Chapter Seven

Confidence Intervals

Confidence Intervals for a Mean

Confidence Intervals for a Proportion

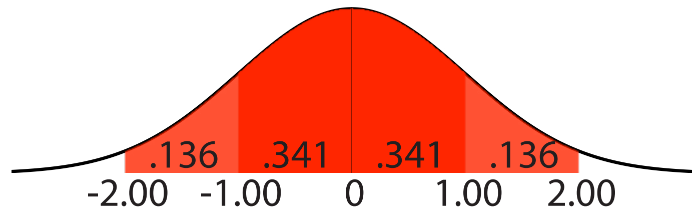
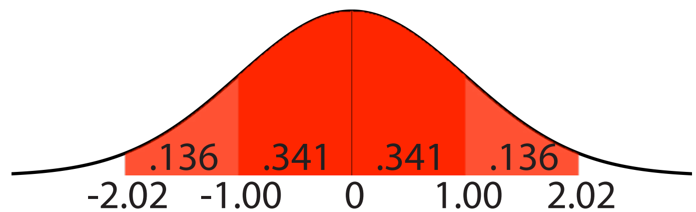
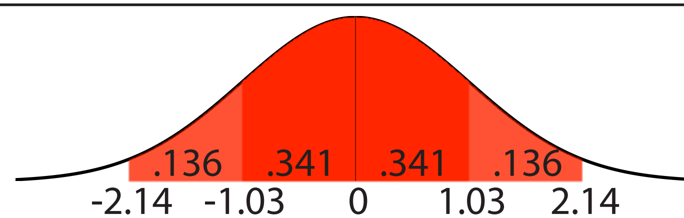
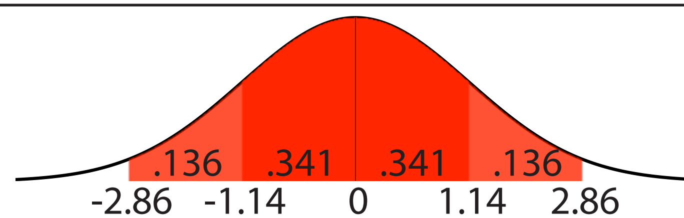
Sample Size Needed for a Specified Margin of Error

The *T* Distribution

The normal (z) distribution relies on knowing the value of the population standard deviation σ. In most cases, this value is not known and is instead estimated by the sample standard deviation s.

When σ is not known, t should be used instead of z. Like z, t is the number of standard deviations above the mean: $t = \frac{x - \bar{x}}{s}$. For sampling distributions, $t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$.

Like the normal distribution, the **t distribution** is mound-shaped, symmetrical, and centered at 0, and in fact it appears the same as a normal distribution. The difference is the scale: The t distribution is more spread out to account for less certainty since σ is unknown. The smaller the sample used to estimate σ, the more uncertainty there is and the more spread out the curve is. Conversely, as n approaches ∞, the t distribution approaches a perfect normal distribution.

Distribution	Curve
Normal (z)	
t, s calculated from a sample of $n = 100$	
t, s calculated from a sample of $n = 20$	
t, s calculated from a sample of $n = 5$	

Degrees of Freedom

A situation's **degrees of freedom** is the number of data values that need to be stated (in addition to the sample statistics) in order for all data values to be known. This essentially is all but one of the values in each sample or category. For example, to know how many boys and how many girls are in a class of 30, you only need to know one of the two quantities to know both of them.

Statistic	Degrees of Freedom	Example	Example *df*
mean	$n - 1$	Three bulldogs have an average weight of 23kg.	$3 - 1 = 2$
multiple means	$n_1 - 1 + n_2 - 1 + \ldots$	Three male bulldogs have an average weight of 23kg, and five female bulldogs have an average weight of 19kg.	$(3 - 1) + (5 - 1) = 6$
quantities by category	$k - 1$	The 70 football players include freshmen, sophomores, juniors, and seniors.	$4 - 1 = 3$
quantities by category within multiple categories	$(k_1 - 1)(k_2 - 1)$	The 70 football players 50 soccer players, and 45 basketball players include 25 freshmen, 39 sophomores, 44 juniors, and 57 seniors.	$(4 - 1)(3 - 1) = 6$

Critical Values

A **critical value** of a distribution, commonly labeled with a subscript zero, is a value that bounds a specific percentage of the distribution. This chapter uses z_0 such that a specific percentage of the normal distribution lies between $-z_0$ and z, and uses t_0 in the same manner. This percentage is the **confidence level**, abbreviated ***c***.

Critical values can be found in the t table by cross referencing the confidence level c with the degrees of freedom df. For normal distributions, $df = \infty$ can be used on the t table. Some examples are shown below.

Distribution	$c = .80$	$c = .90$	$c = .95$	$c = .99$
normal	$z_0 = 1.28$	$z_0 = 1.64$	$z_0 = 1.96$	$z_0 = 2.58$
$t, df = 99$	$t_0 = 1.29$	$t_0 = 1.66$	$t_0 = 1.98$	$t_0 = 1.63$
$t, df = 19$	$t_0 = 1.33$	$t_0 = 1.73$	$t_0 = 2.09$	$t_0 = 2.86$
$t, df = 4$	$t_0 = 1.53$	$t_0 = 2.13$	$t_0 = 2.78$	$t_0 = 4.60$

For example, 80% of the normal curve lies between $z = -1.28$ and $z = 1.28$, and 80% of the t distribution with 4 degrees of freedom lies between $t = -1.53$ and $t = 1.53$.

Confidence Intervals

A **confidence interval** is a range, centered about a sample statistic, that is likely to include the population parameter. It can be found by identifying the critical value and multiplying it by the standard error to find the **margin of error E**, which is how far the confidence interval goes above and below the sample statistic. Alternatively, if a desired margin of error has already been stated, the sample size needed to achieve this margin of error can be calculated.

Statistic	Margin of Error	Confidence Interval	Sample size needed
mean	$E = z_0 \frac{\sigma}{\sqrt{n}}$	$\bar{x} - E < \mu < \bar{x} + E$	$n = (\frac{z_0 \sigma}{E})^2$
proportion	$E = z_0 \sqrt{\frac{pq}{n}}$	$\hat{p} - E < p < \hat{p} + E$	$n = pq(\frac{z_0}{E})^2$

For example, if 106 out of 200 random California voters say they support proposition 89, then $\hat{p} = \frac{106}{200}$ and a 95% percent confidence interval would be $.46 < p < .60$. This allows for the conclusion *We are 95% confident that between 46% and 60% of all California voters support proposition 89.* Givenn our sample data, our best estimate of the population proportion p is 53%, and there is a 95% chance that p is between 46% and 60%.