Jan 16th, 2013

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

sample size

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$E[\hat{\sigma}^2] = \sigma^2 \frac{n-1}{n}\frac{N}{N-1} \qquad E(S^2) = E\left[\frac{n}{n-1}\hat{\sigma}^2\right] = \frac{n}{n-1}\frac{n-1}{n}\frac{N}{N-1}\sigma^2 = \frac{N}{N-1}\sigma^2$$

Usually N is very large which implies $\frac{N}{N-1} \approx 1$

Hence $E[S^2] \approx \sigma^2$

Proposition: $Var(\bar{x}) \approx \frac{S^2}{n}(1 - \frac{n}{N})$

In practice, we denote $S_{\bar{x}}^2 = \frac{S^2}{n}(1 - \frac{n}{N})$. $S_{\bar{x}}$ is called the standard error of $\bar{x}$.

Proof: $Var(\bar{x}) = \frac{\sigma^2}{n}(1 - \frac{n-1}{N-1})$

$$E[S^2] = \frac{N}{N-1}\sigma^2$$

$$E[\frac{S^2}{n}(1-\frac{n}{N})] = \frac{1}{n}(1 - \frac{n}{N})E(S^2) = \frac{1}{n}(1-\frac{n}{N})\frac{N}{N-1}\sigma^2$$

$$= \frac{1}{n}\frac{N-n}{N}\cdot\frac{N}{N-1}\sigma^2 = \frac{1}{n}\frac{N-n}{N-1}\sigma^2$$

$$= \frac{1}{n}\sigma^2(1 - \frac{n-1}{N-1}) = Var(\bar{x}) \quad \blacksquare$$

* Estimating population proportions

Population $\underbrace{x_1, x_2, \cdots x_N}$.

$x_i = 0$ or 1 (support or not support)

$$\frac{1}{N}\sum_{i=1}^{N} x_i = p$$

← population proportion

unknown but constant

population variance of a proportion is $\frac{1}{N}\cdot\sum_{i=1}^{N}(x_i - \bar{x})^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - (\bar{x})^2$

since $x_i = 0$ or 1

$$= \frac{1}{N}\sum_{i=1}^{N} x_i - (\bar{x})^2$$

$$= p - p^2$$

$$= p(1-p)$$

Now suppose we have a sample $X_1, X_2, \cdots X_n$

(0 or 1) $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \hat{p}$ ← sample proportion

Following the same arguments in last lecture $E[\hat{p}]=p$

$$Var(\hat{p}) = \frac{p(1-p)}{n}\left(1-\frac{n-1}{N-1}\right)$$

$$S_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}\left(1-\frac{n-1}{N-1}\right)} \leftarrow \text{Standard error}$$

	Estimate	Variance of Estimate	Estimated variance
μ	$\bar{x}$	$\frac{\sigma^2}{n}\left(1-\frac{n-1}{N-1}\right)$ this is not operational because not knowing σ^2	$\frac{s^2}{n}\left(1-\frac{n}{N}\right)$
p	$\hat{p}$	$\frac{p(1-p)}{n}\left(1-\frac{n-1}{N-1}\right)$	$\frac{\hat{p}(1-\hat{p})}{n}\left(1-\frac{n-1}{N-1}\right)$
σ^2	$\frac{N-1}{N}s^2$	/	/

The normal approximations to the sampling distribution

Thm: for sufficiently large n central limit theorem

$$P\left[\frac{\bar{x}_n-\mu}{\sigma_{\bar{x}}} \le z\right] \approx P[Z \le z] \quad \text{where } Z \sim N(0,1)$$

another notation: $\Phi(z)$

In practice often times $\sigma_{\bar{x}}$ is replaced by $S_{\bar{x}}$

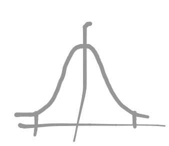
Applications:

Constructing confidence intervals (CI) for μ.

Sps we want to find a random interval $[L_l, L_u]$ such that $P[L_l \le \mu \le L_u] \approx 1-\sigma$

① $P\left[\frac{\bar{x}_n-\mu}{S_{\bar{x}}} \le z\right] \approx \Phi(z)$

$\Phi(z)+\Phi(-z)=1$



② $P\left[\frac{\bar{x}-\mu}{S_{\bar{x}}} \le -z\right] \approx \Phi(-z)$

①−② $P\left[-z \le \frac{\bar{x}_n-\mu}{S_{\bar{x}}} \le z\right] \approx 2\Phi(z)-1$

According to standard normal tables, we can find a z such that
$2\Phi(z)-1 = 1-\alpha \Leftrightarrow \Phi(z) = 1-\frac{\alpha}{2}$

Therefore $P[-z \le \frac{\bar{X}_n-\mu}{S_{\bar{X}}} \le z] \approx 1-\alpha \Leftrightarrow P[\bar{X}_n - zS_{\bar{X}} \le \mu \le \bar{X}_n + zS_{\bar{X}}] \approx 1-\alpha$

Hence $L_l = \bar{X}_n - zS_{\bar{X}}$
$L_u = \bar{X}_n + zS_{\bar{X}}$

7.4 Estimation of a ratio

Population consists of N subjects (persons). For each subject, two characteristics x & y are of interest.
Hence we have in the population

$$\binom{x_1}{y_1}\ \binom{x_2}{y_2} \cdots \binom{x_n}{y_n}$$ of interest is the following quantity

$r = \sum_{i=1}^{N} y_i / \sum_{i=1}^{N} x_i \leftarrow$ population ratio

Obtain sample of size n
record random variables

$$\binom{X_1}{Y_1}, \binom{X_2}{Y_2}, \cdots, \binom{X_n}{Y_n}$$

And we use $R_n = \sum_{i=1}^{n} Y_i / \sum_{i=1}^{n} X_i$ as an estimate of r

P.S. $R_n \neq \sum_{i=1}^{n} \frac{Y_i}{X_i}$ Note that $R_n = \bar{Y}/\bar{X}$

Lemma: $\mathrm{cov}(\bar{X},\bar{Y}) = \frac{\sigma_{xy}}{n}\left(1-\frac{n-1}{N-1}\right)$ → proof of Lemma very similar to the proof of $Var(\bar{X})$. Take home exercise.

where $\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y)$ → population covariance between x and y.
not random

Then $Var(R_n) \approx \frac{1}{\mu_{\bar{X}}^2}(r^2\sigma_{\bar{X}}^2 + \sigma_{\bar{Y}}^2 - 2r\sigma_{\bar{X}\bar{Y}})$

$E[R_n] = r + \frac{1}{n}\left(1-\frac{n-1}{N-1}\right)\frac{1}{\mu_x^2}(r\sigma_x^2 - \rho\sigma_x\sigma_y)$

$\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \longleftarrow$ population correlation between x and y.

Delta Method

Suppose we want to know the distance of $f(X)$ based on the distribution of X
Major tool: Taylor expansion

Sps $f(\cdots)$ is smooth, then we can expand $f(\cdots)$ around μx

Hence $f(t) = f(\mu x) + f'(\mu x)(t-\mu x) + \frac{f''(\mu x)}{2}(t-\mu x)^2$ + small terms

Therefore, $f(X) \approx f(\mu x) + f'(\mu x)(X-\mu x) + \frac{f''(\mu x)}{2}(X-\mu x)^2$

Take expectation of two sides

Then $Ef(X) \approx f(\mu x) + \frac{f''(\mu x)}{2} Var(X)$

Suppose we want to know the distance of $f(X,Y)$ based on the distribution of (X, Y)

Major tool: Taylor expansion

Suppose $f(\cdots)$ is smooth then we can expand $f(\cdots)$ around μ_x, μ_y

$$f(t,s) = f(\mu x, \mu y) + \frac{\partial f(\mu x, \mu y)}{\partial x}(t-\mu x) + \frac{\partial f(\mu x, \mu y)}{\partial y}(s-\mu y)$$

+small terms

Therefore, $f(X,Y) \approx f(\mu x, \mu y) + \frac{\partial f(\mu x, \mu y)}{\partial x}(X-\mu x) + \frac{\partial f(\mu x, \mu y)}{\partial y}(Y-\mu y)$

$$Var(f(X,Y)) = Var\left(\frac{\partial}{\partial x}(X-\mu x) + \frac{\partial}{\partial y}(Y-\mu y)\right) \approx \left(\frac{\partial}{\partial x}\right)^2 Var(X) + \left(\frac{\partial}{\partial y}\right)^2 Var(Y) + 2\left(\frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial y}\right) COV(X,Y)$$

In our case,

$f(x,y) = \frac{x}{y}$

$x = \bar{Y}$ $\qquad \frac{\partial f(x,y)}{\partial x} = \frac{1}{y}$

$y = \bar{X}$ $\qquad \frac{\partial f(x,y)}{\partial y} = -\frac{x}{y^2}$

$$Var(R_n) \approx \left(\frac{1}{\mu x}\right)^2 Var(\bar{Y}) + \left(-\frac{\mu y}{\mu^2 x}\right)^2 Var(\bar{X}) + 2\frac{1}{\mu x}\cdot\left(-\frac{\mu y}{\mu^2 x}\right) cov(\bar{X}, \bar{Y})$$

$$Var(\bar{Y}) = \frac{\sigma^2 y}{n}\left(\frac{N-n}{N-1}\right)$$

$$Var(\bar{X}) = \frac{\sigma_x^2}{n}\left(\frac{N-n}{N-1}\right)$$

$$COV(\bar{X}, \bar{Y}) = \frac{\sigma_{X,Y}}{n}\left(\frac{N-n}{N-1}\right)$$

$$Var(R_n) \approx \left(\frac{1}{\mu^2_x}\sigma_y^2 + \frac{\mu^2_y}{\mu^2_x}\sigma_x^2 - 2\frac{\mu_y}{\mu^3_x}\sigma_{xy}\right)\frac{1}{n}\left(\frac{N-n}{N-1}\right)$$

Note that $r = \frac{\mu_y}{\mu_x}$

$$Var(R_n) \approx \left(\frac{1}{\mu^2_x}\sigma_y^2 + \frac{r^2}{\mu_x^2}\sigma_x^2 - 2r\cdot\frac{1}{\mu_x^2}\sigma_{xy}\right)\frac{1}{n}\left(\frac{N-n}{N-1}\right)$$

$$= \frac{1}{\mu_x^2}\left(\sigma^2 y + r^2\sigma_x^2 - 2r\sigma_{xy}\right)\frac{1}{n}\left(\frac{N-n}{N-1}\right)$$

Note that $\sigma^2_{\bar{Y}} = \sigma_y^2 \frac{1}{n}\left(\frac{N-n}{N-1}\right)$, $\sigma^2_{\bar{X}} = \sigma_x^2 \frac{1}{n}\left(\frac{N-n}{N-1}\right)$

$$\sigma_{\bar{X}\cdot\bar{Y}} = \sigma_{xy}\frac{1}{n}\left(\frac{N-n}{N-1}\right) \quad \blacksquare$$

Sample correlation

$$\hat{\rho}_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)$$

Then we can use $S^2_{R_n} = \frac{1}{\bar{X}^2}\left[S^2_{\bar{Y}} + R_n^2 S^2_{\bar{X}} - 2R_n\hat{\rho}_{XY}S_{\bar{X}}S_{\bar{Y}}\right]$ to approximate the Var(R_n)

Hence a $1-\alpha$ CI for r can be constructed as $R_n \pm Z_{\frac{\alpha}{2}} S_{R_n}$

Critical value for standard normal

7.2, 7.3, 7.4 till page 222