Tutorial 4 - Jainam Jain

Q.1: $\mathrm{Var}(x) = \langle (x - \langle x \rangle)^2 \rangle$

$$= \langle x^2 \rangle - \langle x \rangle^2$$

$$= \langle (\textstyle\sum_i w_i s_i)^2 \rangle - \langle \textstyle\sum_i w_i s_i \rangle^2$$

$$= \langle (\textstyle\sum_i w_i s_i)^2 \rangle - (\textstyle\sum_i w_i \langle s_i \rangle)^2$$

$$= \langle (\textstyle\sum_i w_i s_i)(\textstyle\sum_i w_i s_i) \rangle - (\textstyle\sum_i w_i \langle s_i \rangle)(\textstyle\sum_i w_i \langle s_i \rangle)$$

$$= \langle \textstyle\sum_{i,j} w_i w_j s_i s_j \rangle - \textstyle\sum_{i,j} w_i w_j \langle s_i \rangle \langle s_j \rangle$$

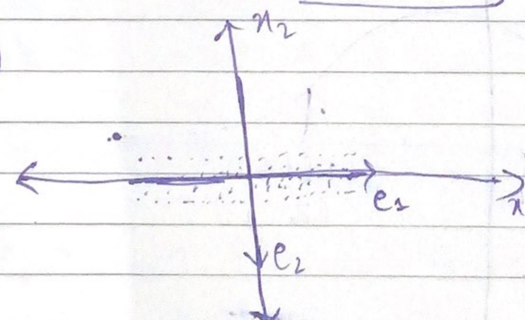
$$= \sum_{i,j} w_i w_j \underbrace{(\langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle)}_{= \mathrm{Var}(s_i) = 1} + \sum_{i,j} w_i w_j \underbrace{(\langle s_i s_j \rangle - \langle s_j \rangle \langle s_i \rangle)}_{0}$$

$$= \sum_i w_i^2$$

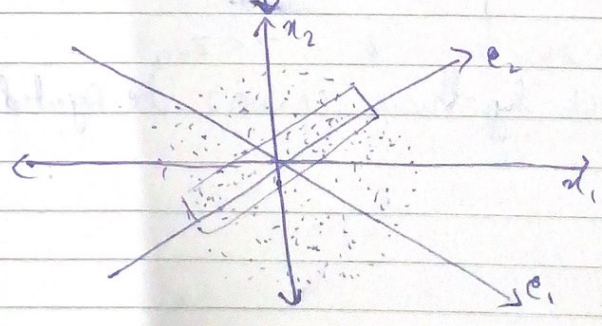
$\therefore$ For ~~that~~ unit variance constraint

$$\boxed{\sum_j w_i^2 = 1}$$

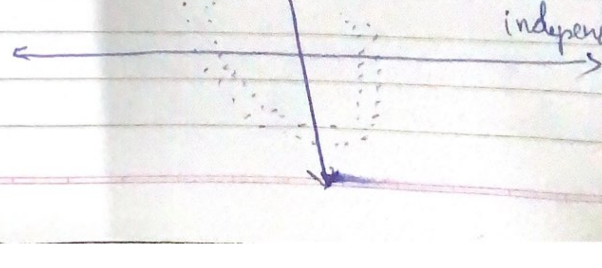
Ex.2: a)



(b)



(c)



$\Rightarrow$ cannot be separated into independent components

Tutorial 5 Jainam Jain

Q.1) Null Hypothesis : $H_0 : p = 0.7$

Alternate Hypothesis : $H_1 : p > 0.7$

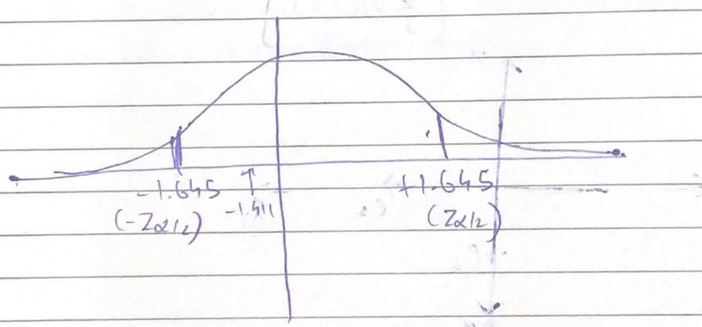
$\alpha = 0.1$, two tailed test

Critical Value $= \pm 1.645 = Z_{\alpha/2}$

Test Statistic $= Z = \dfrac{\hat{p} - p}{\sqrt{pq/n}} \sim N(0,1)$

$\hat{p} = 8/15 = 0.533$

$$Z = \frac{0.533 - 0.7}{\sqrt{0.7 \times 0.3/15}} = -1.411$$



$\because Z > -Z_{\alpha/2}$ & $Z < Z_{\alpha/2}$

the NULL hypothesis CANNOT be Rejected

Q.2) $H_0 : p = 0.6$

$H_1 : p > 0.6$

$\alpha : 0.05$

Critical Value Z_α …

$Z = \dfrac{\hat{p} - p}{\sqrt{pq/n}}$

$\therefore \hat{p} = 70/100$ …

$\therefore Z = \dfrac{0.7 - 0.6 \ldots}{\sqrt{0.6 \times 0.4/ \ldots}}$

$\because Z > Z_{\alpha/2}$

$\therefore$ Reject N…

Q.3) Let the proportio…

Let the proportio…

$\hat{p}_2 = 120/200 = \ldots$

$\hat{p}_p = \dfrac{120 + 240 \ldots}{200 + 500 \ldots}$

$H_0 : p_1 \le p_2$

Given $\alpha = 0.05$ …

$Z_\alpha = Z_{0.05} = \ldots$

$Z = \dfrac{\hat{p}_1 - \hat{p}_2 \ldots}{\sqrt{\hat{p}_p (1 - \hat{p}_p) \ldots}}$

$Z = \dfrac{0.6 - 0.4 \ldots}{\sqrt{0.5143(1 - \ldots}}$

The Rejectio…

$\therefore$ we reject…

Q.2) $H_0 : p = 0.6$
$H_1 : p > 0.6$
$\alpha : 0.05$
Critical Value $Z_{\alpha/2} = \pm 1.96$

$$Z = \frac{\hat{p} - p}{\sqrt{pq/n}}$$

$\therefore \hat{p} = 70/100 = 0.7$

$$\therefore Z = \frac{0.7 - 0.6}{\sqrt{0.6 \times 0.4/100}} = 2.04$$

$\because Z > Z_{\alpha/2}$, it lies in the rejection region
$\therefore$ Reject NULL Hypothesis

Q.3) Let the proportion of Mumbai Voter is p_1
Let the proportion of Surrounding voter is p_2

$\hat{p}_1 = 120/200 = 0.6$ $\hat{p}_2 = 240/500 = 0.48$

$$\hat{p}_p = \frac{120 + 240}{200 + 500} = \underline{0.5143}$$

$H_0 : p_1 \leq p_2$ $H_1 : p_1 > p_2$

Given $\alpha = 0.05$
$Z_\alpha = Z_{0.05} = 1.65$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_p (1 - \hat{p}_p) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$Z = \frac{0.6 - 0.48}{\sqrt{0.5143 (1 - 0.5413) \left(1/200 + 1/500\right)}} = \underline{2.8697}$$

The Rejection Region is $Z > 1.63$
$\therefore$ We reject the NULL Hypothesis

Ex. 4

(a) NULL Hypothesis is

$H_0 : p = 0.2$

Alternate Hypothesis: $H_1 : p > 0.2$

The critical region is in right tail

(b) Null Hypothesis $H_0 : \mu = 3$

Alternate Hypothesis $H_1 : \mu_1 \neq 3$

The critical region is in both tails

(c) Null Hypothesis $H_0 : p = 0.15$

Alternate Hypothesis $H_1 : p < 0.15$

Critical region is in left tail.

(d) Null Hypothesis $H_0 : \mu = 500$

Alternate Hypothesis $H_1 : \mu > 500$

Critical region is in Right tail

(e) $H_0 : \mu = 15$

$H_1 : \mu \neq 15$

The critical region is in both tails.

Q.5) Let μ_1 & μ_2 be population mean for company A and company B

$H_0 : \mu_1 = \mu_2$ $\quad$ $H_1 : \mu_1 \neq \mu_2$

at significance level $\alpha = 0.05$

$$\bar{x}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i = \frac{79.5}{10} = 7.95$$

$$\bar{X}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2i} = \frac{102.6}{10} = \underline{10.26}$$

$$S_1^2 = \frac{1}{n_1 - 1}\left[\sum_{i=1}^{n_1} x_{1i}^2 - n_1 \bar{X}_1^2\right] = \frac{10.865}{9}$$

$$= \underline{1.207}$$

$$S_2^2 = \frac{1}{n_2 - 1}\left[\sum_{i=1}^{n_2} x_{2i}^2 - n_2 \bar{X}_2^2\right] = \frac{2.924}{9} = \underline{0.325}$$

Since sample variances are quite different, we cannot assume that the population variances are equal, so we will use the unpooled t-test

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{S_2^2}{n_2}\right)^2}$$

$$= \frac{\left(\left(\frac{1.207}{10}\right) + \left(\frac{0.325}{10}\right)\right)^2}{\frac{1}{10-1}\left(\frac{1.207}{10}\right)^2 + \frac{1}{10-1}\left(\frac{0.325}{10}\right)^2} = \underline{10.3}$$

Test Statistic.

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$= \frac{7.95 - 10.26}{\sqrt{\frac{1.207}{10} + \frac{0.325}{10}}} = \underline{-5.9}$$

P-value is doubled area under density curve of t-distribution ~~with~~ with 10 degree of freedom, $|t| = |-5.9| = \underline{5.9}$

www.jeevandeep.in
PAGE NO.
DATE: / /

p-value = $2P(T \geq |t|) = 2P(T \geq 5.9)$

$\boxed{p\text{-value} \leq 0.001}$

$\because t_{0.0005}(10) = 4.587$

since $|t| = 5.9$ is greater than $P(T \geq 5.9) < 0.0005$

$\because$ p-value is less than level of significance, we reject NULL hypothesis.

9.1) (a) Since
Pr (

(b) Let M
get
Let
Given
is al
Solv

9.2) (a) By
E

(b) Var (
If

T ≥ 5.4)

587.

is

(T ≥ 5.4)

< 0.0005

nificance, we

Tutorial-6 Jainam Jain

Q.1)(a) Since this is a Bernoulli distribution,

Pr(k tails in the first k tosses, then 1 head)

$= (1-\lambda)^k \lambda$

(b) Let M be the number of the tosses required to get the first head and

let $S = E[M]$

Given that tosses are independent, and expectation is additive: $S = \lambda \times 1 + (1-\lambda) \times (S+1)$

Solving for S gives

$$\cancel{S} = \cancel{\lambda} \times 1 + \cancel{S} + 1 - \lambda S - \cancel{\lambda}$$

$$\therefore \boxed{S = \frac{1}{\lambda}}$$

Q.2)(a) By definition of Variance:

$$E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2]$$
$$= E[X^2] - 2E[XE[X]] + E[X]^2$$
$$= E[X^2] - E[X]^2$$

(b) $Var(X) = E[X^2] - E[X]^2 = 1$

If $Y = a + bX$

$$E[Y^2] = E[(a+bX)^2] = E[a^2 + 2abX + b^2X^2]$$
$$= a^2 + 2abE[X] + b^2E[X^2]$$

$$E[Y] = E[a + bX] = a + bE[X] = a$$

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

Q.3) Let T be event that "Aku predicts that black beauty is a winning horse". Let ~T be event "Aku predicts that" black beauty is not winning horse."

Let W be event that black beauty horse wins.

Let ~W be event that black beauty horse does not wins.

a) Given a horse, probability that it wins is: $P(W) = P(W, T) + P(W, \sim T)$

$= P(W|T)\ P(T) + P(W|\sim T)\ P(\sim T)$

$= 0.99 * 10^{-5} + (1 - 0.99999) \times (1 - 10^{-5})$

$\approx 1.99 \times 10^{-5}$

(b) Probability that Aku correctly predicts the winning horse:

$$P(T|W) = \frac{P(T, W)}{P(W)} = \frac{P(W|T)\ P(T)}{P(W)}$$

$$= \frac{0.99 \times 10^{-5}}{0.99 \times 10^{-5} + (1 - 0.9999) \times (1 - 10^{-5})}$$

$\boxed{P(T|W) \approx 0.497}$

Tutorial

Exercise

Tr

- Travel and
- Foreign pu
- Travel ex
evidence

1) P(fraud =

$\alpha * \sum$ tra

$= \alpha * [P($

$* P(f$

$* P(tra$

$* P(fo$

$* P(t$

$= \alpha [0.01$

$= \alpha [0.00$

=> P(fraud =

$= \alpha * [P(frau$

* P(forei

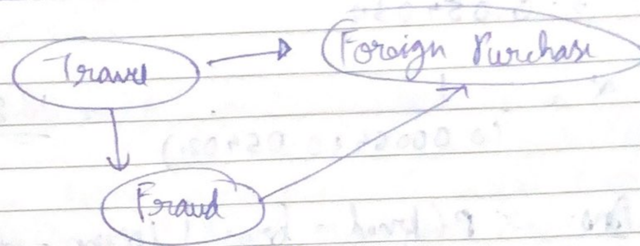
* P(tra

* P(forei

* P(tr

Tutorial 7 Jainam Jain

Exercise



- Travel and fraud both can cause foreign purchase
- Foreign purchase is evidence for fraud.
- Travel explains foreign purchase and hence is evidence against fraud.

1) $P(fraud = true \mid foreign\text{-}purchase = true)$

$$\alpha * \sum_{travel} [P(fraud = true \mid travel) * P(foreign\text{-}purchase = true \mid travel, fraud = true) * P(travel)]$$

$= \alpha * [P(fraud = true \mid travel = true)$
$* P(foreign\text{-}purchase = true \mid travel = true, fraud = true)$
$* P(travel = true) + P(fraud = true \mid travel = false)$
$* P(foreign\text{-}purchase = true \mid travel = false, fraud = true)$
$* P(travel = false)]$

$= \alpha [0.01 \times 0.9 \times 0.05 + 0.002 \times 0.1 \times 0.95]$

$= \alpha [0.00045 + 0.00019] = \underline{0.00064\,\alpha}$

$\Rightarrow P(fraud = false \mid foreign\text{-}purchase = true) =$
$= \alpha * [P(fraud = false \mid travel = true)$
$* P(foreign\text{-}purchase = true \mid travel = true, fraud = false)$
$* P(travel = true) + P(fraud = false \mid travel = false)$
$* P(foreign\text{-}purchase = true \mid travel = false, fraud = false)$
$* P(travel = false)$

$= \alpha [0.99 \times 0.9 \times 0.05 + 0.998 \times 0.01 \times 0.95]$

$= \alpha [0.04455 + 0.009481]$

$= 0.054031$

$$\alpha = \frac{1}{(0.00064 + 0.054031)} = \underline{18.29}$$

Ans: $\therefore P(\text{fraud} = \text{true} \mid \text{foreign-purchase} = \text{true})$

$= \alpha \times 0.00064 = 18.29 \times 0.00064$

$= 0.011$

$= \underline{1.1\%}$

2) $P(\text{fraud} = \text{true} \mid \text{foreign-purchase} = \text{true}, \text{travel} = \text{true}) = \alpha \times 0.00045$

$P(\text{fraud} = \text{false} \mid \text{foreign-purchase} = \text{true}, \text{travel} = \text{true}) = 0.01$

Ans: $\therefore P(\text{fraud} = \text{true} \mid \text{foreign-purchase} = \text{true} \mid \text{travel} = \text{true}) = \underline{1.0\%}$

Tutorials Jainam Jain

Q.1) (1) The first plot is time-series and second plot is sparkline.

(2) 1) There is a weekly cycle in time series
2) High level traffic in first week ~~Feb~~ February.
3) Some days have more than 90 visits meaning they visit site more than once per day or external visitors

Q.2) category axis and value axis.

Q.3) Spikes, upward/downward trend, missing values (gap) in time data.

Q.4) It reduces time/work to interpret the plot by removing non-essential elements.
eg. safety critical situations.

Q.5) → Histograms show distribution while bars used to compare data.

→ Bars can be reordered in bar charts but not in histograms.
→ Histograms plot quantitative data while bar plots plots categorical data.