# CONFIDENCE INTERVALS

## 1. For a population mean, $\theta = \mu$

$-$ large sample ($n > 30$) or normal underlying population and $\sigma$ known, $TT = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \in N(0, 1)$;

x – sample, alpha - significance level, sigma - standard deviation

n = length(x);

z = norminv(1 - alpha/2);

mx = mean(x);

li = mx – sigma/sqrt(n) * z;

ri = mx + sigma/sqrt(n) * z;

CI = [li, ri];

$-$ large sample ($n > 30$) or normal underlying population $TT = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} \in T(n-1)$.

x - sample, alpha – significance level

n = length(x);

t = tinv(1 - alpha/2, n - 1);

mx = mean(x);

s = std(x);

li = mx - s/sqrt(n)*t;

ri = mx + s/sqrt(n)*t;

```
CI = [li, ri];
```

**2.** For a population variance, $\theta = \sigma^2$, for a normal underlying population, $TT = \dfrac{(n-1)s^2}{\sigma^2} \in \chi^2(n-1)$.

x – sample, alpha - significance level

```
n = length(x);

q1 = chi2inv(1 - alpha/2, n - 1);
q2 = chi2inv(alpha/2, n - 1);

v = var(x);

li = (n - 1) * v/q1;
ri = (n - 1) * v/q2;

CI = [li, ri];
```

**3.** For the difference of two population means, $\theta = \mu_1 - \mu_2$, for large samples ($n_1 + n_2 > 40$) or normal underlying populations and independent samples,

$$- \sigma_1, \sigma_2 \text{ known, } TT = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \in N(0,1);$$

x1 – first sample, x2 – second sample, sig1 – first sample standard deviation, sig2 – second sample standard deviation, alpha – significance level

```
n1 = length(x1);
n2 = length(x2);

mx1 = mean(x1);
```

```
mx2 = mean(x2);

z = norminv(1 - alpha/2);

li = mx1 - mx2 - z*sqrt(sig1^2/n1 + sig2^2/n2);
ri = mx1 - mx2 + z*sqrt(sig1^2/n1 + sig2^2/n2);

CI = [li, ri];
```

$$- \sigma_1 = \sigma_2, \text{ unknown}, TT = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \in T(n_1 + n_2 - 2), \text{ where } s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2};$$

x1 – first sample, x2 – second sample, alpha – significance level

```
n1 = length(x1);
n2 = length(x2);

n = n1 + n2 - 2;

mx1 = mean(x1);
mx2 = mean(x2);

vx1 = var(x1);
vx2 = var(x2);

t = tinv(1 - alpha/2, n);

rad = sqrt(1/n1 + 1/n2);
sp = sqrt(((n1 - 1)*vx1 + (n2 - 1)*vx2)/n);

li = mx1 - mx2 - t*sp*rad;
```

```
ri=mx1-mx2+t*sp*rad;

CI = [li, ri];
```

$$- \sigma_1 \neq \sigma_2, \text{ unknown, } TT = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \in T(n), \text{ with}$$

$$\frac{1}{n} = \frac{c^2}{n_1 - 1} + \frac{(1-c)^2}{n_2 - 1} \quad \text{and} \quad c = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

x1 – first sample, x2 – second sample, alpha – significance level

```
n1 = length(x1);
n2 = length(x2);

mx1 = mean(x1);
mx2 = mean(x2);

vx1 = var(x1);
vx2 = var(x2);

c = (vx1/n1)/(vx1/n1 + vx2/n2);
n = 1/(c^2/(n1-1) + (1-c)^2/(n2-1));

t = tinv(1 - alpha/2, n);
rad = sqrt(vx1/n1 + vx2/n2);

li = mx1 - mx2 - t*rad;
ri = mx1 - mx2 + t*rad;
```

```
CI = [li, ri];
```

**4.** For the ratio of two population variances, $\theta = \dfrac{\sigma_1^2}{\sigma_2^2}$, for normal underlying populations and independent samples, $TT = \dfrac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \in F(n_1 - 1, n_2 - 1)$.

x1 – first sample, x2 – second sample, alpha- significance level

```
n1 = length(x1);
n2 = length(x2);

vx1 = var(x1);
vx2 = var(x2);

f1 = finv(1 - alpha/2, n1 - 1, n2 - 1);
f2 = finv(alpha/2, n1 - 1, n2 - 1);

rap = vx1/vx2;

li = rap/f1;
ri = rap/f2;

CI = [li, ri];
```