The T-test (one-sample quantitative data)

Ryan Miller

Introduction

Previously, we introduced the Z-test:

1) State the null hypothesis (a conjecture about the population that would be useful to disprove)
2) Use the null hypothesis (and corresponding null model) to find a Z-value describing the *sample estimate*
3) Locate the Z-value on the Standard Normal curve to find the p-value
4) Use the p-value to make a decision regarding the null hypothesis

Introduction

Previously, we introduced the Z-test:

1) State the null hypothesis (a conjecture about the population that would be useful to disprove)
2) Use the null hypothesis (and corresponding null model) to find a Z-value describing the *sample estimate*
3) Locate the Z-value on the Standard Normal curve to find the p-value
4) Use the p-value to make a decision regarding the null hypothesis

This approach needs to be adjusted for scenarios involving means (since there is added uncertainty induced by estimating an extra parameter), the resulting procedure is known as the T-test

The T-test

- Procedurally, only difference between the T-test and Z-test is the probability distribution used to calculate the p-value
  - When analyzing one-sample *categorical* data, the Z-test compares $z = \frac{\hat{p}-p}{SE}$ to the Standard Normal distribution
  - When analyzing one-sample *quantitative* data, the T-test compares $t = \frac{\bar{x}-\mu}{SE}$ to a t-distribution with $df = n - 1$

Example

- According to national data collected by the Australian government, the mean birthweight of all babies born in Australia is 7.86 lbs
- A hospital in Missouri reports the average birthweight of 112 born there last year was 7.68, with a sample standard deviation of 1.31
- Assuming the Missouri hospital is a representative of all babies born in the US, do these data support the hypothesis that birthweight of US babies is different from that of Australian babies?

Example (solution)

1) $H_0 : \mu = 7.86$ vs $H_A : \mu \neq 7.86$

Example (solution)

1) $H_0 : \mu = 7.86$ vs $H_A : \mu \neq 7.86$
2) Noting that we observed $\bar{x} = 7.68$, and $SE = \frac{s}{\sqrt{n}}$, we calculate $t = \frac{7.68 - 7.86}{1.31/\sqrt{112}} = -1.45$

Example (solution)

1) $H_0 : \mu = 7.86$ vs $H_A : \mu \neq 7.86$
2) Noting that we observed $\bar{x} = 7.68$, and $SE = \frac{s}{\sqrt{n}}$, we calculate $t = \frac{7.68 - 7.86}{1.31/\sqrt{112}} = -1.45$
3) We next must locate $t = -1.45$ on a t-distribution with $df = n - 1 = 111$ using StatKey

Example (solution)

1) $H_0 : \mu = 7.86$ vs $H_A : \mu \neq 7.86$
2) Noting that we observed $\bar{x} = 7.68$, and $SE = \frac{s}{\sqrt{n}}$, we calculate $t = \frac{7.68-7.86}{1.31/\sqrt{112}} = -1.45$
3) We next must locate $t = -1.45$ on a t-distribution with $df = n - 1 = 111$ using StatKey
4) The two-sided p-value is 0.15, so we conclude insufficient evidence to believe the mean birthweight of babies in the US differs from that of Australia

Comparison vs. The Z-test

- As previously mentioned, the only difference between the T-test and Z-test is null distribution
  - We use the T-test to account for the small amount of additional uncertainty introduced when using the sample standard deviation, s, to estimate σ, the standard deviation of the population

Comparison vs. The Z-test

- As previously mentioned, the only difference between the T-test and Z-test is null distribution
  - We use the T-test to account for the small amount of additional uncertainty introduced when using the sample standard deviation, s, to estimate σ, the standard deviation of the population
- Thus, we know the p-value of a T-test will always be higher than that of the corresponding Z-test
  - Here, the p-value is 0.15 comparing our T-value of -1.45 to the $t_{df=111}$ distribution
  - If we compared -1.45 to the Standard Normal curve, we'd get a p-value of 0.148

Conclusion

- The T-test is a modified version of the Z-test that makes the Normal results of CLT suitable for statistical inference on quantitative data
  - You should use the Z-test for hypothesis testing on proportions (categorical data)
  - You should use the T-test for hypothesis testing on means (quantitative data)

Conclusion

- The T-test is a modified version of the Z-test that makes the Normal results of CLT suitable for statistical inference on quantitative data
  - You should use the Z-test for hypothesis testing on proportions (categorical data)
  - You should use the T-test for hypothesis testing on means (quantitative data)
- The next presentation will provide an overview of the assumptions behind both the Z-test and T-test, as well as some alterantive ways to conduct statistical inference when these assumptions are not met.