Large Sample Size Hypothesis Testing

March 25, 2020

1 Common-Large Sample Tests

There will typically be a value θ_0 for the unknown parameter θ which corresponds to the uninteresting pure randomness case, so $\theta = \theta_0$ will be H_0, the null hypothesis. The alternative hypothesis, H_a, that something more interesting is going on, will be of the form $\theta > \theta_0$ (or of the form $\theta < \theta_0$).

The test statistic will simply be $\hat{\theta}$, an estimator for θ.

The rejection region is then of the form $\hat{\theta} > k$ (or of the form $\hat{\theta} < k$) for some value of k. The value of k is chosen based on α, a chosen acceptable probability of having a Type I error. Since we are in the large sample size regime, we will assume, by the Central Limit Theorem, that $\hat{\theta}$ is normally distributed.

Let's look at an example before we make this more formal.

EXAMPLE 1.1. *In a given country the height of females in the population is normally distributed with mean 64 inches and standard deviation of 6 inches. A certain city in this country wants to know if it's population is taller than average. 49 females are randomly chosen from this city their average height is found to be 65 inches. At a level $\alpha = .01$, is there sufficient evidence to accept that the typical height in this city is larger than the rest of the country?*

Solution: We begin by formally writing out the elements of our statistical test. Let θ be the average mean of the female population in this city.

The null hypothesis, H_0, is $\theta = 64$.
The alternative hypothesis, H_a, is $\theta > 64$.
The test statistic is $\hat{\theta}$ the average height of the random sample.
We now need to do a little work to determine the RR. Under the null hypothesis $\hat{\theta}$ is a normal random variable with mean 64 and standard deviation 6, our goal is to find k such that

$$\mathbb{P}(\hat{\theta} > k) = .01 = \alpha.$$

We do this because if H_0 is true, we want the probability that the test statistic lies in the rejection region to be .01 (this is the α in the problem statement). Standardizing our normal random variable we see the above statement is equivalent to

$$\mathbb{P}\left(\frac{\hat{\theta} - 64}{6/\sqrt{49}} > \frac{k - 64}{6/\sqrt{49}} \right) = .01.$$

Since the left side of the inequality inside the probability is a standard normal random variable we need

$$\frac{k-64}{6/\sqrt{49}} = z_{.01} = 2.33.$$

Rearranging leads to

$$k = \frac{6}{7} * 2.33 + 64 \approx 66$$

So RR is $\hat{\theta} > 66$.

Now that we have computed our RR we simply check if our measured test statistic lies in it or not. In this case $65 < 66$, so the test statistic did not land in the RR, therefore there is not sufficient evidence to reject the null hypothesis in favor of the alternative. In other words, we cannot conclude this city is taller than a typical city.

Note in the above example we do not conclude that the city has a normal height, we simply do not conclude it is taller that typical. Which is a somewhat subtle distinction.

We now formalize what was done in the last example. In the large sample-size regime we assume that under the null hypothesis the test statistic is normally distributed. The elements of the statistic test are:

- $H_0 : \theta = \theta_0$

- Either $H_a : \theta > \theta_0$ or $\theta < \theta_0$ or $\theta \neq \theta_0$

- Test statistic: $\hat{\theta}$ (the estimator for θ)

- RR is $\hat{\theta} > z_\alpha \sigma_{\hat{\theta},0} + \theta_0$ or $\hat{\theta} < z_\alpha \sigma_{\hat{\theta},0} + \theta_0$ or $|\hat{\theta} - \theta_0| > z_{\alpha/2} \sigma_{\hat{\theta},0}$ respectively. (NOTE in an earlier version there was an extra θ that didn't make any sense in these formulas.)

When computing the rejection region, α is the given level (acceptable Type I error) and $\sigma_{\hat{\theta},0}$ is the standard deviation of $\hat{\theta}$ under the null hypothesis, (REMEMBER this is not the std dev of an individual sample, but a typically a factor of $1/\sqrt{n}$ smaller, depending on the problem). Sometimes $\sigma_{\hat{\theta},0}$ is not explicitly given and must either be deduced from the sample data or can be computed from information of $\hat{\theta}$ under H_0 (for example the variance of the Binomial(n,p) random variable is $np(1-p)$).

Here are some pictures to keep in mind:

If $H_0 : \theta = \theta_0$, for some fixed θ_0 and $H_a : \theta > \theta_0$, then we should reject H_0 is the test statistic, $\hat{\theta}$, is much larger than θ_0. More precisely, given a level α, we want to find a region such that the probability the test statistic lies in this region when H_0 is true is α. So the picture is:
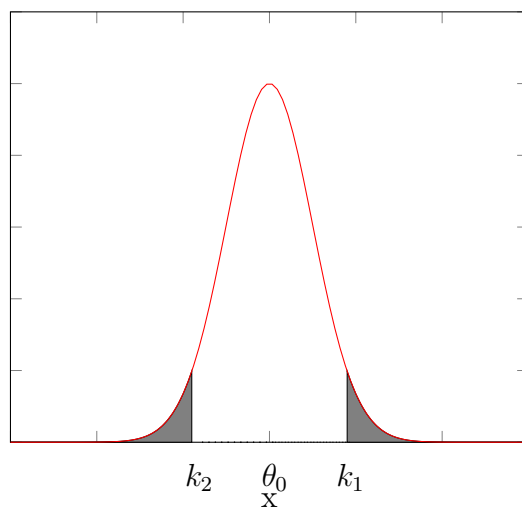
Figure 1.1: The pdf of $\hat{\theta}$ when H_0 holds, with RR
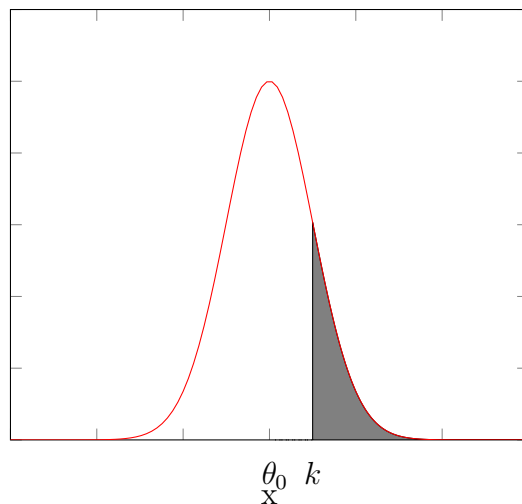
Since we are in the large sample regime, we assume the estimator for θ, $\hat{\theta}$, is normally distributed. When we furthermore assume H_0 holds, then it's mean is at θ_0, as shown in the picture. The area of the shaded region is α. As stated above $k = z_\alpha \sigma_{\hat{\theta},0} + \theta_0$.

If $H_0 : \theta = \theta_0$, for some fixed θ_0 but instead $H_a : \theta \neq \theta_0$, then we should reject H_0 is the test statistic, $\hat{\theta}$, is far away from θ_0, on either side. More precisely, given a level α, we want to find a region such that the probability the test statistic lies in this region when H_0 is true is α. So the picture is:

Figure 1.2: The pdf of $\hat{\theta}$ when H_0 holds, with RR

Since we are in the large sample regime, we assume the estimator for θ, $\hat{\theta}$, is normally distributed. When we furthermore assume H_0 holds, then it's mean is at θ_0, as shown in the picture. The area of each individual shaded region is $\alpha/2$. As stated above $k_1 = \theta_0 - z_{\alpha/2} \sigma_{\hat{\theta},0}$ and $k_2 = \theta_0 + z_{\alpha/2} \sigma_{\hat{\theta},0}$.

Let's do an example to estimate p.

EXAMPLE 1.2. *A campaign wants to know if its candidate is currently favored in an election. To test this it polls 400 constituents and asks is they support the candidate or not, 220 responded that they do. Is there sufficient evidence to support the claim that their candidate is favored? Use $\alpha = .025$.*

Solution: Let p be the proportion of the population that supports the candidate. Then we have:
$H_0 : p = .5$
$H_a : p > .5$
The test statistic $\hat{p}$ is the proportion of those sampled that support the candidate (in this case $\hat{p} = 220/400$).
Now we compute the rejection region, under H_0, $\hat{p}$ is a Binomial(400,.5) random variable divided by 400, which we approximate by a normal random variable with mean $1/2$ and variance $\frac{1/2*1/2}{400} = 1/1600$. The lower bound of RR is then k such that

$$\mathbb{P}\left(\frac{\hat{p} - 1/2}{\sqrt{1/1600}} > \frac{k - 1/2}{\sqrt{1/1600}} \right) = .025$$

or

$$\frac{k - 1/2}{\sqrt{1/1600}} = z_{.025} = 1.96$$

rearranging gives:
$$k = .549$$

So RR is
$$\hat{p} > .549$$

We observe $\hat{p} = .55$. Since this is greater that .549, we reject H_0 in favor of H_a and conclude that at a $\alpha = .025$ level the candidate is favored.

Just like in Chapter 8, we are also interested in determining if one population has a larger mean that the other. From a hypothesis testing point of view this will work very similarly to the last two examples, we just need to replace the estimator for a single mean with the estimator for the difference of means and remember how to compute the standard deviation of this estimator.

This is best seen by an example:

EXAMPLE 1.3. *Two soil types are measured for their shear strength and the following data is collected:*

	Type I	Type II
Sample size	30	35
Sample mean	1.65	1.43
Sample standard deviation	.26	.22

Is there sufficient evidence at a $\alpha = .05$ level the two soils have different shear strengths?

Solution: Let μ_1 be the true mean of the first type and μ_2 be the true mean of the second type. Then $\mu_1 - \mu_2$ is their difference. Then we have:

H_0: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 \neq 0$ (Note: we were not trying to determine if one shear strength is stronger, but if they are different, hence the $\neq$.)

The test statistic is $\hat{\mu}_1 - \hat{\mu}_2$, the difference in the empirical means, (In this case we measured $1.65 - 1.43 = .22$).

Now we determine the RR, we need k such that under H_0

$$\mathbb{P}(|\hat{\mu}_1 - \hat{\mu}_2 - (\mu_1 - \mu_2)| \geq k) = .05 = \alpha.$$

Under H_0, $\mu_1 - \mu_2 = 0$, so that term cancels. Since H_0 contains no information about the variance and we are in the large sample size regime, we use the sample standard deviation for the true standard deviation, to get

$$\sigma^2_{\hat{\mu}_1 - \hat{\mu}_2} = .26^2/30 + .22^2/35$$

(Don't forget to divide by the sample size!)

Since we are in the large sample size regime, we assume that $\hat{\mu}_1 - \hat{\mu}_2$ is approximately normal and standardize it to

$$\mathbb{P}\left(\frac{|\hat{\mu}_1 - \hat{\mu}_2|}{\sqrt{.26^2/30 + .22^2/35}} \geq \frac{k}{\sqrt{.26^2/30 + .22^2/35}}\right) = .05 = \alpha.$$

Using the Z-score we need

$$k = \sqrt{.26^2/30 + .22^2/35} \; z_{.025} = \sqrt{.26^2/30 + .22^2/35} * 1.96 = .118$$

(Recall we're doing a two-sided test, so we divide α by 2)

We measured a difference of $.22 > .118$, so your data lies in the rejection region, and we reject H_0 in favor of H_a, that the shear strengths are different.