The Z-test

February 9, 2017

Contents

- Example 1: (one tailed z-test)
- Example 2: (two tailed z-test)
- Questions
- Answers

The z-test is a hypothesis test to determine if a single observed mean is significantly different (or greater or less than) the mean under the null hypothesis, \( \mu_{hyp} \) when you know the standard deviation of the population. Here’s where the z-test sits on our flow chart.

The assumption is that if the null hypothesis is true, then our observed mean, \( \bar{x} \) is drawn from a normal distribution with mean \( \mu_{hyp} \) and standard deviation equal to the standard error of the mean:

\[
\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}
\]

Where \( n \) is the sample size and \( \sigma_x \) is the population standard deviation.

To conduct the test we convert our observed mean, \( \bar{x} \), to a z-score (standard deviation units):

\[
z = \frac{(\bar{x} - \mu_{hyp})}{\sigma_{\bar{x}}} = \frac{(\bar{x} - \mu_{hyp})}{\frac{\sigma_x}{\sqrt{n}}}
\]

We can then look up the probability of our observed mean under the null hypothesis in the z-table.
Example 1: (one tailed z-test)

The population of all verbal GRE scores are known to have a standard deviation of 8.5. The UW Psychology department hopes to receive applicants with a verbal GRE scores over 210. This year, the mean verbal GRE scores for the 42 applicants was 212.79. Using a value of $\alpha = 0.05$ is this new mean significantly greater than the desired mean of 210?

For this example, the mean under the null hypothesis is $\mu_{hyp} = 210$, the population standard deviation is $\sigma_x = 8.5$, and the observed mean is $\bar{x} = 212.79$.

The standard error of the mean is therefore:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{8.5}{\sqrt{42}} = 1.31$$

To find the probability of finding a mean above 212.79 we convert our observed mean, $\bar{x}$, to a z-score:

$$z = \frac{(\bar{x} - \mu_{hyp})}{\sigma_{\bar{x}}} = \frac{(212.79 - 210)}{1.31} = 2.13$$

This will be a one tailed test because we’re only rejecting $H_0$ if our observed mean is significantly larger than 210. To make our decision we need to find the critical value of $z$, which is the $z$ for which the area above is 0.05. From our z-table, the critical value of $z$ is 1.64:

Our observed value of $z$ is 2.13 which is greater than the critical value of 1.64. We therefore reject $H_0$.

Equivalently, we can calculate the p-value for our observed mean and compare it to alpha. For this one-tailed test, the p-value is the area under the normal distribution above our observed value of z. From the z-table, $p = 0.0167$:

Our p-value is less than alpha (0.05). If the null hypothesis is true, then the probability of obtaining our observed mean or greater is less than 0.05. We therefore reject $H_0$ and state that (in APA format):

The verbal GRE scores of UW Psych applicants ($M = 212.79$) is significantly greater than 210, $z=2.13$, $p = 0.0167$. 
Example 2: (two tailed z-test)

Suppose you start up a company that has developed a drug that is supposed to increase IQ. You know that the standard deviation of IQ in the general population is 15. You test your drug on 36 patients and obtain a mean IQ of 102.96. Using an alpha value of 0.01, is this IQ significantly different than the population mean of 100?

To solve this, we first calculate the standard error of the mean:

\[
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{36}} = 2.5
\]

and then convert our observed mean to a z-score:

\[
z = \frac{(\bar{x} - \mu_{hyp})}{\sigma_{\bar{x}}} = \frac{(102.96 - 100)}{2.5} = 1.18
\]

We then compare our observed value of z to the critical values of z for alpha = 0.01. We are looking for a significant difference, so this will be a two-tailed test. We reject the null hypothesis if our observed mean is both significantly larger or smaller than 100. Our critical values of z are therefore the two values that span the middle 99% of the area under the standard normal distribution. This means that the areas in each of the two tails is \(\frac{0.01}{2} = 0.005\):

The rejection region contains values of z less than -2.58 and greater than 2.58. Our observed value of z falls outside the rejection region, so we fail to reject \(H_0\) and conclude that our drug did not have a significant effect on IQ.

To calculate the p-value we need to find the area under the standard normal distribution beyond our observed value of z and double it. This is because for a two-tailed test we want the probability of obtaining our observed value or more extreme in either direction. This makes sense if you think about what happens if the observed value of z falls exactly on the critical value (2.58 in this example). The area beyond the observed value of z in both the positive direction and the negative direction will add up to alpha (0.01).

For this example, the area above \(z = 1.18\) plus the area below \(z = -1.18\) is 0.1182 + 0.1182 = 0.2364

Since our p-value of 0.2364 is greater than 0.01, we fail to reject \(H_0\) and state that:

The IQ of drug patients (M = 102.96) is not significantly different than 100, \(z=1.18\), \(p = 0.2364\).
Questions
Here are 10 practice z-test questions followed by their answers.

1) Suppose the rain of monkeys has a population that is normally distributed with a standard deviation of 6. Let’s pretend that you sample 9 monkeys from this population and obtain a mean rain of 10.24 and a standard deviation of 7.8419. Using an alpha value of $\alpha = 0.01$, is this observed mean significantly different than an expected rain of 14?

2) Suppose the gravity of ugliest web sites has a population that is normally distributed with a standard deviation of 2. Suppose you sample 19 ugliest web sites from this population and obtain a mean gravity of 31.23 and a standard deviation of 2.0668. Using an alpha value of $\alpha = 0.05$, is this observed mean significantly different than an expected gravity of 32?

3) Suppose the advice of brave nerds has a population that is normally distributed with a standard deviation of 4. For some reason you sample 95 brave nerds from this population and obtain a mean advice of 55.09 and a standard deviation of 4.0094. Using an alpha value of $\alpha = 0.05$, is this observed mean significantly less than an expected advice of 56?

4) Suppose the farts of winters has a population that is normally distributed with a standard deviation of 9. I sample 39 winters from this population and obtain a mean farts of 81.34 and a standard deviation of 10.1131. Using an alpha value of $\alpha = 0.05$, is this observed mean significantly less than an expected farts of 83?

5) Suppose the equipment of nasty iPods has a population that is normally distributed with a standard deviation of 6. You want to sample 59 nasty iPods from this population and obtain a mean equipment of 87.9 and a standard deviation of 5.9234. Using an alpha value of $\alpha = 0.01$, is this observed mean significantly less than an expected equipment of 89?

6) Suppose the happiness of six teams has a population that is normally distributed with a standard deviation of 3. Just for fun, you sample 33 six teams from this population and obtain a mean happiness of 23.95 and a standard deviation of 3.5006. Using an alpha value of $\alpha = 0.01$, is this observed mean significantly different than an expected happiness of 22?

7) Suppose the cash of students has a population that is normally distributed with a standard deviation of 5. You go out and sample 41 students from this population and obtain a mean cash of 67.38 and a standard deviation of 5.6318. Using an alpha value of $\alpha = 0.05$, is this observed mean significantly different than an expected cash of 69?

8) Suppose the span of thin iPhones has a population that is normally distributed with a standard deviation of 10. I sample 96 thin iPhones from this population and obtain a mean span of 58.81 and a standard deviation of 10.756. Using an alpha value of $\alpha = 0.05$, is this observed mean significantly less than an expected span of 59?

9) Suppose the body mass index of Asian food has a population that is normally distributed with a standard deviation of 1. Just for fun, you sample 58 Asian food from this population and obtain a mean body mass index of 58.03 and a standard deviation of 0.9394. Using an alpha value of $\alpha = 0.01$, is this observed mean significantly greater than an expected body mass index of 58?

10) Suppose the taste of easy fathers has a population that is normally distributed with a standard deviation of 2. I’d like you to sample 66 easy fathers from this population and obtain a mean taste of 64.6 and a standard deviation of 2.1469. Using an alpha value of $\alpha = 0.01$, is this observed mean significantly different than an expected taste of 64?
Answers

1) Two tailed z-test
\( H_0 : \mu_x = 14 \)
\( H_A : \mu_x \neq 14 \)
\( \sigma_\bar{x} = \frac{\sigma_x}{\sqrt{n}} = \frac{6}{\sqrt{9}} = 2 \)
\( z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_\bar{x}} = \frac{(10.24 - 14)}{2} = -1.88 \)
\( z_{crit} \) for \( \alpha = 0.01 \) (Two tailed) is \( \pm 2.58 \)
We fail to reject \( H_0 \).
The rain of monkeys (M = 10.24) is not significantly different than 14, \( z=-1.88, p = 0.0601 \).

2) Two tailed z-test
\( H_0 : \mu_x = 32 \)
\( H_A : \mu_x \neq 32 \)
\( \sigma_\bar{x} = \frac{\sigma_x}{\sqrt{n}} = \frac{2}{\sqrt{19}} = 0.4588 \)
\( z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_\bar{x}} = \frac{(31.23 - 32)}{0.4588} = -1.68 \)
\( z_{crit} \) for \( \alpha = 0.05 \) (Two tailed) is \( \pm 1.96 \)
We fail to reject \( H_0 \).
The gravity of ugliest web sites (M = 31.23) is not significantly different than 32, \( z=-1.68, p = 0.0933 \).

3) One tailed z-test
\( H_0 : \mu_x = 56 \)
\( H_A : \mu_x < 56 \)
\( \sigma_\bar{x} = \frac{\sigma_x}{\sqrt{n}} = \frac{4}{\sqrt{95}} = 0.4104 \)
\( z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_\bar{x}} = \frac{(55.09 - 56)}{0.4104} = -2.22 \)
\( z_{crit} \) for \( \alpha = 0.05 \) (One tailed) is \(-1.64 \)
We reject \( H_0 \).
The advice of brave nerds (M = 55.09) is significantly less than 56, \( z=-2.22, p = 0.0133 \).

4) One tailed z-test
\( H_0 : \mu_x = 83 \)
\( H_A : \mu_x < 83 \)
\( \sigma_\bar{x} = \frac{\sigma_x}{\sqrt{n}} = \frac{9}{\sqrt{39}} = 1.4412 \)
\( z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_\bar{x}} = \frac{(81.34 - 83)}{1.4412} = -1.15 \)
\( z_{crit} \) for \( \alpha = 0.05 \) (One tailed) is \(-1.64 \)
We fail to reject \( H_0 \).
The farts of winters (M = 81.34) is not significantly less than 83, \( z=-1.15, p = 0.1247 \).
5) One tailed z-test
\[ H_0 : \mu_x = 89 \]
\[ H_A : \mu_x < 89 \]
\[ \sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{59}} = 0.7811 \]
\[ z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_x} = \frac{(87.9 - 89)}{0.7811} = -1.41 \]
\[ z_{crit} \text{ for } \alpha = 0.01 \text{ (One tailed) is } -2.33 \]
We fail to reject \( H_0 \).
The equipment of nasty iPods (M = 87.9) is not significantly less than 89, \( z=-1.41, p = 0.0795 \).

6) Two tailed z-test
\[ H_0 : \mu_x = 22 \]
\[ H_A : \mu_x \neq 22 \]
\[ \sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{33}} = 0.5222 \]
\[ z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_x} = \frac{(23.95 - 22)}{0.5222} = 3.73 \]
\[ z_{crit} \text{ for } \alpha = 0.01 \text{ (Two tailed) is } \pm 2.58 \]
We reject \( H_0 \).
The happiness of six teams (M = 23.95) is significantly different than 22, \( z=3.73, p = 0.0002 \).

7) Two tailed z-test
\[ H_0 : \mu_x = 69 \]
\[ H_A : \mu_x \neq 69 \]
\[ \sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{41}} = 0.7809 \]
\[ z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_x} = \frac{(67.38 - 69)}{0.7809} = -2.07 \]
\[ z_{crit} \text{ for } \alpha = 0.05 \text{ (Two tailed) is } \pm 1.96 \]
We reject \( H_0 \).
The cash of students (M = 67.38) is significantly different than 69, \( z=-2.07, p = 0.038 \).

8) One tailed z-test
\[ H_0 : \mu_x = 59 \]
\[ H_A : \mu_x < 59 \]
\[ \sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{96}} = 1.0206 \]
\[ z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_x} = \frac{(58.81 - 59)}{1.0206} = -0.19 \]
\[ z_{crit} \text{ for } \alpha = 0.05 \text{ (One tailed) is } -1.64 \]
We fail to reject \( H_0 \).
The span of thin iPhones (M = 58.81) is not significantly less than 59, \( z=-0.19, p = 0.4262 \).

9) One tailed z-test
**H₀ : μₓ = 58**

**Hₐ : μₓ > 58**

\[σₓ = \frac{σ}{\sqrt{n}} = \frac{1}{\sqrt{58}} = 0.1313\]

\[z_{obs} = \frac{\bar{x} - \mu_{hyp}}{σₓ} = \frac{(58.03 - 58)}{0.1313} = 0.23\]

\[z_{crit} \text{ for } α = 0.01 \text{ (One tailed) is 2.33}\]

We fail to reject \(H₀\).
The body mass index of Asian food (M = 58.03) is not significantly greater than 58, \(z=0.23\), \(p = 0.4096\).

**10) Two tailed z-test**

\(H₀ : μₓ = 64\)

**Hₐ : μₓ ≠ 64**

\[σₓ = \frac{σ}{\sqrt{n}} = \frac{2}{\sqrt{66}} = 0.2462\]

\[z_{obs} = \frac{\bar{x} - \mu_{hyp}}{σₓ} = \frac{(64.6 - 64)}{0.2462} = 2.44\]

\[z_{crit} \text{ for } α = 0.01 \text{ (Two tailed) is } ±2.58\]

We fail to reject \(H₀\).
The taste of easy fathers (M = 64.6) is not significantly different than 64, \(z=2.44\), \(p = 0.0148\).