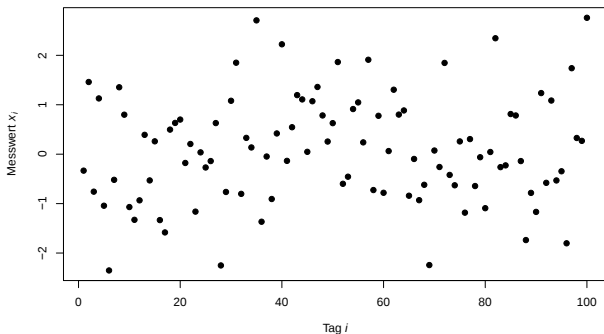


Problematisch:
„Aufhören, wenn es gut aussieht“

Ein simuliertes Experiment

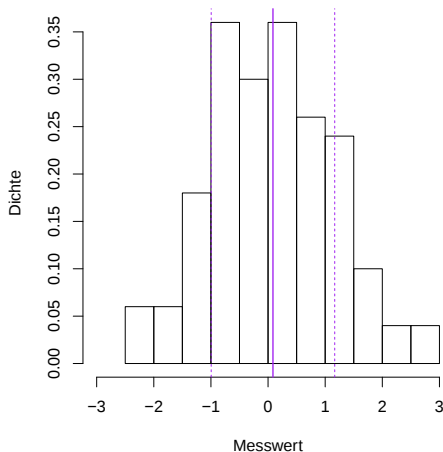
Ein Versuch werde an $n = 100$ aufeinanderfolgenden Tagen unabhängig unter identischen Bedingungen wiederholt,

x_i = Messergebnis am i -ten Tag



(unter der Nullhypothese $\mu = 0$ simulierte Daten, d.h. es gibt in Wirklichkeit keinen Effekt)

Ein simuliertes Experiment



$n = 100$, $\bar{x} = 0.086$, $s/\sqrt{n} = 0.108$, $t = 0.794$, p -Wert ist 0.43
(zweiseitiger t -Test)

„Aufhören, wenn es gut aussieht“

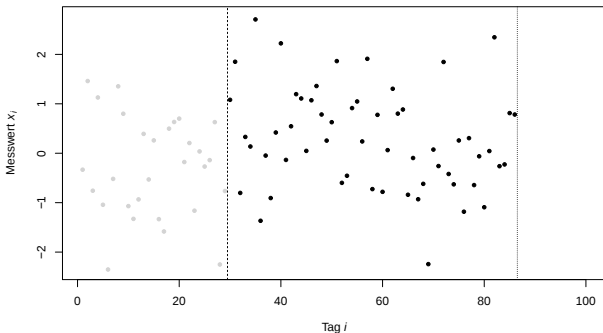
Der Experimentator überlegt am Tag 86:

Der Monat erste war noch eine Übungs- und Kalibrierungsphase,

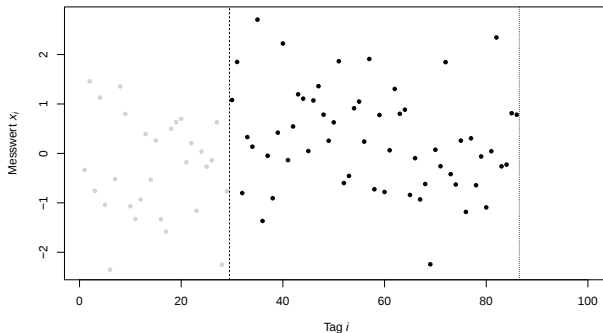
ich lasse einmal die ersten 29 Beobachtungen weg und schaue, was ich dann bis jetzt so habe (57 Beobachtungen)

„Aufhören, wenn es gut aussieht“

Der Experimentator überlegt am Tag 86:
Der Monat erste war noch eine Übungs- und
Kalibrierungsphase,
ich lasse einmal die ersten 29 Beobachtungen weg und
schaue, was ich dann bis jetzt so habe (57 Beobachtungen)

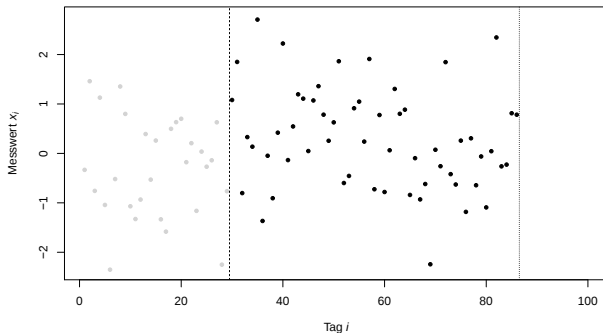


„Aufhören, wenn es gut aussieht“



Mit den $n = 57$ Messwerten $x_{30}, x_{31}, \dots, x_{85}, x_{86}$ ergibt sich $\bar{x} = 0.293$, $s = 1.021$, $s/\sqrt{n} = 0.135$, $t = 2.167$, p -Wert ist 0.035 (zweiseitiger t -Test)

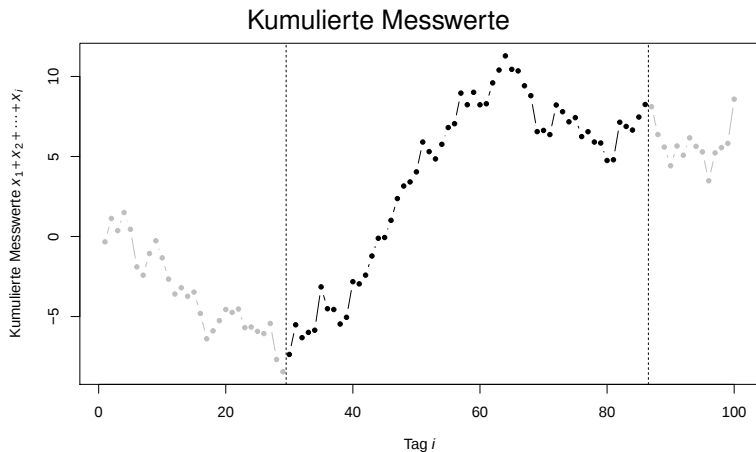
„Aufhören, wenn es gut aussieht“



Mit den $n = 57$ Messwerten $x_{30}, x_{31}, \dots, x_{85}, x_{86}$ ergibt sich $\bar{x} = 0.293$, $s = 1.021$, $s/\sqrt{n} = 0.135$, $t = 2.167$, p -Wert ist 0.035 (zweiseitiger t -Test)

Demnach: Wir sehen scheinbar eine signifikante Abweichung von der 0?
Was ist hier passiert?

„Aufhören, wenn es gut aussieht?!“



Problem des multiplen Testens

Wenn wir den Beginn und die Länge der „richtigen“ Versuchsreihe nicht im vorhinein festlegen, haben wir ein multiples Testproblem vorliegen:

Problem des multiplen Testens

Wenn wir den Beginn und die Länge der „richtigen“ Versuchsreihe nicht im vorhinein festlegen, haben wir ein multiples Testproblem vorliegen:

Angenommen, an jedem Tag $i = 50, 51, \dots, 100$ geht der Experimentator die $i - 50 + 1$ möglichen Messreihen

$$x_1, x_2, \dots, x_{i-1}, x_i$$

$$x_2, x_3, \dots, x_{i-1}, x_i$$

$$\vdots$$

$$x_{i-50+1}, x_{i-50+2}, \dots, x_{i-1}, x_i$$

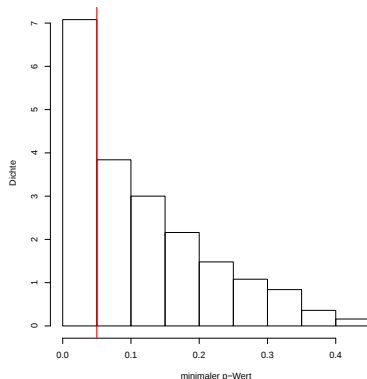
der Länge ≥ 50 , die mit dem heutigen Tag enden, durch und führt mit jeder davon einen (zweiseitigen ein-Stichproben) t -Test zur Nullhypothese $\mu = 0$ aus.

Problem des multiplen Testens

Dann wurden insgesamt $1 + 2 + \dots + 51 = \frac{51 \cdot 52}{2} = 1326$ Tests ausgeführt. Wie wahrscheinlich ist es, dass mindestens einer einen p -Wert < 0.05 liefert?

Problem des multiplen Testens

Dann wurden insgesamt $1 + 2 + \dots + 51 = \frac{51 \cdot 52}{2} = 1326$ Tests ausgeführt. Wie wahrscheinlich ist es, dass mindestens einer einen p -Wert < 0.05 liefert?



500 simulierte Versuchsreihen

W'keit, dass mindestens einer der Tests anschlägt ≈ 0.35