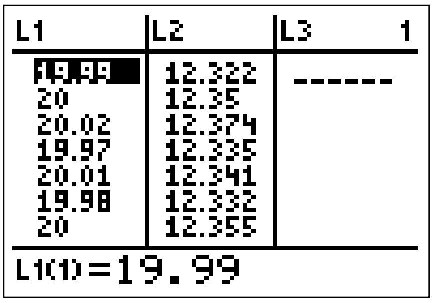
a) **► Gleichung der Regressionsgeraden bestimmen** (10P)

Der Operator „bestimmen“ lässt einen Einsatz des GTR zu. Du kannst die Gleichung der Regressionsgeraden also mit dem GTR berechnen. Gib hierzu zunächst die Daten in den GTR ein und führe anschließend die lineare Regression durch. Achte darauf, dass die Regressionsgerade die Massen in Abhängigkeit von den Längen der Zylinder angeben soll. Wähle also für die Liste der x-Werte die Zeile zur Länge und als Liste für die y-Werte die Zeile zur Masse.

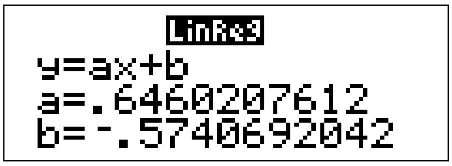
1. Schritt: Daten eingeben

Wechsle mit `STAT` ins Statistik-Menü und wähle `Edit`. Gib hier die Zeile zu „Länge“ in der Liste L_1 und die Zeile zu „Masse“ in der Liste L_2 ein.



2. Schritt: Regression durchführen

Wechsle mit `STAT` erneut ins Statistik-Menü und wähle `CALC`. Hier findest du den Befehl `LinReg(ax+b)`. Wähle ihn aus. Gib im aufgehenden Fenster für XList und YList die Listen L_1 bzw L_2 an und bestätige mit Enter.



Die Regressionsgerade hat die Gleichung $y = 0{,}646x - 0{,}574$.

► Regressionsgerade auswählen und begründen

Es stehen die Geraden g_1 und g_2 als mögliche Regressionsgeraden zur Verfügung, wobei als Grundlage nur die Daten der Zylinder 3, 4 und 5 verwendet wurden. Überlege dir, was die Eigenschaft und die Funktion einer Regressionsgeraden ist:

- Mit einer Regressionsgeraden wird versucht, die **Datenpaare** so gut wie möglich zu **anzunähern**. Es wird also die Gerade gesucht, die von den Punkten, welche die Datenpaare darstellen, **immer möglichst wenig entfernt** ist.
- Die Regressionsgerade unter g_1 und g_2 ist also diejenige, welche die Datenpaare besser annähert.

Du kannst also so vorgehen:

- Berechne die y-Werte der Geraden g_1 und g_2 für die Zylinder 3, 4 und 5.
- Vergleiche die Werte der Geraden mit den echten Werten. „Vergleichen" heißt dabei: Berechne die **Summe der Quadrate der Differenzen**.
- Je kleiner die Summe der Quadrate der Differenzen ist, desto besser die Regressionsgerade.

1. Schritt: y-Werte der beiden Geraden berechnen

Die Geraden geben die Masse der Zylinder in Abhängigkeit von deren Länge an. Die x-Werte sind also in der Zeile zur Länge eingetragen. Setze diese x-Werte der Zylinder 3, 4 und 5 in die Gleichung von g_1 und g_2 ein und berechne die y-Werte:

Gerade g_1:

Zylinder 3 $\quad g_1(20,02) = 0,6 \cdot 20,02 + 0,35 = 12{,}362$

Zylinder 4 $\quad g_1(19,97) = 0,6 \cdot 19,97 + 0,35 = 12{,}332$

Zylinder 5 $\quad g_1(20,01) = 0,6 \cdot 20,01 + 0,35 = 12{,}356$

Gerade g_2:

$g_2(20,02) = 0,8 \cdot 20,02 - 3,65 = 12{,}366$

$g_2(19,97) = 0,8 \cdot 19,97 - 3,65 = 12{,}326$

$g_2(20,01) = 0,8 \cdot 20,01 - 3,65 = 12{,}358$

2. Schritt: Summe der Quadrate der Differenzen berechnen und vergleichen

Du kannst die eben berechneten y-Werte, die tatsächlichen y-Werte und die Quadrate der Differenzen $(g_1(x) - y)^2$ bzw. $(g_2(x) - y)^2$ in einer Tabelle festhalten. Wir führen die Berechnung der Quadrate der Differenzen an zwei Beispielen vor:

Zylinder 3: $x = 20,02, \quad y = 12,374, \quad g_1(20,02) = 12,362, \quad$ und $\quad g_2(20,02) = 12,366.$

Die Quadrate der Differenzen haben immer die Form $(g_1(x) - y)^2$ bzw. $(g_2(x) - y)^2$:

$$(g_1(20,02) - y)^2 = (12,362 - 12,374)^2 = (-0,012)^2 = 0,000144$$
$$(g_2(20,02) - y)^2 = (12,366 - 12,374)^2 = (-0,008)^2 = 0,000064$$

Auf diese Weise kannst du die Tabelle ausfüllen:

	x	y	$(g_1(x) - y)^2$	$(g_2(x) - y)^2$
Zylinder 3	20,02	12,374	0,000144	0,000064
Zylinder 4	19,97	12,335	0,000009	0,000081
Zylinder 5	20,01	12,341	0,000225	0,000289
Summe			0,000378	0,000434

Ein Vergleich zeigt: Die Summe der Quadrate der Differenzen ist bei Gerade g_1 **kleiner** als bei g_2. Also ist die Gerade g_1 die Regressionsgerade.

b) ► **Vertrauensintervall bestimmen** (11P)

Sei X zunächst die Anzahl der Zylinder der Qualitätsstufe I in der Stichprobe. X kann näherungsweise als **binomialverteilte** Zufallsgröße angenommen werden mit $n = 200$ und p unbekannt. Einen ersten Schätzwert für p kannst du über die Angabe ermitteln, dass 80 von 200 Zylindern der Qualitätsstufe I zugeordnet wurden, d.h.:

$$\frac{80}{200} = 0,4.$$

Gesucht ist nun ein Intervall, in dem der tatsächliche Anteil p der Wähler mit einer Wahrscheinlichkeit von 95 % liegt. Einen Ansatz für dieses Problem bieten die σ-Regeln. Diese dürfen angewandt werden, wenn das Laplace-Kriterium $\sigma > 3$ erfüllt ist. Tatsächlich ergibt sich z.B. mit dem Schätzwert $0,4$ für p die Standardabweichung

$$\sigma = \sqrt{200 \cdot 0,4 \cdot (1-0,4)} \approx 6,93 > 3.$$

Selbstverständlich kann dies nur als Näherung gesehen werden. Tendenziell kann aber davon ausgegangen werden, dass die Bedingung $\sigma > 3$ erfüllt ist.

Du kannst also so vorgehen:

- Wähle die σ-Regel, welche eine Aussage über ein 95 %-Konfidenzintervall um den **Erwartungswert** μ macht.
- Bedenke: $\mu = n \cdot p$. Forme den Ausdruck in der σ-Regel also so um, dass er eine Aussage über p macht. Hieraus ergibt sich: $P\left(\left|\frac{X}{n} - p\right| \leq 1,96\sqrt{\frac{p\cdot(1-p)}{n}}\right) \leq 0,95.$
- Löse die Ungleichung nach p auf und berechne so die Grenzen des Intervalls.

1. Schritt: σ-Regel auswählen

Du findest die Regel

$$P(\mu - 1,96\sigma \leq X \leq \mu + 1,96\sigma) \approx 0,95$$

2. Schritt: Ausdruck umformen

Betrachte nur den Ausdruck in Klammern und forme ihn so um, dass er eine Aussage über p macht. Du kennst bereits:

- $n = 200$
- $\sigma = \sqrt{n \cdot p \cdot (1-p)}$
- die relative Häufigkeit $\frac{X}{n} = 0,4$

$$\begin{aligned}
\mu - 1,96\sigma \leq \; & X \leq \mu + 1,96\sigma && \mid \mu = n\cdot p \\
n\cdot p - 1,96\sigma \leq \; & X \leq n\cdot p + 1,96\sigma && \mid \sigma = \sqrt{n\cdot p\cdot(1-p)} \\
n\cdot p - 1,96\cdot\sqrt{n\cdot p\cdot(1-p)} \leq \; & X \leq n\cdot p + 1,96\sqrt{n\cdot p\cdot(1-p)} && \mid :n \\
p - 1,96\cdot\sqrt{\frac{p\cdot(1-p)}{n}} \leq \; & \frac{X}{n} \leq p + 1,96\sqrt{\frac{p\cdot(1-p)}{n}} && \mid -p \\
-1,96\cdot\sqrt{\frac{p\cdot(1-p)}{n}} \leq \; & \frac{X}{n} - p \leq 1,96\sqrt{\frac{p\cdot(1-p)}{n}} && \\
\left|\frac{X}{n} - p\right| \leq \; & 1,96\sqrt{\frac{p\cdot(1-p)}{n}} && \mid \frac{X}{n} = 0,4; \quad n = 200 \\
|0,4 - p| \leq \; & 1,96\sqrt{\frac{p\cdot(1-p)}{200}} &&
\end{aligned}$$

3. Schritt: Ungleichung lösen

Du kannst auf beiden Seiten **quadrieren** und die Ungleichung nach nach p auflösen.

$$\begin{aligned}
|0,4-p| &\leq 1,96\sqrt{\frac{p\cdot(1-p)}{200}} && \mid (\,)^2\\
(0,4-p)^2 &\leq (1,96)^2\cdot\frac{p\cdot(1-p)}{200}\\
0,4^2-2\cdot 0,4\cdot p+p^2 &\leq \frac{3,8416}{200}\cdot p\cdot(1-p)\\
0,4^2-2\cdot 0,4\cdot p+p^2 &\leq \frac{3,8416}{200}p-\frac{3,8416}{200}p^2 && \mid -\frac{3,8416}{200}p+\frac{3,8416}{200}p^2
\end{aligned}$$

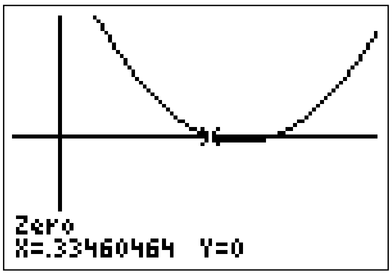
$$p^2+\frac{3,8416}{200}p^2-2\cdot 0,4\cdot p-\frac{3,8416}{200}p+0,4^2\leq 0$$

Fasse den Ausdruck links vom Gleichheitszeichen als Funktionsterm $f(p)$ einer Funktion f auf. Der Graph von f ist eine nach oben geöffnete Parabel.

Gesucht ist der Bereich, in welchem f **negative** Funktionswerte annimmt, d.h. der Bereich, in dem die Parabel unterhalb der x-Achse verläuft. Du kannst diese Ungleichung grafisch lösen:

Zeichne den Graphen von f und berechne mit `2nd → TRACE (CALC) → Zero` die Nullstellen von f. Sie sind die Grenzen deines Intervalls.

Der GTR liefert die Werte $p_1 = 0,3346$ und $p_2 = 0,4692$.



Damit folgt: der tatsächliche Anteil p der Wähler liegt mit einer Wahrscheinlichkeit von 95 % im Intervall $[0,3346\,;\,0,4692]$.

▶ Aussagen begründet entscheiden

Betrachte deinen Ansatz, den du zur Berechnung der Intervallgrenzen verwendet hast:

$$|0,4-p|\leq 1,96\sqrt{\frac{p\cdot(1-p)}{200}}$$

Überlege, an welcher Stelle der Stichprobenumfang n und an welcher Stelle die Sicherheitswahrscheinlichkeit auftreten und betrachte, welchen Einfluss sie auf die Länge des Intervalls haben. Dabei ist wichtig: der Ausdruck $1,96\sqrt{\frac{p\cdot(1-p)}{200}}$ ist der **Radius** des Intervalls. Je größer der Radius, desto größer ist auch die Länge.

Aussage 1: Vergrößerter Stichprobenumfang

Der Stichprobenumfang tritt im **Nenner im Bruch unter der Wurzel** auf und ist im Moment mit der Zahl 200 besetzt. Je größer dieser Wert wird, desto kleiner wird der Bruch an sich. Damit wird auch die gesamte Wurzel kleiner und folglich auch der gesamte Ausdruck auf der rechten Seite der Ungleichung.

Damit folgt: mit wachsendem Stichprobenumfang **verkleinert** sich der Radius des Vertrauensintervalls. Also wird auch die **Länge** des Intervalls kleiner.

Die Aussage ist deshalb **falsch.**

Aussage 2: Größere Sicherheitswahrscheinlichkeit

In unserem Fall liegt die Sicherheitswahrscheinlichkeit bei 95 %. Aufgrund dieses Wertes hast du in der Anlage für die Umgebung $1,96\sigma$ entschieden.

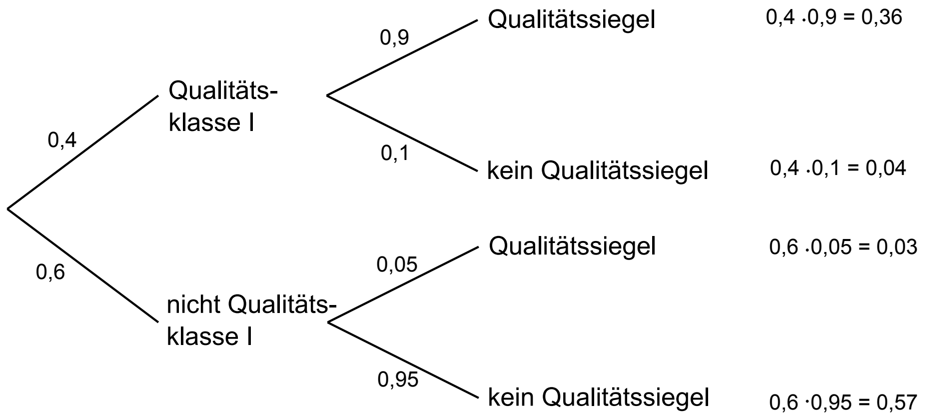
Du siehst, dass die Faktoren (1,96; 2;2; 2,58;...) größer werden mit wachsender Sicherheitswahrscheinlichkeit. Betrachte nun wieder den Radius des Vertrauensintervalls:

Es wird nur der Faktor 1,96 verändert und zwar wird ersetzt durch eine größere Zahl. Dadurch wird die gesamte rechte Seite der Ungleichung **größer**. Da der Radius des Vertrauensintervalls größer wird, wächst auch die Länge des Intervalls. Also ist die Aussage wahr.

c) ► **Wahrscheinlichkeit für richtige Entscheidung berechnen** (9P)

Du kannst die Situation in einem Baumdiagramm grafisch darstellen und dir so einen besseren Überblick über die Zusammenhänge verschaffen. Wir wollen dabei auf der ersten Stufe die Unterscheidung „Qualitätsklasse I – nicht Qualitätsklasse I" machen und auf der zweiten Stufe die Unterscheidung „Qualitätssiegel – kein Qualitätssiegel".

Auf der ersten Stufe ist dabei die Wahrscheinlichkeit für „Qualitätsklasse I" genau 0,4. Also ist ein Zylinder mit einer Wahrscheinlichkeit von 0,6 nicht Qualitätsklasse I.
Auf der zweiten Stufe hat ein Zylinder der Qualitätsklasse I mit einer Wahrscheinlichkeit von 90 % ein Qualitätssiegel . Also hat ein Zylinder der Qualitätsstufe I mit einer Wahrscheinlichkeit von 10 % das Qualitätssiegel nicht.
Zugleich hat ein Zylinder, der nicht Qualitätsklasse I ist, mit Wahrscheinlichkeit 0,05 ein Qualitätssiegel; d.h. mit Wahrscheinlichkeit 0,95 hat solch ein Zylinder kein Qualitätssiegel.



Berechne nun die Wahrscheinlichkeit dafür, dass das Prüfgerät eine richtige Entscheidung trifft. Dies ist dann der Fall, wenn es einem Zylinder der Qualitätsklasse I ein Qualitätssiegel gibt **oder** wenn es einem Zylinder von geringerer Qualität kein Qualitätssiegel gibt.

$$\begin{aligned} P(\text{„richtige Entscheidung"}) &= P(\text{Quali-Klasse I und Quali-Siegel } \textbf{oder} \text{ nicht Quali-Klasse I und kein Quali-Siegel}) \\ &= P(\text{Quali-Klasse I und Quali-Siegel}) + P(\text{nicht Quali-Klasse I und kein Quali-Siegel}) \\ &= 0,36 + 0,57 = 0,93 \end{aligned}$$

Das Prüfgerät trifft mit einer Wahrscheinlichkeit von 93 % eine richtige Entscheidung.

► Wahrscheinlichkeit für Zylinder der Klasse I berechnen

Nun wird der Fall betrachtet, dass ein Zylinder mit einem Qualitätssiegel versehen wurde. Du sollst die Wahrscheinlichkeit dafür berechnen, dass der Zylinder **dann** zur Qualitätsklasse I gehört. Gesucht ist also die Wahrscheinlichkeit dafür, dass der Zylinder zur Qualitätsklasse I gehört, **unter der Bedingung dass** er mit einem Siegel versehen wurde.

Es handelt sich hierbei also um eine bedingte Wahrscheinlichkeit. Für diese gilt:
$P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$.

In unserem Fall ist das Ereignis A: „Ein Zylinder ist Qualitätsklasse I“ und das Ereignis B ist „Ein Zylinder hat ein Siegel“.
Das zweite Ereignis setzt sich dabei aus zwei Ereignissen zusammen:

- Ein Zylinder ist Qualitätsklasse I und hat ein Siegel.
- Ein Zylinder ist nicht Qualitätsklasse I und hat ein Siegel.

Die benötigten Wahrscheinlichkeiten findest du im Baumdiagramm. Einsetzen liefert:

$$\begin{aligned} P(\text{Zylinder ist Klasse I} \mid \text{Zylinder hat Siegel}) &= \frac{P(\text{Zyliner ist Klasse I} \cap \text{Zylinder hat Siegel})}{P(\text{Zylinder hat Siegel})} \\ &= \frac{P(\text{Zylinder ist Klasse I und hat Siegel})}{P(\text{Zylinder hat Siegel})} \\ &= \frac{0,36}{0,36+0,03} = \frac{0,36}{0,39} \approx 0,9231 \end{aligned}$$

Ein Zylinder, der mit einem Siegel versehen wurde, gehört mit einer Wahrscheinlichkeit von etwa 92,31 % zur Qualitätsklasse I.