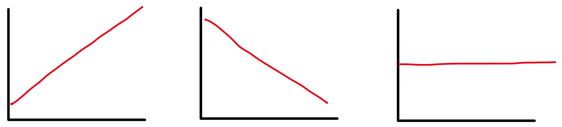
chapter 11 – Correlation

Friday, 2 January 2026 5:02 PM

* Scatter Plot : a graphical representation that displays individual data points that correspond to 2 variables, one on x-axis and the other on y-axis

→ the scatter plot examines the relationship of the 2 variables
↓
correlation ρ



$\rho > 0$: there is a positive linear relationship between the 2 variables

$\rho < 0$: there is a negative linear relationship between the 2 variables

$\rho = 0$: there is no linear relationship between the 2 variables

* $\rho = 0 \Rightarrow$ no linear relationship
* $\rho = 1 \Rightarrow$ Perfect Positive linear relationship
* $\rho = -1 \Rightarrow$ Perfect negative linear relationship

* * *

■ Some important calculations :)

→ the raw sum of squares for $X \Rightarrow \sum_{i=1}^{n} x_i^2$

→ the corrected sum of squares for $X \Rightarrow \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = L_{xx}$

→ the raw sum of squares for $y = \sum_{i=1}^{n} y_i^2$

→ the corrected sum of squares for $y = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = L_{yy}$

→ the raw sum of the cross Product $= \sum_{i=1}^{n} x_i y_i$

→ the corrected sum of cross Product is $\sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n} = L_{xy}$

$= \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$

we are going to study hypotheses related to correlation

→ sample correlation coefficient $r = \frac{L_{xy}}{\sqrt{L_{xx} L_{yy}}}$ → lies between 1 and -1

- $r > 0$: the variables are positively correlated
- $r = 0$: the variables aren't correlated
- $r < 0$: the variables are negatively correlated

Correlated ⇒ there is a linear relationship

we use the correlation coefficient r only if the 2 variables used in calculating it are normally distributed.

* r is the correlation coefficient of a Sample which is a point estimate of the Population correlation coefficient $\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$

$r = \frac{S_{xy}}{S_x S_y}$

* Testing for a correlation coefficient

if $H_0: \rho = 0$ vs $H_1: \rho \neq 0$ we use t-test

⇒ steps :)

① find $r = \frac{L_{xy}}{\sqrt{L_{xx} L_{yy}}}$

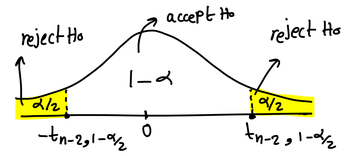
② find the test statistic $t = r\sqrt{\frac{(n-2)}{(1-r^2)}}$ ← this follows t-distribution of (n-2) degree of freedom

Critical value method

→ critical values $-t_{n-2,1-\alpha/2}$, $t_{n-2,1-\alpha/2}$

⇒ t.s > critical or t.s < -critical ⇝ reject H_0

⇒ -critical < t.s < critical ⇝ accept H_0



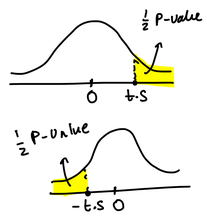
P-value method

T.S ≥ 0 → $2P(t > t.s)$

T.S < 0 → $2P(t < t.s)$

⇒ P-value > α ⇝ accept H_0

⇒ P-value < α ⇝ reject H_0



if $H_0: \rho = \rho_0$ vs $H_1: \rho \neq \rho_0$ we use z-test

⇒ Steps :)

① we find the z transformed correlation coefficient

① find $r = \frac{L_{xy}}{\sqrt{L_{xx} L_{yy}}}$

② find the fisher's z transformation $= \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right)$ ⇒ which could be calculated directly using the table

⚠ important notes :-

- r is a distribution which isn't normally distributed
- we use fisher's z transformation to transform it to an approximately normal distribution
- Z (which is normally distributed) has a mean $Z_0 = \frac{1}{2} \ln\left(\frac{1+\rho_0}{1-\rho_0}\right)$ and a variance $= \frac{1}{n-3}$

③ find the test statistic λ

$\lambda = (Z - Z_0)\sqrt{n-3}$

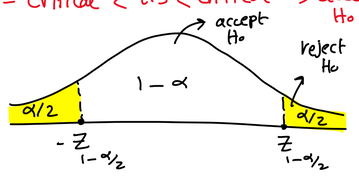
the transformation of r — the mean of Z distribution

Critical value method

→ critical values $-Z_{1-\alpha/2}$, $Z_{1-\alpha/2}$

t.s > critical or t.s < critical → reject H_0

- critical < t.s < critical → accept H_0



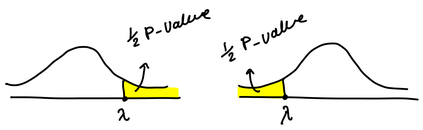
P-value method

$\lambda > 0 \Rightarrow P = 2P(z > t.s)$

$\lambda < 0 \Rightarrow P = 2P(z < t.s)$

P-value > α ⇒ accept H_0

P-value < α ⇒ reject H_0



* note that X and y variables are normally distributed

* * *

Finding the confidence interval for ρ which is (ρ_1, ρ_2)

⚠ important notes :)

in previous chapters we used to find CI for the distribution of means $\bar{X}$ which is normally distributed $\bar{X} \sim N(\mu, \sigma^2/n)$

now to find CI for ρ we will be finding the CI of Z_ρ which is the mean of Zr normal distribution

$Z_r \sim N\left(Z_\rho, \frac{1}{n-3}\right)$, $Z_\rho = \frac{1}{2} \ln\left(\frac{1+\rho}{1-\rho}\right)$

⇒ there isn't a direct way to find ρ (1-α)% CI that is why we find CI for Z_ρ then transform it into CI of ρ

Steps :)

① Z_ρ CI interval (Z_1, Z_2)

$Z_1 = Z - \frac{Z_{1-\alpha/2}}{\sqrt{n-3}}$

$Z_2 = Z + \frac{Z_{1-\alpha/2}}{\sqrt{n-3}}$

Z: the fisher's Z transformation of r

② ρ CI interval (ρ_1, ρ_2)

$\rho_1 = \frac{e^{2Z_1} - 1}{e^{2Z_1} + 1}$

$\rho_2 = \frac{e^{2Z_2} - 1}{e^{2Z_2} + 1}$