

CHALMERS UNIVERSITY OF BEER

---

## VLE-test 2

---

FOR DUMMIES

*Author:*  
G. STARIUS

*Supervisor:*  
Klot JOHAN

13 oktober 2011

# Innehåll

<b>1</b>	<b>s3. Poisson</b>	<b>3</b>
1.0.1	Q1 . . . . .	3
1.0.2	Q2 . . . . .	3
1.0.3	Q3 . . . . .	4
1.0.4	Q4 . . . . .	4
1.0.5	Q6 . . . . .	4
1.0.6	Q7 . . . . .	5
1.0.7	Q8 . . . . .	5
1.0.8	Q9 . . . . .	5
<b>2</b>	<b>s4.1 Point estimation</b>	<b>6</b>
2.0.9	Q1 . . . . .	6
<b>3</b>	<b>s4.2 Confidence intervals</b>	<b>6</b>
3.1	Confidence intervals for mean . . . . .	6
3.1.1	Q1 . . . . .	6
3.1.2	Q4 . . . . .	7
3.1.3	Q6 . . . . .	7
3.1.4	Q9 . . . . .	7
3.1.5	Q11 . . . . .	8
3.1.6	Q12 . . . . .	8
3.2	Confidence intervals for variance . . . . .	8
3.2.1	Q2 . . . . .	8
<b>4</b>	<b>s4.3 Hypothesis testing</b>	<b>9</b>
4.0.2	Q1 . . . . .	9
4.0.3	Q2 . . . . .	9
4.0.4	Q3 . . . . .	10
4.0.5	Q4 . . . . .	10
<b>5</b>	<b>s4.4 Central limit theorem</b>	<b>11</b>
5.0.6	Q1 . . . . .	11
5.0.7	Q3 . . . . .	11
5.0.8	Q4 . . . . .	11
<b>6</b>	<b>s5. Proportions</b>	<b>12</b>
6.1	confidence intervals . . . . .	12
6.2	Sample size . . . . .	12
6.3	approximations . . . . .	12
6.3.1	Q1 . . . . .	13
6.3.2	Q3 . . . . .	13
6.3.3	Q4 . . . . .	14
6.4	Proportions . . . . .	14
6.4.1	Q1 . . . . .	14
6.4.2	Q2 . . . . .	15
6.4.3	Q3 . . . . .	16
6.4.4	Q4 . . . . .	16
6.4.5	Q5 . . . . .	16

6.4.6	Q8 . . . . .	17
6.4.7	Q9 . . . . .	17
<b>7</b>	<b>v6. Linear regression</b>	<b>17</b>

## 1 s3. Poisson

mean = sdev alltid.

### 1.0.1 Q1

$$P(X = k) = (e^{-\lambda} * (\lambda^k)) / k!$$

Poisson kan enbart anta positiva heltal.

X is a Poisson distributed random variable with parameter 2.8 = lambda.

a) What is the probability that X=3?

```
exp(-lambda)
```

```
poisspdf(X, lambda)
```

b) What is the probability that X is at most 1?

```
X=0:1
```

```
sum(poisspdf(X, lambda))
```

c What is the probability that X=-1?

```
poisspdf(X, lambda)
```

d What is the probability that X=1.5?

```
poisspdf(X, lambda)
```

e What is the probability that X<1.5?

```
poisscdf(X, lambda) %notera cdf
```

f What is the expectation of X?

```
lambda
```

g What is the standard deviation of X?

```
sqrt(lambda)
```

### 1.0.2 Q2

Viktigt: oberoende

In a particular hospital it was observed that there is on average 7 cases per month when treating a person in Emergency department requires use of a specialised operation room. You may assume that the incidents occur independently. Since this type of incidents is relatively rare, the number of cases X per day requiring the specialised room can be approximated by a Poisson distributed random variable.

a Since a typical month contains 30.5 days what is the parameter of this Poisson distribution?

```
param = n/k = 7/30.5
```

b What is the probability that there will be no casualty requiring this operation room tomorrow?

```
poisspdf(0,param)
```

c How many days on average in a year (consider a year with 365-days) the room will not be needed? Give the answer rounded to 1 decimal point.

```
poisspdf(0,param) * 365
```

d The hospital has just one such an operation room, so if there is a second casualty in a day, the patient needs to be transferred to another hospital. The local Health Care authorities consider organising a second specialised room in the hospital, but want first to estimate the cost effectiveness of such decision. How likely that there will be at least two cases in a day requiring a specialised room?

```
x = 0:1 1 - sum(poisspdf(x,param))
```

e How many days on average in a 365-day year the second room will be idling? Give the answer rounded to 1 decimal point.

```
x = 0:1 sum(poisspdf(x,param)) * 365
```

### 1.0.3 Q3

$$P(X = k) = (e^{(-\lambda)} * (\lambda^k)) / k!$$

### 1.0.4 Q4

Antag att det i hela världen finns 640 kärnkraftverk. Sannolikheten att det under ett år skall ske minst ett reaktorhaveri i ett slumpmässigt valt kraftverk är en på 57000. Antag att denna sannolikhet inte förändras över tiden och att antalet kraftverk förblir detsamma.

Vad är sannolikheten att det under en 150-års period sker minst ett haveri?

$$\lambda = 1/57000 * 640 * 150$$

```
1 - poisspdf(0,lambda)
```

### 1.0.5 Q6

Tiden det tar för en besiktning hos bilprovningen är exponentielfördelad med genomsnittstiden 21 minuter.

a Vad är sannolikheten för att besikningen tar mindre än 19 minuter?

```
integral i wolframalpha:
```

```
integral((1/m)*(e^(-x/21))) from starttid to sluttid
```

```
integral((1/21)*(e^(-x/21))) from 0 to 19
```

b Vad är sannolikheten för att besikningen tar mer än 16 minuter?

När det står mer än så tar man bara 1 - svaret för mindre än. (Logiskt)

```
integral i wolframalpha:
```

```
(1 - (integral((1/21)*(e^(-x/21))) from 0 to 16))
```

### 1.0.6 Q7

Antag att bilar, som anländer till en vägtull, kan beskrivas med hjälp av en Poissonfördelning med en genomsnittstakt på 2.3 bilar per minut. .

Vad är sannolikheten att det går mellan 25 och 35 sekunder mellan två successiva bils ankomst?

```
integral i wolframalpha:  
integral(2.3*e^(-x*2.3)) x=25/60 to 35/60
```

### 1.0.7 Q8

exponentiell fördelning

Tiden mellan stora jordbävningar (av magnituden 6.5 eller större på Richterskalen) är exponentialfördelad med en konstant parameter  $\lambda$  som skiljer sig från land till land. För Storbritanien är  $\lambda = 0.01$  för exponentialfördelningen.

a Vad är genomsnittstiden mellan stora jordbävningar i Storbritanien?

$$m = \frac{1}{\lambda}$$

b Vad är sannolikheten för att en stor jordbävning inträffar de närmaste 100 åren i Storbritanien?

```
integral i wolframalpha:  
integral(lambda*e^(-x*lambda)) x=0 to 100
```

c Vad är sannolikheten att inga stor jordbävningar inträffar de närmaste 80 åren i Storbritanien?

```
integral i wolframalpha:  
1 - integral(lambda*e^(-x*lambda)) x=0 to 80
```

### 1.0.8 Q9

Antag att  $\xi_1, \xi_2$  är två oberoende Poissonfördelade variabler med väntevärdena 0.6 och 1.6.

Låt  $\eta = \xi_1 + \xi_2$ .

a Beräkna väntevärdet för  $\eta$  .

$$\mu = 0.6 + 1.6$$

b Beräkna standardavvikelsen för  $\eta$ .

$$\sqrt{0.6 + 1.6}$$

c Beräkna sannolikheten att  $\eta$  är större än 2.

```
mu=0.6+1.6  
1-poisscdf(2,mu)
```

## 2 s4.1 Point estimation

### 2.0.9 Q1

Modal betyder det värdet som det finns flest av.

Till exempel 67 av:

66, 67, 65, 65, 66, 70, 66, 65, 69, 68, 67, 67, 68, 67, 69, 65, 68, 67

## 3 s4.2 Confidence intervals

### 3.1 Confidence intervals for mean

$Z_{\frac{\alpha}{2}}$  = vle table  $\rightarrow$

inverse normal both  $\rightarrow$

$p = 1 - \text{graden (typ 90\%)}$ ,

$mean = 0$ ,

$sdev = 1$

#### 3.1.1 Q1

The following random sample is taken from a Normal distribution with standard deviation 0.24 .

1.402, 1.714, 1.063, 1.48, 1.21, 1.505, 1.472

a What is the sample mean?

```
x = [ table ]  
mean(x)
```

b What is the standard deviation of the sample mean of random samples of 7?

```
x = [ table ]  
sqrt((0.24^2)/length(x)) %std dev in question...
```

c Based on the sample, what is the lower 99% confidence limit for the population mean?

$$\hat{x} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

```
x = [ table ]  
mean(x) - Z * stddev/(sqrt(length(x))) %std dev in question...
```

d And what is the upper 99% confidence limit for the population mean?

```
x = [ table ]  
mean(x) + Z * stddev/(sqrt(length(x))) %std dev in question...
```

### 3.1.2 Q4

The sample estimate  $s$  of standard deviation ( $\sigma$ ) is reliable when the sample is big enough. (What is 'enough' depends on the population; but 100 and over is usually enough). In such cases we often use  $s$  for in calculating a confidence interval. A sample of 250 gives a mean of 1.98 and variance  $s^2=0.8$ . Calculate the 85 % confidence interval for the population mean ( $\mu$ ).

Glöm inte roten ur variancen för att få sigma.

$$\mu \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

### 3.1.3 Q6

Antalet timmar som studenter deltidsarbetar är normalfördelat med standardavvikelse 6 timmar. Ett stickprov på 50 studenter togs och stickprovsmedelvärde beräknades till 10.3 timmar.

Hitta nedre gränsen till en 90% konfidensintervall för medeltalet timmar studenter deltidsarbetar

`n = antalet (studenter i detta fall`

`x = stickprovsmedel`

`sigma = standardavvikelse`

$$\text{Gränsvärden} = \hat{x} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

### 3.1.4 Q9

Följande siffror visar temperaturen i grader celsius av reaktorkärnan för en kemisk reaktor, tagna vid olika tillfällen under en månads tid. Temperaturen anses vara normalfördelad.

Katalysatorns effektivitet försämrats snabbt om temperaturen i kärnan överstiger 878 grader celsius. Hitta en 99% konfidensintervall för medeltemperaturen av reaktorkärnan.

870

942

771

862

779

865

895

850

836

777

Vad blir nedre/övre gränsen av intervallet?



```
n = antalet indata i tabellen (10 i detta fall t.ex)
a = [tabellen] %dvs inlagda som vektor...
```

för att få T: table → inverse T → both:  
 p = 1-konfidensintervallsgraden(procent).  
 V = n-1

$$\text{Gränsvärden} = \hat{x} \pm T_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

```
s = std(a) %a = matrix given
x = mean(a)

x = T * (s/sqrt(n))
```

### 3.1.5 Q11

En viss population följer normalfördelningen med variansen 283 ( $\sigma^2$ ) och ett okänt medelvärde. Man vill ta ett stickprov för att få en approximation till medelvärdet. Hur stort stickprov behöver man, så att approximationen inte hamnar längre än 3 från det riktiga medelvärdet:

a med sannolikhet 90%

Glöm inte roten ur variancen för att få sigma.

$$\left( \frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{B} \right)^2$$

$Z_{\frac{\alpha}{2}}$  = vle table →

inverse normal both →

$p = 1 - \text{graden (typ 90\%)}$ ,

$mean = 0$ ,

$sdev = 1$

```
sigma = sqrt(sigma^2) <- sigma^2 finns givet i uppgiften
och kallas variansen
```

```
B = hur stort felet kan bli, 3 i detta fallet t.ex.
```

```
ceil(((Z*sigma)/B)^2)
```

### 3.1.6 Q12

En studie genomfördes där man ville skatta väntevärdet för en normalfördelad variabel med standardavvikelsen 4.7. Efter att ha undersökt 100 enheter så erhöill man konfidensintervallet:

$$87.29 \leq \mu \leq 88.91$$

a Vilken konfidensgrad har intervallet? Ange svaret i heltalsprocent.

```

m = (88.91 - 87.29) / 2
Z = (sqrt(n) * m) / stdavvikelse
table ~ normal both ~ x = z
grad = 1 - X (X = resultatet ur table, notera: inte lilla x utan stora.)

```

b Hur många observationer borde man ha undersökt om man vill ha samma intervalllängd men en 94%-ig konfidensgrad? Avrunda uppåt till närmaste heltal!

```

m = (88.91 - 87.29) / 2

ekvation:
wolframalpha: Z * (stdavvikelse / sqrt(n)) = m

```

## 3.2 Confidence intervals for variance

### 3.2.1 Q2

Kaffemaskin: behöver troligtvis inte repareras.

## 4 s4.3 Hypothesis testing

### 4.0.2 Q1

A herd of cattle is to be tested for a relatively minor disease. If at least one animal in the herd is infected then the herd is declared 'positive' and all animals must be given antibiotics. The test is expensive and therefore it is not possible to test the full herd.

It is known, from previous research, that 11% of animals in the country are infected at any time. How many animals should be tested to ensure that the probability of missing an infection (if it is present at the country-wide level) is less than 0.02?

```

ceil(log(<less than>) / log(1 - p))
ceil(log(0.02) / log(1 - .11))

```

### 4.0.3 Q2

A naturalist is studying the prevalence of beetles in a certain wood. She sets up 12 quadrats (regions of area 1 square metre) in various parts of the wood and counts the number of beetles in each quadrat. The results are as follows:

38 47 38 46 32 30 34 37 49 37 37 39

A five year old study showed that the beetle density in the wood was Normally distributed with a mean of 45 beetles per square metre. Does the new sample indicate any change in the beetle density from its value five years ago?

A test is to be applied with level of significance 1% and null hypothesis  $H_0$ :  $\mu = 45$ . What is the alternative hypothesis?

b Which test should be used?

1 Z-test for the mean  
 2 T-test for the mean  
 3 Z-test for the proportion  
 4 Two-sample Z-test for means  
 5 Two-sample T-test for means  
 6 Two-sample F-test for variances  
 7 IQ-test for the naturalist  
 Enter the number, from 1 through 7, that corresponds to the most satisfactory response.

2, T-test, vi får inget s

c Calculate the corresponding test statistic (Give it to 4 dp).

```
x = [ table ]
sqrt( length(x) ) * (mean(x) - mu)
```

d In relation to this statistic, calculate the p-value for the test (Give your answer to 4 dp if p-value  $\geq 0.1$ ).

```
x = [ table ]
table = T both
t = sqrt( length(x) ) * (mean(x) - mu)
V = n - 1
```

#### 4.0.4 Q3

A dietary expert claims that, on average, UK men are more than 7 'kg' overweight. To test her assertion a random sample of 22 men is assembled and the excess of their actual weight over the 'ideal' weight based on their height and age is recorded (in 'kg') :

8.5	8.5	8.6	8.6	8.9	7.4	7.6	8.8	7.4	7.4	7.6
8.2	8.7	6.1	7.2	8.1	5.8	6.9	7.3	7.2	6.9	6.1

On the assumption that these excess weights are Normally distributed, test whether the dietician's claim is supported by the evidence.

a: A test is to be applied with level of significance 10% and null hypothesis  $H_0 : = 7$ . What is the alternative hypothesis? svaret står i frågan...

b: Calculate the standardised t-statistic needed for the test

$t = ((x_{\text{tak}} - \mu_0) * \sqrt{n}) / s$   $x_{\text{tak}} = \text{mean}(\text{table})$

c: table - t high: t värde från förra  $V = n - 1$

d: är c större än significance

e: pvalue < significance level : Reject  $H_0$

pvalue < significance level : Så stämmer förändringen, resultatet är signifikant

#### 4.0.5 Q4

Till ett slumpmässigt urval av 41 unga vuxna män (mellan 20-30 år gamla) ställdes frågades om hur många minuter sport de tittade dagligen på TV. Medelvärde för urvalet blev 62 minuter. Man visste sedan tidigare att standar-

davvikelsen för sporttittandet var 25 minuter. Testa på 2% signifikansnivå ifall tiden unga män tittar på sport i TV är större än 53 minuter.

a Noll hypotesen är  $H_0 : = 53$  . Vad är den alternativa hypotesen?

b:

$$z = \frac{\sqrt{n}(\hat{x} - \mu_0)}{\sigma}$$

c: table → normal {står i frågan, större? mindre?}:

mean = 0

x = z

sdev = 1

p = X

d och e: Är resultatet signifikant på 2% nivå? (Med andra ord, finns det belägg på att, i snitt, tittar unga män mer än 53 minuter på sport i TV dagligen?) och en tabell

pvalue <significance level : Reject  $H_0$

pvalue <significance level : Så stämmer förändringen, resultatet är signifikant

## 5 s4.4 Central limit theorem

### 5.0.6 Q1

Previous research has estimated that 20% of passengers travelling from the USA to Sweden carry more than their duty-free allowance. A plane carrying 109 passengers arrives at Landvetter airport and customs officers randomly select 13 passengers and search them.

What is the probability that 0 passengers in the sample are carrying more than their limit (to 5 decimal places)?

```
k = random_selected
binopdf(0,k,p)
```

What is the probability that 20 passengers on the plane are carrying more than their limit (to 5 decimal places)?

```
x = number_of_testsubjects
binopdf(x,n,p)
```

What is the probability that the customs officers do not find any passengers in the sample of 13 carrying more than their limit (assuming the officers are very experienced and will always find any undeclared goods)? Answer to 5 decimal places.

```
k = random_selected
binopdf(0,k,p)
```

### 5.0.7 Q3

Beloppet som en slumpmässigt vald kund spenderar per besök på ett nöjesfält är i genomsnitt 324 kronor med en standardavvikelse på 99 kronor.

Hur stor är sannolikheten att en slumpmässigt vald grupp på 65 individer spenderar fler än 20900 kronor på ett besök?

```
table ~ normal high
sdev = sqrt((stdavvikelse^2)/n) = sqrt((99^2)/65) i detta fall
mean = 324 i detta fall
x = 20900/n i detta fall
```

### 5.0.8 Q4

Antag att diametern hos tillverkade muttrar har en standardavvikelse på 0.4 mm. För utvärdering av tillverkningsprocessen genomförs en kvalitetskontroll varvid man tar ut 40 muttrar och mäter deras diameter.

Hur stor är risken att den genomsnittliga diametern i urvalet avviker från diametersnittets väntevärde med mer än 0.1 mm?

$$\mu = 0.4 \quad n = 40 \quad \sigma = d - \mu = 0.1$$

table ~ normal both

$$x = \text{väntevärdet} \quad \text{mean} = 0 \quad sdev = \sigma / \text{roten}(n)$$

## 6 s5. Proportions

$$Z_{\frac{\alpha}{2}} = \text{vle table} \rightarrow$$

inverse normal both  $\rightarrow$

$$p = 1 - \text{graden (typ 90\%)},$$

$$\text{mean} = 0,$$

$$\text{sdev} = 1$$

### 6.1 confidence intervals

```
n = stickprovet
```

```
k = antalet utvalda/svarar ja/ whatever
```

$$\hat{p} = k/n$$

```
Z_alpha/2 = vle table -> inverse normal both ->
```

$$p = 1 - \text{graden (typ 90\%)}, \text{ mean} = 0, \text{ sdev} = 1$$

$$CI = \hat{p} \pm Z_{\alpha/2} * \sqrt{\hat{p} * ((1 - \hat{p})/n)}$$

+ = upper limit

- = lower limit

## 6.2 Sample size

Ett opinionsinstitut behöver uppskatta stödet för ett visst politisk parti och vill därför göra en stickprovsundersökning i en viss valkrets.

I den senaste mätningen fick partiet stöd av en andel på 0.38 i valkretsen.

I den nya undersökningen vill institutet uppnå en konfidsgrad av 99 % samt att resultatet skall vara minst 0.005 nära den korrekta andelen.

Vad behöver den minsta storleken på undersökningen bli för att uppnå detta (förutsatt att resultatet inte kommer att avvika nämnvärt från den tidigare undersökningen). Använd z-värden upp till minst 5 decimaler i dina beräkningar. B = hur nära det ska komma (kan va typ 0.01)

$$B = Z_{\frac{\alpha}{2}} \cdot \frac{\sqrt{\hat{p} \cdot (1 - \hat{p})}}{\sqrt{n}}$$
$$n = \left( Z_{\frac{\alpha}{2}} \cdot \frac{\sqrt{\hat{p} \cdot (1 - \hat{p})}}{B} \right)^2$$

## 6.3 approximations

vle table → normal low

p = sannolikheten för något.. typ 0.24

n = antal av något

y = sannolikhet att något blir något annat, typ x mindre än: 0.21

hitta  $p(X < x)$

$$mean = n \cdot p$$

$$sdev = \sqrt{n \cdot p \cdot (1 - p)}$$

$$x = y \cdot n$$

X = outcomen

### 6.3.1 Q1

The probability of success on any trial of a binomial experiment is 32% .

Use the Normal approximation to the binomial to find the probability that the proportion of successes in a sample of 300 is less than 29 %.

(Do not make use of a "continuity correction" in this question.)

vle table → normal low

p = sannolikheten för något.. typ 0.24

n = antal av något

y = sannolikhet att något blir något annat, typ x mindre än: 0.21

hitta  $p(X < x)$

$$mean = n \cdot p$$

$$sdev = \sqrt{n \cdot p \cdot (1 - p)}$$

$$x = y \cdot n$$

X = outcomen

### 6.3.2 Q3

In the last election 68% of the electorate voted for party P.

a) On the assumption that the electorate have not changed their allegiances what is the probability that in a random sample of 50 electors, 32 (64%) or fewer indicate that they will vote for party P. (Use the binomial distribution).

$x = 0:32$

$\text{sum}(\text{binopdf}(x, 50, 0.68)) \leftarrow$  eftersom det står "or fewer" i frågan!

b) Use the Normal approximation to the binomial distribution to estimate the probability that 32 people or fewer in a random sample of 50 indicate that they will vote for party P. Use a "continuity correction". Correction Table:

IF	use
$P(X=n)$	$P(n - 0.5 < X < n + 0.5)$
$P(X > n)$	$P(X > n + 0.5)$
$P(X \leq n)$	$P(X < n + 0.5)$
$P(X < n)$	$P(X < n - 0.5)$
$P(X \geq n)$	$P(X > n - 0.5)$

källa: <http://www.statisticshowto.com/articles/what-is-the-continuity-correction-factor/>

ingen binomialsak när man har continuity-värden...

$$\text{mean} = n \cdot p(50 \cdot 0.68)$$

$$\text{sdev} = \text{sqrt}(n * p * (1 - p))$$

$$x = 32 \pm \text{correction enligt tabell}$$

c) In a large sample the Normal distribution is a good representation of the probability distribution of the proportion (supporting party P). If the sample size were 200 electors, we would use the Normal distribution with mean 0.68 and what standard deviation?

$$\text{sdev} = \text{sqrt}(n * p * (1 - p)) / n$$

d) On this basis find the probability that 64% or less of the large sample intend to vote for party P. Do NOT make use of a "continuity correction".

vle table  $\rightarrow$  normal low:

$$x = .64$$

$$\text{mean} = p * n = .68 * 200$$

$$\text{sdev} = \text{sqrt}(n * p * (1 - p))$$

$$\text{returner} aX$$

### 6.3.3 Q4

Given a binomial random variable  $X$  with  $n=32$  and  $p=0.5$ , find the (exact) probabilities of the following events.

```
P(X = k).  
binopdf(k,n,p)
```

```
P(X <= k).  
binocdf(k,n,p)
```

Find the Normal approximation to

```
P(X=k). Use the "continuity correction".  
use table for "continuity correction"  
table - normal low:  
mean = n * p  
sdev = sqrt(n * p * (1 - p))
```

## 6.4 Proportions

### 6.4.1 Q1

In the last election 39% of the electorate voted for party P.

a On the assumption that the electorate have not changed their allegiances what is the probability that in a random sample of 50 electors, 18 (36%) or fewer indicate that they will vote for party P. (Use the binomial distribution).

```
k = 0:18  
sum(binopdf(k,n,P))
```

b Use the Normal approximation to the binomial distribution to estimate the probability that 18 people or fewer in a random sample of 50 indicate that they will vote for party P. Use a "continuity correction". Correction Table:

IF	use
$P(X=n)$	$P(n - 0.5 < X < n + 0.5)$
$P(X > n)$	$P(X > n + 0.5)$
$P(X \leq n)$	$P(X < n + 0.5)$
$P(X < n)$	$P(X < n - 0.5)$
$P(X \geq n)$	$P(X > n - 0.5)$

källa: <http://www.statisticshowto.com/articles/what-is-the-continuity-correction-factor/>

*Table - normal*

$mean = n * p$

$sdev = \sqrt{n * p * (1 - p)}$

$x = k \pm correction(+0.5idettafall)$

c In a large sample the Normal distribution is a good representation of the probability distribution of the proportion (supporting party P). If the sample



size were 300 electors, we would use the Normal distribution with mean 0.39 and what standard deviation?

```
mean = p
sdev = sqrt(p*(1-p)/n)
```

d On this basis find the probability that 36% or less of the large sample intend to vote for party P. Do NOT make use of a "continuity correction".  
vle table → normal low:

$$\begin{aligned}x &= .64 * n \\ \text{mean} &= p * n = .68 * 200 \\ \text{sdev} &= \text{sqrt}(n * p * (1 - p))\end{aligned}$$

*returnera X*

#### 6.4.2 Q2

På löpande bandet i en fabrik tillverkar man komponenter med 1.4% chans att en komponent inte klarar kvalitetskontrollen och därför kasseras.

a Ett stickprov på 500 tages för testning. Om felprocenten är fortfarande 1.4%, vad är sannolikheten att alla 500 komponenter klarar kvalitetskontrollen?

```
binopdf(0,n,p)
```

b Vad är sannolikheten att två eller fler komponenter fallerar kvalitetskontrollen?

```
k = 2:n
sum(binopdf(k,n,p))
```

#### 6.4.3 Q3

En stor servicefirma har 295 räkningar som förfaller till betalning denna månad. Av erfarenhet vet man att 9% av alla fakturerade kunder inte betalar förrän efter förfallodagen.

Hur stor är sannolikheten att fler än 9 kunder av de 295, som fått fakturor, betalar efter förfallodagen?

```
x = 0:k
1 - sum(binopdf(x,n,p))
```

#### 6.4.4 Q4

På löpande bandet i en fabrik tillverkar man 250 komponenter per arbetsdag. Antag att idag är 19 av dessa defekta.

a) Ett stickprov på 14 tages för testning. Vad är sannolikheten att alla 14 komponenter klarar kvalitetskontrollen?

$$p = \frac{k}{n} = \frac{19}{250}$$

```
binopdf(0,14,p)
```

b) Vad är sannolikheten att minst 2 men högst 6 komponenter fastnar i kvalitetskontrollen?

$$p = \frac{k}{n} = \frac{19}{250}$$

```
x = 2:6  
sum(binopdf(x,14,p))
```

#### 6.4.5 Q5

På löpande bandet i en fabrik tillverkar man 2500 komponenter per arbetsdag. Antag att idag är 37 av dessa defekta.

a Ett stickprov på 60 tages för testning. Vad är sannolikheten att alla 60 komponenter klarar kvalitetskontrollen?

```
p=37/2500 i detta fall.  
lambda = n*p  
poisspdf(0,lambda)
```

b Vad är sannolikheten att två eller fler komponenter fallerar kvalitetskontrollen?

```
p=37/2500 i detta fall.  
lambda = n*p  
1-poisscdf(1,lambda) % 2 - 1 = 1... daah
```

#### 6.4.6 Q8

Man hade 140 gäster på en bröllopsmiddag och av dessa blev 84 sjuka av matförgiftning.

Av de som blev sjuka åt 67 kyckling.

Av de som inte blev sjuka åt 11 kyckling.

**Bygg träd**

#### 6.4.7 Q9

väntevärde är  $n * p$  variansen är  $np(1-p)$  standardavvikelsen är roten ur variansen

Antag att  $\xi_1$  och  $\xi_2$  är två oberoende binominalfördelade variabler med  $p_1 = p_2 = 0.3$  samt  $n_1 = 6$  och  $n_2 = 4$ .

Låt  $\eta = \xi_1 + \xi_2$ .

a Beräkna väntevärdet för  $\eta$ .

```
n1*p1 + n2*p2
```

b Beräkna standardavvikelsen för  $\eta$ .

```
sqrt(n1*p1*(1-p1)+n2*p2*(1-p2))
```

c Beräkna sannolikheten att är mindre än 4.

```
x = 0:3 (inkludera inte 4, star mindre an...)  
sum(binopdf(x,n1+n2,p))
```

## 7 v6. Linear regression

Given the values of bivariate random variables X and Y.

x	y
3	1.21
3.4	1.12
3.5	0.65
3.7	1.35
3.8	0.9
...	...

Find the equation of the least squares straight line  $y = b_0 + b_1x$  that fits this data.

```
y = b1*x + b0  
  
y = [y-kolumnen]  
x = [x-kolumnen]  
  
regstats(y,x)  
  
choose t.stat[x]  
  
choose last created tstat in workspace  
  
b1 = beta min  
b0 = beta max  
  
SSE = sum((y - b0 - b1*x).^2)  
  
SSE = SY - SSR  
  
SST = SY = sum((y - mean(y)).^2) %variation in y
```

Gradient of regression line that fits the data:

```
y = b1*x + b0  
  
y = [y-kolumnen]  
x = [x-kolumnen]  
  
regstats(y,x)  
  
choose t.stat[x]
```

```
choose last created tstat in workspace
```

```
b1 = beta min %the gradient
```

```
b0 = beta max
```

```
t-statistic = t min
```

```
annat s'a'tt att r'a'kna ut t:
```

```
r * sqrt((n-2)/(1-r^2))
```

Räkna ut  $r \rightarrow$  inte 1.0000:

```
corrcoef(y,x)
```