Lösungsnotizen Aufgabe 1

ACHTUNG: Die Lösungsnotizen sind (teilweise) eher knapp in Bezug auf die in der Aufgabe geforderte Begründung!

a)
- Merkmal Lebenszufriedenheit, feine Skala → metrisches Skalenniveau
- Zwei-Stichproben, Stichproben unabhängig

⇒ (Approximativer) Test auf Erwartungswert-Differenz bei unabhängigen Stichproben mit unbekannter Varianz (Stichprobe entweder groß genug oder Lebenszufriedenheit normalverteilt). (2.4.4 Zwei-Stichproben-t-Test, Variante II)
$H_0 : \mu_B \leq \mu_H \quad H_1 : \mu_B > \mu_H$

b)
- Merkmal Aggressionsscore, normalverteilt
- Zwei abhängige Stichproben

⇒ Tests auf Erwartungswertdifferenz für verbundene (abhängige) Stichproben (2.4.5 Gauß-Test und t-Test für verbundene Stichproben):
hier also: t-Test auf Differenz $D_i = X_{i,\text{nachher}} - X_{i,\text{vorher}}$
$H_0 : \mu_{nD} \geq \mu_{vD} \quad H_1 : \mu_{nD} < \mu_{vD}$

c) Zwei Stichproben t-Test auf die Erwartungswertdifferenz bei unabhängigen Stichproben (nötige Annahme: Körpergröße ist normalverteilt) Varianzen gleich oder nicht? (FS 2.4.4 Zwei-Stichproben-t-Test Variante I oder II)
$H_0 : \mu_M = \mu_J \quad H_1 : \mu_M \neq \mu_J$

d) Tests auf Erwartungswertdifferenz bei abhängigen Stichproben, Varianz unbekannt (2.4.4 Zwei-Stichproben-t-Test Variante II)
$H_0 : \mu_{nI} = \mu_{vI} \quad H_1 : \mu_{nI} \neq \mu_{vI}$

e) Diffferenz von Anteilen (bei unabhängigen Stichproben) ⇒ kein passender Test aus Vorlesung bekannt!!!

f) Differenz von Anteilen bei **abhängigen** Stichproben ⇒ kein passender Test aus Vorlesung bekannt!!!

g) Zusammenhang zwischen nominalen Merkmalen: Chi-Quadrat-Test (2.4.6) $n_i j > 5$.
H_0 : Es gibt keinen Zusammenhang zwischen B und W
H_1 : Es gibt einen Zusammenhang zwischen B und W

h)
- Ordinales Merkmal mit wenigen Ausprägungen
- Zwei unabhängige Stichproben

⇒ Nichtparametrischer Test

Lösungsnotizen Aufgabe 2

Angaben aus Aufgabe 3a (Blatt 12):

- $n = 500$
- 98 von 500 haben Fernsehsendung gesehen $\Rightarrow \bar{X} = \frac{98}{500}$

(1) Inhaltliche Frage: Haben mehr als 18% der Grundgesamtheit (diese ist nicht genau definiert) die Fernsehsendung gesehen?

(2) Statistisches Modell:
$X_1, \ldots, X_n$ i.i.d. Stichprobe, mit $X_i = \begin{cases} 1 & \text{hat Sendung gesehen} \\ 0 & \text{hat Sendung nicht gesehen} \end{cases}$ und $P(X_i = 1) = \pi$

(3) Hypothese statistisch formulieren:

$$H_0: \ \pi \leq \pi_0 = 0.18$$
$$H_1: \ \pi > \pi_0 = 0.18$$

(4) Festlegen eines Signifikanzniveaus: siehe Angabe; hier $\alpha = 0.05$

(5) Festlegen der Testgröße und der kritischen Region

- Anteilswert wird getestet
- Eine Stichprobe
- $n = 500$ groß $\Rightarrow$ aproximativer Test möglich (Normalverteilungsannahme)
- $\Rightarrow$ Approximative Tests für Hypothesen über Anteilswerte (Fall 1):
  - **Testgröße:** $\dfrac{\bar{X} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \overset{a}{\sim} \mathcal{N}(0,1)$ falls $\pi = \pi_0$ gilt.
  - **Kritische Region:** H_0 ablehnen, falls $T \geq z_{1-\alpha}$

(6) Auswerten der Stichprobe:

$$T = \frac{\bar{X} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{\frac{98}{500} - 0.18}{\sqrt{\frac{0.18(1-0.18)}{500}}} \approx 0.931$$

(7) Testentscheidung:
$z_{1-\alpha} = z_{0.95} = 1.64 > 0.931 = T \Rightarrow H_0$ kann nicht abgelehnt werden; d.h. es konnte nicht nachgewiesen werden, dass der Zuschaueranteil der Sendung größer als 18% ist.

Lösungsnotizen Aufgabe 4

(1) Inhaltliche Fragestellung: Gibt es einen Zusammenhang zwischen Vollzeit-Erwerbstätigkeit und Wohnort in Deutschland?

(2) Statistisches Modell

Zwei nominale Merkmale:

- X: Umfang der Erwerbstätigkeit (Voll- oder Teilzeit)
- Y: Wohnort (West- oder Ostdeutschland)

(3) Formulierung der Statistischen Hypothesen

H_0: Die beiden Merkmale sind stochastisch unabhängig
H_1: Die beiden Merkmale sind stochastisch abhängig

formal:

H_0: $p_{ij} = p_{i\bullet} \cdot p_{\bullet j}$ für alle $i = 1, \dots, k$, $j = 1, \dots, l$
H_0: $p_{ij} \neq p_{i\bullet} \cdot p_{\bullet j}$ für mindestens eine ij-Kombination

d.h. $H_0: \; P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$ für alle Paare i, j
gegen $H_1: \; P(X = x_i, Y = y_j) \neq P(X = x_i) \cdot P(Y = y_j)$ für mindestens *ein* Paar i, j

(4) Festlegen des Signifikanzniveaus $\alpha = 0.01$ (siehe Angabe)

(5) Testgröße und kritische Region

Geeigneter Test:
Zusammenhang zwischen zwei nominalen Merkmalen $\Rightarrow$ χ^2-Unabhängigkeitstest

Testgröße:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

Ablehnbereich: H_0 ablehnen, falls $T \geq \chi^2_{1-\alpha}((k-1)(m-1)$

(6) Auswerten der Stichprobe (über die absoluten Häufigkeiten!)

- Erst: Berechnung der erwarteten Besetzungszahlen, falls Unabhängigkeit vorliegt:

  Beobachtete Tabelle h_{ij}:

	Vollzeit ja	Vollzeit nein	
Ostdeutschland	401	343	744
Westdeutschland	699	2456	3155
	1100	2799	3899

  Unabhängigkeitstabelle $\tilde{h}_{ij}$

	Vollzeit	
	ja	nein
Ostdeutschland	201	534
Westdeutschland	890	2265

Beispielhaften Nebenrechnungen:

$$\tilde{h}_{ij} = \frac{h_{i\bullet} \cdot h_{\bullet j}}{n}$$
$$\tilde{h}_{11} = \frac{h_{1\bullet} \cdot h_{\bullet 1}}{n} = \frac{744 \cdot 1100}{3899}$$
$$\tilde{h}_{21} = \frac{h_{2\bullet} \cdot h_{\bullet 1}}{n} = \frac{3155 \cdot 1100}{3899}$$

– Dann Berechung der Teststatistik

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{\left(h_{ij} - \tilde{h}_{ij}\right)^2}{\tilde{h}_{ij}}$$
$$\approx \frac{(401-210)^2}{210} + \frac{(343-534)^2}{534} + \frac{(699-890)^2}{890} + \frac{(2456-2265)^2}{2265}$$
$$\approx 173.72 + 68.32 + 40.99 + 16.11$$
$$\approx 299.14$$

(6) Fällen der Testentscheidung

Vergleich der Testgröße mit dem $\chi^2_{1-\alpha}, (k-1)(l-1)$ Quantil der χ^2– Verteilung:

$\alpha = 0.01$, $k = 2$, $l = 2 \Rightarrow \chi^2_{0.99}, (1) = 6.64$

Also:
$299.14 > 6.64 = q_{0.99,(1)} \Rightarrow H_0$ ablehnen

Inhaltliche Antwort Es gibt einen zum Signifikanzniveau von 1% signifikanten Unterschied zwischen der Anzahl vollzeitig berufstätiger Frauen in West- und Ostdeutschland.