CLARK UNIVERSITY

Inferences for normal samples
Math 218, Mathematical Statistics

D Joyce, Spring 2016

z-confidence intervals and z-tests. Section 7.1.

We first consider the cases when (1) we have a normal distribution with a known variance σ^2, or (2) any distribution with a large sample. We'll next consider the small sample case.

With a large sample, say $n \geq 30$, the central limit theorem says that $\overline{X}$ is approximately normal, and we can use the sample variance S^2 to approximate σ^2 if we don't happen to know σ^2.

In either case, the population mean μ is unknown, but it's estimated with the sample mean $\overline{X}$. This sample mean $\overline{X}$ is (at least almost) normally distributed with mean $\mu_{\overline{X}} = \mu$, variance $\sigma^2_{\overline{X}} = \sigma^2/n$, and standard deviation $\sigma_{\overline{X}} = \sigma/\sqrt{n}$. The normalized sample mean is $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ if σ is known, and if σ is not known, for large n the sample standard deviation s very closely approximates σ, so we can use it instead. Since we know what the distribution of $\overline{X}$ is, we can use that knowledge to construct interval estimates for μ and hypothesis tests about μ.

z-intervals. The confidence intervals in this situation are called z-intervals because they're based on standard normal distributions and Z is the usual symbol for a standard normal random variable.

For historical reasons more than anything else, the confidence level is often taken to be 95%, so the associated level of significance is $\alpha = 0.05$. For a two-sided confidence interval, which is the usual kind, $\alpha/2$ of the area under each end of the standard normal density function is needed, and that means the cut-offs for the standard normal distribution are $\pm z_{\alpha/2}$. Here's a table of the these numbers for common confidence levels for one- and two-sided confidence intervals.

confidence	α	z_α	$\alpha/2$	$z_{\alpha/2}$
90%	0.1	1.28	0.05	1.645
95%	0.05	1.645	0.025	1.96
99%	0.01	2.33	0.005	2.575

You can find other values in table A.3 of the text. Note how you have to read the table backwards.

The two-sided z-interval at the α-level is

$$\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right],$$

while the lower one-sided confidence interval is

$$\left[\overline{x} - z_{\alpha}\frac{\sigma}{\sqrt{n}}, \infty\right),$$

and the upper one-sided confidence interval is

$$\left(-\infty, \overline{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}\right].$$

In the two-sided case, the *margin of error* is the quantity

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Typically, before an experiment is designed, both the α-level and the desired margin of error E are specified. From these, the size of the sample n can be determined. It is

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2.$$

Hypothesis tests for μ. These are summarized in table 7.1 on page 240. They correspond to the three kinds of z-intervals. In the two-sided test, the null hypothesis is $H_0 : \mu = \mu_0$ where μ_0 is a constant. In the lower one-sided test the null hypothesis is $H_0 : \mu \geq \mu_0$, while in the upper one sided test $H_0 : \mu \leq \mu_0$.

The P-values for these three hypothesis tests for μ are summarized in table 7.2.

We'll look at example 7.2 in detail.

t-confidence intervals and t-tests. We now consider the case when n is small where we have a normal distribution but the variance σ^2 is unknown. As we saw in chapter 6, the statistic

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a t-distribution with $n-1$ degrees of freedom.

If $n \geq 30$, the t-distribution is so close to the z-distribution, the z-distribution is used instead. But for $n < 0$ (or so), the difference is sufficient to justify using the t-distribution. In between, it may depend on other things if it makes any difference which test to use.

The confidence intervals and hypothesis tests for a t-distribution with $n-1$ degrees of freedom are the same as those for the z-distribution, except, of course, the distribution is different. In practice this means a different table is consulted. For the two- and one-sided cases for a z-distribution, a standard normal table is used to look up $z_{\alpha/2}$ or z_α, but for a t-distribution with $n-1$ degrees of freedom, a t-table is used to look up $t_{n-1,\alpha/2}$ or $t_{n-1,\alpha}$.

Note that the t-table in our text, table A.4, has a very different format than the z-table, table A.3. To look up $z_{0.05}$, you find what value of z gives 0.95, namely, 1.645, but to look up $t_{9,0.05}$ (9 degrees of freedom, so $n = 10$), you look in line 9 under column $\alpha = 0.05$ to get 1.833. Note how the values in column 0.05 in the t-table approach 1.645 as n gets larger.

Inferences on variances. In 7.3, we switch from making inferences about μ to inferences about σ^2, but only in the case that the population distribution is a normal distribution. The statistical tests will all involve the χ^2 distribution.

When the population distribution is a normal $N(\mu, \sigma^2)$ distribution, then the sample variance S^2 is an estimator of the population variance σ^2, furthermore,

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

is a chi square distribution with $n-1$ degrees of freedom. That's the statistic we'll use for confidence intervals and hypothesis tests for σ^2.

Thus, a two-sided confidence interval for σ^2 at the confidence level $1-\alpha$ is

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

while a lower one-sided confidence interval at the same confidence level is

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

and an upper one-sided confidence interval is

$$\left(-\infty, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha}}\right].$$

These confidence intervals directly give corresponding hypothesis tests.

For a two-sided hypothesis test with $H_0 : \sigma^2 = \sigma_0^2$, reject H_0 if σ_0^2 is not in the first interval. For a lower one-sided hypothesis test with $H_0 : \sigma^2 \geq \sigma_0^2$, reject H_0 if σ_0^2 is not in the second interval. And for an upper one-sided hypothesis test with $H_0 : \sigma^2 \leq \sigma_0^2$, reject H_0 if σ_0^2 is not in the third interval.

Math 218 Home Page at
`http://math.clarku.edu/~djoyce/ma218/`