Statistik und Qualität - Ausarbeitung der zweiten Übung

Stefan Dünser

#### 1) Simuliere die Folge von 100 Münzwürfen mit dem Computer. Wir gehen dabei von einer „fairen“ Münze aus, d.h. von einer Münze, die nach einem Wurf zu je 50 % „Kopf“ oder „Zahl“ zeigt. Erzeuge dazu ein Feld (einen Vektor) mit 100 Komponenten, in den mit je 50 % Wahrscheinlichkeit „0“ („Kopf“) oder „1“ („Zahl“) eingetragen werden. Betrachte die entstehenden Muster. Wie oft treten 3 oder mehrere gleiche Ergebnisse in Serie auf? Wie (statistisch) bedeutsam sind diese Häufungen von Ereignissen?

N <- 100  
Muenzwurf.vec <- floor(runif(N) \* 2)  
Muenzwurf.vec

## [1] 1 1 1 0 0 1 1 0 1 0 1 0 0 1 0 1 0 1 0 0 0 1 0 1 1 1 1 1 0 0 1 0 0 0 1 0 0  
## [38] 1 1 1 1 1 0 0 0 0 0 0 1 1 1 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 0 1 1 1 0 1 1  
## [75] 1 1 0 0 1 1 0 0 1 1 0 0 0 0 1 0 1 0 0 0 0 1 0 1 0 0

# Speichern des zufälligen Ausgangsvektors für die Feststellung der Nuller-Serien  
Muenzwurf0.vec <- Muenzwurf.vec

n <- 12

Im folgenden wird angenommen, dass eine Serie aus maximal 12 Elementen bestehen kann.

Serie.df <- data.frame(Seriengröße = c(1:n), Anzahl\_Zahl = rep(0,n), Anzahl\_Kopf = rep(0,n))  
N1 <- sum(Muenzwurf.vec)  
Muenzwurf1.vec <- Muenzwurf.vec \* c(0,Muenzwurf.vec[-length(Muenzwurf.vec)])  
N2 <- sum(Muenzwurf1.vec + c(Muenzwurf1.vec[-1],0) > 0)  
Serie.df$Anzahl\_Zahl[1] <- N1-N2  
  
i <- 2  
while (i <= n) {  
 Muenzwurf.vec <- Muenzwurf1.vec  
 N1 <- sum(Muenzwurf.vec)  
 Muenzwurf1.vec <- Muenzwurf.vec \* c(0,Muenzwurf.vec[-length(Muenzwurf.vec)])  
 N2 <- sum(Muenzwurf1.vec + c(Muenzwurf1.vec[-1],0) > 0)  
 Serie.df$Anzahl\_Zahl[i] <- N1-N2  
 i <- i + 1  
}  
  
Serie.df

## Seriengröße Anzahl\_Zahl Anzahl\_Kopf  
## 1 1 15 0  
## 2 2 3 0  
## 3 3 3 0  
## 4 4 1 0  
## 5 5 2 0  
## 6 6 0 0  
## 7 7 0 0  
## 8 8 0 0  
## 9 9 0 0  
## 10 10 0 0  
## 11 11 0 0  
## 12 12 0 0

Muenzwurf.vec <- (Muenzwurf0.vec == 0) \* 1  
  
N1 <- sum(Muenzwurf.vec)  
Muenzwurf1.vec <- Muenzwurf.vec \* c(0,Muenzwurf.vec[-length(Muenzwurf.vec)])  
N2 <- sum(Muenzwurf1.vec + c(Muenzwurf1.vec[-1],0) > 0)  
Serie.df$Anzahl\_Kopf[1] <- N1-N2  
  
i <- 2  
while (i <= n) {  
 Muenzwurf.vec <- Muenzwurf1.vec  
 N1 <- sum(Muenzwurf.vec)  
 Muenzwurf1.vec <- Muenzwurf.vec \* c(0,Muenzwurf.vec[-length(Muenzwurf.vec)])  
 N2 <- sum(Muenzwurf1.vec + c(Muenzwurf1.vec[-1],0) > 0)  
 Serie.df$Anzahl\_Kopf[i] <- N1-N2  
 i <- i + 1  
}  
  
Serie.df

## Seriengröße Anzahl\_Zahl Anzahl\_Kopf  
## 1 1 15 9  
## 2 2 3 8  
## 3 3 3 3  
## 4 4 1 2  
## 5 5 2 0  
## 6 6 0 1  
## 7 7 0 0  
## 8 8 0 1  
## 9 9 0 0  
## 10 10 0 0  
## 11 11 0 0  
## 12 12 0 0

i <- 3  
Zahl <- 0  
Kopf <- 0  
SerieGroesserGleich3\_1 <- 0  
SerieGroesserGleich3\_0 <- 0  
  
while (i <= n) {  
 Zahl <- Serie.df$Anzahl\_Zahl[i]  
 Kopf <- Serie.df$Anzahl\_Kopf[i]  
 SerieGroesserGleich3\_1 <- SerieGroesserGleich3\_1 + Zahl  
 SerieGroesserGleich3\_0 <- SerieGroesserGleich3\_0 + Kopf  
 i <- i + 1  
}  
SerieGroesserGleich3\_1

## [1] 6

SerieGroesserGleich3\_0

## [1] 7

SerieGroesserGleich3 <- SerieGroesserGleich3\_0 + SerieGroesserGleich3\_1  
SerieGroesserGleich3

## [1] 13

Die Auswertung der Zufallsdaten hat ergeben, dass es bei den zufälligen Münzwürfen mit identischer Wahrscheinlichkeit insgesamt 13 Serien gibt, die 3 oder mehr Elemente enthält. Dabei entfallen 6 Serien, bei denen der Münzwurf das Ergenbins “Zahl” und 7 bei denen der Münzwurf das Ergebnis “Kopf” ist.  
  
**Statistische Bedeutsamkeit**  
Die Wahrscheinlichkeit, dass beim Münzwurf mit einer fairen Münze dasselbe Ergebnis 3 Mal hintereinander auftritt, beträgt P(“Zahl” und “Zahl” und “Zahl”) = P(1 \* 1 \* 1) = 0,5 \* 0,5 \* 0,5 = (1/2)^3 = 1/8. Bei 4 gleichen Ereignissen hintereinander wäre die Wahrscheinlichkeit nur noch 6,25 %. Kommt es zu einer Häufung von solchen Serien mit drei oder mehr gleichen Münzwürfen hintereinander, gehen Teilnehmer diese Münzwurfspiels davon aus, dass etwas nicht stimmt und dass es sich vermutlich nicht um eine faire Münze handelt, bei des sowohl “Zahl” als auch “Kopf” gleich wahrscheinlich vorkommen.

#### 2) Simuliere die Folge von 100 Würfen eines Würfels mit dem Computer. Wir gehen dabei von einem „fairen“ Würfel aus, bei dem die Zahl „6“ mit einer Wahrscheinlichkeit von 1/6 gewürfelt wird.

**Erzeuge dazu ein Feld (einen Vektor) mit 100 Komponenten, in den mit 1/6 Wahrscheinlichkeit „1“ („Sechs“) oder mit Wahrscheinlichkeit 5/6 „0“ („nicht Sechs“) eingetragen werden. Betrachte die entstehenden Muster. Wie oft treten 3 oder mehrere Sechser in Serie auf?**

X.vec <- 1:100  
Wuerfeln.vec <- floor(runif(length(X.vec),1,7))  
Wuerfeln.vec

## [1] 6 3 3 3 2 2 3 3 5 2 6 1 4 6 5 1 1 2 1 2 2 2 5 1 4 1 4 2 1 1 4 1 1 1 4 5 3  
## [38] 1 6 6 2 3 3 6 2 6 2 4 6 3 5 1 6 6 5 5 6 5 3 3 6 4 5 5 3 5 2 4 3 6 1 1 5 6  
## [75] 5 3 3 3 4 6 2 5 4 4 3 2 6 4 4 6 2 5 5 3 1 6 5 1 2 4

Wuerfelwahrscheinlichkeit.vec <- (Wuerfeln.vec == 6) \* 1  
Wuerfelwahrscheinlichkeit.vec

## [1] 1 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## [38] 0 1 1 0 0 0 1 0 1 0 0 1 0 0 0 1 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 1  
## [75] 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 1 0 0 0 0

Wuerfelwahrscheinlichkeit0.vec <- Wuerfelwahrscheinlichkeit.vec

n <- 5

Es wird angenommen, dass nicht öfter als 5 Mal hintereinander die Zahl 6 gewürfelt wird.

Wuerfelserie.df <- data.frame(Seriengröße = c(1:n), Anzahl\_6 = rep(0,n))  
N1 <- sum(Wuerfelwahrscheinlichkeit.vec)  
Wuerfelwahrscheinlichkeit1.vec <- Wuerfelwahrscheinlichkeit.vec \* c(0,Wuerfelwahrscheinlichkeit.vec[-length(Wuerfelwahrscheinlichkeit.vec)])  
N2 <- sum(Wuerfelwahrscheinlichkeit1.vec + c(Wuerfelwahrscheinlichkeit1.vec[-1],0) > 0)  
Wuerfelserie.df$Anzahl\_6[1] <- N1-N2  
  
  
i <- 2  
while (i <= n) {  
 Wuerfelwahrscheinlichkeit.vec <- Wuerfelwahrscheinlichkeit1.vec  
 N1 <- sum(Wuerfelwahrscheinlichkeit.vec)  
 Wuerfelwahrscheinlichkeit1.vec <- Wuerfelwahrscheinlichkeit.vec \* c(0,Wuerfelwahrscheinlichkeit.vec[-length(Wuerfelwahrscheinlichkeit.vec)])  
 N2 <- sum(Wuerfelwahrscheinlichkeit1.vec + c(Wuerfelwahrscheinlichkeit1.vec[-1],0) > 0)  
 Wuerfelserie.df$Anzahl\_6[i] <- N1-N2  
 i <- i + 1  
}  
Wuerfelserie.df

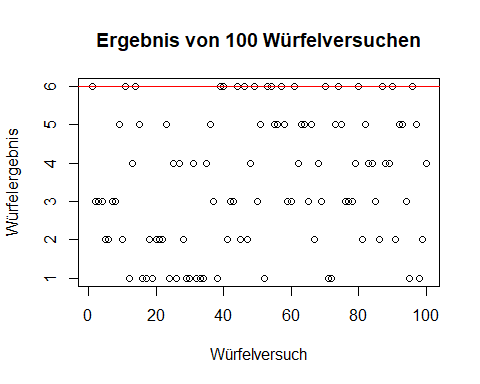
## Seriengröße Anzahl\_6  
## 1 1 14  
## 2 2 2  
## 3 3 0  
## 4 4 0  
## 5 5 0

Sechserserie <- Wuerfelserie.df$Anzahl\_6[3] + Wuerfelserie.df$Anzahl\_6[4] + Wuerfelserie.df$Anzahl\_6[5]  
Sechserserie

## [1] 0

Beim zufälligen Würfelversuch mit einem fairen Würfel tritt eine Serie mit 3 oder mehr 6en hintereinander 0 Mal auf.

plot(X.vec, Wuerfeln.vec, ylim=c(1,6), xlab = "Würfelversuch", ylab = "Würfelergebnis", main = "Ergebnis von 100 Würfelversuchen")  
abline(h=6, col="red")



#### 3) Erzeuge eine Menge von N auf dem Intervall [-1, +1] gleichverteilten Zufallszahlen mit dem Computer. Verwende diese Zufallszahlen um daraus eine Menge M von N/2 Punkten in der Ebene zu erzeugen. Sie liegen alle in einem Quadrat Q zwischen -1 <= x <= +1 und -1 <= y <= +1. Nähere mit diesen Zufallszahlen die Zahl pi mit einer Monte-Carlo-Simulation. Dazu werden jene Zufallszahlen auf Q, die innerhalb eines Kreises mit dem Radius 1 liegen (in Q eingeschriebener Kreis) durch die Gesamtanzahl |M| der Punkte auf Q dividiert (und damit das Verhältnis der Kreisfläche zur Quadratfläche geschätzt). Wie entwickelt sich die Schätzung für pi in Abhängigkeit von N?

library("plotrix")

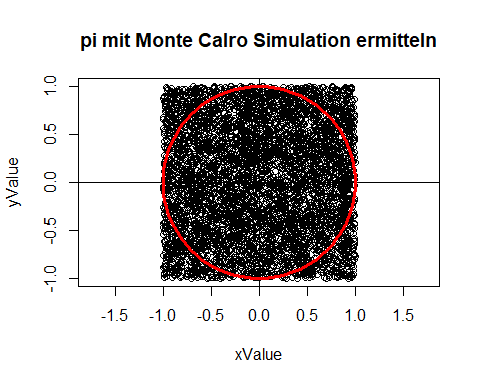
N <- 10  
Sprung <- 100  
MaxN <- 10000  
Pi.vec = rep(0,Sprung)

while (N < MaxN) {  
 Pi\_Random.vec <- runif(N,-1,1)  
 M.df <- as.vector(data.frame(xValue = sample(Pi\_Random.vec, N/2, replace = TRUE), yValue = sample(Pi\_Random.vec, N/2, replace = TRUE)))  
   
 n <- 0  
 j <- 1  
 while (j <= N/2) {  
 D <- sqrt(M.df$xValue[j]^2 + M.df$yValue[j]^2)  
 if(D <= 1){  
 n <- n + 1  
 }  
 j <- j + 1  
 }  
 N <- N + Sprung  
 Pi.vec[N/Sprung] <- (n/N) \* 8  
}

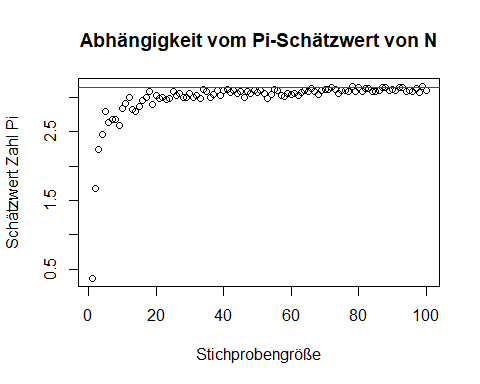
summary(Pi.vec)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.3636 3.0002 3.0726 2.9874 3.1033 3.1645

plot(M.df, main = "pi mit Monte Calro Simulation ermitteln", xlim = c(-1,1), ylim = c(-1,1), asp = 1)  
abline(h = 0)  
abline(v = 0)  
draw.circle(0, 0, 1, border = "red", lwd = 3)

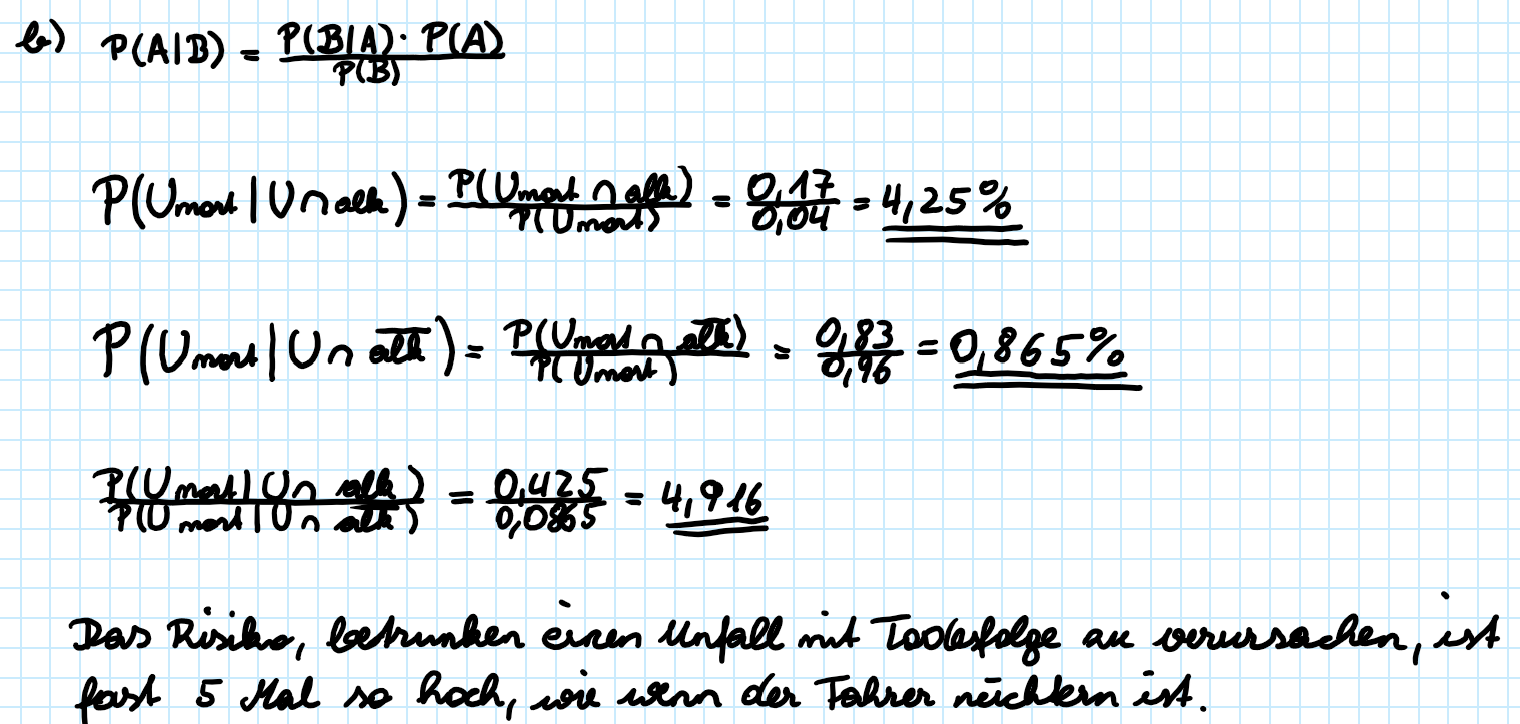


plot(Pi.vec, ylab = "Schätzwert Zahl Pi", main = "Abhängigkeit vom Pi-Schätzwert von N", xlab = "Stichprobengröße")  
abline(h = 3.14159, col = "red")

 Mit steigender Menge an Zufallszahlen, steigt auch die Wahrscheinlichkeit, dass sich mehr Zufallszahlen innerhalb des Kreises befinden relativ zu den Zufallszahlen außerhalb des Kreises im Quadrat. Die Annäherung an Pi ist daher bei steigendem N größer.

#### 4) Aus Alkoholkontrollen nach Verkehrsunfällen schätzt man, dass rund vier Prozent der verunfallten Lenker einen unzulässig hohen Alkoholpegel hatten. Gleichzeitig wurden 17 % der Unfälle mit tödlichen Folgen von alkoholisierten Lenkern verursacht. Alkoholisierte Fahrer sind somit nich die sichereren Autofahrer.

**a) Daraus ergibt sich, dass bei 83 % der Unfälle mit tödlichen Folgen kein Alkohol mit im Spiel war. Sind also alkoholisierte Autofahrer doch die sichereren Fahrer? (Beantworten Sie diese Frage am besten nach Klärung von Punkt b)**  
Nein. Das Risiko, einen Unfall mit Todesfolge in alkoholisiertem Zustand zu verursachen ist fas 5 Mal so hoch wie in nicht alkoholisiertem Zustan. Die Berechnung dazu ist bei b) zu finden.

**b) Sei Umort das Ereignis „Unfall mit tödlichen Folgen“ und P(Umort | U Ç alk) die Wahrscheinlichkeit (das Risiko) für tödliche Folgen bei einem Unfall unter Alkoholeinfluss. Benennen wir gleichermaßen mit P(Umort | U Ç ¬alk) die Wahrscheinlichkeit (das Risiko) in nüchternem Zustand einen Unfall mit tödlichen Folgen zu verursachen. Wie groß ist das Verhältnis der beiden Risiken, d.h. das Verhältnis von P(Umort | U Ç alk) zu P(Umort | U Ç ¬alk) ?** 

#### 5) Die Auswertung von gemeldeten Autounfällen hat für 10.000 beteiligte Insassen folgendes Ergebnis ergeben:  90 der 190 dabei verstorbenen Insassen waren nicht angegurtet. Den größten Teil machen jene 8.910 Personen aus, die angegurtet waren und den Unfall überlebt haben.

**a) Erstelle aus den Angaben eine Kontingenztabelle und vervollständige sie.**

Autounfall.tab <- as.table(rbind(c(100,90,190), c(8910,900,9810), c(9010,990,10000)))  
dimnames(Autounfall.tab) <- list(Gurt = c("angeschnallt", "nicht angeschnallt", "Gesamt"), Leben = c("gestorben", "überlebt", "Gesamt"))  
Autounfall.tab

## Leben  
## Gurt gestorben überlebt Gesamt  
## angeschnallt 100 90 190  
## nicht angeschnallt 8910 900 9810  
## Gesamt 9010 990 10000

Während in der oberen Tabelle die absoluten Zahlen der Autounfälle mit der Untersuchung zwischen angeschnallt und nicht angeschnallt sowie überleben und sterben dargestellt sind, kann man der unteren Tabelle die relativen Werte dieser Untersuchung sehen.

Kontingtab.tab <- as.table(rbind(c(1,0.9,1.9), c(89.1,9,98.1), c(90.1,9.9,100)))  
dimnames(Kontingtab.tab) <- list(Gurt = c("angeschnallt", "nicht angeschnallt", "Gesamt"), Leben = c("gestorben", "überlebt", "Gesamt"))  
Kontingtab.tab

## Leben  
## Gurt gestorben überlebt Gesamt  
## angeschnallt 1.0 0.9 1.9  
## nicht angeschnallt 89.1 9.0 98.1  
## Gesamt 90.1 9.9 100.0

**b) Welcher Anteil der verunfallten Personen war insgesamt angegurtet?**

Verunf\_Angeg <- 8910+(190-90)  
Verunf\_Angeg

## [1] 9010

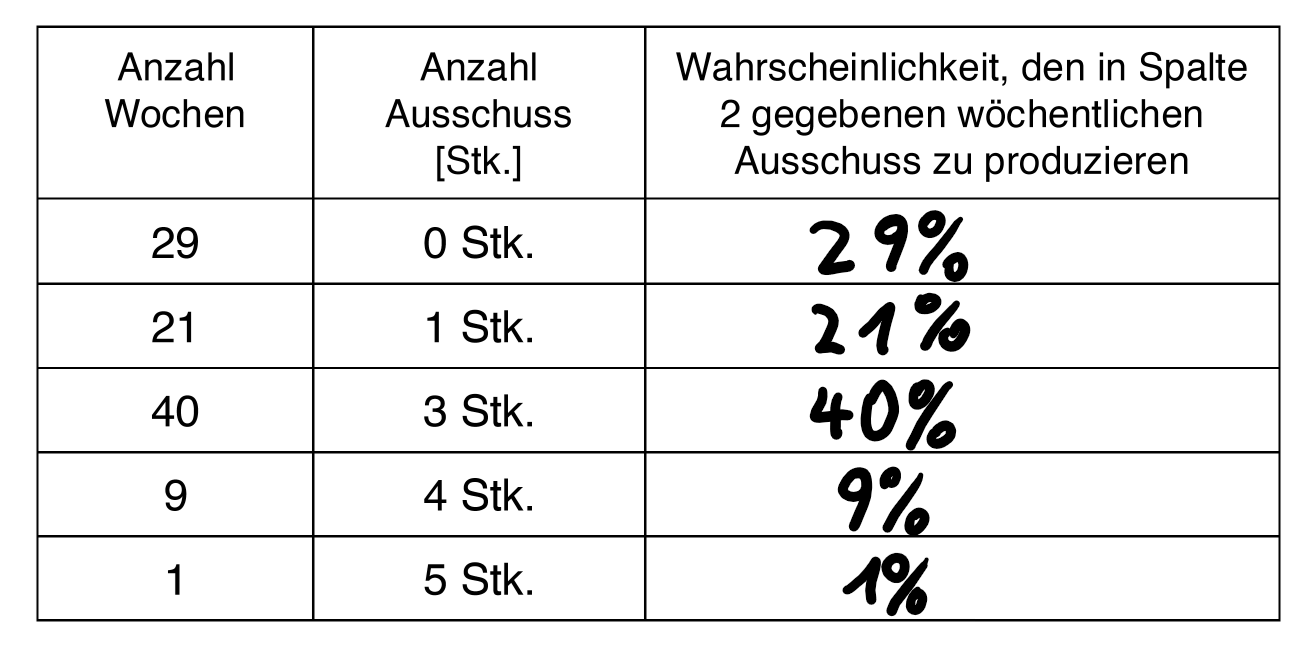
Anteil\_Verunf\_Angeg <- (Verunf\_Angeg/10000)\*100  
Anteil\_Verunf\_Angeg

## [1] 90.1

90.1 % der verunfallten Personen waren angegurtet.  
**c) Sind das Überleben mit und ohne Gurt unabhängige Ereignisse?**  
Da die Wahrscheinlichkeit, mit Gurt zu überleben unterschiedlich zu der Wahrscheinlichkeit, mit Gurt tödlich zu Verunglücken, sind die beiden Ergeinisse abhängig voneinander.  
Ist man angegurtet, überlebt man einen Unfall oder nicht. Zum Beispiel sind 100 Personen bei einem Unfall angegurtet. 98 Personen überleben den Unfall. Um auf 100 Personen zu kommen, müssen folglich zwei Personen sterben. Die Ereignisse sind voneinander abhängig, da die Überlebenden und die Toten gemeinsam wieder die Ausgangsmenge ergeben müssen.  
**d) Um wieviel höher ist das Risiko tödlich zu verunglücken für Insassen ohne Gurt im Vergleich zu angegurteten Insassen?**

angeschnallt <- (190-90) / Verunf\_Angeg  
nichtAngeschnallt <- 90 / (10000-Verunf\_Angeg)  
Risiko <- nichtAngeschnallt/angeschnallt  
Risiko

## [1] 8.190909

Das Risiko, bei einem Unfall zu sterben, wenn man nicht angeschnallt ist, ist 8.1909091-fach höher.  
  
#### 6) In einer Produktion wurde über eine längere Periode mittels 100%-Endkontrolle der pro Woche entdeckte Ausschuss festgehalten.  Berechnung des Erwartungswerts:

mu <- 0\*0.29 + 1\*0.21 + 3\*0.4 + 4\*0.09 + 5\*0.01  
mu

## [1] 1.82

Der Erwartungswert von 1.82 beschreibt die Wahrscheinlichkeit für den Auswurf pro Woche.