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** Bias: $B(\hat{\theta}) = E(\hat{\theta}) - \theta$ unbiased means bias is zero
s^2 is an unbiased estimate of σ^2 but s is not an unbiased estimate of σ.
Use 1.028s to estimate σ for n = 10 and 1.005s for n = 50.

[2] **Consistency** it becomes more likely that $\hat{\theta}$ is close to θ as n becomes large

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

[4] **Sufficiency** 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** the probabilistically most likely estimate

Interval Estimation

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

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

MOST COMMON ESTIMATORS:

A. SINGLE SAMPLE

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

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

Proportion $\hat{p} = \frac{r}{n}$ has 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 $\bar{x}_1 - \bar{x}_2$ 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 $T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

with min(n_1 – 1, n_2 – 1) d.f. (assumes $\sigma_1 \neq \sigma_2$) **

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 $\hat{p} = \frac{r+2}{n+4}$ which brings the estimate closer to 0.5 and away from the extremes at 0 and 1.

** If $s_1 \approx s_2$ use $T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ 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.

Confidence Intervals

Mean μ $\quad \bar{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} = \bar{x} \pm E$ for known σ

estimate sample size $n \approx \left(\dfrac{z_{\alpha/2}\sigma}{E}\right)^2$

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

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

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

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

estimate sample size $n \approx \hat{p}\hat{q}\left(\dfrac{z_{\alpha/2}}{E}\right)^2$ note: if $\hat{p}$ is unknown use $n \approx 0.25\left(\dfrac{z_{\alpha/2}}{E}\right)^2$

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

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

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

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

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)

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