Pivots and t-tests

Patrick Breheny

October 15

Introduction

- As we discussed previously, W.S. Gossett derived the t-distribution as a way of addressing the known small-sample problems of the z-test
- Today, we will discuss Gossett's improved test, known as the t-test, and its associated confidence interval
- Historical note: Gossett's employers had him publish under the pen name "Student" because they didn't want the competition to know how useful his results could be
- Because of this, the t-test is often referred to as "Student's t-test", and the one-sample version we will discuss today is often called the "one-sample t-test" or, when applied to paired data, the "paired t-test"

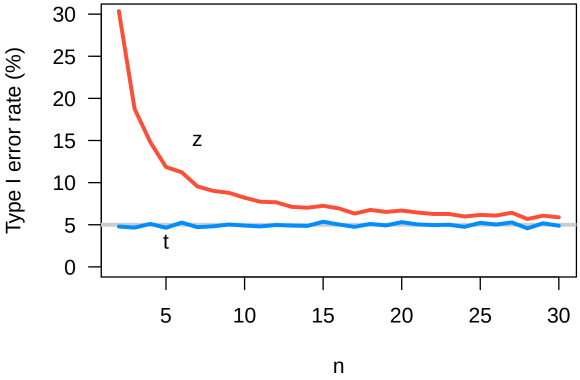
The t-test procedure

- The procedure for carrying out a one-sample t-test is exactly the same as that for the z-test, except for the distribution to which we compare the test statistic
- The test statistic itself is unchanged:

$$t = \frac{\bar{x} - \mu_0}{SE},$$

  where $\mathrm{SE} = s/\sqrt{n}$
- To obtain a p-value, however, we need to calculate tail probabilities based on the CDF of the t distribution with $n-1$ degrees of freedom instead of the standard normal CDF

Does the t-test fix the z-test's problem?



The t-test performs perfectly, although recall that the underlying distribution was indeed normal here

FVC example

- In the cystic fibrosis experiment, the mean difference in FVC reduction (placebo − drug) was 137, with standard deviation 223:
  - #1 $SE = 223/\sqrt{14} = 60$
  - #2 $t = (137 - 0)/60 = 2.28$
  - #3 The area outside ± 2.28 on the Student's curve with 13 degrees of freedom is $2F_t(-2.28|13) = 0.04$, where $F_t(x|\nu)$ is the CDF of the t-distribution with ν degrees of freedom
- Our p-value from the z-test was 0.02, which as the simulations indicated, overstates the evidence against the null hypothesis

z-tests vs. t-tests

- For reasonably large sample sizes (> 50), the z- and t-tests are essentially the same
- However, it is difficult to justify z-tests and z-confidence intervals – as we have seen, their p-values and coverage probabilities are not correct even in the best case scenario of perfectly normal data
- So, in practice, no one uses z-tests for one-sample, continuous data; t-tests, on the other hand, are probably the most common type of statistical test on the planet

Binomial vs. t-tests

- The t-test fixes an important problem with the z-test (correcting for the uncertainty in the sample standard deviation), but its fix is based on the data being normally distributed
- If the sample size is small and the data is skewed, the t-test may be questionable
- On the other hand, the binomial test from earlier made very minimal assumptions about the data (basically, just the assumption of independence between patients)

Binomial vs. t-tests (cont'd)

- Recall that when we used the binomial test, we calculated a p-value of $.06$ (as opposed to the t-test p-value of $.04$)
- Neither one is wrong, they are just two different ways of performing the hypothesis test, and in fact are testing slightly different hypotheses
- Each approach has advantages and disadvantages
    - The binomial test makes fewer assumptions
    - The paired t-test is generally more powerful than the binomial test, assuming its assumptions are met

Confidence intervals based on the t-distribution

- The t-distribution can also be used to construct confidence intervals
- Letting $t_{n-1,\alpha}$ denote the value that contains the middle $100(1-\alpha)$ percent of the t distribution with $n-1$ degrees of freedom,

$$\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$$

  implies that

$$P(\mu \in [\bar{X} \pm t_{n-1,\alpha}\mathrm{SE}]) = 0.95$$

- Note the high degree of similarity with the Wald interval for proportions, although for the t intervals, this isn't just an approximation – the SE really is completely independent of the mean

FVC example: Patients taking drug

- For patients taking the drug in the cystic fibrosis crossover experiment, the mean reduction in FVC was 160, with standard deviation 197
- Let's calculate a 95% confidence interval for the average reduction in lung function that individuals with cystic fibrosis in the population would be likely to experience over a 25-week period, if they took this drug:
  - #1 The standard error is $197/\sqrt{14} = 53$
  - #2 The values ± 2.16 contain the middle 95% of Student's curve with 13 degrees of freedom
  - #3 Thus, my confidence interval is:

$$\begin{aligned}(160 - 2.16 \cdot 53, 160 + 2.16 \cdot 53) \\ = (46, 274)\end{aligned}$$

FVC example: Patients taking placebo

- For patients taking the placebo, the mean reduction in FVC was 296, with standard deviation 297
  - #1 The standard error is $297/\sqrt{14} = 79$
  - #2 The values ± 2.16 still contain the middle 95% of Student's curve with 13 degrees of freedom
  - #3 Thus, my confidence interval is:

$$\begin{aligned}(296 - 2.16 \cdot 79, 296 + 2.16 \cdot 79) \\ = (125, 467)\end{aligned}$$

Comparing drug and placebo

- Note that our two confidence intervals $\{[46, 274]; [125, 467]\}$ overlap quite a bit
- On the surface, this would seem to indicate a lack of evidence that the drug is effective
- However, recall that paired designs are powerful ways to reduce noise; constructing separate confidence intervals does not take advantage of this design
- To assess whether drug is more effective than placebo, we should instead construct a single confidence interval for the **difference** in FVC reduction for each patient

FVC example: Difference between two treatments

- The mean difference in reduction in FVC (placebo − drug) was 137, with standard deviation 223
  - #1 The standard error is $223/\sqrt{14} = 60$
  - #2 Once again, the values ± 2.16 contain the middle 95% of Student's curve with 13 degrees of freedom
  - #3 Thus, the confidence interval is:

$$
\begin{aligned}
&(137 - 2.16 \cdot 60, 137 + 2.16 \cdot 60) \\
&\quad = (7, 267)
\end{aligned}
$$

- This gives us a range of likely values by which taking the drug would slow the decline of lung function in cystic fibrosis patients
- Note that all of the values are positive, indicating benefit from taking the drug, which agrees with the hypothesis test

Pivots: Introduction

- The methods we have discussed today are all derived from the relationship:

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

- We used this to test hypotheses, but even more impressively, it allowed us to directly obtain confidence intervals without having to invert hypothesis tests
- Note: You still *could* invert the t-test to construct confidence intervals, but there's no need – you'd get the same interval either way

Pivotal quantities

- This is an example of what is known in statistics as a pivotal quantity, or pivot
- A *pivotal quantity* is a function of observable data and unobservable parameters whose distribution does not depend on any unknown parameters
- Pivots are the other standard approach to constructing confidence intervals (along with inverting a hypothesis test)

Confidence interval for the variance

- To see another example of how this works, consider a different pivotal quantity that we encountered in the previous lecture:

$$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$$

- This allows us to immediately construct confidence intervals for the variance using the quantiles $\chi^2_{n-1,\alpha/2}$ and $\chi^2_{n-1,1-\alpha/2}$ containing the middle $100(1-\alpha)$ percent of the χ^2_{n-1} distribution:

$$\sigma^2 \in \left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}\right]$$

and of course, by taking square roots, we have an interval for σ

Confidence interval for the variance: Example

- So, for example, the observed sample standard deviation for the decline in FVC scores for the 14 patients on the drug was 197
- Plugging this into the formula on the previous slide (and taking square roots), we obtain the 95% interval $[143, 317]$
- Likewise, for the placebo group ($s = 297$) we have the 95% interval $[215, 479]$

Summary

- Using the t distribution produces hypothesis tests with proper type I error rates and confidence intervals with proper coverage (at least for normally distributed data) by accounting for uncertainty in the estimated standard deviation
- Know how to calculate these tests and intervals
- Whether two confidence intervals overlap does not tell you whether an interval for the difference would include zero
- Pivotal quantities are functions of observable random variables and unknown quantities whose distribution does not depend on the unknown quantities; they are very useful in constructing confidence intervals