Hypothesis Testing with the t-distribution

Ryan Miller

The Null Distribution of a Single Mean

Consider the null hypothesis $H_0 : \mu = \mu_0$, and the following CLT result:

$$\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

- Recall that this approximation is only accurate when σ, the *population* standard deviation of the variable X, *is known*
- Otherwise replacing σ with s, the *sample* standard deviation, introduces additional uncertainty due to the variability in how accurately s estimates σ

The One-sample t-test

- Assuming for a minute that we know σ (something that is never true in practice), we could use the CLT result to derive a z-test involving the test statistic:

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$$

- If σ is estimated by s, William Gosset derived the t-distribution to properly account for the extra variation introduced by estimating σ
- This gives rise to the t-test, which uses the following *test statistic*:

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$$

- Assuming $H_0 : \mu = \mu_0$ is true, the test statistic, t, follows a t-distribution with $df = n - 1$

Practice - z-test vs. t-test

- Radon is a toxic gas and the second leading cause of lung cancer, it is particularly prevalent in the midwest in states including Iowa
- The EPA has a federal action limit of 4 pCi/L, but they advise residents to take action if average levels higher than 0.4 pCi/L are observed
- In the basement of an Iowa home, radon levels are tested on 8 randomly selected dates over the course a month, the measurements are:

$$\{.2, .7, .3, .9, .5, .3, .7, .6\}$$

1. Assuming $\sigma = 0.243$ is known, use a z-test to evaluate whether the average radon level *is higher* than 0.4
2. Using s to estimate σ, use a t-test to evaluate whether the average radon level *is higher* than 0.4
3. How do the p-values from these tests compare?

Practice - Solution

1. $H_0 : \mu = 0.4$ and $H_A : \mu > 0.4$; $z = \frac{0.525-0.4}{0.243/\sqrt{8}} = 1.455$; the one-sided p-value (using N(0,1)) is 0.073
2. $H_0 : \mu = 0.4$ and $H_A : \mu > 0.4$; $t = \frac{0.525-0.4}{0.243/\sqrt{8}} = 1.455$; the one-sided p-value (using $\mathrm{t}(\mathrm{df} = 7)$) is 0.094
3. The t-test has a larger p-value because the uncertainty introduced by estimating σ increases the variability of possible test statistics we might encounter when H_0 is true.

Conditions for the t-test

- In practice, we will *never* use the z-test when testing a single mean
  - In this situation, we'll mostly use the t-test; however, the t-test is based upon normality
- The t-test requires a normally distributed population, *or* a sample size of $n \geq 30$
  - When these conditions aren't met, we should prefer a *randomization test*

Practice - *t*-test vs. Randomization Test

The "Home Prices - Canton" dataset in StatKey (under Randomization Test for a Mean) displays the sale prices of a random sample of 10 homes in Canton, NY

1. Comparing the sample mean and median, do you believe the distribution of sale prices in the population (all homes in Canton, NY) is approximately normally distributed?
2. Use a randomization test to determine whether the mean home price in Canton *is lower* than 200k
3. Use a *t*-test to determine whether the mean home price in Canton *is lower* than 200k
4. How do the results of these tests compare?

Practice - Solution

1. The mean is much larger than the median, so the data are likely from a right-skewed population
2. The one-sided *p*-value is approximately 0.02 (using the randomization distribution)
3. Here $t = \frac{146.8 - 200}{94.998/\sqrt{10}} = -1.77$; leading to a one-sided *p*-value of 0.055
4. These results are very different because the assumptions of the *t*-test are not met

The Two-sample t-test

Recall that for a difference in means, the CLT suggests the following normal approximation:

$$\bar{x}_1 - \bar{x}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

- This approximation can be used when testing the hypothesis: $H_0 : \mu_1 - \mu_2 = 0$
- If σ_1 and σ_2 are estimated from the sample, the test uses a t-distribution with $df = \min(n_1 - 1, n_2 - 1)$
  - This approach to determining df is conservative, software will use the exact value

Practice - Two-sample t-test

- After swimming several world records were broken at the 2008 Olympics, controversy arose regarding whether new swimsuit designs provided an unfair advantage
- In 2010, new international rules were created that regulated swimsuit coverage and material, but does a certain swimsuit really make a swimmer faster?
- The wetsuits data (available on the course website) contains 1500m swim velocities of 12 competitive swimmers with and without a wetsuit

1. Perform a two-sample t-test on swim velocities in the wetsuits data to evaluate whether wearing a wetsuit results in a higher swim velocity than not wearing a wetsuit. You should construct your test statistic by hand, but use Minitab to find the sample means and standard deviations.

Practice - Solution

$H_0 : \mu_{wetsuit} = \mu_{noWetsuit}$ vs. $H_A : \mu_{wetsuit} > \mu_{noWetsuit}$

We observed $\bar{x}_{wetsuit} - \bar{x}_{noWetsuit} = 1.51 - 1.43 = 0.08$

Then, $t = \frac{0.08 - 0}{\sqrt{.14^2/12 + .12^2/12}} = \frac{.08}{.053} = 1.5$

Using a t-distribution with $df = 11$, this test statistic leads to a one-sided p-value of 0.081; thus we conclude that there is marginal evidence that wetsuit is associated with higher swim velocity, but we might not be fully convinced.

Paired Designs

- The way in which the wetsuit data were collected is special, the researchers used a **paired design**
- Paired designs offer a statistical advantage because each subject can serve as their own control, block out variability that is attributable to individual differences unrelated to the wetsuit
  - That is, by analyzing *within subject change* we eliminate a lot of variation in swim velocities
- We can further understand the statistical advantage of a paired design by comparing standard errors:

$$\sqrt{\frac{\sigma_1^2}{n_1}} < \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \sqrt{\frac{\sigma_1^2}{n_1}} + \sqrt{\frac{\sigma_2^2}{n_2}}$$

- A single variable (the paired difference) has less variability than the difference in two variables, and much less variability than the two variables by themselves

The Paired t-test

The **paired t-test** is actually just a one-sample t-test done on the paired differences:

$$t = \frac{\bar{x}_d - \mu_0}{s_d / \sqrt{n_{pairs}}}$$

- $\bar{x}_d$ is the sample average of the paired differences
- μ_d is the average paired difference specified in the null hypothesis (typically zero)
- s_d is the sample standard deviation of the paired difference
- n_{pairs} is the number of pairs in the sample
- This test statistic follows a t-distribution with $df = n_{pairs} - 1$

Practice - Paired *t*-test

For the *Wetsuits data*, use Minitab to:

1. Create a new column containing the paired differences. Then use this column and the "1-sample t . . . " menu to evaluate whether swim velocities are faster wearing a wetsuit
2. Use the "Paired t . . . " menu to conduct a paired *t*-test on the original variables "Wetsuit" and "NoWetsuit" evaluating whether swim velocities are faster wearing a wetsuit
3. Recall that the two-sample *t*-test resulted in a one-sided *p*-value of 0.081, compare this with the results of the paired *t*-test

Practice - Solution

1. On Minitab
2. $H_0 : \mu_d = 0$, $t = 12.3$, p-value is approximately zero
3. The paired test leads to a much smaller p-value, this is because between subject differences are accounted for by the design of the study (leading to a much more precise estimate of the difference that is attributable to the wetsuit)

Reminder - Assumptions for the t-test

- Recall that the t-distribution was derived assuming a normally distributed population
- When the sample size is large, we don't need to worry about this because the sampling distribution of $\bar{x}$ approaches normality
- When the sample size is small, we must check whether it is reasonable to assume the data come from a normal distribution
- If $n_1 < 30$ or $n_2 < 30$, or if there is clear skew in our sample(s), we should prefer a randomization test

Conclusion

Right now you should...

1. Be able to perform t-tests in one and two sample settings
2. Know the assumptions involved with these approaches
3. Understand why the t-distribution is necessary when σ is estimated
4. Know when to conduct a paired t-test, and the advantage it provides over a two-sample t-test

These notes cover parts of Ch 6 from the textbook, I encourage you to read through the chapter and its examples