

How to compute p -values and Cohen's d for z -tests

p -Values:

The p -value of your sample is the lowest alpha level you could have used for your test and still rejected the null hypothesis given your sample. Consider the following example:

You read in a newspaper that the mean SAT score of the nation's high school graduates is 500 with a standard deviation of 110. You think that UCSD freshmen have higher SAT scores on average than that of the nation's high school graduates. To find out, you randomly select 100 UCSD freshmen and find out their SAT scores. On average, they have an SAT score of 533.

Because of the central limit theorem, if the null hypothesis (i.e., that UCSD freshmen SAT scores are no different than those of the nation's high school graduates) is true, your sample mean will have come from an approximately normal distribution with a mean of 500 and standard deviation of 11 (i.e., 11 is the standard error of the mean). To get the p -value for your sample mean of 533, you first convert it to a z -score:

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{533 - 500}{11} = 3$$

Now look up the z -score of your sample mean in the table of z -scores to find out what percentage of the standard normal distribution falls between your sample z -score and positive infinity (since you're doing an **upper tailed test**—i.e., your alternative hypothesis is that the mean UCSD freshmen SAT score is greater than the mean of the null hypothesis). You do this and find out:

$$p(z > 3) = .0013$$

Thus the p -value for your sample is: $p=.0013$

Since the smallest alpha level you could have used and still rejected the null hypothesis with this sample is .0013.

Let's change the scenario a bit:

You read in a newspaper that the mean SAT score of the nation's high school graduates is 500 with a standard deviation of 110. You think that UCSD freshmen might have different SAT scores on average than that of the nation's high school graduates, but you're not sure if they are higher or lower. To find out, you randomly select 100 UCSD freshmen and find out their SAT scores. On average, they have an SAT score of 533.

Now, to test the null hypothesis you would use a **two-tailed test**. This means that we have to double the p -value that we derived from the original scenario to account for the second tail: $p=.0026$

In other words:

$$p(z > 3 \text{ or } z < -3) = .0026$$

Let's change the scenario one last time:

You read in a newspaper that the mean SAT score of the nation's high school graduates is 500 with a standard deviation of 110. You think that UCSD freshmen have lower SAT scores on average than that of the nation's high school graduates. To find out, you randomly select 100 UCSD freshmen and find out their SAT scores. On average, they have an SAT score of 533.

Now, to test the null hypothesis you would use a **lower tailed test**. This means that we have to look up the z -score of the sample mean in the table of z -scores to find out what percentage of the standard normal distribution falls between your sample z -score and negative infinity. You do this and find out:

$$p(z < 3) = 1 - .0013 = .5 + .4987 = .9987$$

Thus the p -value for your sample is: $p = .9987$

Cohen's d :

Cohen's d is a unitless measure of "effect size." In other words, it's a standardized statistic (like percentages and z -scores) for measuring how far the mean of your observations is from the mean of the null hypothesis. Unlike p -values, the magnitude of Cohen's d is not affected by the number of observations. Cohen's d is described on page 299 of the textbook for t -tests. We can also derive it for z -tests using the following formula:

$$d = \left| \frac{\bar{x} - \mu}{\sigma} \right|$$

\bar{x} = sample mean

μ = null hypothesis population mean

σ = null hypothesis population standard deviation

Note that Cohen's d is derived from the population standard deviation of the null hypothesis; it does NOT use the standard error of the mean (i.e., the standard deviation of the sampling distribution of the mean).

The textbook gives the following rules of thumb for interpreting Cohen's d :

Cohen's d	Effect Size
0.20	Small
0.50	Medium
0.80	Large

Let's compute Cohen's d for our first SAT example above:

You read in a newspaper that the mean SAT score of the nation's high school graduates is 500 with a standard deviation of 110. You think that UCSD freshmen have higher SAT scores on average than that of the nation's high school graduates. To find out, you randomly select 100 UCSD freshmen and find out their SAT scores. On average, they have an SAT score of 533.

$$d = \left| \frac{\bar{x} - \mu}{\sigma} \right| = \left| \frac{533 - 500}{110} \right| = 0.30$$

Thus, the difference between your average UCSD student and the national average is relatively small (.3 standard deviations).