

Cross Validated is a question and answer site for people interested in statistics, machine learning, data analysis, data mining, and data visualization. It's 100% free, no registration required.

Sign up ×

Why do I get this p-value doing the Jarque-Bera test in R?

Doing a Jarque Bera test in R I get this result:

```
jarque.bera.test(rnorm(85))

data:  rnorm(85)

X-squared = 1.259, df = 2, p-value = 0.5329
```

Does it mean that the probability to discard the normality hypothesis (it being true) is 53.29%?

If so, why do I get this value if I used a random number from a normal distribution?

hypothesis-testing | statistical-significance | p-value | goodness-of-fit

edited Dec 26 '14 at 15:45

asked Dec 26 '14 at 15:00



will198
154 1 8

- 4 The null hypothesis of the Jarque-Bera test is a joint hypothesis of the skewness being zero and the excess kurtosis being zero. With a p -value > 0.05 , one would usually say that the data are consistent with having skewness and excess kurtosis zero. A high p -value is expected here because you use normally distributed random numbers. It doesn't mean that the data are normally distributed. See [this post](#). – COOLSerdash Dec 26 '14 at 15:10
- 1 I think it's about time to stop referring to anecdotes about normality tests. – Aksakal Dec 26 '14 at 15:27
- 1 @COOLSerdash I know what you mean by "A high p -value is expected here", but for clarity, a high p -value should be just as likely as a low one, in the sense that when H_0 is true, the distribution of p -values should be uniform. Actually the sample size $n = 85$ is too small for the underlying asymptotic approximations for the Jarque-Bera test to work so the reported p -values aren't uniform and shouldn't be used for hypothesis testing (further detail in my answer.) – Silverfish Dec 26 '14 at 20:12

4 Answers

p-value = 0.5329

Does it mean that the probability to discard the normality hypothesis

A p-value is not "the probability to discard the hypothesis". You should review the meaning of p-values. The first sentence of the relevant [wikipedia page](#) should help:

the p-value is the probability of obtaining the observed sample results (or a more extreme result) when the null hypothesis is actually true.

It goes on to say:

If this p-value is very small, usually less than or equal to a threshold value previously chosen called the significance level (traditionally 5% or 1% [1]), it suggests that the observed data is inconsistent with the assumption that the null hypothesis is true

This is quite different from "probability to discard the hypothesis".

is 53.29%?

in any case 0.5329 doesn't convert to 53.29% -- so you should also review how percentages work.

A p-value around 53% is quite consistent with the null hypothesis.

edited Dec 26 '14 at 15:42

answered Dec 26 '14 at 15:35



Glen_b ♦
119k 12 190 406

Your data have come from a normal distribution so the null hypothesis for the Jarque-Bera test (that the population the sample are drawn from has zero skew and zero excess kurtosis) is

actually true. Although we usually call Jarque-Bera a "test for normality", there are other distributions which also have zero skew and zero excess kurtosis (see [this answer](#) for an example), so a Jarque-Bera test can't distinguish them from a normal distribution.

A p -value is

the probability of getting a result as or more extreme than the observed result, **assuming the null hypothesis is true**. It is *not* the probability of rejecting the null hypothesis.

I hope this deals with the "Does it mean that..." aspect of your question. If we see a very small p -value, like 0.001, this means that our observed results would be very improbable if H_0 were true (indeed, highly surprising - something as or more extreme than this we'd only expect to happen 1 time in 1000). This leads us to suspect that H_0 is incorrect. On the contrary, a high p -value is not at all surprising, and although it is not evidence actively in favour of H_0 it certainly does not put H_0 into doubt. In general we consider low p -values as evidence against H_0 , and a lower p -value constitutes stronger evidence. What would lead us to *reject* H_0 ? It's common to set a level of significance, often 5%, and reject H_0 if we observe a p -value lower than the significance level. In your case we would *not* reject H_0 at any sensible level of significance.

When H_0 is true, the p -value will have a continuous uniform distribution between 0 and 1, also known as the rectangular distribution because of the shape of the pdf. This isn't just true for the Jarque-Bera test, and while it isn't quite true for *all* hypothesis tests (consider tests on discrete distributions such as a binomial proportion test or Poisson mean test) "the p -value is equally likely to be anywhere from 0 to 1" is usually a good way of thinking about the p -value under the null.

NB to address a common misconception: just because the null is true does not mean we should expect the p value to be high! There is a 50% chance of it being above 0.5, 50% chance of it being below. If you set your significance level to 5% - that is, you will reject H_0 if you obtain a p value below 0.05 - then be aware this will happen 5% of the time *even if the null is true* (this is why your significance level will be the same as your probability of a Type I error). But there's also a 5% chance of it being between 0.95 and 1, or between 0.32 and 0.37, or between 0.64 and 0.69. I hope this covers the "why do I get this p -value" aspect of your query.

Caution: I have been describing here the ideal situation where the Jarque-Bera test is working well. The test relies on the sample skewness and sample kurtosis being normally distributed - the Central Limit Theorem guarantees this will be asymptotically true in large sample sizes, but **this approximation is not very good in smaller sample sizes**. In fact your $n = 85$ is too small - and so the reported p -values under the null aren't quite uniformly distributed. But if you'd used `rnorm(1000)` instead, my description would have been accurate.

When you refer to the "probability to discard the normality hypothesis (it being true)" you seem to be thinking about the Type I error rate. But you can't see that from just *one* sample, you need to think about the chances of making an incorrect decision across *many* samples. A good way to understand how error rates work is by simulation. Keep running the same R code and you'll keep getting different p values. Make a histogram of those p values and you'll find them approximately equally likely to be drawn anywhere between 0 and 1, so long as you've chosen a large enough n for the Jarque-Bera test to work nicely. If you set your significance level at 5% you'll find that, in the long run, you'll make the Type I error of rejecting the null hypothesis even though it's true (which happens in your simulation when $p < 0.05$) about 5% of the time. If you want to reduce your Type I error rate to 1% then set your significance level to 1%. You might even set it lower. **The problem with doing so is that you make it much harder to reject the null hypothesis when it is false, so you are increasing the Type II error rate.**

Also, if you do want to apply a Jarque-Bera test on a sample size as low as 85, my earlier caution about small sample sizes applies. Since the reported p -values based on the asymptotic distribution will not be uniformly distributed under the null, $p < 0.05$ doesn't occur 5% of the time. So you can't achieve a Type I error rate of 5% simply by rejecting H_0 if the reported $p < 0.05$! Instead, you have to adjust critical values e.g. based on simulation results, as is done in Section 4.1 of Thadewald, T., and H. Buning, 2004, *Jarque-Bera test and its competitors for testing normality - A power comparison*, Discussion Paper Economics 2004/9, School of Business and Economics, Free University of Berlin.

In your simulation you only considered normally distributed data; what if you simulate data that isn't normal instead? In this case we should reject the null hypothesis but you will find you don't always get a p value below 0.05 (or whatever significance level you set) so the Jarque-Bera test results do not give you sufficient evidence to reject. The more powerful the test, the better it is at telling you to reject H_0 in this situation. You will find that you can improve the power of the test by increasing the sample size (whereas when the null was true, changing the sample size makes no difference to the rectangular distribution of the p values - try it! - when the data isn't drawn from a normal population, you'll find low p values become increasingly likely as you increase the sample size). The power of the test is also higher if your data are more blatantly departing from normality - see what happens as you sample from distributions with more extreme skew and kurtosis. **There are alternative normality tests available, and they will have different powers against different types of departure from normality.**

A final word of warning. Be aware that in many practical situations, **we do not really want to run a normality test at all**. Sometimes normality tests can be useful, though - for instance, if you are of a skeptical disposition and want to check whether the "random normal deviates" generated by your statistical software are genuinely normal. You should find that the `rnorm` function in R is fine, however!

edited Dec 27 '14 at 12:53



ssdecontrol

3,865 2 5 36

answered Dec 26 '14 at 16:01



Silverfish

6,254 5 21 64

I made an edit to highlight what I see as the most important sentence; feel free to roll back – ssdecontrol
Dec 27 '14 at 12:28

Also this could use section headings; it's great but it's a wall of text (at least on the iPhone) – ssdecontrol
Dec 27 '14 at 12:31

The other answers are detailed but are concept-heavy. Since the p-value has a mathematical derivation, laying out the math behind it might tighten up the understanding.

Definitions

A *statistic* is any function of data. Typically we assume that observed data is the realization of a random variable or sequence of random variables. Therefore, a statistic is also a random variable. Call the data X , and use x to denote a particular observation of X (in this case, a data set). Call the statistic T and use t to denote the value of the statistic that is computed from a particular x .

This way we can refer to "the event that statistic T takes value t " and attach a probability to it. Many statistics are interesting because they describe complex aspects of the data but follow well-known probability distributions.

The Jarque-Bera test

The Jarque-Bera test is built on a statistic T that has two special properties:

1. t is large if and only if skewness s or kurtosis k , or both, are large.
2. If X follows the normal distribution and the sample size is large, T approximately follows the chi-square distribution with two degrees of freedom.

We know that, in the chi-square distribution with two degrees of freedom, larger values are less probable: $\lim_{t \rightarrow \infty} \Pr(T \leq t | X \sim N) = 0$. So if we observe a large s or a large k , we also observe a large t , and that means that, *if the data is normally distributed*, we have observed a very improbable event. So if we observe a large t , either we have observed a very rare event, or $X \sim N$.

The p value is defined as $p \equiv \Pr(T > t | X \sim N)$.

If t is large, p is small because large t s are unlikely if $X \sim N$. We conduct an hypothesis test by choosing a small value like 0.05 below which p is "too small" for us to believe that it was truly a draw from the chi square distribution, and therefore that X must not be normally distributed.

But then what does this say about a *large* p value? Absolutely nothing. It could be that $X \sim N$ so that $T \sim \chi^2_2$. But it could also be that $X \sim N$ and T follows some other distribution. There just isn't any way to tell. This is why it is never correct to "accept" the hypothesis that $X \sim N$. We can only fail to reject it.

answered Dec 27 '14 at 16:08



ssdecontrol

3,865 2 5 36

You have to read about [hypothesis testing](#) and [Jarque-Bera test](#). It seems that you don't understand either or both of the concepts.

JB test's null hypothesis is that your sample is from normal distribution. The test p-value reflects the probability of accepting the null hypothesis. If it's too low then you reject it. You must set the confidence level, for instance $\alpha = 5\%$, then reject the null if p-value is below this α . In your case p-value is over 50%, which is too high to reject the null. Note, that the hypothesis testing will never tell you to *accept* the null, it may only tell to *reject* it or *not reject* it.

So, your test tells you that it can't reject the hypothesis that your sample is from normal distribution. Which is an expected result, I suppose.

answered Dec 26 '14 at 15:26



Aksakal

11.3k 1 9 33

1 "The test p-value reflects the probability of accepting the null hypothesis." That is simply not true. – ssdecontrol Dec 27 '14 at 12:25

