Sampling Distribution and the Central Limit Theorem

$X \sim N(\mu, \sigma)$

$$z = \frac{x - \mu}{\sigma}$$ (for individual members)

Sampling Distribution of Sample Mean

$\mu_{\bar{x}} = \mu$ (mean of the sample mean $\mu_{\bar{x}}$)

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ (standard deviation of sample mean $\sigma_{\bar{x}}$ = standard error)

Mean of sample size n

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

The Central Limit Theorem

① Samples from normally distributed population
→ sampling distribution → Normal

② Samples from any distribution population whose sample size $n \geq 30$ → Sampling distribution → Normal

$$\therefore \quad z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Population Proportion - p

$X \sim B(n, p)$

$P(X = r) = \binom{n}{r} p^r q^{n-r} \quad , \quad r = 0, 1, \ldots, n$

n = trials

r = success times

p = probability of success

$E = np$ (Expectation)

$$p = \frac{E}{n}$$

$$p = \frac{r}{n} \quad \text{(proportion)}$$

Normal Approximation to Binomial Distribution

$np \geqslant 5 \,,\; nq \geqslant 5 \implies$ Normal

Likewise

$n\hat{p} \geqslant 5 \,,\; n\hat{q} \geqslant 5$

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} \quad \text{(standard error)}$$

$$\therefore \; z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}}$$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Sampling distribution

* $\mu = \frac{\Sigma x_i}{n}$, $\sigma^2 = \frac{\Sigma (x_i - \mu)^2}{n}$

* Sampling distribution table $\Rightarrow$ sample | $\bar{x}$

* Sampling dis of sample mean (prob dis) table

$\bar{x}$ | f | prob $\left(\frac{f_i}{\Sigma f_i} \right)$

* Prob Histogram of sampling distribution

* $\mu_{\bar{x}} = \frac{\Sigma f_i \bar{x}}{\Sigma f_i}$ (or) $\Sigma \bar{x}_i \, p_i \left(p_i = \frac{f_i}{\Sigma f_i} \right)$

$\sigma^2_{\bar{x}} = \frac{\Sigma f_i (\bar{x}_i - \mu_{\bar{x}})^2}{n} = (\Sigma \bar{x}_i^2 \, p_i) - \mu_{\bar{x}}^2$

* Prove $\mu_{\bar{x}} = \mu$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Chapter (6)

Confidence Intervals

c - confidence level

z_c - critical value

c	z_c
0.9	1.645
0.95	1.96
0.99	2.575

$$P(-z_c < z < z_c) = c$$

$$P\left(-z_c < \frac{\bar{x} - \mu}{\underbrace{\sigma/\sqrt{n}}_{\text{std error}}} < z_c\right) = c$$

$$P(-z_c\, \sigma/\sqrt{n} < \underbrace{\bar{x} - \mu}_{\text{sampling error}} < z_c\, \sigma/\sqrt{n}) = c$$

$$P(-\bar{x} - z_c\, \sigma/\sqrt{n} < -\mu < -\bar{x} + z_c\, \sigma/\sqrt{n}) = c$$

$$P(\bar{x} - z_c\, \sigma/\sqrt{n} < \mu < \bar{x} + z_c\, \sigma/\sqrt{n}) = c$$

$$P(\bar{x} - E < \mu < \bar{x} + E) = c$$

$$E = z_c\, \sigma/\sqrt{n} \quad (\text{Margin of Error})$$

critical value & std error

Confidence Interval for μ

$$\bar{x} - E < \mu < \bar{x} + E$$

① σ known

$$E = z_c \frac{\sigma}{\sqrt{n}}$$

② σ unknown

$$E = t_c \frac{s}{\sqrt{n}}$$

$$df = n - 1$$

Confidence Interval for p

$$E = z_c \sqrt{\frac{pq}{n}}$$

but not knowing population pq, use point estimate $p \approx \hat{p}$ and $q \approx \hat{q}$

$$E = z_c \sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

where $n\hat{p} \geq 5$ and $n\hat{q} \geq 5$

$$\hat{p} - E < p < \hat{p} + E$$

Confidence Interval for σ^2

$$P(\chi^2_L < \chi^2 < \chi^2_R) = c$$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

$$P\left(\chi^2_L < \frac{(n-1)s^2}{\sigma^2} < \chi^2_R\right) = c$$

$$\chi^2_L < \frac{(n-1)}{\sigma^2} \qquad \bigg| \qquad \frac{(n-1)s^2}{\sigma^2} < \chi^2_R$$

$$\chi^2_L \sigma^2 < (n-1)s^2 \qquad \bigg| \qquad (n-1)s^2 < \chi^2_R \sigma^2$$

$$\sigma^2 < \frac{(n-1)s^2}{\chi^2_L} \qquad \bigg| \qquad \frac{(n-1)s^2}{\chi^2_R} < \sigma^2$$

$$P\left(\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}\right) = c$$

$\underline{\sigma^2}$

$$\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$

$\underline{\sigma}$

$$\sqrt{\frac{(n-1)\,s^2}{\chi^2_R}} < \sigma < \sqrt{\frac{(n-1)\,s^2}{\chi^2_L}}$$

$$\chi^2_R = \frac{1-c}{2}$$

$$\chi^2_L = \frac{1+c}{2} \qquad (\chi^2_L = 1 - \chi^2_R = 1 - \frac{1-c}{2} = \frac{1}{}$$

$$d_f = n-1$$

Scanned by CamScanner

Testing	One Sample	Two Samples		Others
		Independent	Dependent	
z	(a) μ (σ known) (b) p	(a) μ (σ known) (b) p	-	-
t	μ (σ unknown)	μ (σ unknown)	μ_d	-
χ^2	σ^2	-	-	(a) Goodness-of-fit (b) Independence

One Sample

*** μ (σ known)**

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

*** p**

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

*** μ (σ unknown)**

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$d_f = n - 1$$

*** σ^2**

$$\chi^2 = \frac{(n-1)\, s^2}{\sigma^2}$$

$$d_f = n - 1$$

Two Samples

Independent

* μ (σ known)

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

* p

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}\bar{q}\,(1/n_1 + 1/n_2)}}$$

where

$$\bar{p} = \frac{r_1 + r_2}{n_1 + n_2}$$

$$\bar{q} = 1 - \bar{p}$$

* μ (σ unknown)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

$$df \begin{cases} n_1 + n_2 - 2 \Leftarrow \text{Equal Var} \\ n - 1 \text{ (smaller } n) \quad \text{Inequal Var} \end{cases}$$

$s_{\bar{x}_1 - \bar{x}_2}$

For Equal Variance

$$\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

For Inequal Variance

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Dependent

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$$

where

$$s_d = \sqrt{\frac{\Sigma d^2 - (\Sigma d/n)^2}{(n-1)}}$$

$$\bar{d} = \frac{\Sigma d}{n}$$

χ^2 Testing

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

(a) Goodness-of-fit

$d_f = k - 1$

(b) Independence

$d_f = (r-1)(c-1)$

Alpha-α	Tail	z
0.1	Left	-1.28
	Right	1.28
	Two	± 1.645
0.05	Left	-1.645
	Right	1.645
	Two	± 1.96
0.01	Left	-2.33
	Right	2.33
	Two	± 2.575