Chapter 7, 8, 9

Chapter 7

Statistical Interval Based on Single Sample

sample ?

$$Z = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

7.1

1. Assumption:

   ① Population is normal.

   ② SD "σ" of population is known

   two tails

2. Single Sample CI (Normal)

   $$\boxed{\bar{x} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$$

   A $100(1-\alpha)\%$ CI:

   99% - 2.58

   95% - 1.96

   90% - 1.645

3. Sample size

   $$n = \left(2 \times Z_{\alpha/2} \cdot \frac{\sigma}{w}\right)^2$$ width

7.2.

1. Assumption.

   ① By CLT, we assume population has normal dis

   ② Replace population sd "σ" with sample sd "s"

2. large-sample CI

   $$\bar{x} \pm Z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

   $$Z = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

3. A General large-sample CI

① Properties: A. Approx. normal dis

B. Approx. unbiased.

C. $\sigma_{\hat{\theta}}$ aviliable

② $P(-z_{\alpha/2} < \frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} < z_{\alpha/2}) \approx 1-\alpha$.

4. CI for Population Propotion. [Code √]

let $\tilde{p} = [\hat{p} + z_{\alpha/2}^2/2n] / [1 + z_{\alpha/2}^2/n]$

$$\tilde{p} \pm z_{\alpha/2} \frac{\sqrt{\hat{p}\cdot\hat{q}/n + z_{\alpha/2}^2/4n^2}}{1 + z_{\alpha/2}^2/n}$$

5. One-Sided CI

① large-sample upper/lower bd for μ.

$\mu < \bar{x} + z_{\alpha} \cdot \frac{s}{\sqrt{n}}$. P291

$\mu > \bar{x} - z_{\alpha} \cdot \frac{s}{\sqrt{n}}$ (Calculate sample size from Proportion)

$$n = \frac{2z^2pq_c - z^2w^2 \pm \sqrt{4z^4pq_c(pq_c - w^2) + w^2z^4}}{w^2}$$

$$n \approx \frac{4z^2pq_c}{w}$$

7.3

1. Assumption.

①. n is small.

② μ, σ unknown.

③ Apply t-dis.

2. t-dis

$T = \frac{\bar{x}-\mu}{S/\sqrt{n}}$. with $df = n-1$

CI: $\bar{x} \pm t_{\alpha/2, n-1} \cdot S/\sqrt{n}$.

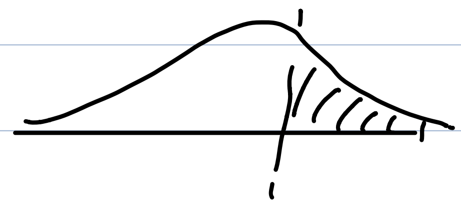
3. Properties of t-dis   P296

① More spread out than normal dis.

② As df ↑, the spread ↘

③ As df → ∞, t-dis → normal-dis

4. $t_{\alpha, v}$ = Area on the right of $t_{\alpha, v}$.



5. A Prediction Interval for a single future value.

$\bar{x} \pm t_{\alpha/2, n-1} \cdot S\sqrt{1+\frac{1}{n}}$

6. Tolerance Interval.   P301

Let $k \in [0, 100]$

A tolerance interval for capturing at least k% of value in population is

$\bar{x} \pm$ tolerance CV $\cdot$ S.

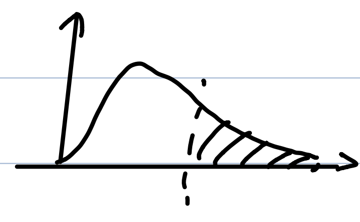
7.4 CI for the Variance & SD of a Normal Dis.

1. Assumption.

① Normal Dis

② θ, v follows Chi^2 Dis

Chi^2-Dis



2. $100(1-\alpha)\%$ CI for σ^2 of normal Dis.

$\left((n-1)S^2/\chi^2_{\alpha/2, n-1},\ (n-1)S^2/\chi^2_{(1-\alpha/2), n-1} \right)$

Chapter 8. Test of Hypothesis Based on single sample

8.1

1. P-value & α-value.

- $\boxed{\begin{array}{l}\text{If P-value} \le \alpha\text{, we reject } H_0 \\ \text{If P-value} > \alpha\text{, we fail to reject } H_0\end{array}}$

- Important Points: (P316)

① P-value is a probability

② P-value is calculated assume H_0 is true.

③ The ↓ P-value, the ↗ evidence against H_0

2. Erros:

- Type I error: reject H_0 but H_0 is true
- Type II error: fail to reject H_0 but H_0 is false.
- α = Type I error.

8.2. Z Test for Hypothesis of μ.

1. ① $H_0: \mu = \mu_0$

$H_a: \mu \ne \mu_0$

② $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

③ find P-value based on z.

H_α	P-value
$H_\alpha: \mu > \mu_0$	Area right of z.
$H_\alpha: \mu < \mu_0$	Area left of z.
$H_\alpha: \mu \neq \mu_0$	$2 \cdot$ Area to right of $\lvert z \rvert$.

④ Compare & Conclude. $\Phi(a)$ means apply a to z-table

2. β and Sample size.

type II error

Alternative Hypo.	Type II error.
$H_a: \mu > \mu_0$	$\beta(\mu') = \Phi\left(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$
$H_a: \mu < \mu_0$	$1 - \Phi\left(-z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$
$H_a: \mu \neq \mu_0$	$\Phi\left(z_{\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) - \Phi\left(-z_{\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$.

Sample size n for $\beta(\mu') = \beta$ at μ':

$$n \begin{cases} \left[\dfrac{\sigma(z_\alpha + z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{for one-tailed test.} \\ \left[\dfrac{\sigma(z_{\alpha/2} + z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{for two-tailed test.} \end{cases}$$

$z_{0.01} = 2.33$ $z_{0.1} = 1.28$.

$z_{0.05} = 1.64$.

8.3: One-Sample t-test

1. $H_0: \mu = \mu_0$

$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ with $df = n-1$

Table A.8

for $t > 0$

for $t < 0$

Alternate Hypo	P-value.
$H_a: \mu > \mu_0$	Area right of t_{n-1}
$H_a: \mu < \mu_0$	Area left of t_{n-1}.
$H_a: \mu \neq \mu_0$	2·Area right of t_{n-1}.

2. β and n.

Skip

8.4. Proportion Test.

1. Procedure

① $H_0: P = P_0$

② $$Z = \frac{\hat{P} - P_0}{\sqrt{P_0(1-P_0)/n}}$$

③

Alternative Hypothesis	P-value.
$H_a: P > P_0$	Area to the right of Z.
$H_a: P < P_0$	Area to the left of Z.
$H_a: P \neq P_0$	$2 \times$ Area to the right of $\lvert Z \rvert$

④ Assumption.

$nP_0 \geq 10$ & $n(1-P_0) \geq 10$

2. β and n.

① β

[code] ✓

Alternative	$\beta(P')$
$H_a: P > P_0$	$\Phi\left[\frac{P_0 - P' + Z_\alpha\sqrt{P_0(1-P_0)/n}}{\sqrt{P'(1-P')/n}}\right.$

$H_a: p < p_0$ $\quad 1 - \Phi\left[\dfrac{p_0 - p' - z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$

$H_a: p \neq p_0$ $\quad \Phi\left[\dfrac{p_0 - p' + z_{\alpha}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right] -$

$z_{0.01} = 2.33$ $\quad z_{0.1} = 1.28$

$z_{0.05} = 1.64$ $\quad \Phi\left[\dfrac{p_0 - p' - z_{\alpha}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$

② n: [code] ✓

$$n \begin{cases} \left[\dfrac{z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p'(1-p')}}{p' - p_0}\right]^2 & \text{one-tailed test} \\ \left[\dfrac{z_{\alpha/2}\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p'(1-p')}}{p' - p_0}\right]^2 & \text{two-tailed test.} \end{cases}$$

3. Small sample p test.

① P-value $= 1 - B(x-1; n, p_0)$?

② $\beta(p') = B(c_\alpha - 1; n, p')$

Chapter 9. Inference Based on Two Sample

9.1. Z Test and CI for difference

1. Assumption:

① $X_1 \cdots X_m \in$ Dis with μ_1, σ_1^2

② $Y_1 \cdots Y_n \in$ Dis with μ_2, σ_2^2

③ X and Y are independent.

2. $E(\bar{X}-\bar{Y}) = \mu_1-\mu_2$.

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}$$

3. Test Procedure.

① Null Hypothesis: $H_0: \mu_1-\mu_2=\Delta_0$

② Test statistic:

$$Z = \frac{\bar{x}-\bar{y}-\Delta_0}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}$$

Alternative	P-value.
$H_a: \mu_1-\mu_2>\Delta_0$	Area to the right of z.
$H_a: \mu_1-\mu_2<\Delta_0$	Area to the left of z.
$H_a: \mu_1-\mu_2\neq\Delta_0$	$2\times$ Area to the right of $\|z\|$.

4. β and Sample Size.

Alternative	$\beta(\Delta') = P(\text{Type II err})$
$H_a: \mu_1-\mu_2>\Delta_0$	$\Phi\left(z_\alpha - \frac{\Delta'-\Delta_0}{\sigma}\right)$.
$H_a: \mu_1-\mu_2<\Delta_0$	$1-\Phi\left(-z_\alpha - \frac{\Delta'-\Delta_0}{\sigma}\right)$
$H_a: \mu_1-\mu_2\neq\Delta_0$	$\Phi\left(z_{\alpha/2}-\frac{\Delta'-\Delta_0}{\sigma}\right)-\Phi\left(-z_{\alpha/2}-\frac{\Delta'-\Delta_0}{\sigma}\right)$

$\sigma = \sqrt{\sigma_1^2/m+\sigma_2^2/n}$.

4. large Sample Test.

① Assumption:

By CLT, $\boxed{m>40,\ n>40.}$

$$Z = \frac{\bar{x}-\bar{y}-\Delta_0}{\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}}}.$$

4. CI for $\mu_1-\mu_2$. or $\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}$

$100(1-\alpha)\%$:

$$\bar{x}-\bar{y} \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}}$$

9.2 Two-Sample t Test & CI

1. t-statistic

$$T=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}}}$$ [Code √]

$$\nu=\frac{\left(\frac{S_1^2}{m}+\frac{S_2^2}{n}\right)^2}{\frac{(S_1^2/m)^2}{m+1}+\frac{(S_2^2/n)^2}{n-1}}$$ (round down to integer)

[Code] √

2. CI for two Sample

two-sample t CI for $\mu_1-\mu_2$ $100(1-\alpha)\%$

$$\bar{x}-\bar{y} \pm t_{\alpha/2,\nu}\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}}$$

3. two-sample t test

$H_0 = \mu_1-\mu_2=\Delta_0$

Test Statistic:

$$t=\frac{\bar{x}-\bar{y}-\Delta_0}{\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}}}$$

$$t=\frac{\bar{x}-\bar{y}-\Delta_0}{S_p\sqrt{\frac{1}{m}+\frac{1}{n}}}$$ for pooled variance

$df = m+n-2$.

Alternative Hypo	P-value
$H_a: \mu_1-\mu_2>\Delta_0$	Area right of t.
$H_a: \mu_1-\mu_2<\Delta_0$	Area left of t
$H_a: \mu_1-\mu_2\neq\Delta_0$	$2\times$ Area right of $\lvert t\rvert$.

4. Pooled t-Procedures

① Assumption:

- Two populations are normal distribution
- Two populations have equal variance

② Pooled variance. [Code √]

$$S_p^2 = \frac{m-1}{m+n-2} S_1^2 + \frac{n-1}{m+n-2} S_2^2$$

⇒ two-sample t CI for $\mu_1 - \mu_2$ $100(1-\alpha)\%$

$$\bar{x} - \bar{y} \pm t_{\alpha/2}, S_p \cdot \sqrt{\frac{1}{m} + \frac{1}{n}}$$

9.3. Analysis of Paired Data.

1. Assumption.

① N independent selected pairs.

② $E(X_i) = \mu_1$ & $E(Y_i) = \mu_2$

③ Let $D_n = X_n - Y_n$.

④ D assumed to be normally distributed. with μ_D and σ_D^2

[use minitab]

2. Paired t Test

① $H_0 : \mu_D = \Delta_0$

② $t = \frac{\bar{d} - \Delta_0}{S_D / \sqrt{n}}$

. $\bar{d} = \bar{X} - \bar{Y}$

$S_d = \frac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{n}$

$df = n-1$

③ Alternative Hypo — P-value.

$H_a : \mu_D > \Delta_0$ — Area right of t.

$H_a : \mu_D < \Delta_0$. — Area left of t

$H_a : \mu_D \neq \Delta_0$ — $2 \times$ Area to right of $|t|$.

3. Paired t CI

$$T = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$$

Then the paired t CI for μ_D is

$$\bar{d} \pm t_{\alpha/2, n-1} \cdot S_D / \sqrt{n}.$$

4. Paired Data and Two-sample t Procedure.

$$V(\bar{X} - \bar{Y}) = \frac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{n}.$$

Difference between two-sample t and paired-t test.

① Two sample t: Groups are independent.
i.e. samples from different population.

② Paired-t test: Groups are dependent.
i.e. samples from same population.

9.4. Inference Concerning a Difference Btw Population Proportions.

1. Proposition.

Let $\hat{p}_1 = X/m$, $\hat{p}_2 = Y/n$ ($X \sim Bin(m, p_1)$, $Y \sim Bin(n, p_2)$).

Then $E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$.

$$V(\hat{p}_1 - \hat{p}_2) = \frac{p_1 q_1}{m} + \frac{p_2 q_2}{n} \quad (q = 1-p)$$

2. Large Sample Procedure.

① $H_0: p_1 - p_2 = 0$

② $Z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{m}+\frac{1}{n}\right)}}$ $\rightarrow$ $\hat{p} = \dfrac{X+Y}{m+n} = \dfrac{m}{m+n}\cdot\hat{p}_1 + \dfrac{n}{m+n}\cdot\hat{p}_2$

Alternate Hypo	P-value
$H_a: p_1 - p_2 > 0$	Area right of z.
$H_a: p_1 - p_2 < 0$	Area left of z
$H_a: p_1 - p_2 \neq 0$	$2 \times$ Area right of $\lvert z \rvert$.

③ Assumption

$\boxed{m\hat{p}_1,\ m\hat{q}_1,\ n\hat{p}_2,\ n\hat{q}_2 > 10.}$

④ Type II error and Sample size. [Code √]

Alternate Hype	$\beta(p_1, p_2)$
$H_a: p_1 - p_2 > 0$	$\Phi\left[\dfrac{z_\alpha\sqrt{\bar{p}\bar{q}\left(\frac{1}{m}+\frac{1}{n}\right)} - (p_1 - p_2)}{\sigma}\right]$
$H_a: p_1 - p_2 < 0$	$1 - \Phi\left[\dfrac{z_\alpha\sqrt{\bar{p}\bar{q}\left(\frac{1}{m}+\frac{1}{n}\right)} - (p_1 - p_2)}{\sigma}\right]$

$H_a: p_1 - p_2 \neq 0$ $\quad \Phi\left[\frac{z_{\alpha/2}\sqrt{\bar{p}\,\bar{q}_e\left(\frac{1}{m}+\frac{1}{n}\right)} - (p_1 - p_2)}{\sigma}\right] -$

$$- \Phi\left[\frac{-z_{\alpha/2}\sqrt{\bar{p}\,\bar{q}_e\left(\frac{1}{m}+\frac{1}{n}\right)} - (p_1 - p_2)}{\sigma}\right]$$

where $\sigma = \sqrt{\frac{p_1 q_{e1}}{m} + \frac{p_2 q_{e2}}{n}}$

$\bar{p} = (mp_1 + np_2)/(m+n)$

$q_e = (m q_{e1} + n q_{e2})/(m+n)$.

- Sample size.

For $m=n$, the level α test has type II error β with $p_1 - p_2 = d$.

$$n = \frac{\left[z_\alpha\sqrt{(p_1+p_2)(q_{e1}+q_{e2})/2} + z_\beta\sqrt{p_1 q_{e1} + p_2 q_{e2}}\right]^2}{d^2}$$

② large Sample CI

- A CI for $p_1 - p_2$ with CI $100(1-\alpha)\%$ is.

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{p_1 q_{e1}}{m} + \frac{p_2 q_{e2}}{n}}$$

(* $mp_1, mq_{e1}, np_2, nq_{e2} \geq 10$)