Technische Hochschule Köln
Fakultät für Wirtschafts- und Rechtswissenschaften
Prof. Dr. Arrenberg
Raum 221, Tel. 39 14
jutta.arrenberg@th-koeln.de

Übungen zur Vorlesung QM III

Chi-Quadrat-Anpassungstest

Aufgabe 17.1

Wir möchten überprüfen, ob der nachfolgende Datensatz $y_1, y_2, \ldots, y_{250}$ aus einer Normalverteilung stammt. Zunächst standardisieren wir die Werte, d.h. wir berechnen das arithmetische Mittel $\overline{y}$ und die Standardabweichung s_y und transformieren den ursprünglichen Datensatz wie folgt:

$$x_i = \frac{y_i - \overline{y}}{s_y}$$

Die Werte $x_1, x_2, \ldots, x_{250}$ klassieren wir wie folgt:

$x_{j-1}^* < x \leq x_j^*$	n_j
$x \leq -2$	8
$-2 < x \leq -1$	31
$-1 < x \leq 0$	89
$0 < x \leq 1$	90
$1 < x \leq 2$	26
$2 < x$	6
$\sum$	$n = 250$

Überprüfen Sie mit dem Chi-Quadrat-Anpassungstest zum Niveau $\alpha = 0{,}05$, ob der standardisierte Datensatz $x_1, x_2, \ldots, x_{250}$ aus einer Normalverteilung $\mathsf{N}(0;1)$ stammt.

Lösung zu Aufgabe 17.1

H_0: „Der standardisierte Datensatz stammt aus der Normalverteilung $\mathsf{N}(0;1)$“ gegen
H_1: „nicht H_0“

Wir berechnen zunächst die Normalverteilungs-Wahrscheinlichkeiten p_j der $I = 6$ Klassen:

$p_1 = F_U(-2) = 0{,}023$

$p_2 = F_U(-1) - F_U(-2) = 0{,}159 - 0{,}023 = 0{,}136$

$p_3 = F_U(0) - F_U(-1) = 0{,}5 - 0{,}159 = 0{,}341$

Aus Symmetrie-Gründen gilt für die NV: $p_1 = p_6, p_2 = p_5, p_3 = p_4$. Somit ergibt sich die folgende Arbeitstabelle:

$x^*_{j-1} < x \leq x^*_j$	n_j	p_j	$n \cdot p_j$
$x \leq -2$	8	0,023	5,75
$-2 < x \leq -1$	31	0,136	34
$-1 < x \leq 0$	89	0,341	85,25
$0 < x \leq 1$	90	0,341	85,25
$1 < x \leq 2$	26	0,136	34
$2 < x$	6	0,023	5,75
$\sum$	$n = 250$	≈ 1	250

Da alle unter H_0 erwarteten absoluten Häufigkeiten $n \cdot p_j$ größer gleich fünf sind, ist die Faustregel des Chi-Quadrat-Anpassungstests erfüllt.

Es ergibt sich folgender Wert der empirischen Teststatistik:

$$\frac{(8-5{,}75)^2}{5{,}75} + \frac{(31-34)^2}{34} + \frac{(89-85{,}25)^2}{85{,}25} + \frac{(90-85{,}25)^2}{85{,}25} + \frac{(26-34)^2}{34} + \frac{(6-5{,}75)^2}{5{,}75} =$$

$0{,}8804348 + 0{,}26470588 + 0{,}16495601 + 0{,}26466276 + 1{,}88235294 + 0{,}0186957 = 3{,}467982$

Der Freiheitsgrad der Chi-Quadrat-Verteilung beträgt $df = I - 1 = 6 - 1 = 5$. Der obere 5%-Punkt der Chi-Quadrat-Verteilung mit fünf Freiheitsgraden beträgt 11,070.

(*Hinweis: Leider wurde in der 2. Auflage in der Tabelle auf Seite 295 dieser Wert falsch gerundet, sodass der Wert gemäß dieser Tabelle 11,071 beträgt. Sorry! Dieser Fehler wurde in der 3. Auflage korrigiert.*)

Da gilt: $3{,}467982 < 11{,}070$, wird die Nullhypothese nicht abgelehnt; d.h. der standardisierte Datensatz stammt aus der Normalverteilung $\mathsf{N}(0;1)$.

Bei Universität Münster lässt sich kostenlos die Statistik- und Mathe-Software R herunterladen:

https://cran.uni-muenster.de/

Um mit der Software R den Chi-Quadrat-Anpassungstest (englisch: Chi-Squared goodness-of-fit test) durchzuführen, lauten die R-Befehle wie folgt:

```
ni=c(8,31,89,90,26,6)
pi=c(0.023,0.136,0.341,0.341,0.136,0.023)
chisq.test(ni,p=pi)
```

Die Ausgabe von R lautet wie folgt:

```
X-squared = 3.486
df=5
p-value = 0.6282
```

Der sortierte standardisierte Datensatz lautet wie folgt:

−2.511893933	−2.488170111	−2.460524009	−2.365268908	−2.284130946
−2.222593137	−2.131613722	−2.041842841	−1.908615391	−1.900150597
−1.873983058	−1.811123100	−1.803865735	−1.737075280	−1.683882999
−1.620904742	−1.609392886	−1.577635059	−1.532179945	−1.529012007
−1.513325951	−1.491562691	−1.461627114	−1.452164366	−1.442991564
−1.435975378	−1.429240336	−1.421500850	−1.398754823	−1.371594583
−1.338013657	−1.333678983	−1.324827848	−1.300899585	−1.235372293
−1.212945867	−1.142357086	−1.135192383	−1.122033175	−0.973012826
−0.945822079	−0.942355469	−0.917408777	−0.908796958	−0.902587512
−0.899090470	−0.788570418	−0.768013741	−0.765212027	−0.765145201
−0.763202858	−0.747429776	−0.705322466	−0.683821175	−0.682540214
−0.680417812	−0.679967524	−0.668135118	−0.667716721	−0.667538717
−0.650964183	−0.648539067	−0.645019329	−0.632260083	−0.627535897
−0.607642382	−0.605891482	−0.567431836	−0.548462408	−0.545750740
−0.544864187	−0.541246617	−0.504292732	−0.503406657	−0.484430635
−0.477417617	−0.456595967	−0.428572974	−0.426801356	−0.418457590
−0.407839414	−0.397079204	−0.367638285	−0.355310498	−0.350398950
−0.343001807	−0.340814405	−0.336364169	−0.326042660	−0.325658595
−0.309869878	−0.308157521	−0.307332808	−0.306001947	−0.286225991
−0.285147992	−0.274290306	−0.271007380	−0.270388232	−0.245611450
−0.242031934	−0.236910675	−0.216235923	−0.196623444	−0.186454174
−0.176993129	−0.173084324	−0.173033250	−0.166561340	−0.164942343
−0.163871247	−0.160000548	−0.157228547	−0.143412551	−0.140261909
−0.133758893	−0.104828187	−0.088991508	−0.088021695	−0.084510660
−0.077806946	−0.073809116	−0.063990098	−0.042461863	−0.031161616
−0.028325958	−0.019150434	−0.003241333	0.014679833	0.023859723
0.053889160	0.057496872	0.075105773	0.091305188	0.099597580
0.108509257	0.109229625	0.112318304	0.117414060	0.119541098
0.127924075	0.169572885	0.169832380	0.170638590	0.171033380
0.174367888	0.181187839	0.190214610	0.192815486	0.203102465
0.213262404	0.218693980	0.229975439	0.257006591	0.271441726
0.272195449	0.276612681	0.287421342	0.295049643	0.304455791
0.324928313	0.325304472	0.336697332	0.348303978	0.350965378
0.354386796	0.355071079	0.357750033	0.379028096	0.379762347
0.401897316	0.406586661	0.463532078	0.466788255	0.469150079
0.485753993	0.509357964	0.529882213	0.535531895	0.565252381
0.567671695	0.574131607	0.580092683	0.593222652	0.606554173
0.614739649	0.650003716	0.683488722	0.697179382	0.700719326
0.721302112	0.724188211	0.726967144	0.730297296	0.733222624
0.752499673	0.752994714	0.756640188	0.789153366	0.811962841
0.820106683	0.829098330	0.833675288	0.843424192	0.849668068
0.857642467	0.866334529	0.879221146	0.880852910	0.908630257
0.927555112	0.938069747	0.949139685	0.953204480	0.959029016

0.963405838	0.980236878	0.982890327	1.029983616	1.094581526
1.128140450	1.144325606	1.151177633	1.185653206	1.193295947
1.219686535	1.223057672	1.229303595	1.241577930	1.261305908
1.281923429	1.297212004	1.332858343	1.334515821	1.342473962
1.350370299	1.4025452396	1.439592253	1.513909570	1.664960984
1.692338346	1.858200198	1.873225876	1.875953813	2.031091798
2.196402934	2.214682077	2.596641972	2.622791810	2.786857039

Technische Hochschule Köln
Fakultät für Wirtschafts- und Rechtswissenschaften
Prof. Dr. Arrenberg
Raum 221, Tel. 39 14
jutta.arrenberg@th-koeln.de

Übungen zur Vorlesung QM III

Chi-Quadrat-Anpassungstest
Arbeitsblatt

Beispiel 1:
Vor der Pandemie ergab sich in den Präsenzklausuren von QM III ungefähr folgender Notenspiegel:

Note	sehr gut	gut	befriedigend	ausreichend	durchgefallen
Anteil	12,8%	48,0%	14,3%	9,7%	15,2%

Prüfen Sie mit einem Test zum Niveau 0,05, ob sich der Notenspiegel aufgrund der online-Lehre in den Remote-Prüfungen während der Pandemie nicht verändert hat. Die erste Remote-Prüfung mit 288 Studierenden in QM III ergab folgenden Notenspiegel:

Note	sehr gut	gut	befriedigend	ausreichend	durchgefallen	Summe
Anzahl	153	108	19	5	3	288

Lösung:
X = Note
$p_1 = 0{,}128$ = Anteil von „sehr gut“
$p_2 = 0{,}480$ = Anteil von „gut“
$p_3 = 0{,}143$ = Anteil von „befriedigend“
$p_4 = 0{,}097$ = Anteil von „ausreichend“
$p_5 = 0{,}152$ = Anteil von „durchgefallen“

n_i = Anzahl der Studierenden in der Remote-Prüfung mit Note i
$n = 288$

Note	sehr gut	gut	befriedigend	ausreichend	durchgefallen	Summe
p_i	0,128	0,480	0,143	0,097	0,152	1
n_i	153	108	19	5	3	288
$n \cdot p_i$	36,9	138,2	41,2	27,9	43,8	288

Die kleinste erwartete Häufigkeit beträgt 27,9. Somit ist die Faustregel $n \cdot p_i \geq 5$ erfüllt.
$df = I - 1 = 5 - 1 = 4$
Der obere 5%-Punkt der Chi-Quadrat-Verteilung mit vier Freiheitsgraden beträgt 9,488.

$$\begin{aligned}\chi^2_{\text{emp.}} &= \frac{(153-36{,}9)^2}{36{,}9} + \frac{(108-138{,}2)^2}{138{,}2} + \frac{(19-41{,}2)^2}{41{,}2} + \frac{(5-27{,}9)^2}{27{,}9} + \frac{(3-43{,}8)^2}{43{,}8} \\ &= 365{,}9 + 6{,}6 + 11{,}9 + 18{,}8 + 38{,}0 \\ &= 441{,}2 \geq 9{,}488\end{aligned}$$

d.h. Ablehnung von H_0; d.h. der Notenspiegel in der Remote-Prüfung entspricht nicht dem Notenspiegel in der Präsenz-Prüfung.