Q #1)

a) 1 – 0.9212 = 0.0788
b) 1 – 0.7455 = 0.2545
c) As calculated above for matched printed case, the probability to fail for experts and novices are 0.0788 and 0.2545 respectively. The number of expert and novice participants are equal, then according to the guide notation:
$p(A \cap D) = p(D) \times p(A|D) = 0.5 \times 0.0788 = 0.0394$
$p(A \cap \overline{D}) = p(\overline{D}) \times p(A|\overline{D}) = 0.5 \times 0.2545 = 0.1273$
Therefore; the participant more likely to be a novice.

Q #2) The following table could be constructed. The blue numbers are given data and red ones are calculated:

	U	U'	
+	50	9	59
-	50	891	941
	100	900	1000

a) $p(+|U) = \frac{50}{100} = 0.5$
b) $p(-|\overline{U}) = \frac{891}{900} = 0.99$
c) The same approach gives: $\frac{50}{59} = 0.8475$. Applying Bayesian rule:

$$p(U|+) = \frac{p(+|U)p(U)}{p(+|U)p(U) + p(+|\overline{U})p(\overline{U})} = \frac{0.5 \times 0.1}{0.5 \times 0.1 + 0.01 \times 0.9} = 0.8475$$

Q #3) For the first element, we have n_1 choices to take sample from first set. In the second run we have n_2 choices for set 2. Now, consider we draw the first element from the first set, we have n_2 choice in second run. If we drew the second element of the first set we had n_2 choice for second run and so on. Therefore we have n_1n_2 possible choice in first and second run. With same argument we would have $(n_1n_2)n_3$ possible choices for third run. It could be extended to set k as $(n_1n_2 \ldots n_{k-1})n_k$.

Q #4) For the first draw we have N choices. For the second draw one element is out of the set, so we have N-1 choices. Therefore totally N x (N-1) different choices for two choices are available. For the third draw with same argument [N x (N-1)] x (N-2) choices are available. Therefore, for the kth draw we have N x (N-1) x … x (N-k+1) choices are available. Rewriting in factorial for it gives N! / (N-k)!

Q #5) If we divide a set into k subset; let A be the number of way we can have this subsets with n_1, n_2, …, n_k elements. Each subset also have its own possibilities as $n_i!$. Therefore the total possibility would be $A.n_1!.n_2!. \dots . n_k!$. Otherwise this total possibility is N!. Equalizing these two ways gives us $A = N! / (n_1!.n_2!. \dots . n_k!)$.

Q #6) This is like previous theorem; just divide into two subset with n elements and the rest of the elements from original set. $A = N! / (n_1!.n_2!)$ where $n_1 = n$ an $n_2 = N-n$. Therefore; $A = N! / (n!.(N-n)!)$.

Q #7)

a)

```
y = c(.09, .30, .37, .2, .04)
sum(y)
[1] 1
```

b) p(3U4) = p(3) + p(4) = 0.20 + 0.04 = 0.24

c) p(y<=2) = p(1) + p(2) = 0.09 + 0.30 = 0.39

Q #8)

a) Each probability value in the table is between zero and one. The summation for the column is exactly equal to unit.

b) p(y>=10) = p(10) + p(11) + … + p(20) = 0.02*4 + 0.01*5 + 0.005*2 = 0.14

c)

```
p = c(17, 10, 11, 11, 10, 10, 7, 5, 3, 2, 2, 2, 2, 2, 1, 1, 1, 1,
1, 0.5, 0.5) / 100
y = c(0:20)
names(p) <- as.character(y)
Mu = sum(y * p)
variance = sum( (y - Mu)^2 * p)
Mu
➔ 4.655
variance
➔ 19.85597
```

d) According to Chebshev's rule at least 75% of data laid in the interval of $\mu - 2\sigma$ and $\mu - 2\sigma$

```
SD = sqrt(variance)
Mu - 2 * SD
➔ -4.257009
Mu + 2 * SD
➔ 13.56701
Therefore the interval is [0, 13.57].
```

Q #9)

a)

```
dbinom(10, size = 25, prob = 0.7)
```

```
➔ 0.001324897
```

b)

```
sum(dbinom(1:5, size = 25, prob = 0.7))
➔ 3.457443e-07
It is almost zero.
```

c)

```
y = c(1:25)
py = dbinom(y, size = 25, prob = 0.7)
Mu = sum( y * py)
➔ 17.5
variance = sum( (y - Mu)^2 * py)
SD = sqrt(variance)
➔ 2.291288
```

d) It means that if take 25 samples from this set, the expectation value for the total numbers we got is 17.5 that means 70% of 25 person (or 17.5 person) more likely are foreign PhD. The standard deviation shows us the confidence level of our prediction. For example 75% of data are in the interval of 17.5-2*2.9 and 17.5+2*2.9. It means that with 75% of confidence we can claim that the foreign PhDs are between these amounts.

Q #10) The distribution is multidimensional distribution.

a)

```
dmultinom(rep(5,10), 50, rep(0.1,10))
4.912046e-07
```

b) For this part we can consider it as a binomial distribution which has probability of 0.1 for success and 0.9 for failure:

```
dbinom(0,50,0.1) + dbinom(1,50,0.1)
0.03378586
```

Q #11)

a) It is geometric distribution. For success p = 0.45 + 0.15 = 0.60; Then

$$p(y) = (0.4)(0.6)^{y-1}$$

b) For geometric we have $E(y) = \frac{1}{p} = \frac{1}{0.4} = 2.5$

c) $p(y = 1) = (0.4)(0.6)^{1-1} = 0.4$

d) $p(y > 2) = 1 - p(y \leq 2) = 1 - p(y = 1) - p(y = 2)$

$p(y > 2) = 1 - (0.4)(0.6)^{1-1} - (0.4)(0.6)^{2-1} = 0.36$

Q #12) The distribution is hyper geometric:

a) $E(y) = \frac{nr}{N} = \frac{(10)(8)}{(209)} = 0.382775$

b) $p(y=4)=\frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}=\frac{\binom{8}{4}\binom{201}{6}}{\binom{209}{10}}=0.0001688$

Q #13)

a) $E(y)=V(y)=\lambda=0.03$
b) A;kfal;fk
c) $p(y>3)=1-p(y\leq 3)=1-p(y=1)-p(y=2)-p(y=3)$

$$=1-\frac{0.03^1e^{-0.03}}{1!}-\frac{0.03^2e^{-0.03}}{2!}-\frac{0.03^3e^{-0.03}}{3!}=0.0974456$$

Q #14)

a) $\int_{-\infty}^{+\infty}f(y)dy=1$

$.\int_{-\infty}^{+\infty}c(2-y)dy=\int_0^1 c(2-y)dy=c\left(2y-\frac{1}{2}y^2\right)\Big]_0^1=\frac{3}{2}c=1$

$.\ c=\frac{2}{3}$

b) $F(y)=\int\frac{2}{3}(2-y)dy=\frac{2}{3}\left(2y-\frac{1}{2}y^2\right)$
c) $F(0.4)=\frac{2}{3}\left(2*0.4-\frac{1}{2}0.4^2\right)=0.48$
d) $p(0.1<y<0.6)=F(0.6)-F(0.1)=\frac{2}{3}\left(2*0.6-\frac{1}{2}0.6^2\right)-\frac{2}{3}\left(2*0.1-\frac{1}{2}0.1^2\right)=0.55$

Q #15)

a) $\mu=\int_{-\infty}^{+\infty}f(y)ydy=\int_{-5}^{+5}\frac{3}{500}(25-y^2)ydy=\frac{3}{500}\left(25y^2-\frac{1}{3}y^3\right)\Big]_{-5}^{5}$

$$=\left(\frac{3*25*25}{500}-\frac{125}{500}\right)-\left(\frac{3*25*25}{500}-\frac{-125}{500}\right)=0$$

b) $\sigma^2=\int_{-\infty}^{+\infty}f(y)(y-\mu)^2dy=\int_{-5}^{+5}\frac{3}{500}(25-y^2)y^2dy=\frac{3}{500}\left(\frac{25}{3}y^3-\frac{1}{5}y^5\right)\Big]_{-5}^{5}$

$$=\frac{3}{500}\left(\frac{25}{3}5^3-\frac{1}{5}5^5\right)-\frac{3}{500}\left(\frac{-25}{3}5^3-\frac{-1}{5}5^5\right)=\frac{3*2}{500}\left(\frac{25}{3}5^3-\frac{1}{5}5^5\right)$$

$$=\frac{3*2}{500}\left(\frac{1}{3}-\frac{1}{5}\right)5^5=\frac{3*2*2*5^5}{500*15}=5$$

c) $\mu=\frac{0}{60}=0$ and $\sigma^2=\frac{5}{60^2}=0.0014$
d) $\mu=0\times 60=0$ and $\sigma^2=5\times 60^2=18000$

Q #16)

a) 1 - pnorm(45, mean = 50, sd = 3.2)
```
0.9409149
```
b) pnorm(55, mean = 50, sd = 3.2)

0.9409149

c) pnorm(52, mean = 50, sd = 3.2) - pnorm(51, mean = 50, sd = 3.2)
   0.1113448

Q #17)

a) pnorm(700, mean = 605, sd = 185) - pnorm(500, mean = 605, sd = 185)
   0.4110396
b) pnorm(500, mean = 605, sd = 185) - pnorm(400, mean = 605, sd = 185)
   0.1512568
c) pnorm(850, mean = 605, sd = 185)
   0.9073023
d) qnorm(0.9, mean = 605, sd = 185)

   842.087