POINT AND INTERVAL ESTIMATION – Prof. Richard B. Goldstein

θ	is an unknown population parameter
$\hat{\theta}$	is a point estimator based upon the known sample data
[A, B]	is a confidence interval estimate – A and B are based upon the sample

EXAMPLES

[1] μ is the population mean

various estimates include $\bar{x}$, $\tilde{x}$, $\dfrac{x_{(1)} + x_{(n)}}{2} = \dfrac{L + H}{2}$

[2] σ^2 is the population variance

two estimates are $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$ and $V = \dfrac{\sum (x_i - \bar{x})^2}{n}$

DESIRABLE PROPERTIES

[1] **Unbiased**

<u>Bias</u>: $B(\hat{\theta}) = E(\hat{\theta}) - \theta$ unbiased means bias is zero

If the population is normally distributed, then all of the above estimators for μ are unbiased. For the variance σ^2 the estimate s^2 is unbiased and $E[V] = \dfrac{n-1}{n}\sigma^2$

These last two results are based upon

$$E\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = E\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right] = \sum_{i=1}^{n}\left(\mathrm{Var}(x_i) + E[x_i]^2\right) - n\left(\mathrm{Var}(\bar{x}) + E[\bar{x}]^2\right)$$

$$= \sum_{i=1}^{n}(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) = n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2 = (n-1)\sigma^2$$

Is $s = \sqrt{s^2}$ an unbiased estimate of σ? No! [for proof see Kendall ref.]

$$E(s) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\frac{n-1}{2}}\,\Gamma\left(\frac{n-1}{2}\right)}\sigma = A_n\sigma$$

N	A_n	$\approx 1 - \frac{1}{4n-4}$
2	0.797884	0.750000
10	0.972650	0.972222
50	0.994911	0.994898
100	0.997478	0.997475

[2] **Consistency**

As $n \to \infty$ $\hat{\theta}$ converges in probability converges to θ. That is, $\lim_{n\to\infty} P\left\{\left|\hat{\theta}-\theta\right| < \varepsilon\right\} = 1$ for any $\varepsilon > 0$ which means that it becomes more and more likely that $\hat{\theta}$ is within $\pm\varepsilon$ of θ as n becomes large no matter what ε is chosen.

[3] **Efficiency**

If $\sigma^2_{\hat{\theta}_1} < \sigma^2_{\hat{\theta}_2}$, then $\hat{\theta}_1$ is a more efficient estimator than $\hat{\theta}_2$

For $n > 2$, $\bar{x}$ is more efficient than $\tilde{x}$: $\sigma^2_{\bar{x}} = \frac{2}{\pi}\sigma^2_{\tilde{x}} \approx 0.637\sigma^2_{\tilde{x}}$ which means that a sample mean from 637 observations is just as accurate as a median from 1000 observations from a normal population. A more efficient estimate than the median is mid-quartile given as $\frac{Q_1 + Q_3}{2}$ which has an efficiency of 0.808

Absolute Efficiency / Rao-Cramer Lower Bound

$$\text{Var}(Z) \geq \frac{1}{nE\left[\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2\right]}$$ where Z is an unbiased estimate

For example, $\bar{x}$, is an unbiased estimate of μ and $f(x;\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$$\ln f(x;\mu) = -\frac{1}{2}\ln\left(2\pi\sigma^2\right) - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\frac{\partial \ln f(x;\mu)}{\partial\mu} = \frac{-1}{2\sigma^2}2(x-\mu)(-1) = \frac{(x-\mu)}{\sigma^2}$$

$$E\left[\left(\frac{\partial \ln f(x;\mu)}{\partial\mu}\right)^2\right] = E\left[\left(\frac{x-\mu}{\sigma^2}\right)^2\right] = \frac{1}{\sigma^4}E\left[(x-\mu)^2\right] = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$$

Therefore, the R-C bound is σ^2/n which is the variance of $\bar{x}$ and that makes the sample mean the most efficient estimator of the population mean.

[4] **Sufficiency**

The joint probability density function can be factored as to not depend upon Θ. It should use all of the sample data information.

[5] **Resistance**

A resistant estimator is one that is not influenced by the presence of outliers. For example, the median or mid-quartile resists the influence of outliers more than does the mean.

[6] Maximum Likelihood Estimate

This is a method by which given a particular distribution one can find an estimator. The estimator will be the one with the largest probability.

$$L(x_1, x_2, ..., x_n; \theta) = f(x_1, x_2, ..., x_n; \theta) = f(x_1, \theta) f(x_2, \theta) ... f(x_n, \theta)$$

Example $\quad f(x, \lambda) = \dfrac{e^{-\lambda} \lambda^x}{x!}$

$$L(x_1, x_2, ..., x_n, \lambda) = \prod_{i=1}^{n} f(x_i, \lambda) = \frac{e^{-n\lambda} \lambda^{\sum x_i}}{\prod x_i!}$$

$$\Lambda = \ln L = -n\lambda + \sum x_i \ln(\lambda) - \ln \prod x_i!$$

$$\frac{\partial \Lambda}{\partial \lambda} = \frac{\partial \ln L}{\partial \lambda} = -n + \sum \frac{x_i}{\lambda} = 0$$

$$\Rightarrow \hat{\lambda} = \frac{\sum x_i}{n} = \overline{x}$$

Suppose in each of five hours the number of customers arriving are tabulated as 7, 12, 9, 10, and 16. What value should λ be? Use $\lambda = \overline{x} = 10.8$

Interval Estimation

Also known as confidence intervals: $P[\theta \in [A, B]] = 1 - \alpha$

Here the A and B are determined by the estimate of θ given by $\hat{\theta}$

Example:

$$P\left[\overline{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

where σ is assumed to be known.

MOST COMMON ESTIMATORS:

A. SINGLE SAMPLE

Mean $\overline{x}$ has a normal distribution for large n given as $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

and a Student-t distribution $T = \frac{\overline{x} - \mu}{s/\sqrt{n}}$ with n – 1 d.f

Proportion $\hat{p} = \frac{r}{n}$ has approximately a normal distribution $N\left(p, \sqrt{\frac{pq}{n}}\right)$*

Variance s^2 has a Chi-square distribution $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$ with n – 1 d.f.

B. TWO SAMPLES

Means $\overline{x}_1 - \overline{x}_2$ has a normal distribution for large n_1 and n_2 given as

$$N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

for small samples $T = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

where $s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$ with $n_1 + n_2 - 2$ d.f.**

Proportions $\hat{p}_1 - \hat{p}_2$ has a normal distribution $N\left(p_1 - p_2, \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}\right)$

Variances $\frac{s_1^2}{s_2^2}$ has an F distribution with $n_1 - 1$ d.f. in numerator and $n_2 - 1$ d.f. in denominator

* A better estimate of **p** is $\hat{p} = \frac{r+1}{n+2}$ (Laplace method) which brings the estimate closer to 0.5 and away from the extremes at 0 and 1.

** The above assumes $s_1 \approx s_2$. Otherwise, use the T distribution with $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ as the denominator with $\min(n_1 - 1, n_2 - 1)$ d.f.

Confidence Intervals

Mean μ $\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ for known σ

$\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$ for unknown σ

μ_1-μ_2 $\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ known σ_1 and σ_2

$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2,df}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ unknown σ_1 and σ_2 where df = min(n_1 - 1, n_2 - 1)

Proportion p $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$ is the standard Wald formula

Replace $\hat{p}$ by $\frac{r+2}{n+4}$ or $\frac{r+\frac{z^2}{2}}{n+z^2}$ for the adjusted Wald formula

p_1-p_2 $\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$ can also be adjusted as above

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

$\frac{\sigma_1^2}{\sigma_2^2}$ $\left[\frac{s_1^2}{s_2^2}\frac{1}{F_{\alpha/2,n_1-1,n_2-1}}, \frac{s_1^2}{s_2^2}F_{\alpha/2,n_2-1,n_1-1}\right]$

References:

The Advanced Theory of Statistics – Volume II by M.G. Kendall & A. Stuart (Hafner/Macmillan Publishing), Chapter 17, problem 17.6.

Mathematical Methods of Statistics – Harald Cramér (my thesis advisor's thesis advisor) (Princeton University Press)

Introduction to Statistical Analysis – Wilfrid Dixon & Frank Massey, Jr. (McGraw Hill Publishing)

Probability and Statistics – Kevin J. Hastings (Addison-Wesley)

Confidence Interval Around a Completion Rate – Jeff Sauro
http://www.measuringusability.com/wald.htm

Estimators
http://en.wikipedia.org/wiki/Estimator