chapter four

	Mean	Proportion
Central limit theorem	$\mu_{\bar{x}} = \mu$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	$\mu_{\hat{p}} = p$ $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$
Z-Score	$\frac{\bar{x}-\mu}{\sigma}$	$\frac{\hat{p}-\mu_{\hat{p}}}{\sigma_{\hat{p}}}$
Confidence interval	$\mu \in \bar{x} \mp Z_{1-\alpha/2}\, \sigma/\sqrt{n}$	$p \in \hat{p} \mp Z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}$
Margine error	$\delta_{\mu} = Z_{1-\frac{\alpha}{2}}\, \sigma/\sqrt{n}$	$\delta_{p} = Z_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1-\hat{p})/n}$
Sample size	$n = (\sigma Z_{1-\frac{\alpha}{2}} / \delta_{\mu})^2$	$n = \hat{p}(1-\hat{p})(Z_{1-\alpha/2} / \delta_p)^2$ $n = 0.25 (Z_{1-\alpha/2} / \delta_p)^2$
Confidence interval length		$L = 2\delta_p \rightarrow 2 Z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}$ $n = (Z_{1-\alpha/2} / L)^2$

Hypothesis testing for the population mean:

1. Identify the hypothesis:

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq, >, < \mu_0$

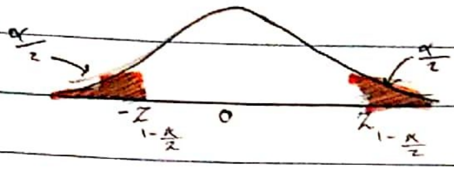
2. Test statistic:

$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

3. critical region or P-value:

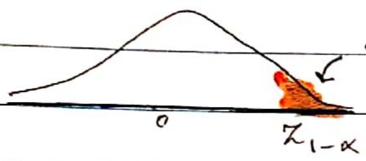
• critical region:

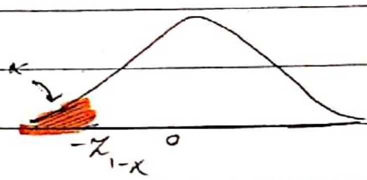
i. $H_1: \mu \neq \mu_0$



ii. $H_1: \mu > \mu_0$

iii. $H_1: \mu < \mu_0$





• P-value:

i. $H_1: \mu \neq \mu_0 \rightarrow 2P(Z > |Z_0|)$

ii. $H_1: \mu > \mu_0 \rightarrow P(Z > Z_0)$

iii. $H_1: \mu < \mu_0 \rightarrow P(Z < -|Z_0|)$

4. Decision:

we reject H_0 when it falls in the critical region or α is greater than the p-value.

Hypothesis testing for the population proportion:

1. Identify the hypothesis:

$H_0: p = p_0$

$H_1: p \neq, <, > p_0$

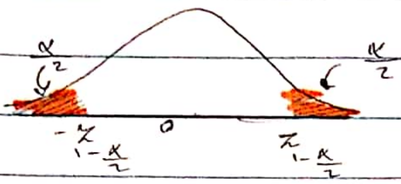
2. Test statistic:

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

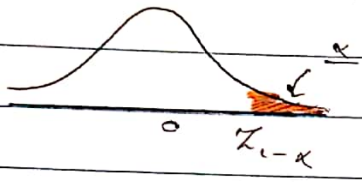
3. critical region or p-value:

• Critical region:

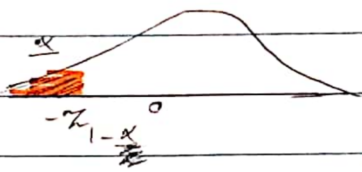
i. $H_1: p \neq p_0$



ii. $H_1: p > p_0$



iii. $H_1: p < p_0$



• P-value:

i. $H_1: p \neq p_0 \rightarrow 2P(Z > |Z_0|)$

ii. $H_1: p > p_0 \rightarrow P(Z > Z_0)$

iii. $H_1: p < p_0 \rightarrow P(Z < -|Z_0|)$

4. Decision:

the same as before.

chapter five

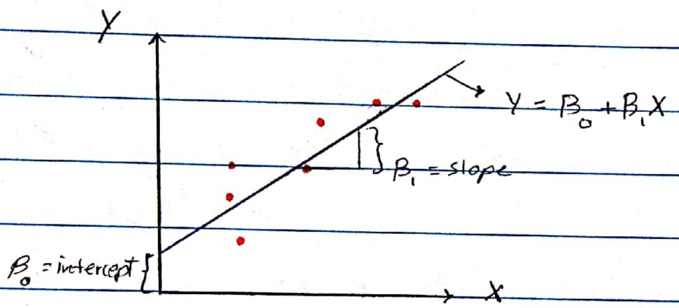
- **Pearson's correlation coefficient**: Also called Product moment correlation coefficient

$$r = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y_i-\bar{y})^2}}, \qquad r = \frac{n\sum x_i y_i - [\sum x_i][\sum y_i]}{\sqrt{n\sum x_i^2 - [\sum x_i]^2}\sqrt{n\sum y_i^2 - [\sum y_i]^2}}$$

- **Simple linear regression line of a population**: $Y = a + bX + \varepsilon$

where, ε is a normal random variable with $E(\varepsilon) = 0$.

a is the Y-intercept and b is the slope.



- **The estimated regression line of a ~~py~~ sample**: $\hat{Y} = \hat{a} + \hat{b}X$

$$\hat{b} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}, \qquad \hat{b} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - [\sum x_i]^2}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}, \qquad \hat{a} = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n\sum x_i^2 - [\sum x_i]^2}$$

- **Coefficient of determination**:

$$r^2 = \frac{SSR}{SS_{tot}}, \qquad r^2 = 1 - \frac{SSE}{SS_{tot}}$$

- **Total sum of squared deviations (Total variation)**: $SS_{tot} = \sum(y_i - \bar{y})^2$,

$SS_{tot} = SSR + SSE$

- **Sum of squared regression error (Explained variation)**: $SSR = \sum(\bar{y} - \hat{y}_i)^2$

- **Sum of squared error or residuals (Unexplained variation)**: $SSE = \sum(y_i - \hat{y}_i)^2 = \sum \varepsilon_i^2$