

Übung 1

Gegeben sind folgende Werte!

x_i^*	3	4	5	6	7	8	9	10
h_i	10	15	30	30	25	20	15	5

Aufgabe:

- Bestimme
- den empirischen Mittelwert von x (\bar{x})
 - den empirischen Median von x (\tilde{x})
 - die empirische Varianz von $S_n^2(x)$ und
 - den Wert $F(9)$!

Schritt 1 - Erweiterung der Tabelle!

Damit uns das Rechnen einfacher fällt, können wir bereits hier an dieser Stelle die Tabelle erweitern. Die Tabelle wird um die relativen Häufigkeiten (f_i) und aufsummierten relativen Häufigkeiten (F_i) erweitert. Wir rechnen f_i aus, in dem wir h_i durch n (Stichprobenumfang) teilen. Den Stichprobenumfang erhalten wir, in dem wir die Zeile h_i aufsummieren.

x_i^*	3	4	5	6	7	8	9	10	
h_i	10	15	30	30	25	20	15	5	n = 150
f_i	0,066	0,1	0,2	0,2	0,166	0,133	0,1	0,023	$\frac{h_i}{n}$
F_i	0,066	0,166	0,366	0,566	0,732	0,865	0,965	1	
$x_i^* \cdot h_i$	30	45	150	180	150	160	135	50	940

Für die Berechnung des Mittelwerts können wir hier schon Vorarbeit leisten, in dem wir $x_i^* \cdot h_i$ rechnen.

Lösung zu Aufgabe a)

Nun können wir den empirischen Mittelwert mit Hilfe der absoluten Häufigkeiten ermitteln. Wir wenden dafür folgende Formel an:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^s h_i \cdot x_i^*$$

Das sieht wild aus, ist es aber gar nicht. Die Summe aus $x_i^* \cdot h_i$ haben wir nämlich schon in der Tabelle berechnet. Wir können die Formel also mit Zahlen füllen und schon haben wir das Ergebnis:

$$\bar{x} = \frac{1}{150} \cdot 940 = 6,26\bar{6}$$

Lösung zu Aufgabe b)

Hier müssen wir zu erst schauen, ob der Stichprobenumfang n gerade oder ungerade ist. Je nach dem müssen wir eine andere Formel für die Berechnung des Medians nehmen. Unser Stichprobenumfang beträgt n=150. Ist also eine gerade Zahl. Wir könnten den Median jetzt mit der Formel

$\tilde{x} = \left[x\left(\frac{n}{2}\right) + x\left(\frac{n}{2} + 1\right) \right] * \frac{1}{2}$ berechnen. Einfacher geht's, wenn wir den Wert einfach in der Tabelle

ablesen. Dies ist nur bei einem geraden Stichprobenumfang möglich. Wir suchen uns der Zeile Fi die Spalte aus, wo der Wert erstmals die 0,5 überschreitet. In unserem Fall ist das die Spalte $x_i^* = 6$. Damit haben wir auch schon die Lösung - $\tilde{x} = 6$.

Lösung zu Aufgabe c)

Zur Berechnung der Varianz benötigen wir wieder eine Formel.

$$Sn^2(x) = \frac{1}{n} \sum_{i=1}^n hi * x_i^{*2} - \bar{x}^2$$

Achtung: hier Aufpassen, dass $-\bar{x}^2$ gehört nicht mit ins Summenzeichen. Dieser Wert wird zum Schluss einmal abgezogen. Die Rechnung sieht dann so aus:

$$Sn^2(x) = \frac{1}{150} (10 * 3^2 + 15 * 4^2 + 30 * 5^2 + 30 * 6^2 + 25 * 7^2 + 20 * 8^2 + 15 * 9^2 + 5 * 10^2) - 6,266^2$$

$$Sn^2(x) = \frac{1}{150} (6380) - 39,271$$

$$Sn^2(x) = 42,533 - 39,271$$

$$Sn^2(x) = 3,262$$

Lösung zu Aufgabe d)

Hier suchen wir den Wert F(9). Diesen finden wir wieder in der Tabelle. Wir schauen einfach in der Spalte x_i^* . Hier muß der Wert 9 auftauchen. Jetzt lesen wir den Wert in der Zeile Fi ab. F(9)= 0,965

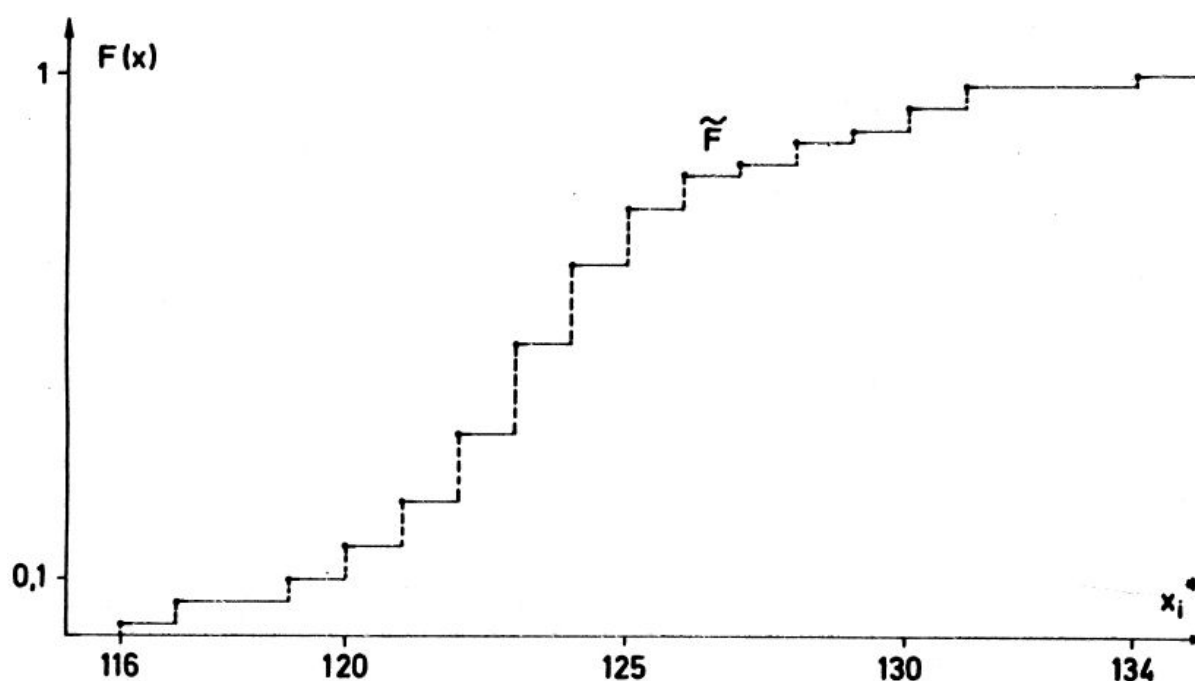
Übung 2

Empirische Verteilungsfunktion – Vorwärtsrechnung

Bei der sog. Vorwärtsrechnung ist eine Tabelle in der nachfolgenden Form gegeben:

Körpergröße x_i^*	absolute Häufigkeit h_i	absolute Summenhäufigkeit H_i	relative Häufigkeit f_i	relative Summenhäufigkeit F_i
116	1	1	0,02	0,02
117	2	3	0,04	0,06
118	0	3	0	0,06
119	2	5	0,04	0,10
120	3	8	0,06	0,16
121	4	12	0,08	0,24
122	6	18	0,12	0,36
123	8	26	0,16	0,52
124	7	33	0,14	0,66
125	5	38	0,10	0,76
126	3	41	0,06	0,82
127	1	42	0,02	0,84
128	2	44	0,04	0,88
129	1	45	0,02	0,90
130	2	47	0,04	0,94
131	2	49	0,04	0,98
132	0	49	0	0,98
133	0	49	0	0,98
134	1	50	0,02	1,00
	$n = 50$		1,00	1,00

Aus dieser Tabelle werden die x_i^* und die passenden F_i -Werte abgelesen und in ein Koordinatensystem eingetragen.



Bei der Rückwärtsrechnung ist das Koordinatensystem gegeben und es müssen die Werte abgelesen und in die Tabelle eingetragen werden.

Übung 3

Ein Filialunternehmen will den Zusammenhang zwischen Jahresumsatz und Ladenverkaufsfläche untersuchen. In einem bestimmten Jahr liefern die $n = 12$ Filialen folgende Daten:

i	x_i	y_i	x_i^2	y_i^2	$x_i * y_i$
1	0,31	2,93			
2	0,98	5,27			
3	1,21	6,85			
4	1,29	7,01			
5	1,12	7,02			
6	1,49	8,35			
7	0,78	4,33			
8	0,94	5,77			
9	1,29	7,68			
10	0,48	3,16			
11	0,24	1,52			
12	0,55	3,15			
Σ					

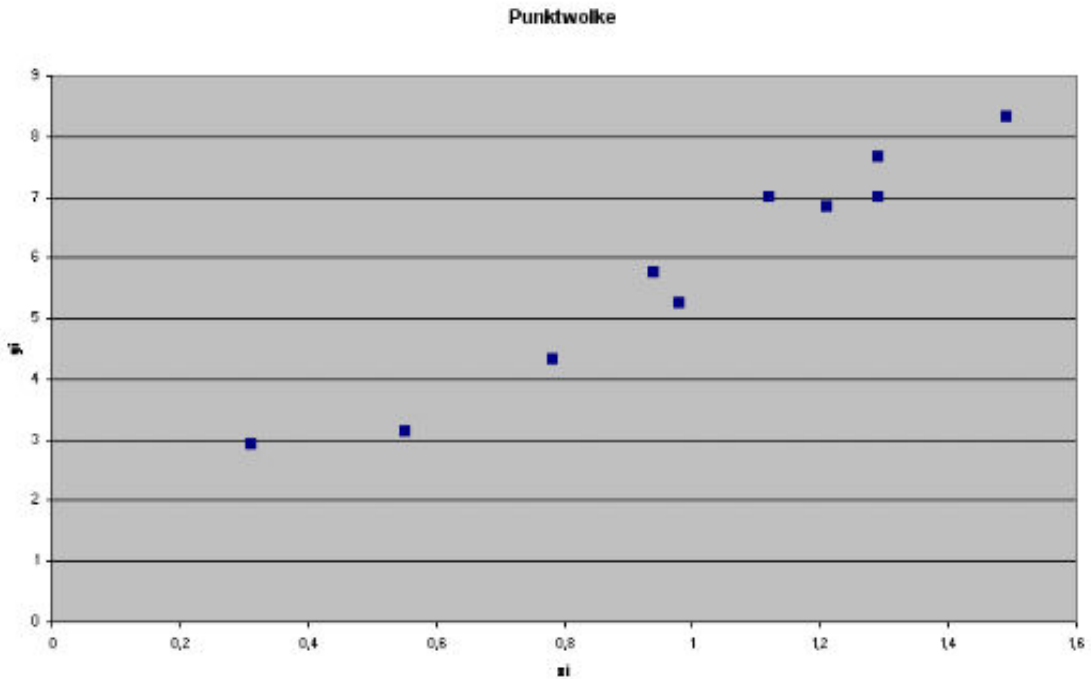
Hierbei bezeichnet i fortlaufend die Nummern der Filialen, x_i die Verkaufsfläche in Tsd. qm und y_i die Jahresumsätze in Mio. DM.

Aufgabe:

- Man trage die Beobachtungspunkte $x_i; y_i$ in ein xy-Koordinatensystem ein und entscheide, ob annähernd linear oder quadratischer Zusammenhang besteht.
- Ausgehend von a) berechne man eine geeignete Kleinste-Quadrate-Regressionsfunktion
- zeichne die in b) ermittelte Regressionsfunktion.
- Man schätze den Jahresumsatz einer Filiale mit 3000 qm Verkaufsfläche

Lösung Aufgabe a)

Das ist wirklich ganz einfach. Wir brauchen nur ein Koordinatensystem zeichnen und dann die Werte aus der Tabelle eintragen. Das ganze nennt sich dann Punktwolke. Ein i zusammen mit den x_i und y_i -Werten ergibt immer einen Punkt. z.B. $i_1(2,93/0,30)$



Wenn man diese Punkte eingetragen hat, kann man auch ganz leicht erkennen, ob ein linearer oder ein quadratischer Zusammenhang besteht. Wir sehen hier einen linearen Zusammenhang.

Wir können das Ganze jetzt aber auch rechnerisch lösen. Das geht mit Hilfe des Korrelationskoeffizienten. Dafür gibt es wieder eine Formel.

$$r_{xy} = \frac{(\sum x_i * y_i) - (n * \bar{x} * \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 * \sum (y_i - \bar{y})^2}}$$

Die sieht jetzt auf dem ersten Blick vielleicht etwas verwirrend aus. Aber keine Sorge, so schlimm ist das gar nicht.

Als erstes schauen wir uns noch mal die Tabelle an und rechnen die „leeren“ Spalten aus.

i	x_i	y_i	x_i^2	y_i^2	$x_i * y_i$
1	0,31	2,93	0,0961	8,5849	0,9083
2	0,98	5,27	0,9604	27,7729	5,1646
3	1,21	6,85	1,4641	46,9225	8,2885
4	1,29	7,01	1,6641	49,1401	9,0429
5	1,12	7,02	1,2544	49,2804	7,8624
6	1,49	8,35	2,2201	69,7225	12,4415
7	0,78	4,33	0,6084	18,7489	3,3774
8	0,94	5,77	0,8836	33,2929	5,4238
9	1,29	7,68	1,6641	58,9824	9,9072
10	0,48	3,16	0,2304	9,9856	1,5168
11	0,24	1,52	0,0576	2,3104	0,3648
12	0,55	3,15	0,3025	9,9225	1,7325
\sum	10,68	63,04	11,4058	384,666	66,0307

Im zweiten Schritt müssen wir jetzt die Mittelwerte von x und y ausrechnen.

Mittelwert von x:

$$\bar{x} = \frac{1}{n} \sum x_i$$

$$\bar{x} = \frac{1}{12} * 10,68 \quad \text{Anmerkung: den Wert 10,68 können wir aus der Tabelle ablesen.}$$

$$\bar{x} = 0,89$$

$$\bar{y} = \frac{1}{n} \sum y_i$$

$$\bar{y} = \frac{1}{12} * 63,04 \quad \text{Anmerkung: den Wert 63,04 können wir aus der Tabelle ablesen.}$$

$$\bar{y} = 5,25\bar{3}$$

Jetzt haben wir alle Werte, um r_{xy} auszurechnen.

$$r_{xy} = \frac{(\sum x_i * y_i) - (n * \bar{x} * \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 * \sum (y_i - \bar{y})^2}}$$

$$r_{xy} = \frac{(66,0307) - (12 * 0,89 * 5,25\bar{3})}{\sqrt{(11,4058 - 12 * 0,89^2) * (384,666 - 12 * 5,25\bar{3}^2)}}$$

$$r_{xy} = \frac{(66,0307) - (56,10559)}{\sqrt{(1,9006) * (53,4958)}}$$

$$r_{xy} = \frac{9,9251}{10,08335844}$$

$$r_{xy} = 0,9843$$

Jetzt wissen wir, dass der Korrelationskoeffizient 0,9843 ist. Wenn der Wert zwischen -1 und +1 liegt, besteht ein linearer Zusammenhang. Also haben wir hier einen linearen Zusammenhang.

Lösung Aufgabe b)

Gesucht ist eine Funktion nach diesem Muster $\hat{y} = a + b * x$

$$\text{„b“ berechnen wir mit der Formel: } b = \frac{\sum x_i * y_i - n * \bar{x} * \bar{y}}{\sum x_i^2 - n * \bar{x}^2}$$

Die benötigten Werte können wir aus der Tabelle entnehmen oder haben diese bei der Berechnung des Korrelationskoeffizienten bereits berechnet.

$$b = \frac{66,0307 - 12 * 0,89 * 5,25\bar{3}}{11,4058 - 12 * 0,89^2}$$

$$b = \frac{9,9251}{1,9006}$$

$$b = 5,222$$

Jetzt müssen wir noch a ausrechnen. Dafür nehmen wir die Funktion $a = \bar{y} - b * \bar{x}$

Auch hier finden wir die Werte in unseren vorherigen Rechnungen:

$$a = 5,25\bar{3} - 5,222 * 0,89$$

$$a = 0,605753$$

Jetzt nur noch die errechneten Werte in die Funktion einsetzen.

$$\hat{y} = 0,605753 + 5,222 * x$$

Lösung Aufgabe c)

Damit wir die Regressionsfunktion zeichnen können, müssen wir mit der Funktion

$\hat{y} = 0,605753 + 5,222 * x$ zwei Werte für \hat{y} berechnen. Dafür setzen wir einfach zwei beliebige Werte für x ein ($x=0$; und $x=1$). Wir erhalten dann zwei Punkte (P1 (0,6/0) und P2 (5,8/1)). Diese Werte einfach in Das Koordinatensystem eintragen und mit einem Lineal verbinden.

Lösung Aufgabe d)

Um den Jahresumsatz einer Filiale mit 3.000qm Verkaufsfläche zu schätzen nutzen wir auch wieder die Funktion $\hat{y} = 0,605753 + 5,222 * x$. Hier brauchen wir jetzt für das x die qm-Zahl einsetzen. Aber Vorsicht. In der Tabelle sind die qm in tausend angegeben. D.h. wir müssen hier nur 3 für das x einsetzen.

$$\hat{y} = 0,605753 + 5,222 * 3$$

$$\hat{y} = 16,271$$

Antwortsatz: Der geschätzte Umsatz einer Filiale mit 3.000qm Verkaufsfläche beträgt 16,271 Mio. DM.

Übung 4

Ein Umschlagbetrieb benutzt bei einem Dauerauftrag für eine Teilstrecke täglich einen bestimmten Zug der Bundesbahn und anschließend ein Schiff einer Binnenreederei. Der zuständige Abteilungsleiter des Umschlagbetriebes weiß aus Erfahrung, dass in 50% aller Fälle der Zug verspätet ist, in 30% aller Fälle der Zug und das Schiff verspätet, in 10% aller Fälle das Schiff verspätet ist, unter der Bedingung, dass der Zug nicht verspätet ist.

Aufgabe:

Wie groß ist die Wahrscheinlichkeit, dass der Zug verspätet ist, unter der Bedingung, dass das Schiff nicht verspätet ist?

1. Schritt: Ereignisdefinition

Z= Zug verspätet

S= Schiff verspätet

2. Schritt. gegebene Werte in mathematische Ausdrücke umwandeln

„...dass in 50% aller Fälle der Zug verspätet ist...“

$$W(Z) = 0,5$$

„...in 30% aller Fälle der Zug und das Schiff verspätet...“

$$W(Z \cap S) = 0,3$$

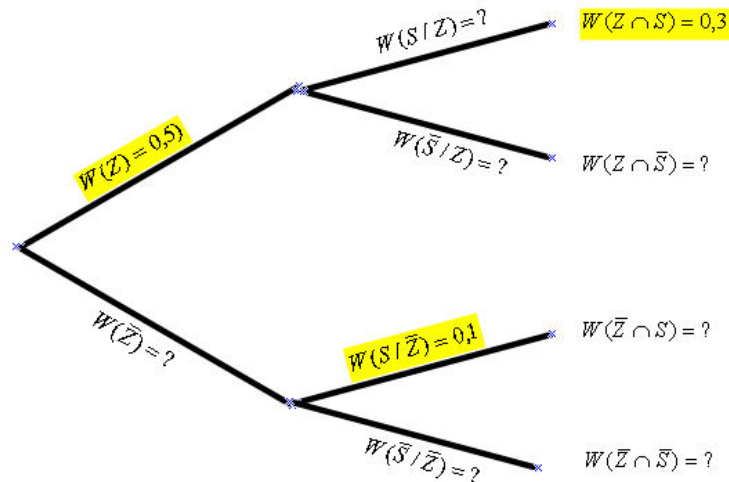
„...in 10% aller Fälle das Schiff verspätet ist, unter der Bedingung, dass der Zug nicht verspätet ist...“

$$W(S \cap \bar{Z}) = 0,1$$

3. Schritt: gesucht?

$$W(\bar{S} \cap Z) = ?$$

4. Schritt: gegebene Werte in Entscheidungsbaum eintragen



Die übrigen Werte können wir jetzt mit Logik berechnen.

Wir wissen, dass in 50% aller Fälle der Zug verspätet ist. ($W(Z) = 0,5$). Dann wissen wir also auch, dass in 50% aller Fälle der Zug **nicht** verspätet ist – also $W(\bar{Z}) = 0,5$.

Wir wissen, dass $W(Z) = 0,5 \cdot W(S|Z) = ? = W(Z \cap S) = 0,3$ ist.

Also gilt auch: $W(S|Z) = ? = W(Z \cap S) = 0,3 / W(Z) = 0,5 = W(S|Z) = 0,6$

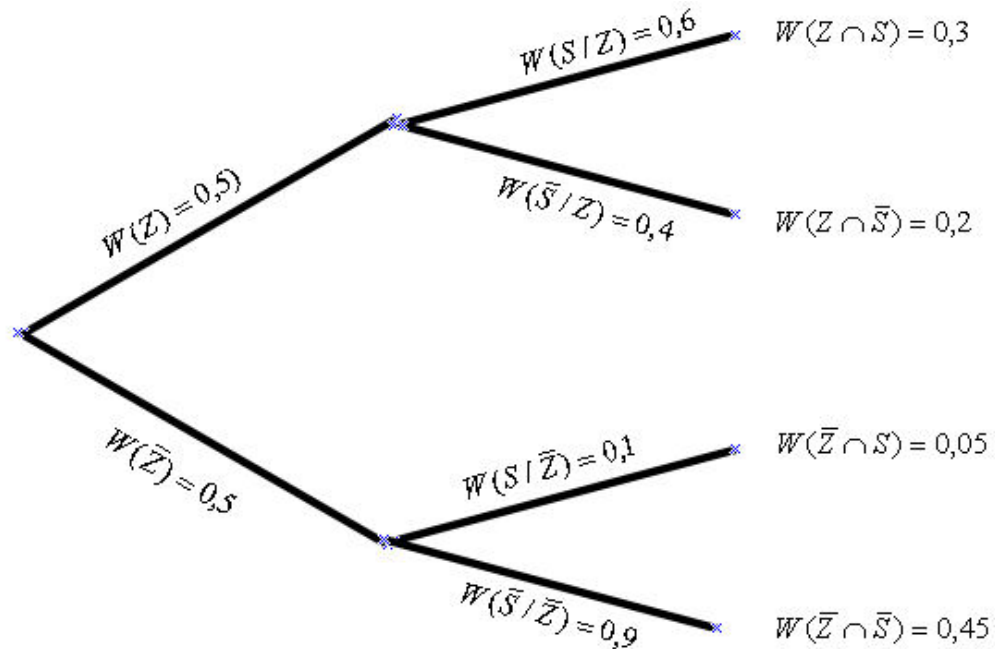
Wir wissen, dass $W(S|Z) = 0,6 + W(\bar{S}|Z) = ? = 1$ sein muss.

Also gilt: $W(\bar{S}|Z) = 0,4$

Wir wissen, dass $W(Z) = 0,5 \cdot W(\bar{S}|Z) = 0,4 = W(Z \cap \bar{S}) = ?$ gilt.

Also: $W(Z \cap \bar{S}) = 0,2$

Mit dieser Technik können wir alle Werte des Baumes bestimmen. Eingetragen sieht das dann so aus:



5. Schritt: Berechnung des gesuchten Ausdrucks

Wir suchen „Wie groß ist die Wahrscheinlichkeit, dass der Zug verspätet ist, unter der Bedingung, dass das Schiff nicht verspätet ist?“

Mathematisch ausgedrückt: $W(Z/\bar{S}) = ?$

Das können wir leider so nicht in dem Entscheidungsbaum ablesen. Hier hilft uns aber die Formelsammlung weiter. Auf Seite 25 steht unter dem Buchstaben „C“:

$W(A/B) = \frac{W(A \cap B)}{W(B)}$. Dort setzen wir jetzt einfach unsere Buchstaben ein.

$$W(Z/\bar{S}) = \frac{W(Z \cap \bar{S})}{W(\bar{S})}$$

Den Ausdruck $W(\bar{S})$ können wir leider so nicht ablesen. Aber wir können ihn berechnen in dem wir $W(\bar{S}) = W(Z \cap \bar{S}) + W(\bar{Z} \cap \bar{S})$ einfach addieren. Dann erhalten wir $W(\bar{S}) = 0,2 + 0,45 = 0,65$
Jetzt können wir unsere gesuchte Wahrscheinlichkeit berechnen:

$$W(Z/\bar{S}) = \frac{0,2}{0,65} = 0,3077.$$

Antwortsatz:

Die Wahrscheinlichkeit, dass der Zug verspätet ist, unter der Bedingung, dass das Schiff nicht verspätet ist liegt bei 30,77 %.

Übung 5 - Approximation

Fallbeispiel 7: In einem Industriebetrieb fallen im Jahresdurchschnitt 1.000.000 Buchungen an. Der Anteil der Fehlbuchungen betrage 0,75%. Bei einer Betriebsprüfung werden 2000 Buchungen zufällig ausgewählt und überprüft. Die Buchführung wird als ordnungsgemäß angesehen, wenn der Prozentsatz der fehlerhaften Buchungen 1% nicht übersteigt. Man berechne die Wahrscheinlichkeit dafür, dass die Ordnungsmäßigkeit der Buchführung verworfen wird.

1. Schritt: Informationen aus Sachverhalt entnehmen und Zufallsvariable definieren

Anzahl Buchungen = Grundgesamtheitsumfang $N = 1.000.000$

ausgewählte Buchungen = Stichprobenumfang $n = 2.000$

Anteil Fehlbuchungen = $\theta = 0,75\% = 0,0075$

Zufallsvariable definieren; x = Anzahl der Fehlbuchungen in der Stichprobe

2. Schritt: Was wird gesucht?

„...die Wahrscheinlichkeit dafür, dass die Ordnungsmäßigkeit der Buchführung verworfen wird“ = Anteil der Fehlbuchungen in der Stichprobe $> 1\%$

1% von 2000 = 20

Mathematisch ausgedrückt: $W(x > 20)$

3. Schritt: Wie ist die Zufallsvariable (ZV) verteilt?

a) diskret oder stetig?

da x abzählbar ist (der Wert muss zwischen 0 und 2000 liegen) ist die Zufallsvariable diskret verteilt.

b) Anzahl der Merkmale der ZV

x kann nur den Werte ordnungsgemäß oder nicht ordnungsgemäß gebucht annehmen. Daher ist das zweiseitige Merkmal gegeben

c) Ist die ZV binomial oder hypergeometrisch verteilt?

Bei der Stichproben werden mit „ziehen ohne zurücklegen“ ausgewählt, somit ist die sie hypergeometrisch verteilt

4. Schritt:

$$W(X = x) = \frac{\binom{M}{x} * \binom{N-M}{n-x}}{\binom{N}{n}} \quad \binom{N}{n} = \binom{1.000.000}{2.000} = \text{rechnerisch nicht lösbar!}$$

Aus diesem Grund können wir die Lösung nicht mit dieser Formel ausrechnen. Wir müssen jetzt überprüfen, ob eine Approximation möglich ist.

5. Schritt: Überprüfung, ob Approx.-Regel erfüllt sind

a) hypergeometrisch \approx binomial

Approx.-Regel $= \frac{n}{N} < 0,05$ für unser Beispiel gilt: $= \frac{2.000}{1.000.000} = 0,002 < 0,05$

Somit ist x näherungsweise binomial verteilt.

b) binomial \approx normal verteilt

Approx.-Regel $= n * \theta * (1 - \theta) > 9$

für unser Beispiel gilt: $= 2000 * 0,0075 * 0,9925 = 14,88 > 9$

Somit ist x näherungsweise normal verteilt

6. Schritt: Wahrscheinlichkeit berechnen

Bevor wir die Wahrscheinlichkeit berechnen können, müssen wir erst noch γ und σ ermitteln.

$$\gamma = n * \theta = 2000 * 0,0075 = 15$$

$$\sigma^2 = n * \theta * (1 - \theta) = 14,8875$$

$$\sigma = \sqrt{14,8875} = 3,8584$$

Als letzten Schritt müssen wir jetzt noch die Stetigkeitskorrektur für $W(Z \geq 20)$ beachten. Hierbei werden einfach 0,5 auf die 20 addiert.

Achtung jetzt wird es ein bisschen Unübersichtlich:

Die eigentliche Wahrscheinlichkeit, die wir suchen ist ja $W(x > 20,5)$. Würden wir jetzt Standardisieren, können wir den Wert nicht aus der Tabelle ablesen, da die nur Werte bis 3 berücksichtigt. Wir müssen also das Gegenereignis suchen:

$$= 1 - W(x > 20,5)$$

$$= 1 - W\left(Z \geq \frac{20 - \gamma}{\sigma}\right)$$

$$= 1 - W\left(Z \geq \frac{20,5 - 15}{3,8584}\right)$$

$$= 1 - W\left(Z \geq \frac{5,5}{3,8584}\right) = 1,425 \approx 1,43$$

$$= 1 - (FN(1,43) \rightarrow \text{Tabelle})$$

$$1 - 0,9236 = 0,0764 = 7,64\%$$

Antwortsatz:

Bei Prüfung dieser Art wird in annähernd 7,64% der Fälle die Ordnungsmäßigkeit der Buchführung verworfen!

Übung 6 – Konfidenzintervall

Fallbeispiel 10: fast food-Restaurantkette

Auf Stichprobenbasis wird der durchschnittliche Wochenumsatz γ der $N = 529$ Filialen einer landesweit operierenden fast food-Restaurantkette geprüft. Eine diesbezügliche Stichprobe von $n = 25$ Filialen ergab die nachfolgenden Wochenumsätze x_i^* mit absoluten Häufigkeiten h_i :

Wochenumsatz x_i^* (in Tsd. DM)	14,3	15,3	15,5	14,7	16,6
Häufigkeit h_i	4	5	6	?	7

Eine Häufigkeit fehlt. Die können wir aber berechnen. Wir wissen ja, dass $n = 25$ ist. Jetzt addieren wir die Häufigkeiten einfach zusammen. Kommen wir auf 22. Fehlen noch 3 bis zur 25. Also kommt hier der Wert 3 rein!

Aufgabe a) Schätzen Sie den durchschnittlichen Wochenumsatz γ aller Filialen mit einem geeigneten Konfidenzintervall auf einem Signifikanzniveau von 5%..

Grundsätzlich gilt: $\bar{x} - t * \hat{\sigma}_x \leq \gamma \leq \bar{x} + t * \hat{\sigma}_x$

Wir müssen also $\hat{\sigma}_x$, \bar{x} und t berechnen, damit wir einen Wert für γ bekommen.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^s h_i * x_i^*$$

$$\bar{x} = \frac{1}{25} * (4 * 14,3 + 5 * 15,3 + 6 * 15,5 + 3 * 14,7 + 7 * 16,6)$$

$$\bar{x} = 15,48$$

$$t = 1 - \frac{\alpha}{2} = 1 - \frac{0,05}{2} = 0,975 = \text{Tabelle} = 2,06; \nu = 25 - 1 = 24$$

$$\hat{\sigma}_x = \frac{S}{\sqrt{n}} * \sqrt{\frac{N-n}{n}} \quad \sqrt{\frac{N-n}{n}} \text{ kann entfallen, da } \frac{n}{N} = \frac{25}{529} = 0,047 \leq 0,05$$

Wir müssen jetzt „S“ berechnen!

$$Sn_{(x)}^2 = \frac{1}{n} \sum_{i=1}^s h_i * x_i^{*2} - \bar{x}^2$$

$$Sn_{(x)}^2 = \frac{1}{25} * (4 * 14,3^2 + 5 * 15,3^2 + 6 * 15,5^2 + 3 * 14,7^2 + 7 * 16,6^2) - 15,48^2$$

$$Sn_{(x)}^2 = \frac{1}{25} * 6007,1 - 239,6304$$

$$Sn_{(x)}^2 = 0,6536$$

Synchronisieren: (Mist hier fehlt etwas in meinen Aufzeichnungen; Wieso muss man hier synchronisieren?)

$$Sn - 1_{(x)}^2 = \frac{n}{n-1} * Sn_{(x)}^2 = \frac{25}{25-1} * 0,6536 = 0,6808$$

$$S = \sqrt{Sn_{(x)}^2} = \sqrt{0,6808} = 0,8251$$

$$\hat{\sigma}_x = \frac{S}{\sqrt{n}} = \frac{0,8251}{5} = 0,16502$$

$$\bar{x} - t * \hat{\sigma}_x \leq \gamma \leq \bar{x} + t * \hat{\sigma}_x$$

$$15,48 - 2,06 * 0,16502 \leq \gamma \leq 15,48 + 2,06 * 0,16502$$

$$15,14 \leq \gamma \leq 15,8199$$

Antwortsatz: Der durchschnittliche Wochenumsatz liegt bei einem Signifikanzniveau von 5% zwischen 15.140 und 15.8199 DM.

Aufgabe b) Testen Sie vor dem Hintergrund obiger Angaben und Resultate die Hypothese des Veräußerers der Restaurantkette, der durchschnittliche Wochenumsatz γ_0 sei 16. Fälln Sie eine Entscheidung (Ablehnung oder Nichtablehnung der Hypothese) auf der Grundlage einer einseitigen und einer zweiseitigen Fragestellung! Welche ist hier angemessen?

Anmerkung: In der Klausur kommt nur die zweiseitige Fragestellung dran!!!!

Gesucht. Konfidenzintervall für γ bei unbekannter Grundgesamtheitsvarianz σ^2 und kleinem Stichprobenumfang ($n \leq 30$)

Hypothese: $\gamma_0 = \gamma = 16$

Bei zweiseitiger Fragestellung stellt das Konfidenzintervall den Annahmebereich für eine Hypothese da!

Da wir in Aufgabe a9 einen Konfidenzintervall von $15,14 \leq \gamma \leq 15,8199$ errechnet hatten, kommen wir zu dem Ergebnis $\gamma = 16 \neq$ Konfidenzintervall somit Hypothese ablehnen.

Da die Aufgabe b) wichtig für die Klausur ist, hier noch ein Lösungsweg über den Parametertest

Nullhypothese= $H_0 = \gamma = \gamma_0 = 16$

Alternativhypothese= $H_A = \gamma \neq \gamma_0 = 16$

1. Schritt: Berechnung der Prüfgröße (siehe Anhang D- Forsam)

$$T = \frac{\bar{x} - \gamma_0}{\frac{\sigma}{\sqrt{n}}} = \frac{15,48 - 16}{\frac{0,8251}{5}} = -3,15$$

2. Schritt: kritische Größe (ablesen aus Tabelle S.45)

$$z_c = z_{1-\frac{\alpha}{2}} \rightarrow t_c = t_{1-\frac{\alpha}{2}}; \nu = n - 1 \rightarrow t_c = 0,975; \nu = 24 \rightarrow \text{Tabelle} = 2,06$$

3. Schritt: Testentscheidung

$z_c < |z| \rightarrow H_0$ ablehnen
 erfüllt, somit ist H_0 abzulehnen. (bei $\alpha = 0,05$)
 $2,06 < |3,15|$

Übung 7 –

Fallbeispiel 11: Verkehrszählung (handschriftliche Notiz des Stadtkämmerers von Kalkhausen)

Aufgabe a) Rekonstruieren Sie den durchgeführten Test und bestimmen Sie damit die Varianz des täglichen Verkehrsaufkommens

Nullhypothese: $H_0 = \gamma = \gamma_0 = 10.000$

Alternativhypothese: $H_A \neq \gamma = \gamma_0 = 10.000$ bei $\alpha = 0,05$

Grundsätzlich gilt: $\bar{x} - z * \theta_x \leq \gamma \leq \bar{x} + z * \theta_x$

Hier wird die Version „Gerade eben“ berechnet. Dies findet man auf der linken Seite des Konfidenzintervalls. Also gilt $\gamma = \bar{x} - z * \theta_x$

$\bar{x} = 9608$; $z = z_{1-\frac{\alpha}{2}} = z_{0,975} = \text{Tabelle} = 1,96$

$$\theta_x = \frac{\theta}{\sqrt{n}} = \frac{\theta}{5} =$$

$$\gamma = \bar{x} - z * \theta_x$$

$$10.000 = 9608 - 1,96 * \theta_x$$

$$\theta_x = (10000 - 9608) * \frac{5}{1,96} = 1000$$

Gesucht ist die Varianz – also θ^2

$\theta^2 = 1000 * 1000 = 1.000.000$ Anmerkung: Varianz ohne Bedeutung – Je höher der Wert, desto höher die Schwankung

Aufgabe b) Geben Sie zu einem Konfidenzniveau von 0,95 mit Hilfe der Stichprobe ein Konfidenzintervall für den wahren Erwartungswert γ an.

Grundsätzlich gilt: $\bar{x} - z * \theta_x \leq \gamma \leq \bar{x} + z * \theta_x$

Hier wird die Version „Gerade noch“ berechnet. Dies findet man auf der rechten Seite des Konfidenzintervalls. Also gilt $\gamma = \bar{x} + z * \theta_x$ $\bar{x} = 9608$

$$\bar{x} - z * \theta_x \leq \gamma \leq \bar{x} + z * \theta_x$$

$$9608 - 1,96 * \frac{1000}{5} \leq \gamma \leq 9608 + 1,96 * \frac{1000}{5}$$

$$9216 \leq \gamma \leq 10000$$

bei $\alpha = 0,05$

Übung 8

Fallbeispiel 13: Unternehmensberatung

In einem Wirtschaftsunternehmen ist die Stelle eines Assistenten der Geschäftsleitung zu besetzen. Der in Frage kommende Kreis von Bewerbern setzt sich aus Hochschulabsolventen mit sechs Studienabschlüssen $i = 1, \dots, 6$ zusammen.

Um eine Fehlbesetzung der Stelle zu vermeiden, wird ein Personalberatungsunternehmen mit der Aufgabe betraut, den in Frage kommenden Bewerberkreis hinsichtlich möglicher Signifikanzen bestimmter Studienabschlüsse zu prüfen. Hierzu werden 90 Kandidaten in einem umfangreichen Prüfverfahren auf die Eignungsgrade "tauglich" und "nicht tauglich" geprüft. Auf Grund der nachfolgend tabellierten Ergebnisse kommt das Personalberatungsunternehmen zu dem Ergebnis, dass alle Bewerber gleich tauglich sind, d. h. die sechs Studienabschlüsse im Kreis der tauglichen Bewerber gleich verteilt sind.

Studienabschluss =i	1	2	3	4	5	6
Anzahl tauglich h_i	19	13	14	12	17	15
f_i^e						

Aufgabe a) Formulieren Sie eine geeignete Null- und Alternativhypothese.

Nullhypothese: Die sechs Studienabschlüsse im Kreis der tauglichen Bewerber sind gleich verteilt.

mathematisch ausgedrückt: $f_i^e = \frac{1}{6}$

Alternativhypothese: Die sechs Studienabschlüsse im Kreis der tauglichen Bewerber sind **nicht** gleich verteilt.

Aufgabe b) Testen Sie diese Hypothese auf einem Signifikanzniveau von $\alpha = 0,05$

Wir erweitern die Tabelle:

Studienabschluss =i	1	2	3	4	5	6
Anzahl tauglich h_i	19	13	14	12	17	15
f_i^e	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$h_i^e = n * f_i^e$	15	15	15	15	15	15

n=90

1. Schritt: Prüfgröße

$$\chi^2 = \sum \frac{(h_i^0 - h_i^e)^2}{h_i^e}$$

$$\chi^2 = \frac{(19-15)^2}{15} + \frac{(13-15)^2}{15} + \frac{(14-15)^2}{15} + \frac{(12-15)^2}{15} + \frac{(17-15)^2}{15} + \frac{(15-15)^2}{15} = 2,2\bar{6}$$

2. Schritt: kritische Größe

$$\chi_c^2 = \chi_{1-\alpha}^2; \nu = K - m - 1 \quad \text{Anmerkung: } m = \text{Anzahl geschätzter Parameter} - \text{hier: } 0$$

$$\chi_c^2 = \chi_{1-0,05}^2; \nu = 6 - 0 - 1 \quad \text{Tabelle} = 11,07$$

$$\chi_c^2 = \chi_{0,95}^2; \nu = 5$$

3. Schritt: Testentscheidung

$$\chi^2 \leq \chi_c^2 \Rightarrow 2,26 \leq 11,07 \quad H_0 \text{ kann nicht abgelehnt werden!}$$

Übung 9

Fallbeispiel 12: Lernerfolge im Internet

Zwei Gruppen von Studenten werden repräsentativ ausgewählt. Die eine Gruppe lernt Statistik in einer herkömmlichen Vorlesung mit Übungen; die zweite Gruppe lernt den Stoff mit Hilfe einer Lernmaschine im Internet. Im Abschlußtest, der für beide Gruppen identisch ist, werden die folgenden mittleren Punktzahlen mit den angegebenen Varianzen erzielt:

Vorlesegruppe	Internetgruppe
$n = 21$	$n = 11$
$\bar{x} = 41$	$\bar{x} = 44,5$
$S_{n-1}^2 = 20$	$S_{n-1}^2 = 10$

Prüfen Sie auf dem 5%-Niveau, ob sich die beiden Gruppen im Lernerfolg, gemessen an der mittleren Punktzahl, unterscheiden.

Nullhypothese: die mittleren Punktzahlen unterscheiden sich nicht! $\Delta\gamma = \gamma_1 - \gamma_2 = 0$

Alternativhypothese: die mittleren Punktzahlen unterscheiden sich! $\Delta\gamma = \gamma_1 - \gamma_2 \neq 0$

1. Prüfgröße (Forsam Seite 58)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s * \sqrt{\frac{n_1 + n_2}{n_1 * n_2}}} \quad \text{mit } s = \sqrt{\frac{(n_1 - 1) * S_1^2 + (n_2 - 1) * S_2^2}{n_1 + n_2 - 2}}$$

$$s = \sqrt{\frac{(21-1) * 20 + (11-1) * 10}{21+11-2}} = 4,082482905$$

$$t = \frac{41 - 44,5}{4,082482905 * \sqrt{\frac{41 + 44,5}{41 * 44,5}}} = -2,303$$

2. kritische Größe (Forsam Seite 47)

$$z_c = t_c$$

$$t_c = t_{1-\frac{\alpha}{2}} = t_{0,975}; \nu = n_1 + n_2 - 2 = 21 + 11 - 2 = 30$$

Tabelle = 2,04

3. Testentscheidung

$$|T| = |-2,303| = 2,303 \geq 2,04 = t_c \quad H_0 \text{ abgelehnt!}$$

Antwortsatz: Die beiden Gruppen unterscheiden sich im Lernerfolg, gemessen an der mittleren Punktzahl bei einem Signifikanzniveau von $\alpha = 0,05$!