Übungen über die Weihnachtsferien

(1) (a)

$$
\begin{aligned}
\mu(X) &= 0\cdot\frac{1}{10}+1\cdot\frac{2}{10}+2\cdot\frac{3}{10}+3\cdot\frac{3}{10}+4\cdot\frac{1}{10}=\frac{21}{10},\\
\sigma^2(X) &= \left(\frac{21}{10}\right)^2\frac{1}{10}+\left(\frac{11}{10}\right)^2\frac{2}{10}+\left(\frac{1}{10}\right)^2\frac{3}{10}+\left(\frac{9}{10}\right)^2\frac{3}{10}+\left(\frac{29}{10}\right)^2\frac{1}{10}=\frac{177}{100},\\
\text{also } \sigma(X) &= \frac{1}{10}\sqrt{177},\ \text{das ist etwa } 1.33.
\end{aligned}
$$

(b) Hier ist eine Tabelle der wesentlich benötigten Werte der Verteilungsfunktion:

a	0	1	2	3	4
$F_X(a)$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{6}{10}$	$\frac{99}{10}$	1

Man erhält eine Treppe (die man zeichnen sollte), Stufenhöhe ist 0 bei $a<0$, $\frac{1}{10}$ zwischen 0 und (unter) 1, usw.

(c) $F_X(2.7)=\frac{6}{10}$.

(2) (a) $F_X(11)-F_X(7)=P(8\le X\le 11)=\sum_{k=8}^{11}\binom{30}{k}\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{30-k}=0.557\,1$ (gerundet).

(b) $F_X(5)=\sum_{k=0}^{5}\binom{30}{k}\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{30-k}=0.0355$ (gerundet). Das sollte nicht allzu sehr überraschen, da wir uns mit dem Bereich von 0 bis 5 noch weit unter dem Erwartungswert bewegen.

(c)

$$
\begin{aligned}
F_X(11)-F_X(7) &\approx \Phi_{0,1}\left(\frac{11.5-10}{\sqrt{10\cdot 2/3}}\right)-\Phi_{0,1}\left(\frac{7.5-10}{\sqrt{10\cdot 2/3}}\right)\\
&\approx \Phi_{0,1}(0.58)-\Phi_{0,1}(-0.97)=0.553.
\end{aligned}
$$

Beachten Sie, wie genau das ist trotz des noch recht bescheidenen $n=30$. Ohne die Korrektur erhielte man den sehr ungenauen und schon unbrauchbaren Wert 0.484! Es sei noch einmal darauf hingewiesen, dass man natürlich bei Nutzung eines Rechners, der Normalverteilung zu verwenden gestattet, den ganzen Ausdruck $\Phi_{0,1}\left(\frac{11.5-10}{\sqrt{10\cdot 2/3}}\right)-\Phi_{0,1}\left(\frac{7.5-10}{\sqrt{10\cdot 2/3}}\right)$ auswertet und nicht etwa die Ausdrücke in der Klammer mit sehr ungenauen Dezimalzahlen wiedergibt und dann auf diese ungenauen Werte die Normalverteilung anwendet - das geschah hier mit Rücksicht auf die Verfügbarkeit allein der Tabelle.

(d) Man zieht 30 Kugeln mit Zurücklegen aus einer Urne, von deren Kugeln genau 1/3 rot sind, und $X=$ Anzahl der gezogenen roten Kugeln. (Ob in der Urne drei Kugeln sind, davon genau eine rot, oder 30, davon 10 rot (usw., usw.), das ist völlig gleichgültig. Das spielt nur eine Rolle beim Ziehen ohne Zurücklegen. Wir könnten natürlich auch mit dem Computer 30 Zufallszahlen im Bereich [0,1] ziehen und dann $X=$ Anzahl der Fälle, in denen die gezogene Zahl zwischen 0 und 1/3 liegt, bilden. $F_X(11)-F_X(7)$ ist dann die Wahrscheinlichkeit dafür, mindestens 8 und höchstens 11 rote Kugeln zu ziehen. $F_X(5)$ ist die Wahrscheinlichkeit dafür, höchstens 5 rote Kugeln zu ziehen.

(3) (a) Eine Mehrheit für 'Ja' entsteht mit Wahrscheinlichkeit $\sum_{k=4}^{7}\binom{7}{k}\frac{1}{2^7}=0.5$, wie man auch sofort hätte erraten sollen.

(b) Bei zwei für 'Ja' Entschlossenenen und fürnf 'Zufallswählern' ist die Wahrscheinlichkeit für 'Ja'-Gruppenentscheidung: $\sum_{k=2}^{5}\binom{5}{k}\frac{1}{2^5} = 0.812\,5$. Sie ist also überraschend stark angestiegen. (So können entschlossene Minderheiten unentschlossene Mehrheiten majorisieren.)

(4) (a) $P(X = 10) = \frac{1}{\binom{30}{10}} = \frac{1}{30\,045\,015}$, das ist so gut wie 0.

(b) $P(X = 5) = \frac{\binom{10}{5}\binom{20}{5}}{\binom{30}{10}} = 0.13$ (gerundet).

(5) Die Wahrscheinlichkeit für lauter zufällig korrekte Antworten ist $\left(\frac{1}{4}\right)^5 = \frac{1}{1024}$. n soll so gewählt werden, dass $\left(1 - \frac{1}{1024}\right)^n \le \frac{1}{2}$, lösen wir die Gleichung, so erhalten wir $\frac{\ln(2)}{\ln(1024)-\ln(1023)} = 709.436$, also genügt $n = 710$.

(6) Empirisches Maß für Aufgabenschwierigkeit: $s(a)$ = relative Häufigkeit, mit der die Aufgabe a gelöst wurde. Für Fähigkeit der Person: $f(A)$ = Anteil der von Person A korrkekt gelösten Aufgaben. Diese beiden liegen jedenfalls sehr nahe. Natürlich könnte man in beiden Fällen auch an absolute Häufigkeiten denken, aber dann hat man die speziellen Verhältnisse dieses Testumfanges (hinsichtlich der Zahl der Aufgaben sowie der Personen) unnötig mit hineingebracht. Die relativen Häufigkeiten eignen sich dagegen zu unmittelbaren Vergleichen mit andern entsprechenden Tests.

(7) (a) $P(I) = 0.1 \cdot 0.95 + 0.9 \cdot 0.05 = 0.14$. ('$I$' steht für: 'Ißt Pralinen am Abend'.)

(b) $P\left(\ddot{A}|I\right) = \frac{P(I|\ddot{A})P(\ddot{A})}{P(I)} = \frac{0.95 \cdot 0.1}{0.14} = 0.68$ (gerundet), also immerhin hat man etwa 2/3 Wahrscheinlichkeit für vorausgehenden Ärger ('$\ddot{A}$') bedingt durch I.

(8) Die Bedingung lautet $2.58\frac{\sigma(X)}{\sqrt{30}} \le 2$, also $\sigma(X) \le 2\sqrt{30}/2.58 = 4.246$. (aufgerundet).

(9) (a) Grenze des gewünschten einseitigen Vertrauensintervalls ist $5 - 2.33\sqrt{\frac{9}{100} + \frac{1.5^2}{100}} = 4.22$.

(b) Aus dem Resultat von (a) kann man unmittelbar ablesen, dass die fragliche Hypothese auf einem Niveau kleiner als 0.01 verworfen werden kann. Genauer ergibt sich: Wir lösen die Gleichung $4 + z\sqrt{\frac{9}{100} + \frac{1.5^2}{100}} = 5$ und erhalten abgerundet $z = 2.98$, also $\alpha = \Phi_{0,1}(-2.98) = 0.001\,44$, das ist also noch deutlich besser als 0.01, nahe einem Tausendstel.

(10) Zu den Daten

Stunden der Beschäftigung mit Kunst vor dem Kurs	70	70	65	85	85	75	60	75	89	90
Stunden der Beschäftigung mit Kunst nach dem Kurs	110	85	75	82	100	92	83	91	99	105

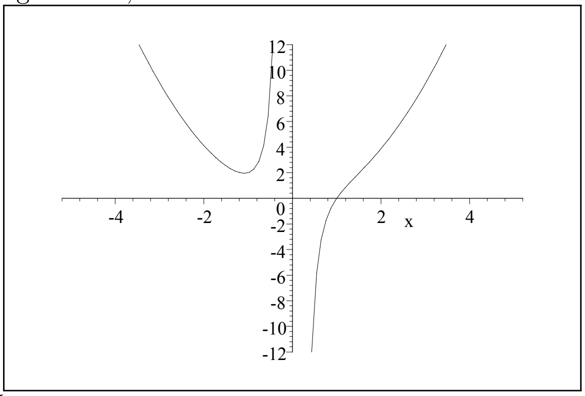
.

Zur ersten Frage: Die Folge der Differenzwerte ist 40,15,10,−3,15,17,23,16,10,15. Das ergibt $\overline{d} = 15.8$ und $s(D) = 10.861$ (aufgerundet). Das gewünschte Vertrauensintervall ist also $15.8 \pm 2.26 \cdot \frac{10.861}{\sqrt{10}}$, also $[8.04; 23.6]$. Zur zweiten Frage: Wir lösen $10 + t\frac{10.861}{\sqrt{10}} = 15.8$, das ergibt $t = 1.689$. Der von uns verwandten Tabelle kann man nur entnehmen, dass auf einem (schlechten) Niveau zwischen 0.05 und 0.1 zu verwerfen ist. Genauer bekommt man mit Computerhilfe $\alpha = 0.0733$. Im Vergleich zur vorigen Aufgabe: Dort hatten wir den Fall der 'unabhängigen Stichproben', so dass die Streuung der Differenz der Mittelwertsvariablen zu bilden war, hier ist die Streuung der Mittelwertsvariablen der Differenzvariablen zu bilden. Außerdem ist hier zwingend t zu verwenden, was aber auch anders als im Falle der letzten Aufgabe, bei der allerdings Verwendung der Normalverteilung zumal bei den recht hohen Stichprobenumfängen in Ordnung war, (Problem der Freiheitsgrade) völlig komplikationslos verläuft, da wir nur noch mit $\overline{D}$ zu tun haben.

(11) (a) Grenze des Verwerfungsbereiches ohne Messfehlerberücksichtigung: $100 + 2.33\frac{10}{\sqrt{100}} = 102.33$. (Die angegebene Hypothese würde also auf dem Niveau $\alpha = 0.01$ verworfen, sobald ein Stichprobenmittel über diesem Wert erschiene.)

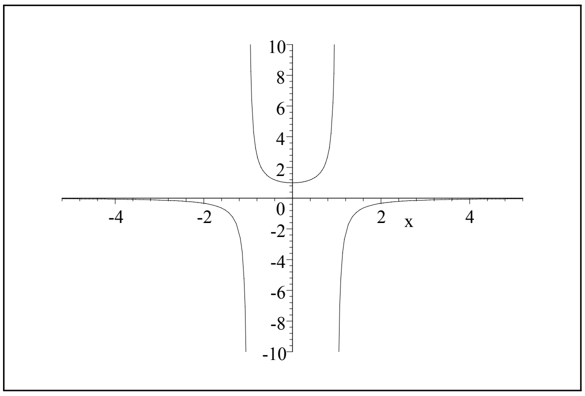
(b) Wird wie angenommen statt X tatsächlich stets $X+2$ gemessen, so läge die korrekte Vertrauensgrenze natürlich bei 104.33, wieder zum Niveau $\alpha = 0.01$. Aber wenn man den Bias der Messung nicht ahnt und bei der Grenze von (a) bleibt, so ergibt sich das tatsächliche Niveau, auf dem man (unbemerkt) arbeitet, durch Lösen von $102 + z = 102.33$, also $z = 0.33$, $\alpha = 1 - \Phi_{0,1}(0.33) = 0.37$. Das wird also abenteuerlich schlecht. Diese Aufgabe sollte Ihnen hinreichend demonstrieren, dass man bei unbemerkt fehlerhafter Messung (dazu gehört vor allem der Fall eines Bias der Stichprobe, und der passiert sehr leicht!) mit noch so guter und korrekter anschließender Statistik nichts mehr retten kann, sondern hoffnungslos verloren ist.

(12) (a) $x^2 - \frac{1}{x^3}$ (Es gibt keine Symmetrie. Bei $x = 0$ dominiert der zweite Term und bewirkt den Pol mit Vorzeichenwechsel, für große $|x|$ dominiert dagegen der erste Summand.) Das erzwingt, wie auch eine gröbere Skizze als diese zeigen sollte, ein lokales Minimum im Bereich $x < 0$.



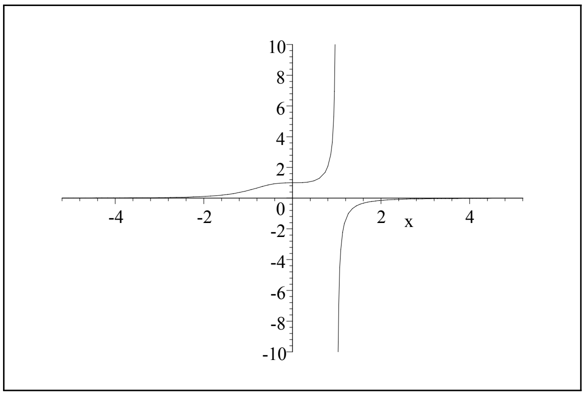
Wir haben $\frac{d}{dx}\left(x^2 - \frac{1}{x^3}\right) = \frac{2x^5+3}{x^4}$, und das wird Null genau an der Stelle $x = \sqrt[5]{-3/2}$. Dort liegt also das erwähnte Minimum, und es gibt also keine weiteren lokalen Extrema.

(b) Zu $\frac{1}{1-x^2}$: Gerade Funktion, Pole bei ± 1, Dominanz des Nenners für große $|x|$. Bereits ohne Ableitung ist zu sehen, dass es bei $x = 0$ ein lokales Minimum gibt, und stark zu vermuten, dass keine weiteren existieren. Die Ableitung bestätigt dies: $\frac{d}{dx}\frac{1}{1-x^2} = \frac{2x}{(1-x^2)^2}$, einzige Nullstelle ist $x = 0$, und dort macht die Ableitung auch den erwarteten Vorzeichenwechsel. Hier ist der Graph (die $x-$ Achse wird nirgends berührt!):

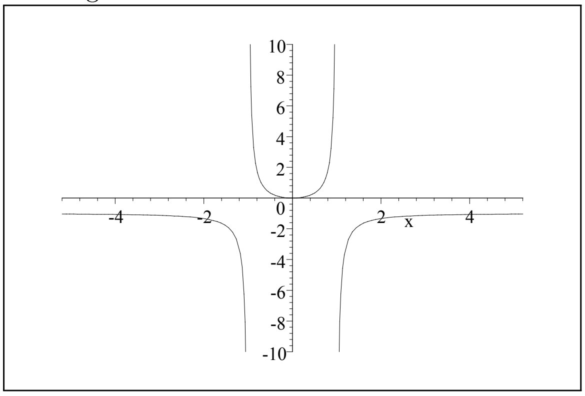


Man kommt zu einer groben Skizze, indem man lediglich die Vorzeichen bei den Polen anschaut, dazu die Vorzeichen der Werte in den von den Polen bestimmten Abschnitten, schließlich das Verhalten für große $|x|$.

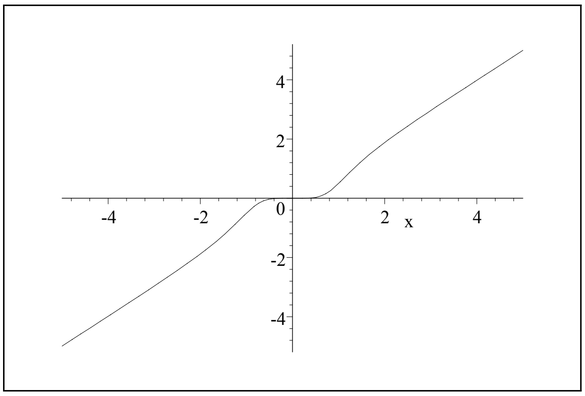
Zu $\frac{1}{1-x^3}$: Keine Symmetrie, Pol bei $x = 1$, links davon stets positives und rechts davon stets negatives Vorzeichen, gegen Null für große $|x|$. Damit ist der Graph schon grob qualitativ klar, allerdings zeigt die erste Ableitung $\frac{d}{dx}\frac{1}{1-x^3} = \frac{3x^2}{(1-x^3)^2}$ eine (einzige) Nullstelle bei $x = 0$, jedoch ohne Vorzeichenwechsel der Ableitung. Es gibt also eine Flachstelle bei $x = 0$ (Sattel). Diese Feinheit wäre uns ohne die Ableitung vermutlich entgangen, sie ist durchaus ein qualitatives Merkmal des Graphen:



Zu $\frac{x^2}{1-x^2}$: Gerade Funktion, Nullstelle bei $x = 0$ (dort offenbar lokales Minimum), Pole bei $x = \pm 1$. Vorzeichenverteilung klar, dazu das Verhalten für große $|x|$: Asymptotische Näherung an die Gerade $y = -1$. Die Ableitung ist kaum nötig, aber wir rechnen sie doch aus und finden mit $\frac{d}{dx}\frac{x^2}{1-x^2} = \frac{2x}{(1-x^2)^2}$, dass es wirklich nur das erwähnte Extremum gibt. Übrigens zeigt die Ableitung auch, dass die Steigung für große $|x|$ gegen Null geht. Hier ist noch das Bild:



Zu $\frac{x^5}{1+x^4}$: Keine Symmetrie, kein Pol (Nenner nirgends Null), einzige Nullstelle bei $x = 0$, asymptotische Annäherung an die Gerade $y = x$ für große $|x|$. Vorzeichenverteilung klar, also findet man ohne weiteres folgende Skizze, wenn man noch überlegt, dass der Nulldurchgang flach verlaufen sollte (Nenner ist um $x = 0$ näherungsweise konstant 1, Zähler x^5, und das hat einen bekannten Sattel), aber die Ableitung zeigt dies spätestens: $\frac{d}{dx}\frac{x^5}{1+x^4} = x^4\frac{5+x^4}{(1+x^4)^2}$, einzige Nullstelle bei $x = 0$, dort kein Vorzeichenwechsel, also Sattel. Keine Extrema.



(13) Wir wissen, was Nullstellen und was Pole bei gebrochen rationalen Funktionen produziert, also werden wir der Einfachheit halber greifen zu $\frac{x+2}{(x-1)(x-2)}$.