1) Consider 2 random variables X and Y with joint PMF given in table.

Let $Z = E[X|Y]$

	$y=0$	$y=1$
$X=0$	$1/5$	$2/5$
$X=1$	$2/5$	0

a) Find the marginal PMFs of X and Y.

$P_X(0) = \frac{1}{5} + \frac{2}{5} = \frac{3}{5}$ $\quad$ $P_X(1) = \frac{2}{5} + 0 = 2/5$

$P_Y(0) = \frac{1}{5} + \frac{2}{5} = \frac{3}{5}$ $\quad$ $P_Y(1) = \frac{2}{5} + 0 = \frac{2}{5}$

$P(x) \cdot P(y) = P(x \cap y) \Rightarrow$ independent!

$P_X(0) \cdot P_Y(0) \stackrel{?}{=} P(X=0 \cap Y=0)$

$\frac{3}{5} \cdot \frac{3}{5} \neq \frac{1}{5}$ so, they are dependent!

b) Find the conditional PMF of X given $Y=0$ and $Y=1$.

So, $P_{X|Y}(x|0)$ and $P_{X|Y}(x|1)$

If $Y=0 \Rightarrow$ $P_{X|Y}(0|0) = \frac{P_{X|Y}(0,0)}{P_Y(0)} = \frac{1/5}{3/5} = \frac{1}{3}$

So, $P_{X|Y}(1|0) = 1 - \frac{1}{3} = \frac{2}{3}$

If $Y=1 \Rightarrow$ $P_{X|Y}(0|1) = \frac{P_{X|Y}(0,1)}{P_Y(1)} = \frac{2/5}{2/5} = 1$

So, $P_{X|Y}(1|1) = 1 - 1 = 0$ ✓

c) Find the PMF of Z. $\swarrow Z = E[X|Y]$

We note that Y can take 2 values: 0 ve 1. Thus, the random variable $Z = E[X|Y]$ can take two values as it is a function of Y. Specifically,

$$Z = E[X|Y] = \begin{cases} E[X|Y=0] & \text{if } Y=0 \\ E[X|Y=1] & \text{if } Y=1 \end{cases}$$

$$E[X|Y=0] = x_0 \, P(X=0|Y=0) + x_1 \cdot P(X=1|Y=0)$$
$$= 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}$$

$$E[X|Y=1] = x_0 \; P(X=0|Y=1) + x_1 \cdot P(X=1|Y=1)$$
$$= 0 \cdot 1 + 1 \cdot 0 = 0$$

$P(Y=0) = \frac{3}{5}$ $\quad$ $P(Y=1) = \frac{2}{5}$

$$Z = E[X|Y] = \begin{cases} \frac{2}{3}, & \text{with prob. } \frac{3}{5} \\ 0, & \text{with prob } \frac{2}{5} \end{cases}$$

So we can write

$$P_Z(z) = \begin{cases} \frac{3}{5}, & \text{if } z = \frac{2}{3} \\ \frac{2}{5}, & \text{if } z = 0 \\ 0, & \text{otherwise} \end{cases}$$

d) Find $E[Z]$ and check that $E[X] = E[Z]$

$$E[Z] = \frac{2}{3} \cdot \frac{3}{5} + 0 \cdot \frac{2}{5} = \frac{2}{5}$$

$$E[X] = \frac{2}{5}$$

So, $E[X] = E[Z]$ ✓

e) Find $Var(Z)$

$$Var(Z) = E[Z^2] - (E[Z])^2$$

$$= E[Z^2] - \frac{4}{25} = \frac{4}{15} - \frac{4}{25} = \frac{8}{75}$$ //

$$E[Z^2] = \frac{4}{9} \cdot \frac{3}{5} + 0 \cdot \frac{2}{5} = \frac{4}{15}$$

$z^2 \quad P(z)$

Conditional Expectation $\Rightarrow$ $E[X \mid Y = y] = \sum_{\text{for } x_i} x_i \cdot P(x_i \mid y)$ ✓

Covariance
~~Correlation~~ of 2 random variables;

$$\operatorname{cov}(X,y) = E\left[(X-E[X])\cdot(y-E[y])\right] = E[Xy] - E[X]\,E[y]$$

If $\operatorname{cov}(X,y)=0$, X and y are uncorrelated!

Correlation coefficient:

$$\rho_{xy} = \operatorname{Corr}(X,y) = \frac{\operatorname{Cov}(X,y)}{\sigma_x \sigma_y} = \frac{\operatorname{Cov}(X,y)}{\sqrt{\operatorname{Var}(x)\cdot\operatorname{Var}(y)}}$$

Discrete

Ex: Let X and y be discrete random variables with joint mass function defined by $f_{X,y}(x,y) = \frac{1}{4}$,

$(x,y) \in \{(0,0),(1,1),(1,-1),(2,0)\}$

X \ y	-1	0	1
0	0	1/4	0
1	1/4	0	1/4
2	0	1/4	0

$$f_x(x) = \begin{cases} \frac{1}{4} & , x=0,2 \\ \frac{1}{2} & x=1 \end{cases}$$

$$f_y(y) = \begin{cases} \frac{1}{4} & , y=-1,1 \\ \frac{1}{2} & , y=0 \end{cases}$$

① Covariance

$\operatorname{cov}(X,y) = E[Xy] - E[X]\cdot E[y]$

$E[X] = \sum_{x=0}^{2} f_x(x)\cdot x = 0\cdot\frac{1}{4} + 1\cdot\frac{1}{2} + 2\cdot\frac{1}{4} = 1$

$E[y] = \sum_{y=-1}^{1} y\cdot f_y(y) = -1\cdot\frac{1}{4} + 0\cdot\frac{1}{2} + 1\cdot\frac{1}{4} = 0$

$$E[XY] = \sum_x \sum_y x \cdot y \cdot F_{x,y}(x,y)$$

$$= 0 \times 0 \times \frac{1}{4} + 1 \times 1 \times \frac{1}{4} + 1 \times -1 \times \frac{1}{4} + 2 \times 0 \times \frac{1}{4} = 0$$

$$Cov_{f_{x,y}}[x,y] = E[xy] - E[x] \cdot E[y]$$
$$= 0 - 1 \cdot 0 = 0$$

✓ correlation = zero! ✓

Continuous

Ex: Let X and Y be continuous random variables with joint pdf

$f_{x,y}(x,y) = 3x \quad 0 \leq y \leq x \leq 1$ and zero otherwise.

Step1: Marginal pdfs 'ler bulunmalı!

✓ $f_x(x) = \int_{-\infty}^{+\infty} f_{x,y}(x,y) \cdot dy = \int_0^x 3x \, dy = 3xy \Big|_0^x = \underline{\underline{3x^2}}, \quad 0 \leq x \leq 1$

y'ye göre integral

✓ $f_y(y) = \int_{-\infty}^{+\infty} f_{x,y}(x,y) \cdot dx = \int_y^1 3x \, dx = \frac{3x^2}{2}\Big|_y^1 = \frac{3}{2} - \frac{3y^2}{2} = \underline{\underline{\frac{3}{2}(1-y^2)}}, \quad 0 \leq y \leq 1$

x'e göre integral

Step2: Expected value

$$X \begin{cases} E_{f_x}[x] = \int_{-\infty}^{+\infty} x \cdot f(x) \cdot dx = \int_0^1 \overbrace{x \cdot 3x^2}^{3x^3} \cdot dx = \frac{3}{4} \cdot x^4 \Big|_0^1 = \frac{3}{4} \\ E_{f_x}[x^2] = \int_{-\infty}^{+\infty} x^2 \cdot f(x) \cdot dx = \int_0^1 \underbrace{x^2 \cdot 3x^2}_{3x^4} \cdot dx = \frac{3}{5} \cdot x^5 \Big|_0^1 = \frac{3}{5} \end{cases}$$

$$Y\left\{\begin{aligned}
E_{f_Y}[Y] &= \int_{-\infty}^{+\infty} y \cdot f_Y(y) \cdot dy = \int_0^1 y \cdot \frac{3}{2}(1-y^2) \cdot dy = \int_0^1 \frac{3}{2}(y-y^3)\, dy \\
&= \left[\frac{3}{2}\left(\frac{y^2}{2}-\frac{y^4}{4}\right)\right]\Bigg|_0^1 = \frac{3}{2}\left(\frac{1}{2}-\frac{1}{4}\right) = \frac{3}{8} \\
E_{f_Y}[Y^2] &= \int_0^1 \frac{3}{2}(y^2-y^4) \cdot dy = \left[\frac{3}{2}\left(\frac{y^3}{3}-\frac{y^5}{5}\right)\right]\Bigg|_0^1 = \frac{3}{2}\left(\frac{1}{3}-\frac{1}{5}\right) = \frac{1}{5}
\end{aligned}\right.$$

$$Cov(x,y) = E[XY] - E[X] \cdot E[Y]$$

$$E[XY] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x \cdot y \cdot f_{XY}(x,y) \cdot \underline{dy} \cdot dx = \int_0^1 \int_0^x xy \cdot 3x \cdot dy\, dx$$

$$= \int_0^1 \underbrace{\left\{\int_0^x y \cdot dy\right\}}_{\frac{y^2}{2}\Big|_0^x = \frac{x^2}{2}} 3x^2 dx = \int_0^1 \frac{x^2}{2} \cdot 3x^2 \cdot dx = \frac{3}{2} \cdot \frac{x^5}{5}\Bigg|_0^1 = \frac{3}{10}$$

İlk önce x'e göre integral alıp devam etsen de aynı sonucu verecek!

$$Cov(x,y) = \frac{3}{10} - \frac{3}{8} \times \frac{3}{4} = \frac{3}{160}$$

$$Corr[X,Y] = \frac{Cov[X,Y]}{\sqrt{Var[X]\ Var[Y]}} = \frac{Cov[X,Y]}{\sigma_x \cdot \sigma_y} = \frac{\frac{3}{160}}{\sqrt{\frac{3}{80} \cdot \frac{19}{320}}} = \boxed{0{,}397} \bigstar$$

$$Var[X] = E[X^2] - (E[X])^2$$

$$= \frac{3}{5} - \frac{9}{16} = \frac{3}{80}$$

$$Var[Y] = E[Y^2] - (E[Y])^2$$

$$= \frac{1}{5} - \left(\frac{3}{8}\right)^2 = \frac{19}{320}$$

Two-Sided Unknown Population Variance

Confidence Interval and Hypothesis Testing

Ex =) It is assumed that the mean systolic blood pressure is $\mu = 120$ mm Hg. In the Honolulu Heart Study, a sample of $n = 100$ people had an average systolic blood pressure of 130.1 mmHg with a standard deviation of 21.21 mm Hg. Is the group significantly different (with respect to systolic blood pressure) from the regular population?

Solution =) $H_0: \mu = 120$ (because there is no specific direction)

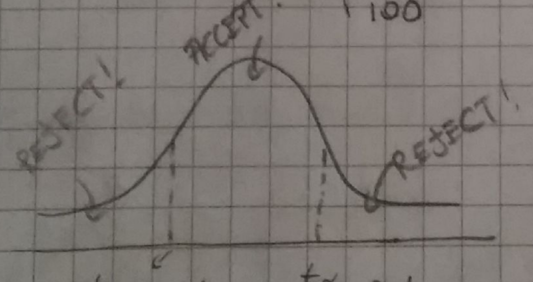
$H_A: \mu \neq 120$

$$T = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad \text{(general formula)}$$

For testing null hypothesis;

$H_0: \mu = \mu_0$ => so, $T = \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ (average, mean, deviation)

$$t_0 = \frac{130.1 - 120}{\frac{21.21}{\sqrt{100}}} = 4,762 \checkmark$$



$-t_{\alpha/2, n-1}$ $\quad t_{\frac{\alpha}{2}, n-1}$

$-t_{0,025, 99}$ $\quad t_{0,025, 99}$

$-1,984$ $\quad 1,984$

$\alpha = 0,05$ olsun...

two-sided old. için $\underbrace{t_{\alpha/2, n-1}}$ değerine bakılır (tablodan)

$t_{0,025, 99} = 1,984$

$t_0 > t_{0,025, 99}$ => $4,762 > 1,984$ => null hypothesis is in the reject region!

So, The group is significantly different from the regular population!

{ population variance unknown → t
" " known → z }

Confidence Interval hesaplaması...

By this way, the decision to reject the null hypothesis is consistent with the one you would make using a 95% confidence interval. Using the data, a 95% CI for the mean μ is

$$= \bar{x} \pm t_{0,025,\,99}\left(\frac{s}{\sqrt{n}}\right)$$

$$= 130.1 \pm 1,984\left(\frac{21.21}{\sqrt{100}}\right) = 130.1 \pm 4,21$$

α That is, we can be 95% confident that the mean systolic blood pressure of the Honolulu population is between 125.89 and 134.31 mm Hg!

z-distribution

Sampling from a normally distributed population – variance known

A simple random sample of 10 people from a certain population has a mean age of 27. Can we conclude that the mean age of the population is not 30? The variance is known to be 20. Let $\alpha = 0,05$

$n = 10$

$\bar{x} = 27$

$\sigma^2 = 20$ ✓ (variance known!)

$\alpha = 0,05$

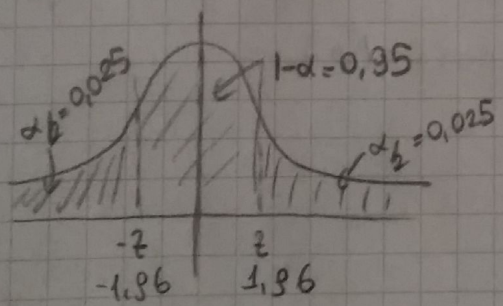
Hypotheses;

$H_0: \mu = 30$
$H_1: \mu \neq 30$

$\Big\}$ two-sided / two-tailed

$$z_0 = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{27 - 30}{\frac{\sqrt{20}}{\sqrt{10}}} = \frac{-3}{\sqrt{2}} = -2,12 \checkmark$$

$\alpha = 0,05$ ama, two-sided



Tablodan $\frac{0,95}{2} = 0,475$ olan değere bakılır!

$z = 1,96$ ✓

$z_0 < -1,96 \Rightarrow$ Reject H_0!

we conclude that μ is not 30!

Confidence Interval

A CI will show that the calculated value of z does not fall within the boundaries of the interval!

$$\bar{X} \pm z \cdot \sigma/\sqrt{n}$$

$$27 \pm 1{,}96 \cdot \sqrt{\frac{20}{10}} = (24{,}228,\ 29{,}772) \checkmark$$

α Same example as one tail test

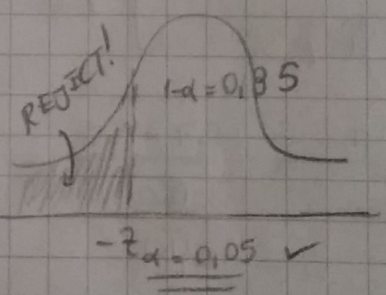
$n=10$ $\sigma^2=20$

$\bar{x}=27$ $\alpha=0{,}05$

Hypotheses ⇒ $H_0: \mu=30$

$H_A: \mu \neq 30$

$$z_0 = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{27-30}{\sqrt{\frac{20}{10}}} = -2{,}12 \checkmark$$



Tablodan 0,95 olan yere bakıyorum!

$-z_{\alpha=0{,}05} = -1{,}65$

$z_0 < z$ ⇒ Reject H_0!

Conclusion ⇒ $\mu < 30$ ✓

MSE (Mean Square Error)

$$\hat{\theta}_1 \sim N(\underbrace{1{,}13\,\theta}_{\text{mean}},\ \underbrace{0{,}02\,\theta^2}_{\text{variance}})$$

$$\hat{\theta}_2 \sim N(1{,}05\,\theta,\ 0{,}07\,\theta^2)$$

$$\hat{\theta}_3 \sim N(1{,}24\,\theta,\ 0{,}005\,\theta^2)$$

which distribution has better quality?

$$MSE(\hat{\theta}) = Var + Bias^2$$

For $\hat{\theta}_1 \Rightarrow \quad 0{,}02\,\theta^2 + (1{,}13\theta - \theta)^2 = 0{,}0369\,\theta^2$

For $\hat{\theta}_2 \Rightarrow \quad 0{,}07\,\theta^2 + (1{,}05\theta - \theta)^2 = 0{,}0725\,\theta^2$

For $\hat{\theta}_3 \Rightarrow \quad 0{,}005\,\theta^2 + (1{,}24\theta - \theta)^2 = 0{,}0626\,\theta^2$

If MSE is smaller, its quality is better!

So, quality order;

$$\hat{\theta}_2 > \hat{\theta}_3 > \hat{\theta}_1$$

⇒ which point estimate would you prefer to estimate θ?