Summary

Friday, 23 January 2026 4:31 PM

Things we care to make conclusions about

The mean of a sample in a population

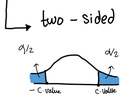
Estimation → CH6

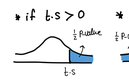
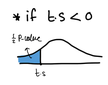
- Point estimation
  - $\bar{X}$ is an unbiased estimator of μ
  - S^2/n is not an unbiased estimator of σ^2
  - $S/\sqrt{n}$ is not an unbiased estimator of σ
- interval estimation → by finding the confidence interval
  - for a population of known variance: $\bar{X} \pm Z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}}$
  - for a population with unknown variance and $n > 200$: $\bar{X} \pm Z_{1-\alpha/2} * \frac{S}{\sqrt{n}}$
  - for a population with unknown variance and $n \leq 200$: $\bar{X} \pm t_{n-1,\,1-\alpha/2} * \frac{S}{\sqrt{n}}$

Hypothesis testing ← CH7

$H_0: \mu = \mu_0$

$H_1: \mu > \mu_0$, $\mu < \mu_0$ → one-sided; $\mu \neq \mu_0$ → 2-sided

- Left-sided
  - 
  - C.V $= -Z_{1-\alpha}$
  - C.V $= -t_{n-1,\,1-\alpha}$
- right-sided
  - 
  - C.V $= Z_{1-\alpha}$
  - C.V $= t_{n-1,\,1-\alpha}$
- two-sided
  - 
  - C.V $= -Z_{1-\alpha/2}$ and $Z_{1-\alpha/2}$
  - C.V $= -t_{n-1,\,1-\alpha/2}$ and $t_{n-1,\,1-\alpha/2}$

* if t.s > 0: 
* if t.s < 0: 

* we multiply by 2

the test statistic used in all cases is $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

or S if the variance of the Population was unknown and we use the t-test

notes :-

* P-value $> \alpha$ is always accept H_0
* P-value $< \alpha$ is always reject H_0
* when we reject H_0 then the test is significant
* when we accept H_0 then the test is not significant
* the length of the confidence interval $L = 2 Z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}}$ or $2 t_{n-1,\,1-\alpha/2} * \frac{S}{\sqrt{n}}$
* to determine the sample size n: $n = \left(2 * Z_{1-\alpha/2} * \frac{\sigma}{L}\right)^2$ or $\left(2 * t_{n-1,\,1-\alpha/2} * \frac{S}{L}\right)^2$ → we must round the answer up ⚠
* if the confidence level wasn't specified then we use $\alpha = 0.05$ in the

The proportion (Prevalence) of a sample in a population

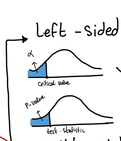
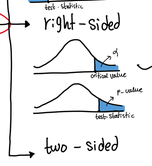
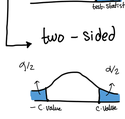
Estimation → CH6

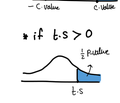
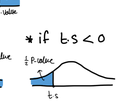
- Point estimation
  - $\hat{P}$ is an unbiased estimator of P
  - $\frac{\hat{P}\hat{Q}}{n}$ is not an unbiased estimator of $\frac{PQ}{n}$
  - $\sqrt{\frac{\hat{P}\hat{Q}}{n}}$ is not an unbiased estimator of $\sqrt{\frac{PQ}{n}}$
- interval estimation → by finding the confidence interval: $\hat{P} \pm Z_{1-\alpha/2} * \sqrt{\frac{\hat{P}\hat{Q}}{n}}$

Hypothesis testing ← CH7

$H_0: P = P_0$

$H_1: P < P_0$, $P > P_0$ → one-sided; $P \neq P_0$ → 2-sided

- Left-sided
  - 
  - C.V $= -Z_{1-\alpha}$
- right-sided
  - 
  - C.V $= Z_{1-\alpha}$
- two-sided
  - 
  - C.V $= -Z_{1-\alpha/2}$ and $Z_{1-\alpha/2}$

* if t.s > 0: 
* if t.s < 0: 

* we multiply by 2

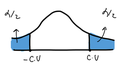
the test statistic used in all cases is $Z = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0 Q_0}{n}}}$

The difference between the means of 2 Samples → CH8

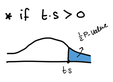
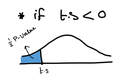
Paired t-test → $H_0: \Delta = 0$, $H_1: \Delta \neq 0$

the 2 sets of data are of the same sample of people at different times

→ the test statistic is $\frac{\bar{d}}{S_d/\sqrt{n}}$



the C.V = $-t_{n-1,\,1-\alpha/2}$ and $t_{n-1,\,1-\alpha/2}$

* if t.s > 0: 
* if t.s < 0: 

* we multiply by 2

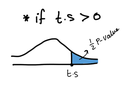
→ the confidence interval: $\bar{d} \pm t_{n-1,\,1-\alpha/2} * \frac{S_d}{\sqrt{n}}$

Un-Paired t-test → $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$

for 2 different samples of people which could be of the same population or of different population



the critical values $-t_{n_1+n_2-2,\,1-\alpha/2}$ and $t_{n_2+n_1-2,\,1-\alpha/2}$

* if t.s > 0: 
* if t.s < 0: 

→ the test statistic is $\frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

if the variance of the population is unknown then we estimate it using the Pooled variance $\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$

→ the confidence interval: $(\bar{X}_1 - \bar{X}_2) \pm t_{n_2+n_1-2,\,1-\alpha/2} * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

$\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_2+n_1-2}}$

* note that if the confidence interval contains 0 then we accept H_0 ⚠
* if the confidence interval didn't contain 0 then we reject H_0

The difference between the proportion of 2 Samples → CH10

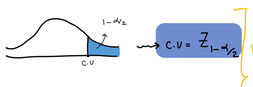
normal theory method (to test wether there is a difference between the 2 proportions or not)

⇒ the test statistic

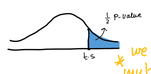
- corrected: $\frac{|\hat{P}_1 - \hat{P}_2| - \left(\frac{1}{2n_1} + \frac{1}{2n_2}\right)}{\sqrt{P_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ (estimated using the Pooled Proportion)
- uncorrected: $\frac{\hat{P}_1 - \hat{P}_2}{\sqrt{P_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$P_0 = \frac{n_1\hat{P}_1 + n_2\hat{P}_2}{n_1+n_2} = \frac{x_1 + x_2}{n_1+n_2}$

$q_0 = 1 - P_0$



C.V = $Z_{1-\alpha/2}$



* we have to multiply by 2

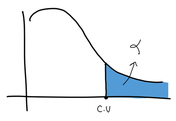
notice that we are only working on the right side because t.s is (+)

contingency table method (to test wether there is a relationship between 2 variables or not)

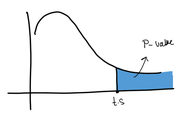
$\Rightarrow E_{ij} = \frac{\text{row margin of the cell} * \text{column margin of the cell}}{\text{grand total}}$

⇒ the test statistic

- 2x2 table — yates-corrected test: $\chi^2 = \sum \frac{(|O_{ij} - E_{ij}| - 0.5)^2}{E_{ij}}$ (we will be calculating the sum of 4 terms)
- RxC table: $\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ (we will be calculating the sum of n terms)



C.V = $\chi^2_{(R-1)(C-1),\,1-\alpha}$



(*) the goodness-of-fit test → whether the distribution is normal or not

⇒ the test statistic depends on the given information in the question

⇒ the critical value is $\chi^2_{g-k-1,\,1-\alpha}$

- g: number of groups
- k: number of parameters of a distribution (2 for normal unless something else is in the question)

The relationship between 2 variables → CH11

they are normally distributed

* $\rho = 0$ ⇒ no linear relationship
* $\rho = 1$ ⇒ there is a Perfect Positive relationship
* $\rho = -1$ ⇒ there is a Perfect negative relationship

- ρ the correlation of the Population
- r the correlation of the sample
  - $r > 0$ Positively Correlated
  - $r = 0$ no Correlation
  - $r < 0$ negatively Correlated

$\frac{L_{xy}}{\sqrt{L_{xx} * L_{yy}}}$

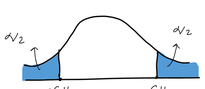
- Corrected sum of Cross Product $= \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$
- Corrected sum of squares of y $= \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$
- Corrected sum of Squares of X $= \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$

Testing for the correlation

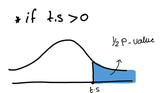
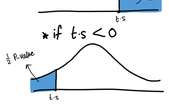
$H_0: \rho = 0$, $H_1: \rho \neq 0$ ⇒ to test wether there is a relationship (correlation) or not

Test statistic = $r\sqrt{\frac{n-2}{1-r^2}}$

$r = \frac{L_{xy}}{\sqrt{L_{xx} * L_{yy}}}$



C.V = $-t_{n-2,\,1-\alpha/2}$ and $t_{n-2,\,1-\alpha/2}$

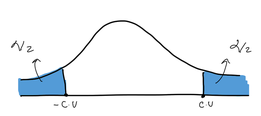
* if t.s > 0: 
* if t.s < 0: 

* we have to multiply by 2

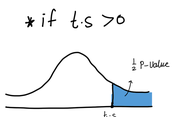
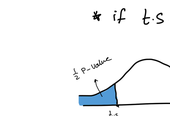
$H_0: \rho = \rho_0$, $H_1: \rho \neq \rho_0$ → to test wether the correlation equals a certain value or not

test statistic = $\lambda = \left((Z - Z_0) * \sqrt{n-3}\right)$

- Z: the transformation of r — using the table, or $Z = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right)$
- Z_0: the mean $= \frac{1}{2} \ln\left(\frac{1+\rho_0}{1-\rho_0}\right)$



C.V = $-Z_{1-\alpha/2}$ and $Z_{1-\alpha/2}$

* if t.s > 0: 
* if t.s < 0: 

* we have to multiply by 2

the confidence interval (ρ_1, ρ_2)

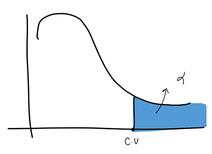
① find (Z_1, Z_2): $Z - \frac{Z_{1-\alpha/2}}{\sqrt{n-3}}$ and $Z + \frac{Z_{1-\alpha/2}}{\sqrt{n-3}}$ (the transformation of r)

② find (ρ_1, ρ_2): $\frac{e^{2Z_1} - 1}{e^{2Z_1} + 1}$ and $\frac{e^{2Z_2} - 1}{e^{2Z_2} + 1}$

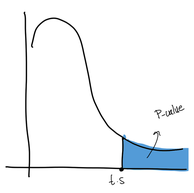
Comparing the means of samples to decide if at least one is different ANOVA → CH12

⇒ the test statistic = $\frac{\text{Between MS}}{\text{within MS}}$

- Between SS/(k-1): $\sum_{i=1}^{k} n_i \bar{y}_i^2 - \frac{\left(\sum_{i=1}^{k} n_i \bar{y}_i\right)^2}{n}$
- within SS/(n-k): $\sum_{i=1}^{k} (n_i - 1) S_i^2$



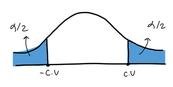
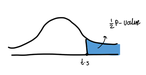
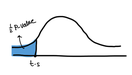
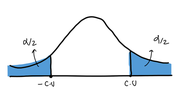
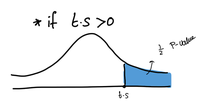
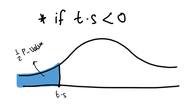
C.V = $F_{(k-1),(n-k),\,1-\alpha}$ (d_1, d_2)



Comparing the means of specific groups among all groups → CH12

⚠ in chapter 8 we only compare 2 groups

in chapter 12 we compare 2 groups among many groups

- if the variance of the Population is known
  - ⇒ the test statistic = $\frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$
  - 
  - C.V = $-Z_{1-\alpha/2}$ and $Z_{1-\alpha/2}$
  - * if t.s > 0: 
  - * if t.s < 0: 
  - * we have to multiply by 2
- if the variance is unknown
  - ⇒ test statistic = $\frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$
  - Pooled variance estimated as within ms $= \frac{\text{within SS}}{n-k} = \sum_{i=1}^{k} \frac{(n_i - 1) S_i^2}{n-k}$
  - 
  - C.V = $-t_{n-k,\,1-\alpha/2}$ and $t_{n-k,\,1-\alpha/2}$
  - * if t.s > 0: 
  - * if t.s < 0: 
  - * we have to multiply by 2

the confidence interval of $\mu_1 - \mu_2$: $(\bar{Y}_1 - \bar{Y}_2) \pm t_{n-k,\,1-\alpha/2} * \sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

* Important rules :-

→ whenever we use the critical value method

- t.s in α ⇒ reject H_0
- t.s outside α ⇒ Accept H_0

→ P-value $> \alpha$ ⇒ accept H_0

→ P-value $< \alpha$ ⇒ reject H_0