

1. Aufgabenblatt vom Montag, den 23. Oktober 2017 zur Vorlesung

Induktive Statistik für Soziologinnen und Soziologen (Mariana Nold)

Thema: Zufallsvorgang, Inferenzschluss und Wahrscheinlichkeitsverteilungen

Abgabe: Freitag 10. November 2017

Wichtige Definitionen und Begriffe aus den ersten beiden Vorlesungen:

- Zufallsvorgang
- Zufallsvariable
- Einfache Zufallsstichprobe
- Wahrscheinlichkeitsverteilung
- Binäre Variable und Bernoulliverteilung
- Binomialverteilung

1. Verteilung des Taschengeldes beschreiben

(18 Punkte)

Der Datensatz `taschengeld.dta` enthält die Ergebnisse einer fiktiven Befragung von 100 Grundschulkindern aus Jena zu folgenden Merkmalen:

**Aufgabe
abgeben!**

- id: Zahl zwischen 1 und 100 zur Identifizierung der Person
 - TG: monatliches Taschengeld in Euro
 - G: Geschlecht (0 : Junge, 1 : Mädchen)
 - WSp: Antwort auf die Frage, ob Wassersport zu den drei Lieblingssparorten gehört (0 : nein, 1 : ja)
- (a) Nennen Sie für jedes der vier Merkmale das Messniveau.
(4 Punkte)
- (b) Nutzen Sie die Übersichtstabellen der 1. Vorlesung um geeignete Lage- und Streuungsmaße zur Beschreibung des Merkmals TG auszuwählen.
Berechnen Sie diese Maße und interpretieren Sie sie inhaltlich.
Nutzen Sie den STATA-Befehl
`. summarize TG.`
(4 Punkte)
- (c) Ein Inferenzschluss ist der Rückschluss von Eigenschaften einer Stichprobe auf Eigenschaften einer Grundgesamtheit, wobei man die Irrtumswahrscheinlichkeit angeben kann. Nehmen Sie an, ihre Hypothese lautet: "Der Mittelwert des Taschengeldes von Grundschulkindern in Jena ist höher als 11 Euro."
Welche zwei Möglichkeiten gibt es, beim Inferenzschluss einen Fehler zu machen?
(2 Punkte)

- (d) Nutzen Sie den STATA-Befehl
- ```
. graph box TG, over(G)
```
- um den Boxplots der Jungs und den Boxplot der Mädchen zu vergleichen.  
Fassen Sie ihren Eindruck mit Bezug auf die Quartile in zwei bis drei Sätzen zusammen. Nutzen Sie für ihre Beschreibung den STATA-Befehl:

```
. bysort G: summarize TG, detail
```

um die Zahlenwerte der Quartile zu erfahren.  
(4 Punkte)

(e) Als Modellverteilung zur Approximation der empirischen Verteilung der Taschengeldes wir die Normalverteilung mit  $\hat{\mu} = 12.35$  und  $\hat{\sigma} = 2.71$  verwendet. Die Abbildungen 1 zeigt die Dichte und Verteilungsfunktion. In welcher Beziehung stehen die beiden Funktionen? Antworten Sie in zwei bis drei Sätzen.  
(4 Punkte)

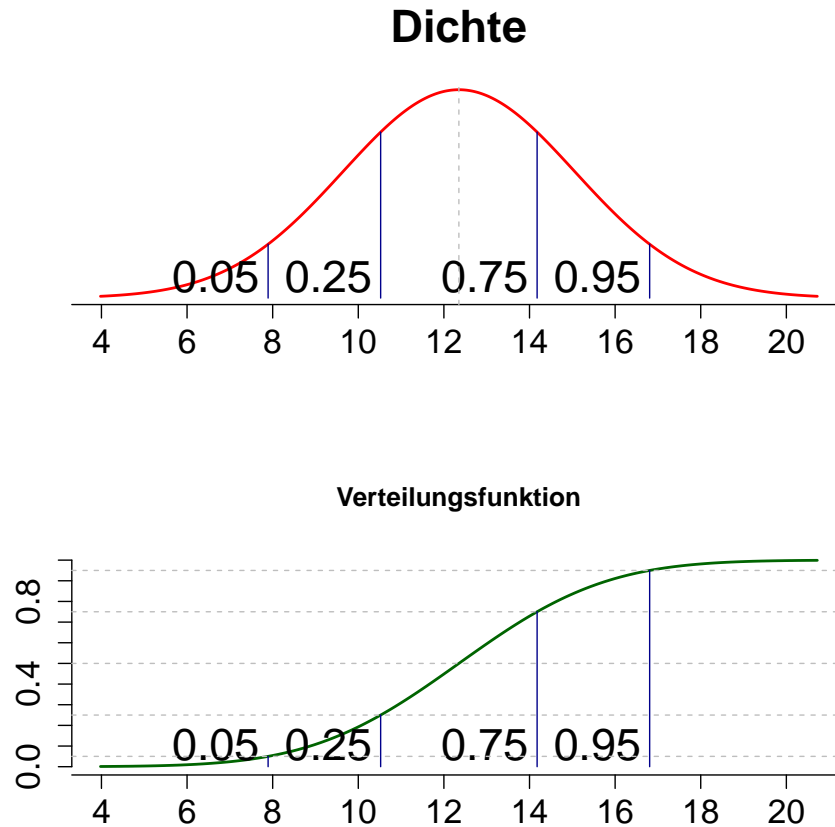


Abbildung 1: Dichte und Verteilungsfunktion der Normalverteilung

## 2. Wiederholung: Prozentsatzdifferenz in $2 \times 2$ -Kreuztabellen

(siehe Übungsblatt 5, Aufgabe 19 aus dem letzten Semester)

Sie sollen die Antwort auf die Frage, ob Wassersport zu den drei Lieblingsspararten gehört (WSP 0 : nein, 1 : ja), als abhängige Variable auffassen und das Geschlecht (G 0 : Junge 1 : Mädchen). Im Folgenden wird die unabhängige Variable mit  $X$  und die abhängige Variable mit  $Y$  bezeichnet.

|          |      | X               |                 | $\Sigma$        |
|----------|------|-----------------|-----------------|-----------------|
|          |      | Junge           | Mädchen         |                 |
| Y        | nein | $h_{1,1}$       | $h_{1,2}$       | $h_{1,\bullet}$ |
|          | ja   | $h_{2,1}$       | $h_{2,2}$       | $h_{2,\bullet}$ |
| $\Sigma$ |      | $h_{\bullet,1}$ | $h_{\bullet,2}$ |                 |

Tabelle 1:  $2 \times 2$ -Kreuztabelle für die Merkmale Geschlecht (X) und Lieblingssportart Wassersport (Y)

- (a) Gemäß der in den Sozialwissenschaften üblichen Konvention: Welcher der beiden folgenden **STATA**-Befehle erzeugt, die richtige Tabelle?

`. tab G WSp` oder `. tab WSp G`

Schreiben Sie Tabelle in Analogie zur Tabelle 1 mit Zahlenwerten gefüllt auf.

- (b) Berechnen Sie die bedingte relative Häufigkeit (= bedingte Anteilswerte) von  $Y|X$  und interpretieren Sie das Ergebnis.

(Nutzen Sie dazu entweder den **STATA**-Befehl `. tab WSp G, col nofreq` oder `tab G WSp, col nofreq`).

- (c) Berechnen und interpretieren Sie die Prozentsatzdifferenz (bei spaltenbezogener Prozentuierung). Die Prozentsatzdifferenz (bei spaltenbezogener Prozentuierung) ist definiert als:

$$d_{Y|X} = 100 \cdot \left( \frac{h_{1,1}}{h_{\bullet,1}} - \frac{h_{1,2}}{h_{\bullet,2}} \right)$$

- (d) Berechnen Sie mit Hilfe des **STATA**-Befehl

`. tab WSp G, chi2`

den  $\chi^2$ -Koeffizienten und erklären Sie die inwiefern der  $\chi^2$ -Koeffizienten die bei Unabhängigkeit erwarteten und die tatsächlichen beobachteten Häufigkeiten vergleicht.

## 3. Zentrale Definitionen

(10 Punkte)

Aufgabe  
abgeben!

- (a) Nennen Sie die Definition des Zufallsvorgangs und geben Sie ein Beispiel dazu an.  
(2 Punkte)

- (b) Mit Bezug auf Ihr Beispiel in Teilaufgabe a): Wie lautet die entsprechende Zufallsvariable?  
(2 Punkte)

- (c) Nennen Sie die Definition der binären Variable und geben Sie ein Beispiel dazu an.  
(2 Punkte)

- (d) Suchen Sie ein Beispiel für einen realen Prozess, den man mit Hilfe der Binomialverteilung modellieren kann. Begründen Sie ihre Auswahl in zwei Sätzen.  
(4 Punkte)