Classic Inference

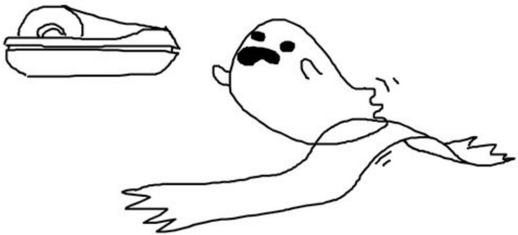
Javier E. Flores

March 8, 2019

Sixth Sense?

In 1973, the results of a study investigating after-death encounters among individuals in Greater Los Angeles were published in the *Journal for the Scientific Study of Religion*.

Being the paranormal fanatic that I am, I've personally sought out encounters of this kind and even have a still image of a ghost caught on tape!



Sixth Sense?

Anyway, one goal of this 1973 study was to estimate the prevalence of after-death encounters among individuals surveyed.

Of the 434 respondents, the authors found that $193/434 = 44\%$ claimed to have encountered - either through a seance, dream, or some other way - someone who had passed away.

If we were the consulting statistician for this study back in 1973, we would have a few tools that would prove useful to the investigating authors.

Single Proportion

If the authors were interested in obtaining an interval estimate of the prevalence, we would be able to generate a bootstrap distribution and compute confidence intervals.

- Last time, we learned that normal approximations to bootstrap distributions may also be used to obtain similar intervals.

If the authors were interested in testing whether the majority of individuals have had after-death encounters, we would know to perform a randomization test.

- We also learned that we may perform z-tests, which rely on normal approximations to the randomization distribution.

In this lecture, we'll learn of analytic expressions for the standard errors of several sample statistics - beginning with the single proportion.

Single Proportion

Earlier I mentioned that the 1973 study found that 193/434 people claimed to have had an after-death encounter. This proportion is our **sample statistic** and is represented by $\hat{p}$. (i.e. $\hat{p} = 0.44$)

The **parameter of interest** for this study is p, the proportion of individuals who have had an after-death encounter.

With a size of 434, our sample is sufficiently large to use a normal approximation. Using this approximation, we compute a 95% confidence interval for p:

$$\hat{p} \pm 1.96SE$$

While we could use the standard error of a generated bootstrap distribution to obtain the SE, it would be much more convenient to use some analytic expression instead...

Single Proportion

Fortunately for us, statisticians of days past put in the time to come up with such an expression for us.

These clever statisticians found that the standard error of the sample proportion, $\hat{p}$, is:

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

From this we see that the standard error is a function of the population parameter, p, and sample size, n!

Question: What happens to the standard error here as n gets larger?

Question: How do we use this expression when we don't even know p?!?

Single Proportion SE

Remember, our sample statistic $\hat{p}$ is **unbiased** for the population parameter p.

Given this fact, the $P\%$ confidence interval estimate of p is:

$$\hat{p} \pm z_{crit}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

where you'll recall that z_{crit} is the critical value defining the cutoffs which contain the middle $P\%$ of the standard normal distribution.

When performing a z-test, we are assuming some null distribution. Therefore, rather than use $\hat{p}$ in lieu of p for the SE, we use the null hypothesized value, p_0:

$$z_{test} = \frac{\text{Sample Statistic} - \text{Null Value}}{SE} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Practice

Recall that in the after-death encounter study, 193/434 people claimed to have had an encounter. With your group:

1) Calculate the 95% confidence interval estimate of p.
2) Conduct a hypothesis at the $\alpha = 0.1$ level investigating whether the proportion is less than 50% (use Minitab to calculate the p-value using the standard normal distribution).
3) Compare your results to bootstrapping/randomization in StatKey.

Solution

95% Confidence Interval:

$$\hat{p} = 0.44; \quad SE = \sqrt{\frac{0.44(1-0.44)}{434}} = 0.024$$

$$0.44 \pm 1.96 * 0.024 = (0.393, 0.487)$$

Hypothesis Test: $H_0 : p \geq 0.50 \quad H_A : p < 0.50$

$$z_{test} = \frac{0.44 - 0.50}{\sqrt{\frac{0.50(1-0.50)}{434}}} = -2.500 \qquad \text{p-value} = 0.006$$

We reject H_0 at the $\alpha = 0.1$ level. There is strong evidence that the majority of individuals have not had an after-death encounter in the Greater Los Angeles area.

Practice #2

In a study investigating the survival of premature babies, researchers at Johns Hopkins found that 0/29 babies born at 22 weeks survived at least 6 months. Assuming a sample size of 29 is sufficiently large for a normal approximation,

1) Calculate the 95% confidence interval estimating the survival proportion of babies born at 22 weeks.
2) Do you believe that this interval actually has 95% coverage? Why or why not?
3) Suppose a new sample was obtained and 1/29 babies were found to have survived. Calculate the 95% confidence interval. Is the resulting interval practically reasonable?

Solution

95% Confidence Interval (#1):

$$\hat{p} = 0; \quad SE = \sqrt{\frac{0(1-0)}{29}} = 0$$

$$0 \pm 1.96 * 0 = (0, 0)$$

This interval has a width of zero - it's no better than a point estimate! This "interval" definitely does not have 95% coverage.

95% Confidence Interval (#2):

$$\hat{p} = 0.034; \quad SE = \sqrt{\frac{0.034(1-0.034)}{29}} = 0.034$$

$$0.034 \pm 1.96 * 0.034 = (-0.033, 0.101)$$

This interval suggests a negative chance of survival. That makes no sense!

When Things Break Down

Aside from requiring a sufficiently large sample, the standard error formulas for a single proportion assume that the true population parameter is not close to 0 or 1.

The previous example demonstrates what happens when this assumption is violated - things stop making sense!

One way to check beforehand whether you can expect the standard error formulas to work reasonable well is to compute $n * p$ and $n * (1 - p)$. Be sure to use $\hat{p}$ or p_0 in place of p depending on whether confidence intervals or hypothesis testing is of interest.

If $n * p \geq 10$ and $n * (1 - p) \geq 10$, you can expect the formulas to work well.

Summary

We can conduct statistical inference on a single proportion p using the sample estimate $\hat{p}$ and standard error (SE) given by:

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

Given that we don't actually know the value of p, we substitute it in the formula for either $\hat{p}$ or p_0 depending on whether we are constructing a confidence interval or performing a hypothesis test.

This normal approximation works well only when $n * p \geq 10$ and $n * (1-p) \geq 10$.

Otherwise, we can still use the bootstrapping and randomization methods we've learned previously.

Single Mean

As we saw for a single sample proportion, there exists an expression for the standard error of a single sample mean:

$$SE = \frac{\sigma}{\sqrt{n}},$$

where σ represents the population standard deviation and n is the size of your random sample.

Similar to the formula for the standard error of a single proportion, we see a dependency of this standard error on the size of our sample.

Question: Given that we typically don't know σ, what would be a reasonable thing for us to do in order to use this formula?

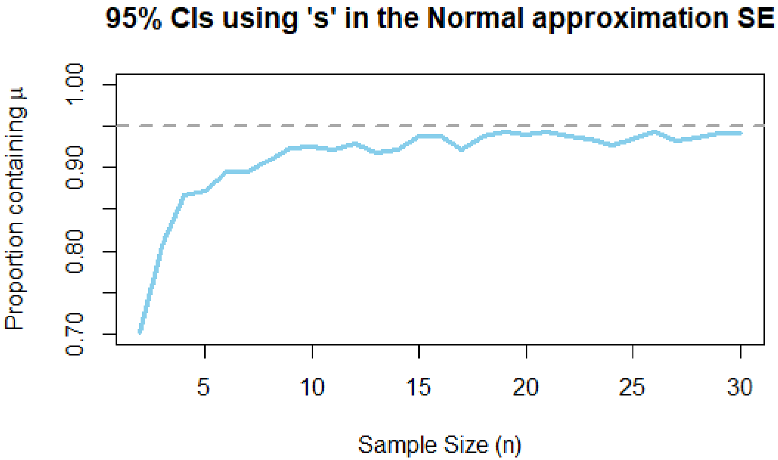
Single Mean

A natural substitute for σ might be the sample standard deviation, s.

Just as we know that the sample mean, $\bar{x}$, is unbiased for the population mean, μ, so too is the sample standard deviation , s, for the population standard deviation, σ.

However in using s to replace σ, things don't work out exactly as expected...

No s Bueno


95% CIs using 's' in the Normal approximation SE
Proportion containing μ
0.70
0.80
0.90
1.00
5
10
15
20
25
30
Sample Size (n)

Single Mean

From the previous figure we see that, when we use s in place of σ, the resulting inference is biased.

Our 95% confidence interval, which should have 95% of all constructed intervals contain the target parameter (μ), does not demonstrate this confidence level.

With smaller samples in particular, the true 'coverage' rate of our intervals is often much lower!

William Gosset

Credit for discovering the flaw in using s in lieu of σ goes to William Gosset, who was a chemist who worked for Guinness Brewing in the 1890s.

While I am probably stating the obvious, statistical computing wasn't exactly around at this time so the fact that Gosset was able to discover this problem without simulation is truly remarkable!

Gosset's experiences at Guinness are what prompted him to investigate the statistical validity of small sample results.

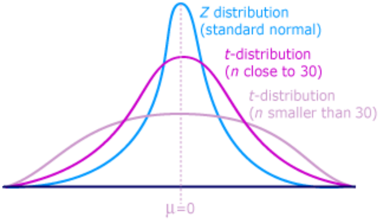
In his investigation, Gosset derived a modified distribution known as the **t-distribution** which proved to lead towards a solution.

The Problem

The issue with using s in the computation of the standard error is that by doing so, we assume that the actual standard error is known rather than estimated.

Making this assumption underestimates the actual degree of uncertainty we have about the population.

Gosset showed that when s is used the statistic $\frac{\bar{x}-\mu}{s}$ does not follow a standard normal distribution, but rather a (Student's) t-distribution.



t-distribution

As implied by the previous figure, the shape of the t-distribution depends on the sample size.

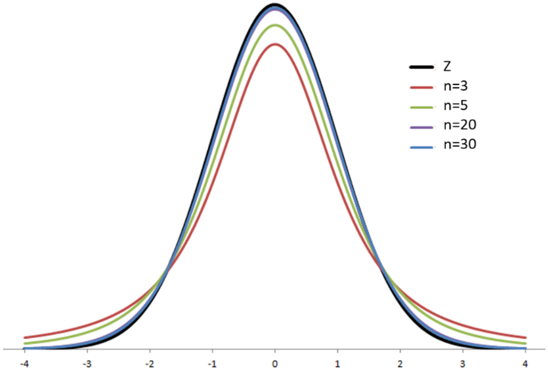
More precisely the t-distribution depends on the sample size through the **degrees of freedom**, df, which is defined as the sample size minus 1 (i.e. $df = n - 1$).

The t-distribution assumes that the original population is normal.

The t-distribution is relatively **robust** to this assumption so we often don't worry about it unless there are substantial deviations from normality that are observed.

t-distribution

As the sample size/degrees of freedom increase, the t-distribution approaches the standard normal distribution.



t-distribution

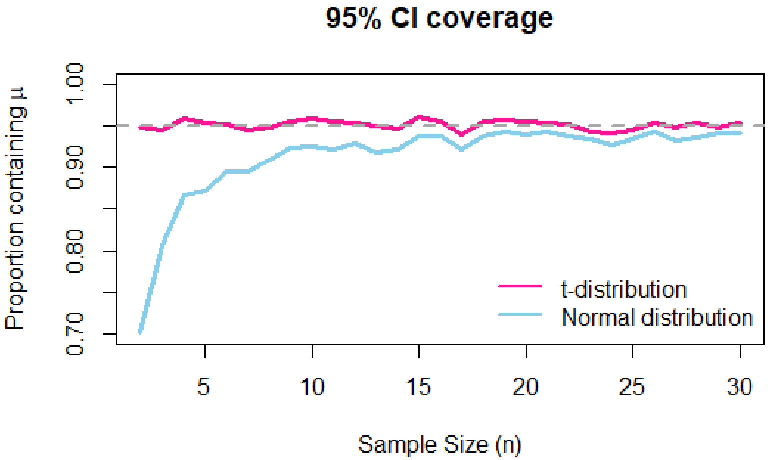
With smaller sample sizes, the tails of the t-distribution are much thicker than the tails of the standard normal distribution.

This is a consequence of the fact that, with smaller samples, there is greater uncertainty in using s as an estimator of σ.

With larger samples, this uncertainty diminishes which is why the t-distribution becomes closer to the standard normal. A sample size of 30 is usually sufficient in size to treat the t-distribution as normal.

Areas under the t-distribution and confidence interval critical values can be found using either Minitab or Statkey.

t-distribution



Practice

Use the "Arsenic in Chicken" dataset in Statkey to:

1) Find a 95% confidence interval estimate for the mean amount of arsenic using the t-distribution.
2) Find a 95% confidence interval estimate for the mean amount of arsenic using the normal distribution.
3) Test whether the population mean differs from 80 using the t-distribution. Report your p-value and conclusion.
4) How would your p-value differ if you performed a z-test? (Don't actually perform a z-test, just discuss what you think would happen)

Solution

1) $91 \pm 2.571 * \frac{23.47}{\sqrt{6}} = (66.37, 115.63)$

2) $91 \pm 1.96 * \frac{23.47}{\sqrt{6}} = (72.22, 109.78)$

3) $H_0 : \mu = 80; \quad H_A : \mu \neq 80$

$$t_{test} = \frac{91 - 80}{\frac{23.47}{\sqrt{6}}} = 1.15$$

Using a t-distribution with 5 degrees of freedom (since $n = 6$), we compute the two-sided p-value and obtain 0.302. We fail to reject the null hypothesis and conclude that there is insufficient evidence to support the claim that the mean arsenic level differs from 80 ppm.

4) Since the t-distribution is characterized by thicker tails relative to the normal distribution, we'd expect the p-value of a z-test to give a smaller p-value. (You can confirm this yourself: the p-value you should obtain is 0.250)

Minitab

The t-test performed in the previous example may also be done in Minitab:

First, enter or copy/paste the quantitative data of interest into a Minitab column.

Next, select **Stat** -> **Basic Statistics** -> **One-sample t-test**.

Finally, select the variable corresponding to the column of data and enter the hypothesized null value.

Note: Assuming the data appear somewhat normally distributed, **always** use the t-distribution when testing with small sample sizes. The results from a $t(df = 5)$ distribution are substantially different relative to the standard normal.

Practice

Use the "Home Prices - Canton" dataset in Statkey to answer the following questions:

1) Are these data from a normally distributed population? Does the randomization distribution (using $\mu_0 = 200$) appear skewed?

2) Is the mean home price in Canton less than 200k? ($H_0 : \mu \geq 200$) Perform one-sided hypothesis tests using the t-distribution and the randomization method. How do these results compare?

Solution

1) The data appears right skewed and non-normal, largely due to the pricier homes included in the data. The randomization distribution is also right skewed.

2) Using the randomization test, we obtain a p-value of 0.019. In contrast the t-distribution p-value is 0.055. Given that the t-test assumes normality (which we clearly do not have), we should not rely on the p-value it provides.

Note: This analysis may have been performed in Minitab by selecting "Options" on the one-sample t-test menu and specifying the appropriate one-sided alternative hypothesis.

Summary

Data	Parameter	Statistic	SE
One categorical variable	p	$\hat{p}$	$\sqrt{\frac{p(1-p)}{n}}$
One quantitative variable	μ	$\bar{x}$	$\frac{s}{\sqrt{n}}$
One categorical variable with groups	$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	TBD
One quantitative variable with groups	$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	TBD

A Taste of Power

In each of the standard error formulas, we've seen a dependence on sample size.

Specifically, we see that the standard error decreases with an increase in sample size.

In hypothesis testing, we also know that the test statistics used are dependent on these standard errors.

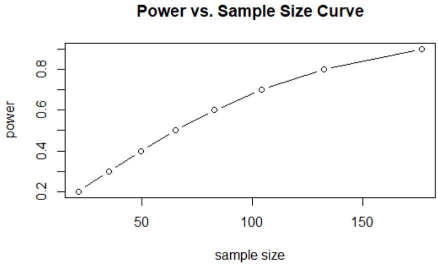
Aside from the sample size we know the test statistics are dependent on the **effect size**, which is the difference between your sample statistic and null value.

All this considered, one natural question is: "How likely is it that I reject my null hypothesis given a certain effect and sample size?"

A Taste of Power

The **power** of a test is the probability of rejecting a false null hypothesis (this is good).

Over our next lab, we'll be investigating power thoroughly.



Wrap-Up

Right now, you should...

- Know when and how to perform z and t tests on single proportions and means.
- Know when and how to construct $P\%$ confidence intervals for single proportions and means.
- Know the assumptions and limitations of each approach (i.e. normal and t-approximations)

These notes cover sections 6.1 and 6.2 of the textbook. Please read through the section and its examples along with any links provided in this lecture.