Bookmark this page

Statistics Package Exercise: Testing for a Linear Relationship

Learning Objective: In a given context, carry out the appropriate inferential method for comparing relationships and draw the appropriate conclusions.

Learning Objective: Specify the null and alternative hypotheses for comparing relationships.

The purpose of this activity is to give you guided practice in testing whether the data provide evidence of a significant linear relationship, and in verifying that the basic conditions under which the results of such a test are reliable are met.

Recall the example from the previous activity:

A method for predicting IQ as soon as possible after birth could be important for early intervention in cases such as brain abnormalities or learning disabilities. It has been thought that greater infant vocalization (for instance, more crying) is associated with higher IQ. In 1964, a study was undertaken to see if IQ at 3 years of age is associated with amount of crying at newborn age. In the study, 38 newborns were made to cry after being tapped on the foot and the number of distinct cry vocalizations within 20 seconds was counted. The subjects were followed up at 3 years of age and their IQs were measured.

- R • StatCrunch • TI Calculator • Minitab • Excel

R Instructions
To open R with the data set preloaded, right-click here and choose "Save Target As" to download the file to your computer. Then find the downloaded file and double-click it to open it in R.

The data have been loaded into the data frame

```
baby
```

. Enter the command

```
baby
```

to see the data. The two variables in the data frame are

```
cry.count
```

and

```
IQ
```

.

We would now like to test whether the observed (weak-to-moderate) linear relationship between cry count and IQ is significant (in other words, we would like to carry out the "t-test for the slope" for this example)

Learn By Doing (1/1 point)

State the appropriate hypotheses, and explain (to someone who has never studied statistics) in plain words (without using the words "significant") what is being tested.

Your Answer:

Ho: there is no relationship between cry count and IQ, i.e. increase in cry count doesn't affect value/s of IQ
Ha: there IS a relationship, i.e. change in value of cry count will affect IQ

Our Answer:
The appropriate hypotheses (for the t-test for the slope) in this case are: Ho: There is no linear relationship between cry count and IQ. Ha: There is a significant linear relationship between cry count and IQ. What we are trying to assess using this test is whether there is evidence that the (weak-to-moderate) linear relationship that was observed between cry count and IQ in the sample we took also exists among infants in the general population.

Resubmit Reset

Learn By Doing (1/1 point)

Verify that the general conditions that allow you to reliably use the test in 1 are met. Comment: Note that not all the information you need is given in the problem description. In particular, no information is given about how the babies were selected for the study, which is important for assessing whether we can assume that the observations are independent. To get this information, click here and read the first couple of pages of the paper that describes the study.

Your Answer:

1. should be linear relationship (let's assume for now)
2. independence, i.e. randomly sampled (we can assume they are)
3. no crazy outliers (we can assume none)
4. large enough sample size (goods, n=38)

Our Answer:
Let's start with the issue of independence: According to the paper, the 38 infants were a sample of infants born at the Long Island Jewish Hospital at the time of the study. Given this information, we can assume that the observations are independent (even though it doesn't explicitly say that the sample was random). The only obvious way that the independence assumption would be violated is if twins (or triplets ...) were included in the study, which we can assume was not the case. Comment: Even though the sample was chosen from a specific hospital (rather than from the "entire population of infants"), unless the infants who were born in this specific hospital at the time of the study were systematically different from infants in general in some relevant way, it will still be reasonable to attempt to generalize the results to infants in general. As for the other conditions: The relationship seems linear (even though it is moderately weak), the scatterplot does not display any alarming outliers, and the sample size is reasonably large (n = 38).

Resubmit Reset

Now carry out the test.

- R
- StatCrunch
- TI Calculator
- Minitab
- Excel

R Instructions
To carry out the test, enter the following command into R:

- `model=lm(baby$IQ~baby$cry.count)`
- `summary(model)`

The results include the following:

- Estimates and standard errors of the intercept and slope
- t-Test statistic and p-value to test the two sided alternative $\beta_1 \neq 0$ as well as a test for the intercept not equal to zero, which we generally ignore
- The R^2 value (multiple R-squared), which is equivalent to the correlation squared, and measure of the proportion of variability in the response explained by the model
- The results of an equivalent test for the significance of the slope using an ANOVA approach

---

Learn By Doing (1/1 point)

Give the p-value of the test and state your conclusions in context.

Your Answer:

p-value = 0.0124

There is strong evidence that a linear relationship is present between cry and IQ score.

Our Answer:
RStatCrunch TI CalculatorMinitabExcel R Here is the R output: The p-value of the test is 0.012. The small p-value (in particular, smaller than .05) tells us that it would be quite unlikely (1.2% chance) to get data like those observed just by chance. The data, therefore provide enough evidence for us to conclude that there is a significant linear relationship among infants between vocalization right after birth (as measured by cry count) and IQ at age 3. StatCrunch Here is the StatCrunch output: The p-value of the test is 0.012. The small p-value (in particular, smaller than .05) tells us that it would be quite unlikely (1.2% chance) to get data like those observed just by chance. The data, therefore provide enough evidence for us to conclude that there is a significant linear relationship among infants between vocalization right after birth (as measured by cry count) and IQ at age 3. TI Calculator Here is the

output: The p-value of the test is 0.012. The small p-value (in particular, smaller than .05) tells us that it would be quite unlikely (1.2% chance) to get data like those observed just by chance. The data, therefore provide enough evidence for us to conclude that there is a significant linear relationship among infants between vocalization right after birth (as measured by cry count) and IQ at age 3. Minitab Here is the Minitab output: The p-value of the test is 0.012. The small p-value (in particular, smaller than .05) tells us that it would be quite unlikely (1.2% chance) to get data like those observed just by chance. The data, therefore provide enough evidence for us to conclude that there is a significant linear relationship among infants between vocalization right after birth (as measured by cry count) and IQ at age 3. Excel Here is the Excel output: The p-value of the test is 0.012. The small p-value (in particular, smaller than .05) tells us that it would be quite unlikely (1.2% chance) to get data like those observed just by chance. The data, therefore provide enough evidence for us to conclude that there is a significant linear relationship among infants between vocalization right after birth (as measured by cry count) and IQ at age 3.

Resubmit Reset