The Z-test

January 11, 2020

Contents

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

Happy birthday to Karen Lupita Sanchez and Alexandra Lauren Kiselman!

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. Looking at our $z$-table for $\alpha = 0.05$:

<table>
<thead>
<tr>
<th>$z$</th>
<th>Area between mean and $z$</th>
<th>Area beyond $z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.62</td>
<td>0.4474</td>
<td>0.0526</td>
</tr>
<tr>
<td>1.63</td>
<td>0.4484</td>
<td>0.0516</td>
</tr>
<tr>
<td><strong>1.64</strong></td>
<td><strong>0.4495</strong></td>
<td><strong>0.0505</strong></td>
</tr>
<tr>
<td>1.65</td>
<td>0.4505</td>
<td>0.0495</td>
</tr>
<tr>
<td>1.66</td>
<td>0.4515</td>
<td>0.0485</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
</tbody>
</table>

You can see that the critical value of $z$ is 1.64:

![z-table](image)

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:
You can see that our p-value is $p = 0.0166$.

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, $z = 2.13$, $p = 0.0166$.

We could also use the 'pnorm' function R to calculate this p-value. Remember, we need to divide the population standard deviation ($\sigma = 8.5$) by the square root of $n$ ($\sqrt{42}$):

$$1 - \text{pnorm}(212.79, 210, 8.5/\sqrt{42})$$

[1] 0.01670141

**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 0.05, 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 \( \alpha = 0.05 \). 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 either significantly larger or smaller than 100. Our critical values of \( z \) are therefore the two values that span the middle 95% of the area under the standard normal distribution. This means that the areas in each of the two tails is \( \frac{0.05}{2} = 0.025 \):

<table>
<thead>
<tr>
<th>( z )</th>
<th>Area between mean and ( z )</th>
<th>Area beyond ( z )</th>
</tr>
</thead>
<tbody>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>1.94</td>
<td>0.4738</td>
<td>0.0262</td>
</tr>
<tr>
<td>1.95</td>
<td>0.4744</td>
<td>0.0256</td>
</tr>
<tr>
<td><strong>1.96</strong></td>
<td><strong>0.4750</strong></td>
<td><strong>0.0250</strong></td>
</tr>
<tr>
<td>1.97</td>
<td>0.4756</td>
<td>0.0244</td>
</tr>
<tr>
<td>1.98</td>
<td>0.4761</td>
<td>0.0239</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
</tbody>
</table>

Which corresponds to a critical value of \( z = 1.96 \).

The rejection region contains values of \( z \) less than -1.96 and greater than 1.96. 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 (1.96 in this example). The area beyond the observed value of \( z \) in both the positive direction and the negative direction will add up to \( \alpha (0.05) \).
<table>
<thead>
<tr>
<th>z</th>
<th>Area between mean and z</th>
<th>Area beyond z</th>
</tr>
</thead>
<tbody>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>1.16</td>
<td>0.3770</td>
<td>0.1230</td>
</tr>
<tr>
<td>1.17</td>
<td>0.3790</td>
<td>0.1210</td>
</tr>
<tr>
<td><strong>1.18</strong></td>
<td><strong>0.3810</strong></td>
<td><strong>0.1190</strong></td>
</tr>
<tr>
<td>1.19</td>
<td>0.3830</td>
<td>0.1170</td>
</tr>
<tr>
<td>1.20</td>
<td>0.3849</td>
<td>0.1151</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
</tbody>
</table>

For this example, the area above $z = 1.18$ plus the area below $z = -1.18$ is $0.119 + 0.119 = 0.238$

Since our p-value of 0.238 is greater than 0.05, we fail to reject $H_0$ and state that:

The IQ of regular drug patients ($M = 102.96$) is not significantly different than 100, $z=1.18$, $p = 0.238$.

To do this in R, we need to be sure to double our p-value since this is a two-tailed test. Here’s how to calculate the p-value in one step. Note the `'2*(1-pnorm...)`’, which doubles the p-value:

```
2*(1-pnorm(102.96,100,15/sqrt(36)))
```

```
[1] 0.2364131
```
Questions

Here are 15 practice z-test questions followed by their answers, including how to use R to find p-values.

1) Suppose the clothing of men has a population that is normally distributed with a standard deviation of 9. Your stats professor asks you to sample 94 men from this population and obtain a mean clothing of 76.47 and a standard deviation of 9.1267. Using an alpha value of $\alpha = 0.01$, is this observed mean significantly less than an expected clothing of 79?

2) Suppose the friendship of sponges has a population that is normally distributed with a standard deviation of 7. You go out and sample 23 sponges from this population and obtain a mean friendship of 20.02 and a standard deviation of 6.1289. Using an alpha value of $\alpha = 0.01$, is this observed mean significantly different than an expected friendship of 18?

3) Suppose the health of athletes has a population that is normally distributed with a standard deviation of 5. For your first year project you sample 80 athletes from this population and obtain a mean health of 97.45 and a standard deviation of 4.9931. Using an alpha value of $\alpha = 0.05$, is this observed mean significantly different than an expected health of 96?

4) Suppose the scenery of politicians has a population that is normally distributed with a standard deviation of 6. Just for fun, you sample 74 politicians from this population and obtain a mean scenery of 10.24 and a standard deviation of 6.4714. Using an alpha value of $\alpha = 0.05$, is this observed mean significantly different than an expected scenery of 11?

5) Suppose the leisure of chickens has a population that is normally distributed with a standard deviation of 5. I go and sample 106 chickens from this population and obtain a mean leisure of 73.06 and a standard deviation of 5.3508. Using an alpha value of $\alpha = 0.01$, is this observed mean significantly different than an expected leisure of 72?

6) Suppose the evil of brothers has a population that is normally distributed with a standard deviation of 6. Your friend gets you to sample 49 brothers from this population and obtain a mean evil of 81.88 and a standard deviation of 6.516. Using an alpha value of $\alpha = 0.01$, is this observed mean significantly different than an expected evil of 81?
7) Suppose the leisure of elbows has a population that is normally distributed with a standard deviation of 6. Your boss makes you sample 64 elbows from this population and obtain a mean leisure of 26.66 and a standard deviation of 5.5607. Using an alpha value of $\alpha = 0.01$, is this observed mean significantly greater than an expected leisure of 25?

8) Suppose the smell of elbows has a population that is normally distributed with a standard deviation of 8. I’d like you to sample 35 elbows from this population and obtain a mean smell of 90.24 and a standard deviation of 8.0088. Using an alpha value of $\alpha = 0.05$, is this observed mean significantly greater than an expected smell of 89?

9) Suppose the frequency of dollars has a population that is normally distributed with a standard deviation of 1. I go and sample 15 dollars from this population and obtain a mean frequency of 71.02 and a standard deviation of 0.8996. Using an alpha value of $\alpha = 0.05$, is this observed mean significantly different than an expected frequency of 72?

10) Suppose the IQ of shocking PhDs has a population that is normally distributed with a standard deviation of 9. Your boss makes you sample 36 shocking PhDs from this population and obtain a mean IQ of 8.87 and a standard deviation of 8.6377. Using an alpha value of $\alpha = 0.05$, is this observed mean significantly less than an expected IQ of 12?

11) Suppose the damage of long response times has a population that is normally distributed with a standard deviation of 10. Your friend gets you to sample 26 long response times from this population and obtain a mean damage of 30.51 and a standard deviation of 8.9765. Using an alpha value of $\alpha = 0.05$, is this observed mean significantly greater than an expected damage of 28?

12) Suppose the IQ of misty spleens has a population that is normally distributed with a standard deviation of 3. You ask a friend to sample 106 misty spleens from this population and obtain a mean IQ of 29.88 and a standard deviation of 3.0712. Using an alpha value of $\alpha = 0.01$, is this observed mean significantly less than an expected IQ of 30?

13) Suppose the body mass index of anxious candy bars has a population that is normally distributed with a standard deviation of 6. Let’s pretend that you sample 102 anxious candy bars from this population and obtain a mean body mass index of 14.96 and a standard deviation of 5.9752. Using an alpha value of $\alpha = 0.01$, is this observed mean significantly less than an expected
body mass index of 17?

14) Suppose the friendship of laboratory rats has a population that is normally distributed with a standard deviation of 8. You get a grant to sample 23 laboratory rats from this population and obtain a mean friendship of 76.2 and a standard deviation of 9.3261. Using an alpha value of $\alpha = 0.01$, is this observed mean significantly different than an expected friendship of 79?

15) Suppose the happiness of angry grad students has a population that is normally distributed with a standard deviation of 4. I go and sample 21 angry grad students from this population and obtain a mean happiness of 80.97 and a standard deviation of 4.3132. Using an alpha value of $\alpha = 0.05$, is this observed mean significantly different than an expected happiness of 82?
Answers

1) One tailed z-test
\[ \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{9}{\sqrt{94}} = 0.9283 \]
\[ z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_{\bar{x}}} = \frac{(76.47 - 79)}{0.9283} = -2.73 \]
\[ z_{crit} \text{ for } \alpha = 0.01 \text{ (One tailed) is } -2.33 \]

We reject \( H_0 \).

The clothing of men (\( M = 76.47 \)) is significantly less than 79, \( z=-2.73, p = 0.0032 \).

# Using R:
\[
\text{pnorm}(76.47,79,9/\sqrt{94})
\]
\[ [1] \ 0.003210468 \]

2) Two tailed z-test
\[ \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{7}{\sqrt{23}} = 1.4596 \]
\[ z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_{\bar{x}}} = \frac{(20.02 - 18)}{1.4596} = 1.38 \]
\[ z_{crit} \text{ for } \alpha = 0.01 \text{ (Two tailed) is } \pm 2.58 \]

We fail to reject \( H_0 \).

The friendship of sponges (\( M = 20.02 \)) is not significantly different than 18, \( z=1.38, p = 0.1676 \).

# Using R: since our observed mean is greater than the mean for \( H_0 \):
\[
2*(1-pnorm(20.02,18,7/\sqrt{23}))
\]
\[ [1] \ 0.1663768 \]

3) Two tailed z-test
\[ \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{5}{\sqrt{80}} = 0.559 \]
\[ z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_{\bar{x}}} = \frac{(97.45 - 96)}{0.559} = 2.59 \]
\[ z_{crit} \text{ for } \alpha = 0.05 \text{ (Two tailed) is } \pm 1.96 \]

We reject \( H_0 \).

The health of athletes (\( M = 97.45 \)) is significantly different than 96, \( z=2.59, p = 0.0096 \).
# Using R: since our observed mean is greater than the mean for H0:
\[
2\times (1 - \text{pnorm}(97.45, 96, 5/\sqrt{80}))
\]
\[
[1] 0.009491096
\]

4) Two tailed z-test
\[
\sigma x = \frac{\sigma x}{\sqrt{n}} = \frac{6}{\sqrt{74}} = 0.6975
\]
\[
z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma x} = \frac{(10.24 - 11)}{0.6975} = -1.09
\]
\[
z_{crit} \text{ for } \alpha = 0.05 \text{ (Two tailed)} \text{ is } \pm 1.96
\]

We fail to reject \( H_0 \).

The scenery of politicians (\( M = 10.24 \)) is not significantly different than 11, \( z = -1.09, p = 0.2757 \).

# Using R: since our observed mean is less than the mean for H0:
\[
2\times \text{pnorm}(10.24, 11, 6/\sqrt{74})
\]
\[
[1] 0.2758771
\]

5) Two tailed z-test
\[
\sigma x = \frac{\sigma x}{\sqrt{n}} = \frac{5}{\sqrt{106}} = 0.4856
\]
\[
z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma x} = \frac{(73.06 - 72)}{0.4856} = 2.18
\]
\[
z_{crit} \text{ for } \alpha = 0.01 \text{ (Two tailed)} \text{ is } \pm 2.58
\]

We fail to reject \( H_0 \).

The leisure of chickens (\( M = 73.06 \)) is not significantly different than 72, \( z = 2.18, p = 0.0293 \).

# Using R: since our observed mean is greater than the mean for H0:
\[
2\times (1 - \text{pnorm}(73.06, 72, 5/\sqrt{106}))
\]
\[
[1] 0.02905986
\]

6) Two tailed z-test
\[
\sigma x = \frac{\sigma x}{\sqrt{n}} = \frac{6}{\sqrt{49}} = 0.8571
\]
\[
z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma x} = \frac{(81.88 - 81)}{0.8571} = 1.03
\]
\[
z_{crit} \text{ for } \alpha = 0.01 \text{ (Two tailed)} \text{ is } \pm 2.58
\]
We fail to reject $H_0$.

The evil of brothers ($M = 81.88$) is not significantly different than 81, $z=1.03$, $p = 0.303$.

# Using R: since our observed mean is greater than the mean for $H_0$:

$$2*(1-pnorm(81.88,81,6/sqrt(49)))$$

[1] 0.3045775

7) One tailed z-test

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{6}{\sqrt{64}} = 0.75$$

$$z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_{\bar{x}}} = \frac{(26.66-25)}{0.75} = 2.21$$

$z_{crit}$ for $\alpha = 0.01$ (One tailed) is 2.33

We fail to reject $H_0$.

The leisure of elbows ($M = 26.66$) is not significantly greater than 25, $z=2.21$, $p = 0.0136$.

# Using R:

$$1-pnorm(26.66,25,6/sqrt(64))$$

[1] 0.01343734

8) One tailed z-test

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{8}{\sqrt{35}} = 1.3522$$

$$z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_{\bar{x}}} = \frac{(90.24-89)}{1.3522} = 0.92$$

$z_{crit}$ for $\alpha = 0.05$ (One tailed) is 1.64

We fail to reject $H_0$.

The smell of elbows ($M = 90.24$) is not significantly greater than 89, $z=0.92$, $p = 0.1788$.

# Using R:

$$1-pnorm(90.24,89,8/sqrt(35))$$

[1] 0.1795733

9) Two tailed z-test

$$2*(1-pnorm(81.88,81,6/sqrt(49)))$$

[1] 0.3045775
\[ \sigma \bar{x} = \frac{\sigma x}{\sqrt{n}} = \frac{1}{\sqrt{15}} = 0.2582 \]

\[ z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma \bar{x}} = \frac{(71.02 - 72)}{0.2582} = -3.80 \]

*z_{crit} for \( \alpha = 0.05 \) (Two tailed) is \( \pm 1.96 \)

We reject \( H_0 \).

The frequency of dollars (\( M = 71.02 \)) is significantly different than 72, \( z = -3.8, p = 0.0001 \).

# Using R: since our observed mean is less than the mean for \( H_0 \):

\[
2 \times \text{pnorm}(71.02,72,1/\sqrt(15))
\]

[1] 0.0001473321

10) One tailed z-test

\[ \sigma \bar{x} = \frac{\sigma x}{\sqrt{n}} = \frac{9}{\sqrt{36}} = 1.5 \]

\[ z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma \bar{x}} = \frac{(8.87 - 12)}{1.5} = -2.09 \]

*z_{crit} for \( \alpha = 0.05 \) (One tailed) is \(-1.64 \)

We reject \( H_0 \).

The IQ of shocking PhDs (\( M = 8.87 \)) is significantly less than 12, \( z = -2.09, p = 0.0183 \).

# Using R:

\[
\text{pnorm}(8.87,12,9/\sqrt(36))
\]

[1] 0.01845914

11) One tailed z-test

\[ \sigma \bar{x} = \frac{\sigma x}{\sqrt{n}} = \frac{10}{\sqrt{26}} = 1.9612 \]

\[ z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma \bar{x}} = \frac{(30.51 - 28)}{1.9612} = 1.28 \]

*z_{crit} for \( \alpha = 0.05 \) (One tailed) is \( 1.64 \)

We fail to reject \( H_0 \).

The damage of long response times (\( M = 30.51 \)) is not significantly greater than 28, \( z = 1.28, p = 0.1003 \).

# Using R:

\[
1 - \text{pnorm}(30.51,28,10/\sqrt(26))
\]
12) One tailed z-test
\[
\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{3}{\sqrt{106}} = 0.2914
\]
\[
z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_{\bar{x}}} = \frac{(29.88 - 30)}{0.2914} = -0.41
\]
\[
z_{crit} \text{ for } \alpha = 0.01 \text{ (One tailed) is } -2.33
\]
We fail to reject \( H_0 \).

The IQ of misty spleens (M = 29.88) is not significantly less than 30, \( z = -0.41, p = 0.3409 \).

# Using R:
```r
pnorm(29.88,30,3/sqrt(106))
```

[1] 0.3402338

13) One tailed z-test
\[
\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{6}{\sqrt{102}} = 0.5941
\]
\[
z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_{\bar{x}}} = \frac{(14.96 - 17)}{0.5941} = -3.43
\]
\[
z_{crit} \text{ for } \alpha = 0.01 \text{ (One tailed) is } -2.33
\]
We reject \( H_0 \).

The body mass index of anxious candy bars (M = 14.96) is significantly less than 17, \( z = -3.43, p = 0.0003 \).

# Using R:
```r
pnorm(14.96,17,6/sqrt(102))
```

[1] 0.0002975568

14) Two tailed z-test
\[
\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{8}{\sqrt{23}} = 1.6681
\]
\[
z_{obs} = \frac{\bar{x} - \mu_{hyp}}{\sigma_{\bar{x}}} = \frac{(76.2 - 79)}{1.6681} = -1.68
\]
\[
z_{crit} \text{ for } \alpha = 0.01 \text{ (Two tailed) is } \pm 2.58
\]
We fail to reject \( H_0 \).
The friendship of laboratory rats (M = 76.2) is not significantly different than 79, z=-1.68, p = 0.093.

# Using R: since our observed mean is less than the mean for H0:
2*pnorm(76.2,79,8/sqrt(23))
[1] 0.09324153

15) Two tailed z-test
\[ \sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{21}} = 0.8729 \]
\[ z_{obs} = \frac{\bar{x} - \mu_{H_0}}{\sigma_x} = \frac{(80.97 - 82)}{0.8729} = -1.18 \]
\[ z_{crit} \text{ for } \alpha = 0.05 \text{ (Two tailed) is } \pm 1.96 \]

We fail to reject H0.

The happiness of angry grad students (M = 80.97) is not significantly different than 82, z=-1.18, p = 0.238.

# Using R: since our observed mean is less than the mean for H0:
2*pnorm(80.97,82,4/sqrt(21))
[1] 0.2379949