Nikhil Sawlani

1) $\mu = 77.0$, $\sigma = 3.4$, Find probability.

a) 72.6, less than 72.6.

Soln → $Z = \frac{X - \mu}{\sigma}$ — x - value to be standardized, μ - mean, σ - Std deviation.

→ $Z = \frac{72.6 - 77}{3.4} = -1.294$, Z-value of $(-1.294) =$

→ pnorm(−1.294) or by z table

→ + 0.09783 or 0.0985

b) greather than 88.5

→ $1 - Z(88.5)$

→ $1 - Z\left(\frac{88.5 - 77}{3.4}\right) = \cancel{2.3823}$ $1 - Z(3.38)$

$= 1 - 0.99964$

→ $\cancel{Z(-2.3}$ $= 0.00036$

c) Between 81 & 84.

$\therefore \frac{84-77}{3.4} = 2.05$, $Z(2.05) \rightarrow 0.9798$ — a

& $\frac{81-77}{3.4} = 1.17$, $Z(1.17) \rightarrow 0.8790$ — b

$\therefore a - b = 0.1008$

d) Between 56 & 92.

$\therefore \frac{92-77}{\sigma} = \frac{92-77}{3.4} = 4.4$, $Z(4.4) =$ pnorm(4.4) — a

& $\frac{56-77}{3.4} = -6.17$, $Z(-6.17) =$ pnorm(−6.17) — b

$\therefore a - b = 0.9999946 \approx 1$

2) $\mu = 145$, $sd = 12 = 6$

ⓐ $\rightarrow$ percentage between 130 & 160 data points.

- $z = \frac{X - \mu}{6} = \frac{130 - 145}{12} = \frac{-15}{12} = -1.25$

For negative value, $1 - z(-1.25) =$ 0.10565 — (a)

- $z = \frac{160 - 145}{12} = 1.25$, $z(1.25) =$ 0.8944 — (b)

- $\therefore$, $b - a = 0.8944 - 0.1056 = 0.7888$

ⓑ $\rightarrow$ less than 130 minutes.

$\rightarrow \frac{130 - 145}{12} = \frac{-15}{12} = -1.25 = 0.10565$

$\boxed{3} \rightarrow$ a) For any question answer can be True or false.

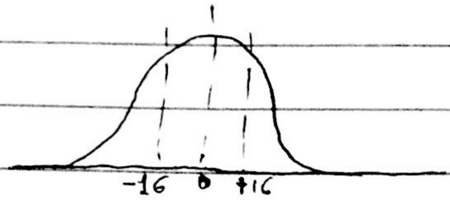
. For question 1, 2, ... 15, probability of true or false i.e. $\frac{1}{2}$ for True or $\frac{1}{2}$ for false.

for 5 questions $\rightarrow \left(\frac{1}{2}\right)^5$

b) $\rightarrow$ for any 5 questions $\rightarrow \left(\frac{1}{2}\right)^5 \; {}^{15}C_5$

$$= \left(\frac{1}{2}\right)^5 \times \frac{15!}{10! \times 5!}$$

4) 68% between 35 & 42



$$Z = \frac{X-\mu}{\sigma}$$

35 42 — one standard deviation.

Here, we have 68% area between 35 & 42 i.e. lowest point & highest point.

$\therefore -1 = \frac{35-\mu}{\sigma}$ & $1 = \frac{42-\mu}{\sigma}$

$-\sigma = 35-\mu$ — (a) & $\sigma = 42-\mu$ — (b)

Now, multiply a & b.

$-\cancel{\sigma} = 35-\mu$ — (a)

$\cancel{\sigma} = 42-\mu$ — (b)

$35-\mu + 42-\mu = 0$

$77 = 2\mu$

$\boxed{\mu = 38.5}$

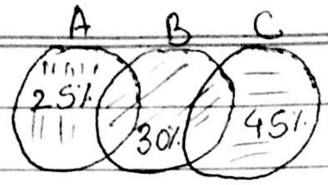
Now, Substituting μ value in z formula — (a)

$-\sigma = 35-38.5$

$\sigma = 3.5$

$\therefore$ Mean = 38.5 & Sd = 3.5

5)



Here, we need to find what percent of A is set of B.

$\therefore \frac{A}{B} = \frac{25}{30} = \frac{5}{6} = 0.833 \approx 83.3\%$

6) $P(T/U) = 82\%$, $P(T \cap D) = 62\%$, $P(D \cap T) = ?$

$P(D \cap T) = \frac{P(T \cap D)}{P(T/U)} = \frac{62\%}{82\%} = 75.60\%$

7) All three heads $\rightarrow$

a) we have only one combination of such case.

HHH $\rightarrow$ $\therefore \frac{1}{8}$, as sample space is $2^3 = 8$ for 3 coins.

b) Exactly one head $\rightarrow$ $3/8$, (HTT, THT, TTH)

c) $P(H/(2\text{ more heads})) = 4/8 = 1/2$

8) $\rightarrow$ $\mu_1 = 1200$, $\mu_2 = 1800$

$Var_1 = 90000$, $Var_2 = 160000$

$X_1 = ?$, $X_2 = ?$, $Sd_1 = 300$, $Sd_2 = 400$

$P(X_1) > P(X_2')$.

Here, to calculate probability of claims filed by 2nd agent is lower, so $\rightarrow$ $X_1 - X_2$

1) $N_1 = 1000$ → given

$\mu_1 = 270$ seconds

$\sigma = 100$ seconds

Variance = 10000

$N_2 = ?$

$\mu_2 = ?$

$\sigma_2 = 10$ seconds

$Variance_2 = 100$

Soln→

$Var = npq$

$10000 = npq$

Here, np~mean so we can replace np by mean

$10000 = 270q$

$\therefore q = 37.03$

& $p = \frac{\text{Mean or np}}{N_1}$

$P = \frac{270}{10000}$

$P = 0.027$

for N_2,

$V_2 = N_2 P_2 q_2$

$100 = N_2 \, (0.027)(37.03)$

$N_2 \approx 100$

∴ difference between 2 random variable

$$\rightarrow \phi\left(\frac{\mu_1 - \mu_2}{(\sqrt{\sigma_1^2 + \sigma_2^2})}\right)$$

we have given in question, Mean1, Mean 2, Sd1 & Sd2 & Variance 1 & var2.

~~Ar~~

$$\therefore \phi\left(\frac{1200 - 1800}{\sqrt{90000 + 160000}}\right) = \frac{-600}{\sqrt{250000}} = \frac{-600}{500} = -1.2$$

Zvalue of -1.2 is → pnorm(-1.2) → 0.1150697

Answer → 11.5%

(9) →

10) $P(S) = 60\%$ S - Sensitivity

$P(S_p) = 99\%$, $P(SP') = 1\%$ (No drug use) Sp - Specificity

$P(D) = 5\%$

let Employee test positive be ε.

$P(\varepsilon/D)$ i.e. Sensitive $= 60\%$

$P(\varepsilon/D')$ = Specificity $= 99\%$ (all employee under test will be under 99%)

find → Employees test positive, that actually use drugs →

$$P(D/\varepsilon) = \frac{P(\varepsilon|D)\ P(D)}{P(\varepsilon)}$$ → whole probability

$$= \frac{\overbrace{P(\varepsilon/D)}^{\text{Sensitive}}\ P(D)}{P(\varepsilon/D)\ P(D) + \underbrace{P(\varepsilon/D')}_{\text{Not Sensitive}}\ P(D')}$$

$$= \frac{60 \times 5}{60\times5 + 1 \times 99} = 0.76$$

11) two bags solution →

y = year from which bag came 1994.

ε → yellow ~~colour~~ color of bag

Find →

$$P(y/\varepsilon) = \frac{P(\varepsilon/y)\,P(\varepsilon)}{P(\varepsilon)}$$

$$= \frac{P(\varepsilon/y)\,P(\varepsilon)}{P(\varepsilon/y)P(\varepsilon) + P(\varepsilon/y')\,P(\varepsilon')} \quad \text{— (a)}$$

So, we have possible outcomes from the years as yellow & green exist in both the years.

If we take yellow & green from 1994 then

$y \rightarrow 20\%$, $G \rightarrow 10\%$.

& for 1996, $y \rightarrow 14\%$, $G \rightarrow 20\%$.

putting value for 1994,

$$= \frac{20 \times 20 \times 1/2}{20 \times 20 \times \cancel{\tfrac{1}{2}} + 14 \times 10 \times \cancel{\tfrac{1}{2}}}$$

— Here, 1/2 is the probability for bag from any 2 ~~bags~~ years.

$$= \frac{400}{400 + 140} = \frac{40\cancel{0}}{54\cancel{0}}$$

= 0.74 is the probability that we have chances of yellow bag from 1994.

12.1 -

```
A<-read.csv("WMT_1.CSV")
A
mean(A$Close)
sd(A$Close)
View(A)
```

12.2 -

```
B<-read.csv("SHLD.CSV")
B
C<-data.frame(B)
class(C)
mean(B$Close)
sd(B$Close)
View(B)
```

12.3 -

The coefficient of variation(CV) talks about dispersion of data in any data sets.

CV function in R is a part of raster library.

Therefore,

```
library(raster)
cv(A$Close)
cv(b$Close)
```

12.4 -

CV helps to identify dispersion in data. Ex as in real life scenarios we can find CV values for any given share from historical data. Lesser the CV higher the reliability.

12.5 - A for Walmart, B for Kmart, C for DJIA

```
cv(c$Close) = 4.6741236
cv(A$Close) = 2.938513
cv(B$Close) = 3.543339
```

CV is minimum for A then B then C, which implies A has minimum dispersion and A has minimum risk, which makes A more attractive to any investor.

```
A<-read.csv("WMT_1.CSV")
A
mean(A$Close)*100
sd(A$Close)*100
```

13.1

```
install.packages('pastecs')
library('pastecs')
```

```
stat.desc(a$orders)
---------------------------------------------
    library(e1071)
library(ggplot2)
a<-read.csv("Prob_Assignment_Dataset.CSV")
summary(a$orders)
skewness(a$orders)
kurtosis(a$orders)
plot(a$site,type="l")
plot(a$visits~a$orders)
boxplot(a$visits~(a$platform))
boxplot(a$orders~(a$platform))
boxplot(a$orders~a$new_customer)
boxplot(a$visits~a$new_customer)
```

Bivariate Analysis

14.1

```
ggplot(a, aes(x = factor(a$visits),a$orders)) +
    geom_point()
```

14.2

```
ggplot(a, aes(x = factor(a$site),a$orders)) +
    geom_point()
```

14.3

```
ggplot(a, aes(x = factor(a$platform),a$orders)) +
    geom_point()+geom_smooth()
```

14.4

```
ggplot(a, aes(x = factor(a$platform),a$visits)) +
    geom_point()
```