Chapter 12 – One-Sample z and t Hypothesis Tests

Return to HT (Hypothesis Testing), considered in Ch. 6, where we used the Sign Test; here, we use z- and t-tests. These latter tests are generally more powerful when the underlying parent population is (nearly) Normal.

Printer Cartridge Example (pp.333-7). The CEO will switch to the new cartridge provided it is significantly better than the old one – in the sense the new average MTF (mean time to failure) is appreciably higher than the old mean of 30,000 pgs. The risks of Type I and II errors are spelled out on p.334, so the burden of proof falls upon us to establish that a change is in order. Thus, the null hypothesis is H_0: $\mu = 30{,}000$ and the alternative hypothesis is H_A: $\mu > 30{,}000$. We know from past history that σ is about 7500. Management wants overwhelming proof, so we choose $\alpha = 0.01$ (1% level of significance). We'll take a sample of size n = 100 new cartridges, and since $z_{0.99}$ = 2.326 and hence $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 750$, the Rejection Region (RR) is:

➔ All values of $\bar{x}$ such that $\bar{x} > 30{,}000+2.326*750 = 31{,}744.5$ pgs.

Next, we gather our data. In our sample, we obtain: n = 100, $\bar{x}$ = 31,500 pages, and s = 7000. Notice that our $\bar{x}$ does not fall in the RR, so our decision is to *retain (fall to reject) the null hypothesis*. We have not convinced the CEO to switch to the new cartridge.

Additional Notes:

(1) Instead of finding the RR, the more common approach is to find the p-value (which here is p = Pr{Z > 2} = 0.02275), compare it with α, and decide (as above);

(2) The data here are failure times, which are notoriously skewed, but since our sample size is so large (100), the Normal calculations are justified here by the Law of Averages and our Chapter 11 results;

(3) Had we used the t test approach (see below), instead of Z = 2, our test statistic would have been t = 1500/700 = 2.14 (since $\frac{s}{\sqrt{n}}$ = 700), and the p-value would have been p = Pr{t_{99} > 2.14} = 0.01741 → same conclusion since p > α;

(4) In this instance, we could not have committed a Type I error (why?), but we may have committed a Type II error (why?) – the latter possibly since we chose a small α or (more likely) a small n … which brings us back to power.

Power Reprise – Recall that power calculations play the "what if" game of wonder what is the probability of rejecting the null hypothesis (H_0) when in fact the alternative hypothesis (H_A) is true. But H_A being true means that μ > 30,000, which is rather vague. So, we have to find the power at a specific value of μ; let's do so at μ = 31000. Power(31000) = Pr{$\bar{x}$ > 31,744.5 given μ = 31000, σ = 7500} = Pr{Z > 0.993} = 0.1604. With n = 100, the power (ability to see a difference when one exists) is only 16.04%.

Had the sample size instead been n = 400, then the RR would be {all values of $\bar{x}$ such that $\bar{x}$ > 30,000+2.326*375 = 30,872.5 pgs}, and Power(31000) = Pr{Z > -0.341} = 0.6333 = 63.33%. So, keeping all other things equal, increasing the sample size increases the power.

12.2. One- and Two-Sided HT's – In the above example, we were only interested in *improvements* in the production process, so H_A was of the one-tailed form, μ > 30,000. When we are interested in looking for a difference, H_A needs to be two-tailed and α and the RR needs to be divided in two. Thus, in the above example with α = 1%, the important Z values are $z_{0.005}$ = -2.5758 and $z_{0.995}$ = 2.5758 (Table 3).

Printer Cartridge Example (pp.333-7). Three cards with one success means that H_0: π = ⅓ (random guessing) and H_A: π ≠ ⅓ (ESP). We need extremely overwhelming proof here so choose α = 0.001. Let's

use the p-value approach here (see text p.341 for RR approach). We'll run the experiment n = 150 times for the ESP candidate. Then, assuming random guessing, the number correctly guessed (K) ~ Binomial(n = 150, π = ⅓), so μ = 50 and σ = $\sqrt{150\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)}$ = 5.7735. Since n is so large, we can use the NA (with CC).

Now we gather our data, and observe k = 64 (not much higher than μ = 50). Here, p = 2*Pr{K ≥ 64} ≈ 2*Pr{Z > (63.5 - 50)/5.7735} = 0.0194. Since p > α, we retain H_0 - we fail to reject the claim of random guessing and no ESP. This does not prove that the person does not have ESP, just that for the chosen α level, it has not been established.

<u>12.3. The one-sample t-test</u> – When we use s in place of σ in the calculation of the TS, and (a) the sample size is small but the parent population is Normal, (b) the sample size is large (e.g. over 40), we use the t-test instead of the z-test.

<u>Ex. 12.5 – Return to Exercise 6.9 (p.175, where we used the Sign Test)</u>
In this before-and-after study, we again look at the differences: previously, we looked at the signs only, now we will look at the actual (Before – After) values. Here, H_0: μ = 0 and H_A: μ > 0, and let's choose α = 0.05 (5%). Here n = 5, so we must assume these differences come from a Normal distribution.

Now for the data: the differences are 5, 5, 4, 0, and 1, so $\bar{x}$ = 3, s = 2.3452, and the TS is t_4 = 2.8604. The p-value is 0.02296 from a computer, or 0.02 < p < 0.05 from Table 4. Since p < α, reject H_0 at the 5% level and conclude that the average pulse rate has dropped after the physical fitness program.

<u>12.4. Which is the Null and which is the Alternative?</u>
*** The null states the status quo (no change) and contains the '=' sign;**
*** The alternative is usually what we want to show.**