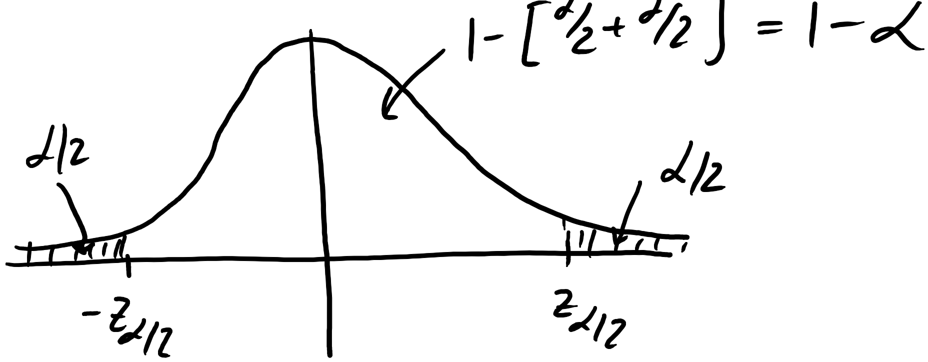
Review confidence intervals

want estimate on population parameter or proportion (e.g. μ, p).

use sample of size n to compute estimate.

Supply α-value. Then construct an <u>interval</u> which contains population parameter with probability $1-\alpha$.



$$P(-z_{\alpha/2} < z < z_{\alpha/2}) = 1-\alpha = P(|z| < z_{\alpha/2})$$

$$= 2P(0 < z < z_{\alpha/2})$$

where z is standard normal random variable.

When can we make use of this? large scale estimation

(1) large samples (large n). $n \geq 30$.

Then $\bar{x}$ is approx normal by CLT and

$\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ so that

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$ is std. normal

If σ (population std deviation) is not known, replace with $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$ sample std deviation

(2) When population is known to be normal (then any samples drawn from it are normally distributed).

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

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

Expand inequality $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2} \Rightarrow \bar{x} - \mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

$\Rightarrow \mu > \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} > -z_{\alpha/2} \Rightarrow \bar{x} - \mu > -z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \Rightarrow \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

confidence interval for μ:

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

μ is contained in this interval with probability $1-\alpha$ for $\alpha \in (0,1)$. The smaller the value of α, the larger the interval.

Small scale estimation (small random sample)

$n < 30$, CLT does not apply

Assume we do not know that population is definitely normal. no basis to assume $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

Replace std normal with student-t distribution

$z_{\alpha/2} \to t_{\alpha/2, n-1}$ — two parameters (α, degrees of freedom)

$\frac{\bar{x}-\mu}{s/\sqrt{n}} \sim$ T-distribution

$df = n-1$ (which appears in $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$).

small sample confidence interval for μ:

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

Ex) printer manufacturer tests $n=15$ printheads and calculates the following statistics:

$\bar{x} = 1.23$ mil chars $\quad s = .27$ mil chars

before printhead fails

Form a 99% confidence interval for the mean number of characters printed.

$\Rightarrow$ $n = 15 < 30$, know nothing about population distribution

$\Rightarrow$ use student-t distribution

$\alpha = 0.01 \Rightarrow \alpha/2 = 0.005$

$t_{0.005,14} = t_{\alpha/2,n-1} = 2.977$ (see table)

$$\bar{x} \pm t_{\alpha/2,n-1}\left(\frac{s}{\sqrt{n}}\right) = 1.23 \pm 2.977\left(\frac{.27}{\sqrt{15}}\right)$$

$$= 1.23 \pm .21$$

$\Rightarrow (1.02 < \mu < 1.44)$

Thus, the manufacturer can be 99% sure that the mean print life is at least 1.02 mil characters.

Note: as $n \rightarrow 30$, std-t distribution approaches std. normal.

Estimation of proportions

For X binomial r.v. and n large

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \to N(0,1) \quad \text{(std. normal)}$$

as $n \to \infty$. (Recall $\mu_X = np$; $\sigma_X = \sqrt{npq}$)

p success probability. $q = 1-p$ failure probability.

Insert into $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha$

$$\Rightarrow P\left(-z_{\alpha/2} < \frac{X-np}{\sqrt{np(1-p)}} < z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow \frac{X-np}{\sqrt{np(1-p)}} > -z_{\alpha/2} \Rightarrow X - np > -z_{\alpha/2}\sqrt{npq}$$

$$X > np - z_{\alpha/2}\sqrt{npq}$$

$$\Rightarrow \frac{X}{n} > p - z_{\alpha/2}\sqrt{\frac{npq}{n^2}} \Rightarrow \frac{X}{n} > p - z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

notice that $\frac{X}{n}$ is the proportion of successes in n samples.

Similarly, $X < np + z_{\alpha/2}\sqrt{npq}$ gives

$$\underbrace{\frac{X}{n}}_{\hat{p}} < p + z_{\alpha/2}\sqrt{\frac{pq}{n}} \Leftarrow \text{replace } \sqrt{\frac{pq}{n}} \text{ by } \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$\Rightarrow$ denote $\frac{x}{n}$ by $\hat{p}$ and write confidence interval for p:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$(1-\alpha)\cdot 100\%$ confidence interval for p. (population proportion of successes)

Ex) 136 of 400 people given a flu vaccine experienced discomfort. Construct 95% interval for true proportion of patients experiencing discomfort from the vaccine.

$n = 400$ (large scale, confidence interval applies)

$$\hat{p} = \frac{136}{400} = 0.34$$

95% CI $\Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$

$z_{\alpha/2} = z_{0.025} = 1.96$ (from std normal table)

$$\Rightarrow 0.34 - 1.96\sqrt{\frac{0.34(0.66)}{400}} < p < 0.34 + 1.96\sqrt{\frac{0.34(0.66)}{400}}$$

$\Rightarrow 0.29 < p < 0.39$ (95% CI).

Estimation of variances and chi-square distribution

Thm If $\bar{X}$ and s^2 are the sample mean and sample variance of a sample of size n from a normal population with mean μ and std deviation σ then:

$$\frac{(n-1)s^2}{\sigma^2}$$ follows a chi-square distribution with $n-1$ degrees of freedom.

chi-square distribution

If $X_1, X_2, \ldots, X_n$ are independent standard normal random variables ($X_i \sim N(0,1)$ iid.) then $Y = \sum_{i=1}^{n} X_i^2$ has a chi-square distribution of degree n.

$\mu(Y) = n$ and $\sigma^2(Y) = 2n$