Ue Statistik u. Wahrscheinlichkeitsth. f. Inf.	Di 12:00 – 18:00
107.251 W 2004|5 http://www.statistik.tuwien.ac.at/RISueb/	**11.Blatt**
Werner GURKER Tel.: 58801–10724 Spr.: Di u. Do 11–12 e–mail: W.Gurker@tuwien.ac.at	25. Jän. 2005

1. Prüfen Sie mittels W–Netz für die Daten von **Tabelle 2.1**, wahlweise für `GE=M` oder `GE=W`, ob die Körpergrößen einer Normalverteilung folgen. Wie kann man – im positiven Fall – dem Netz Schätzwerte für μ und σ entnehmen?

   Hinweis: Zur Vermeidung bzw. Verringerung von Verzerrungen trägt man üblicherweise nicht die Punkte $\big(x_{(i)}, i/n\big)$, $i = 1, \ldots, n$, im Netz auf, sondern die Punkte:

   $$\left(x_{(i)}, \frac{i - 0.5}{n}\right), \quad i = 1, \ldots, n$$

   Ein Nebeneffekt dieser Vorgangsweise besteht darin, daß auf diese Weise auch der letzte Punkt $x_{(n)}$ eingetragen werden kann (wäre sonst wegen $\Phi^{-1}(1) = \infty$ nicht möglich).

   (Vorgefertigte Normalnetze zum Ausdrucken gibt es auf der Ü–Homepage und an anderen Stellen im Internet; `R`–User nehmen `qqnorm` und `qqline`.)

2. Um zu prüfen, ob ein Schaltkreis fehlerfrei ist, wird er mit einer Folge von 100 Inputs betrieben. Am Ausgang werden X Einser (und $100-X$ Nullen) beobachtet. Bei einem fehlerfreien Schaltkreis kann man 50 Einser erwarten. Die Nullhypothese („Schaltkreis fehlerfrei“) wird verworfen, falls $X \notin (40, 60)$.

   (a) Wie groß ist bei diesem Test die Fehlerwahrscheinlichkeit 1. Art?

   (b) Wie groß ist die Fehlerwahrscheinlichkeit 2. Art, wenn der Schaltkreis im Mittel tatsächlich 45 Einser liefert?

3. Die folgende Tabelle enthält einige (empirische) Kenngrößen für die Körpergrößen von **Tabelle 2.1**:

GE	n	$\overline{x}$	s
M	39	182.51	6.54
W	30	168.27	5.23

   Testen Sie mit $\alpha = 5\%$, ob die Varianzen der Körpergrößen von männlichen und weiblichen Studierenden gleich sind, d.h. testen Sie:

   $$\mathcal{H}_0 : \ \sigma_M^2 = \sigma_W^2 \quad \text{gegen} \quad \mathcal{H}_1 : \ \sigma_M^2 \neq \sigma_W^2$$

   Gehen Sie dabei von normalverteilten Beobachtungen aus.

4. Fortsetzung von **Bsp 3**: Ermitteln Sie ein 95%–Konfidenzintervall für die Differenz $\Delta = \mu_M - \mu_W$ der mittleren Körpergrößen von männlichen und weiblichen Studierenden. Gehen Sie dabei von normalverteilten Beobachtungen (mit gleicher Varianz) aus.

   Hinweis: Eine passende Pivotgröße für die obige Situation wurde in der Vorlesung angegeben.

5. Eine Untersuchung von 320 Familien mit fünf Kindern ergab die folgende Häufigkeitsverteilung für die Anzahl von Knaben (K) und Mädchen (M):

Anzahl K|M	5K|0M	4K|1M	3K|2M	2K|3M	1K|4M	0K|5M
Anzahl Familien	18	56	110	88	40	8

Prüfen Sie mittels Chiquadrat–Anpassungstest, ob die Daten mit der Hypothese gleichwahrscheinlicher Knaben– und Mädchengeburten übereinstimmen, d.h. prüfen Sie $H_0 : X \sim B_{n,p}$ mit $n = 5$ und $p = \frac{1}{2}$. Nehmen Sie zuerst $\alpha = 5\%$ und dann $\alpha = 1\%$. Kommentieren Sie die Ergebnisse.

6. Prüfen Sie mittels Chiquadrat–Anpassungstest (mit $\alpha = 5\%$) für die Daten von **Tabelle 2.1**, wahlweise für `GE=M` oder `GE=W`, ob die Körpergrößen einer Normalverteilung folgen.

   Hinweis: Zunächst sind die beiden Parameter μ und σ (plausibel) zu schätzen. Anschließend ist eine Klasseneinteilung vorzunehmen; dafür gibt es zahlreiche Möglichkeiten. Eine komfortable Einteilung ergibt sich, wenn man (unter H_0) *gleichwahrscheinliche* Klassen nimmt. Nehmen Sie beispielsweise 5 Klassen:

$$(-\infty, \widehat{x}_{1/5}],\ (\widehat{x}_{1/5}, \widehat{x}_{2/5}],\ \ldots,\ (\widehat{x}_{4/5}, \infty)$$

   wobei $\widehat{x}_p = \widehat{\mu} + z_p\, \widehat{\sigma}$.

1. Im (Normal–) Wahrscheinlichkeitsnetz ist die Verteilungsfunktion einer Normalverteilung eine Gerade. Wie die Abbildungen zeigen, ist dies für `GE=W` recht gut erfüllt, die Normalverteilung als Datenmodell somit nicht auszuschließen. Bei `GE=M` ist die Anpassung deutlich schlechter („Schlangenlinie"), trotzdem kann auch hier eine Normalverteilung nicht ganz ausgeschlossen werden.

   Die Punkte wurden nach der Anleitung im *Hinweis* eingetragen. Die eingezeichnete Gerade entspricht dabei der *plausiblen Schätzung* der Verteilungsfunktion, d.h. der Funktion:

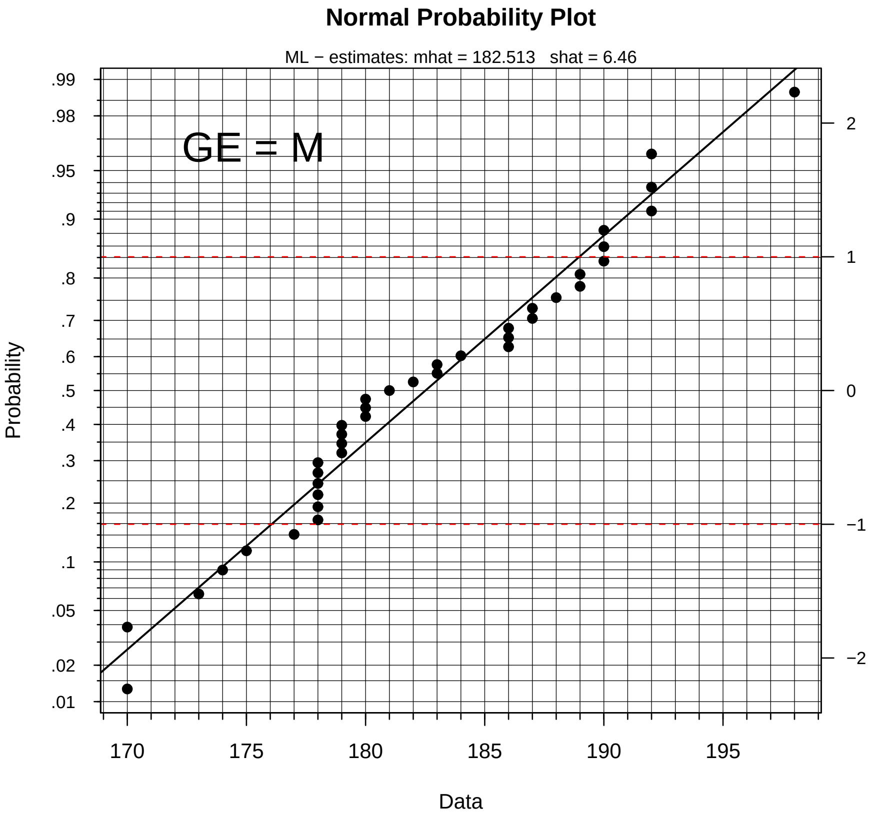
$$\widehat{F}(x) = \Phi\left(\frac{x-\widehat{\mu}}{\widehat{\sigma}}\right) \quad \text{mit} \quad \widehat{\mu} = \overline{x} \quad \text{und} \quad \widehat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}$$

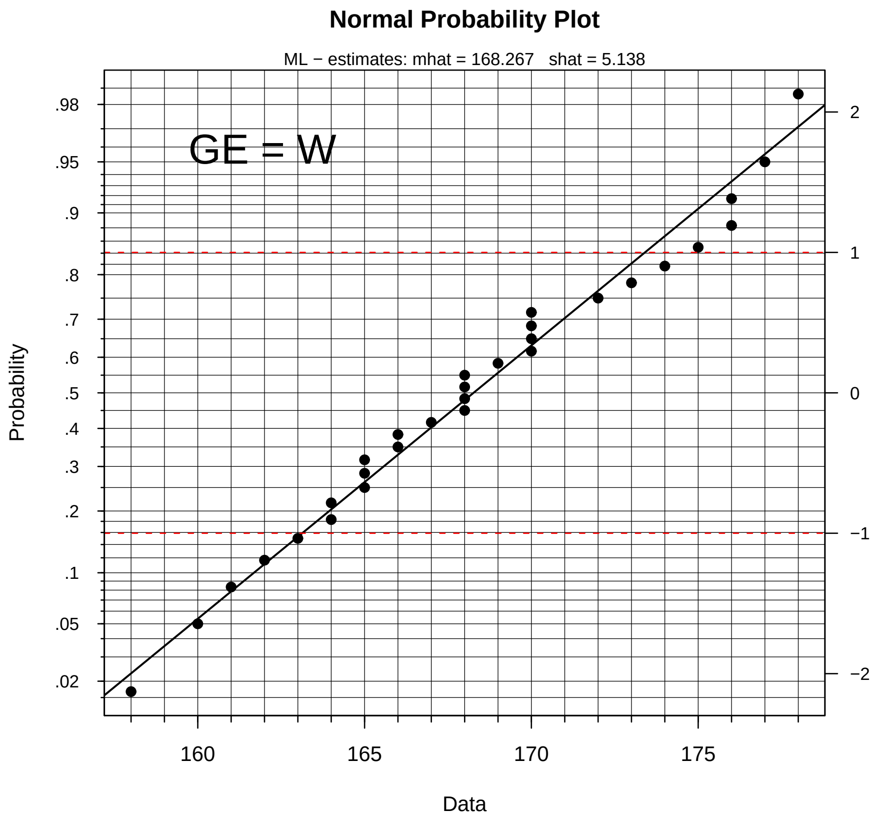
   Einen Schätzwert für μ kann man beim Schnittpunkt der 50%–(bzw. 0–) Linie mit der eingezeichneten Geraden entnehmen:

$$\mu_M \approx 182.5; \quad \mu_W \approx 168.3$$

   Ein Schätzwert für σ ergibt sich als Differenz zwischen dem Schnittpunkt der 84.13%– (bzw. 1–) Linie mit der eingezeichneten Geraden und der Schätzung für μ (oder als Differenz zwischen dem entsprechenden Schnittpunkt der 15.87%–(bzw. −1–) Linie und der Schätzung für μ):

$$\sigma_M \approx 189 - 182.5 = 6.5; \quad \sigma_W \approx 173.5 - 168.3 = 5.2$$





2. (a) Ein Fehler 1. Art liegt vor, wenn eine richtige Nullhypothese verworfen wird; hier gilt mit $X \sim B_{100,p}$:

$$\alpha = W(X \leq 40|p = 0.5) + W(X \geq 60|p = 0.5) = 2\,W(X \leq 40|p = 0.5)$$
$$= 2\sum_{i=0}^{40}\binom{100}{i}(0.5)^{100} \approx 2\Phi\left(\frac{40.5-50}{5}\right) = 2\big[1-\Phi(1.9)\big] = 0.057$$

(b) Ein Fehler 2. Art liegt vor, wenn eine falsche Nullhypothese angenommen (nicht verworfen) wird:

$$\beta(0.45) = W(41 \leq X \leq 59|p = 0.45) = \sum_{i=41}^{59}\binom{100}{i}(0.45)^i(0.55)^{100-i}$$
$$\approx \Phi\left(\frac{59.5-45}{\sqrt{24.75}}\right) - \Phi\left(\frac{40.5-45}{\sqrt{24.75}}\right) = 0.815$$

3. F–Test auf Gleichheit der Varianzen (normalverteilte Beobachtungen):

$$F_{38,29;\,0.025} = \frac{1}{F_{29,38;\,0.975}} = \frac{1}{1.972} < \frac{s_M^2}{s_W^2} = \frac{(6.54)^2}{(5.23)^2} = 1.564 < F_{38,29;\,0.975} = 2.038$$

$\Longrightarrow \quad \mathcal{H}_0$ nicht verwerfen

4. Die Konstruktion eines Konfidenzintervalls für $\Delta = \mu_M - \mu_W$ beruht auf der folgenden Pivotgröße (Vsn.: normalverteilte Beobachtungen, Varianzen gleich):

$$Z = \frac{\overline{X}_M - \overline{X}_W - \Delta}{S_g \sqrt{\dfrac{1}{n_M} + \dfrac{1}{n_W}}} \sim t_{n_M + n_W - 2} \quad \text{mit} \quad S_g^2 := \frac{(n_M - 1)\, S_M^2 + (n_W - 1)\, S_W^2}{n_M + n_W - 2}$$

Einschließen von Z in zwei passende t–Quantile (ÜW $= 1 - \alpha$) und Auflösen der Doppelungleichung nach Δ:

$$\overline{X}_M - \overline{X}_W \pm t_{n_M + n_W - 2;\, 1-\alpha/2}\, S_g \sqrt{\frac{1}{n_M} + \frac{1}{n_W}}$$

Konkret:

$$s_g^2 = \frac{38 \cdot (6.54)^2 + 29 \cdot (5.23)^2}{38 + 29} = 36.10$$

$$182.51 - 168.27 \pm \underbrace{t_{67,\, 0.975}}_{= 1.996} \sqrt{36.10} \sqrt{\frac{1}{39} + \frac{1}{30}} = [11.32,\ 17.15]$$

5. X = Zahl der Knaben|Mädchen (in Familien mit 5 Kindern)

Einfacher Chiquadrat–Test: $\mathcal{H}_0 : X \sim B_{5,\, 0.5}$ (gegen $\mathcal{H}_1 : X \not\sim B_{5,\, 0.5}$)

Wahrscheinlichkeiten (unter $\mathcal{H}_0$): $w_i = \binom{5}{i}\left(\frac{1}{2}\right)^5 = \binom{5}{i}\frac{1}{32}, \quad i = 0, 1, \ldots, 5$

K|M	H	w	nw	$(H - nw)^2/nw$
5|0	18	1/32	10	6.40
4|1	56	5/32	50	0.72
3|2	110	10/32	100	1.00
2|3	88	10/32	100	1.44
1|4	40	5/32	50	2.00
0|5	8	1/32	10	0.40
Summe	320	1	320	11.96

Freiheitsgrade: $r - 1 = 6 - 1 = 5$

$11.96 > \chi^2_{5;\, 0.95} = 11.07; \quad 11.96 < \chi^2_{5;\, 0.99} = 15.09$

Kommentar: Im ersten Fall wird die Nullhypothese (knapp) verworfen, im zweiten Fall (kleinere Wahrscheinlichkeit eines Fehlers 1. Art, Test ist „konservativer") jedoch nicht. Insgesamt kann man sagen, daß die Hypothese gleichwahrscheinlicher Knaben– und Mädchengeburten zumindest zweifelhaft ist; weitere Untersuchungen sind notwendig.

6. X = Körpergröße

Zusammengesetzter Chiquadrat–Test: $\mathcal{H}_0 : X \sim N(\mu, \sigma^2)$

Plausible Schätzwerte (vgl. VO): $\widehat{\mu} = \overline{x}$, $\widehat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$

Klasseneinteilung: $(-\infty, \widehat{x}_{1/5}]$, $(\widehat{x}_{1/5}, \widehat{x}_{2/5}]$, $\ldots$, $(\widehat{x}_{4/5}, \infty)$ (mit $\widehat{x}_p = \widehat{\mu} + z_p\,\widehat{\sigma}$)

(1) Männer: $\overline{x} = 182.51$, $s = 6.46$

Klassen	H	$\widehat{w}$	$n\,\widehat{w}$	$(H - n\widehat{w})^2/n\widehat{w}$
$(-\infty, 177]$	6	0.2	7.8	0.42
$(177, 181]$	13	0.2	7.8	3.47
$(181, 184]$	5	0.2	7.8	1.01
$(184, 188]$	5	0.2	7.8	1.01
$(188, \infty)$	10	0.2	7.8	0.62
Summe	39	1.0	39	6.51

Freiheitsgrade: $r - s - 1 = 5 - 2 - 1 = 2$

$6.51 > \chi^2_{2,\,0.95} = 5.99 \quad\Longrightarrow\quad \mathcal{H}_0$ verwerfen

(2) Frauen: $\overline{x} = 168.27$, $s = 5.14$

Klassen	H	$\widehat{w}$	$n\,\widehat{w}$	$(H - n\widehat{w})^2/n\widehat{w}$
$(-\infty, 164]$	5	0.2	6	0.17
$(164, 167]$	7	0.2	6	0.17
$(167, 170]$	6	0.2	6	0.00
$(170, 173]$	5	0.2	6	0.17
$(173, \infty)$	7	0.2	6	0.17
Summe	30	1.0	30	0.67

Freiheitsgrade: $r - s - 1 = 5 - 2 - 1 = 2$

$0.67 < \chi^2_{2,\,0.95} = 5.99 \quad\Longrightarrow\quad \mathcal{H}_0$ nicht verwerfen