0088 | 0.0035 |

* 1. I would recommend using the False Discovery Rate p-value adjustment. This ensures that the False discovery rate is ≤ 0.05 for α = 0.05. It seems like this would be an appropriate error rate to control for. We don’t want to “discover” that there is a difference between the two treatments if there is in fact no true difference.
  2. After making this adjustment, and reviewing all the other p-values, I am having trouble drawing any particular conclusions. 7 of the 13 tests came back with non-significant results. 6 of the tests determined that there was a significant difference between the two treatments. I think that I would say that these data are inconclusive, and would recommend that the researchers review the process of data collection and look at areas of the study that could be standardized.

Appendix:

/\* Code for problem 1a \*/

data temp;

z = -5.2135;

df = 8;

pval = 2 \* probt(z, df);

proc print data = temp;

title1 'P-value from test statistic';

run;

/\* Code for problem 1a \*/

proc power;

twosamplemeans

dist = normal

nulldiff = 0

stddev = 15

alpha = 0.05

sides = 2

ntotal = 20

meandiff = 15 to 25 by 0.5

power = .

;

title1 'Power Calculation - Problem 1';

plot x = effect min = 15 max = 25;

run;

/\* Code for problem 2b \*/

data pvalues;

input raw\_p;

datalines;

0.350697

0.88816

0.005715

0.000684

0.010598

0.850057

0.651455

0.000005

0.075073

0.006792

0.9034

0.949922

0.0008

;

run;

proc print data = pvalues;

title1 'P-values';

run;

proc multtest pdata = pvalues bonferroni sidak holm hochberg fdr;

title1 'Adjusted P-values';

run;