PRE-READING DOCUMENT 2

Statistics 251
Confidence Intervals and Hypothesis Testing

Lab 7

1 Confidence Intervals

In statistics, we always deal with numbers like the average or mean of some variable, e.g. the average weight a beam can handle. These numbers, however, are based on the data we've collected and so one sample dataset could differ from another, especially if the sample sizes are small, and so there's always variability.

This is where confidence intervals come in. If we repeat the experiment over and over again (for a large number of times) and construct a 95% confidence interval for each run of the experiment, then 95% of all such confidence intervals should contain the true value of μ

1.1 Z test

For example, say we have a sample of n observations $x_1, x_2, \cdots, x_n$, and we've assumed that the observations came from a normal distribution with a mean value μ and variance σ^2, and so $X \sim N(\mu, \sigma^2)$. **Keep in mind that** μ

and σ^2 are the true values and come from the population, whereas $x_1, x_2, \cdots, x_n$ are a smaller sample of the population. Now let's say that we are trying to find a 95% C.I. for μ, which we do by using the sample mean $\bar{X}$. First, though, We have to standardize $\bar{X}$ so that it follows the *standard normal distribution*, which we can read from a table.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

When we look for a 95% C.I., we are looking at a *two-tailed* probability of α $= 0.05$ under the standard normal density curve. Keeping this in mind and the fact that a normal distribution is symmetrical about 0, then:

$$P\left(Z_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < Z_{1-\frac{\alpha}{2}}\right) = 0.95$$

or equivalently

$$P\left(-Z_{1-\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < Z_{1-\frac{\alpha}{2}}\right) = 0.95$$

Looking up a Z-value with $\alpha/2 = 0.025$, and $1 - (\alpha/2) = 0.975$, we find the probability to be:

$$P\left(-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$$

and by rearranging the probability, we see that:

$$-1.96 \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$-\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

Therefore a 95% C.I. for μ is:

$$\left(\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Now that we know the general formula for a 95% CI, or any 100(1-α)% CI, let's put it to practice. Let's say that that we want to compute a 95% CI for the true average μ, where $\sigma = 2.0$, n=31, and $\bar{x} = 80.0$, then the interval is:

$$\bar{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}} = 80.0 \pm 1.96 \cdot \frac{2.0}{\sqrt{31}} = 80.0 \pm 0.7 = (79.3, 80.7)$$

1.2 t test

Now this is good for the case when we know the true population variance, but what happens when we don't know what it is? To deal the case of an unknown σ^2, we use the sample variance s^2 and the student t-distribution.

All the steps are the same, except that when we try to figure out our z-value, we actually need to figure out a t-value instead, which is done in the exact same manner as the z-value except that it is based on $(n-1)$ degrees of freedom. This will compensate for the fact that we don't know what σ^2 is by making our CI slightly larger. Therefore, a 95% CI for a sample with unknown σ^2 is:

$$\left(\bar{x} - t_{1-\frac{\alpha}{2},(n-1)} \cdot \frac{s}{\sqrt{n}}\ ,\ \bar{x} + t_{1-\frac{\alpha}{2},(n-1)} \cdot \frac{s}{\sqrt{n}}\right)$$

2 Hypothesis Testing

A statistical hypothesis is a claim about the value of a parameter, e.g. $\mu = 0.75$, is the true average diameter of a PVC pipe. In any hypothesis-testing problem, there are two contradictory hypotheses to consider. One hypothesis may claim that $\mu = 0.75$ and the other $\mu \neq 0.75$. The objective is to decide, based on sample information, which of the two hypothesis is right.

Now that we have a confidence interval, how does hypothesis test relate to it? The first claim is called the null hypothesis $\mathrm{H_o}$ and the other is the alternative hypothesis $\mathrm{H_a}$. When we are testing, we always assume that the null is true, and then based on the sample data, we will either reject the null hypothesis, in favor of the alternative hypothesis, or fail to reject the null hypothesis. **We do NOT accept the null hypothesis, we can only fail to reject it.** This is just like the system when Judge says that some defendant is *guilty* when it can be proved that he did the crime for which he was accused of. But when his crime cannot be proved, Judge does not say he is *innocent*, he says, based on evidence (which is sample data in our case), he is *NOT guilty* (fail to prove him guilty). *Just because we could not prove that he did the crime, does not mean that he did not do it.* Think about it. This is the basis of hypothesis testing.

A question that you may ask now is, how we determine whether or not we reject or fail to reject the null hypothesis? First, we need to determine our significance level of the test, α, which by tradition are $\alpha = 0.10$, 0.05, or 0.01 (the most common being 0.05). The smaller the α value, the more stringent the test to reject or fail to reject the null hypothesis is.

For practical purposes, there are two types of tests that we will deal with. If μ is the true value and μ_o is the postulated value, for which we are testing, then the possible one-sided and two-sided tests are:

Now that we know what hypothesis testing is, how do we perform the test? This is where confidence interval comes in. **If the corresponding CI excludes μ_o then we reject the null hypothesis. Alternatively, if the**

One-Sided Test	Two-Sided Test
$\mathrm{H_o}$: $\mu \geq \mu_o$, $\mathrm{H_a}$: $\mu < \mu_o$	$\mathrm{H_o}$: $\mu = \mu_o$, $\mathrm{H_a}$: $\mu \neq \mu_o$
$\mathrm{H_o}$: $\mu \leq \mu_o$, $\mathrm{H_a}$: $\mu > \mu_o$	

corresponding CI contains μ_o then we fail to reject the null hypothesis. To calculate the size of the CI, we need to find the z-value or t-value, which is based on the α value. For a two-sided test, the z- and t-values are the same as we determined before; that is, our confidence interval is:

$$\left(\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Please note that for a CI based on a t-value, just replace it with the z-value, by replacing σ with sample standard deviation s, there is NO DIFFERENCE IN THE STRUCTURE. On the other hand, for a single or one-sided test, the z-value is $z_{1-\alpha}$, but the confidence interval is no longer symmetrical like for a two-sided test. For the cases where the alternative hypothesis is $\mathrm{H_a}$: $\mu < \mu_o$, the CI is:

$$\left(-\infty \, , \, \bar{x} + z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

and for cases where the alternative is $\mathrm{H_a}$: $\mu > \mu_o$, the CI is:

$$\left(\bar{x} - z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \, , \, \infty\right)$$

However, this is just one way of reaching conclusion of a hypothesis testing. Other equivalent way is to calculate the value of statistic, and compare that value with tabulated value.

3 Example

Now that we are familiar with the test illustrations, in this section, we will see an example of how we deal with setting hypothesis. Say, the claim is that the average support rating is at least 50% (data shown in the appendix A along with R code), therefore we are testing (based on sample data - that is, a small representative part of the population):

Statement 1: whether or not the true average support rating (from the population) is more than or equal to 50 (as the word "at least" is used), or

Statement 2: if we rephrase the question, we want to test whether or not the true average support rating (from the population) is less than 50.

Three things to notice here:

1. Both the statements are opposite to each other

2. As both statements are contradictions to each other - both cannot be true at the same time. We have to decide which one is ok.

3. One intuitive solution of this problem is to see whether the average of the sample is more than or equal to 50 or not. If it is more than 50, then, we might want to say that the claim is substantiated. However, this will not work, as the inference is always about the population values (here average), which we did not observe. We only had a small sample out of the whole population. Therefore, the average we are having are for sample data, not the population data and hence cannot be directly used to say anything about population. This is why we are doing this hypothesis testing which is a probabilistic way to solve this problem.

Statistically speaking, the (Statement 1) is called Null hypothesis, $\mathrm{H_o}$ and the (Statement 2) is called Alternative hypothesis, $\mathrm{H_a}$. For this problem, they are as follows:

$$\mathrm{H_o}: \mu \geq 50, \ \mathrm{H_a}: \mu < 50$$

Note: Notice that the *strict inequality* is in the $\mathrm{H_a}$. This is a trick to recognize which one should be $\mathrm{H_a}$ (that means, it only can be one of these: $<$ or $>$ -

when the question tells that the population value is "greater than" or "less than" some particular value) and the opposite to that would be H_o (where we have the *equality, with or without any other inequality sign with it*: that means, it only can be one of these: $\leq$ or $=$ or $\geq$ - when the question tells that the population value is "at most", or "equal" or "at least" some particular value).

As we stated previously, knowing the sample mean is not enough because of variability, and so we must create a confidence interval to cover this variability. Now, because we don't know what the population variance is, since we only have a sample, we must use the t-test to perform our hypothesis testing.

Note: In the class it was taught that there might be three cases for choosing the test statistic:

1. When the sample size is sufficiently large and the population variance is unknown, they should construct the z-interval instead of the t-interval.
2. The t-interval will be used when the underlying distribution of the variable of interest is normal and the sample size is less than 20.
3. With large enough sample size, the sample variance is a close estimate of population variance, and hence the t-curve is close to the z-curve, and hence the t-interval can be constructed as an approximation to Z-interval (both intervals will be almost exactly the same - therefore will not make much difference which one we do).

The conclusion about this problem can be obtained as follows (all of which always agree, therefore doing it in any one method will be sufficient):

1. Since $\mu = 50$ is contained within the confidence interval $(-\infty, 50.76343)$, we *Fail To Reject the null hypothesis*. **Remember, we do not accept hypotheses, we fail to reject them**.

2. The reported test statistic value is t = - 0.9511. There were n = 49 data points, making the df = n-1 = 48. Now, we compare the t statistic value with t tabulated value for $\alpha = 0.05$, which is -1.677224. However, during the exam, you might need to find this from a table given to you. So, it is adviced that you learn to look up the table). Now, as -1.677224 (tabulated value) < - 0.9511 (calculated value), we again fail to reject null hypothesis.

Now, what happens if the government got a little over confident and believe that their popularity was in fact exactly 50% of the voters (now using two sided test is appropriate)? Similarly, we'd use the t-test for our hypothesis test except that this time we are testing:

$$\mathrm{H_o}\!: \mu = 50,\ \mathrm{H_a}\!: \mu \neq 50$$

This time, our confidence interval is: $(46.98631, 51.13369)$ and so since $\mu = 50$ is [redacted] contained within the 95% CI, we can not reject the null hypothesis of their claim that the average support rating is 50%.

Note: As the sample size is large, students could do a Z-test on the same external data to check that the constructed intervals are almost the same as that of t. This is not done here, leaving this as an exercise for the curious students.

A Computer Output

```
> x
52 49 40 57 49 50 37 50 57 42 42 52 40 53 58 42 44 51
56 64 54 49 51 48 55 59 28 53 55 57 55 56 46 57 46 51
43 29 52 50 43 39 56 48 52 49 51 41 48 47
```

```
> mean(x)
[1] 49.06
> sd(x)
[1] 7.296658

> t.test(x,mu=50,alternative='less')

        One Sample t-test
data:  x t = -0.9109, df = 49, p-value = 0.1834 alternative
hypothesis: true mean is less than 50 95 percent confidence
interval:
     -Inf 50.79004
sample estimates: mean of x
    49.06

> t.test(x,mu=50,alternative='two.sided')

        One Sample t-test
data:  x t = -0.9109, df = 49, p-value = 0.3668 alternative
hypothesis: true mean is not equal to 50 95 percent confidence
interval:
 46.98631 51.13369
sample estimates: mean of x
    49.06
```