

Ein echter Fall

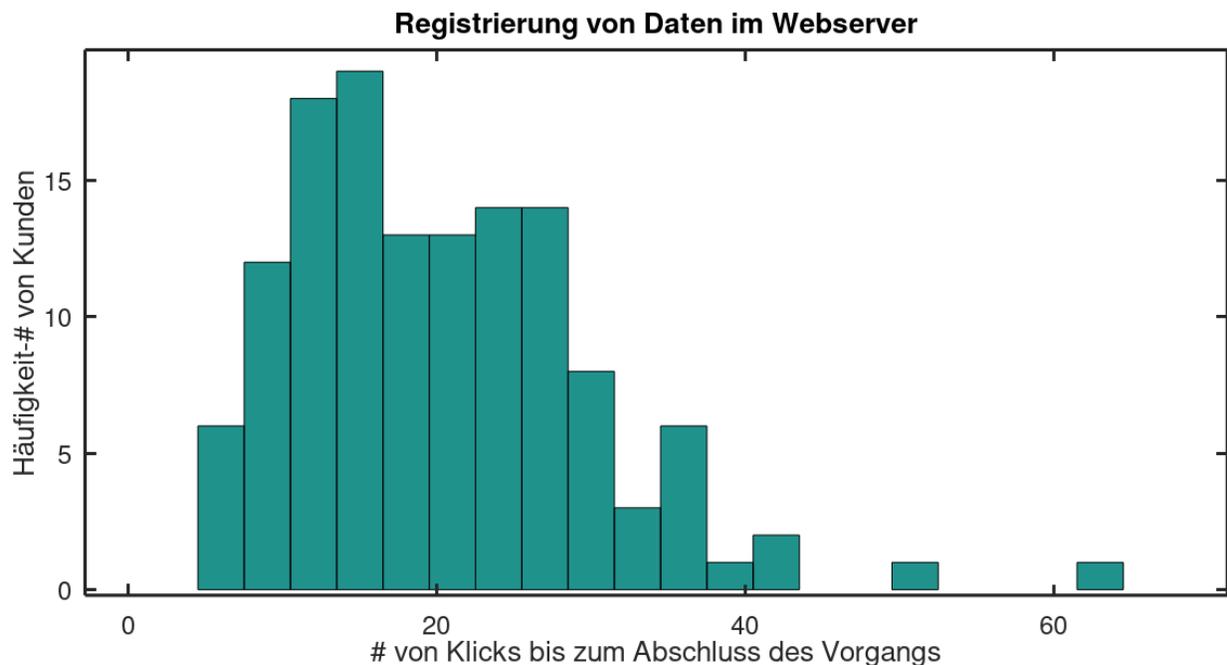
Zusammenhang: Webserver registriert die Daten von Kunden, einige haben den Prozess bereits beendet, andere haben noch fehlende Daten.

Ziel: Wählen die Liste der Kunden aus, um sie daran zu erinnern, den Prozess der Datenregistrierung abzuschließen.

Nummer von Kunden: mehr als 150

Eingangsdaten: Anzahl der Klicks, die ein Benutzer zum Speichern von Daten und zum erfolgreichen Abschluss des Prozesses gemacht hat. (131 Kunden = 131 Daten)

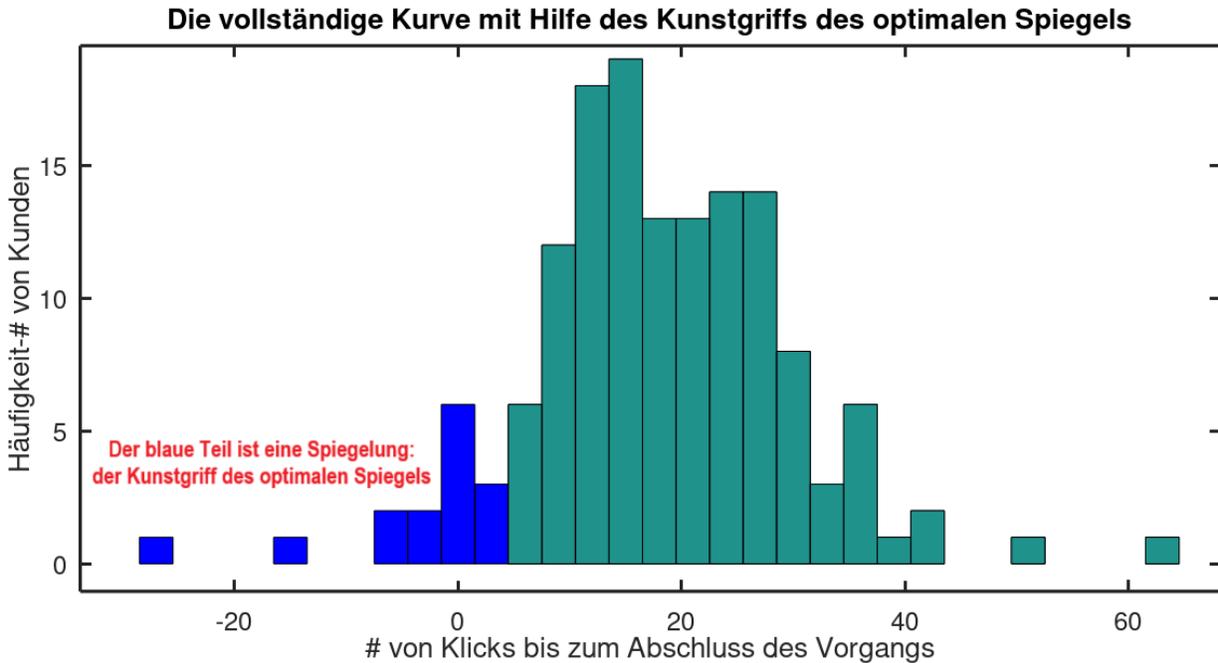
Das Histogramm der Daten:



Es ist zu erkennen, dass die Daten für Werte unter 5 abgeschnitten werden, was sinnvoll ist, da es keine negative Anzahl von Klicks geben kann und es eine minimale Anzahl von Klicks gibt, um den Prozess zu beenden.

Die Lösung

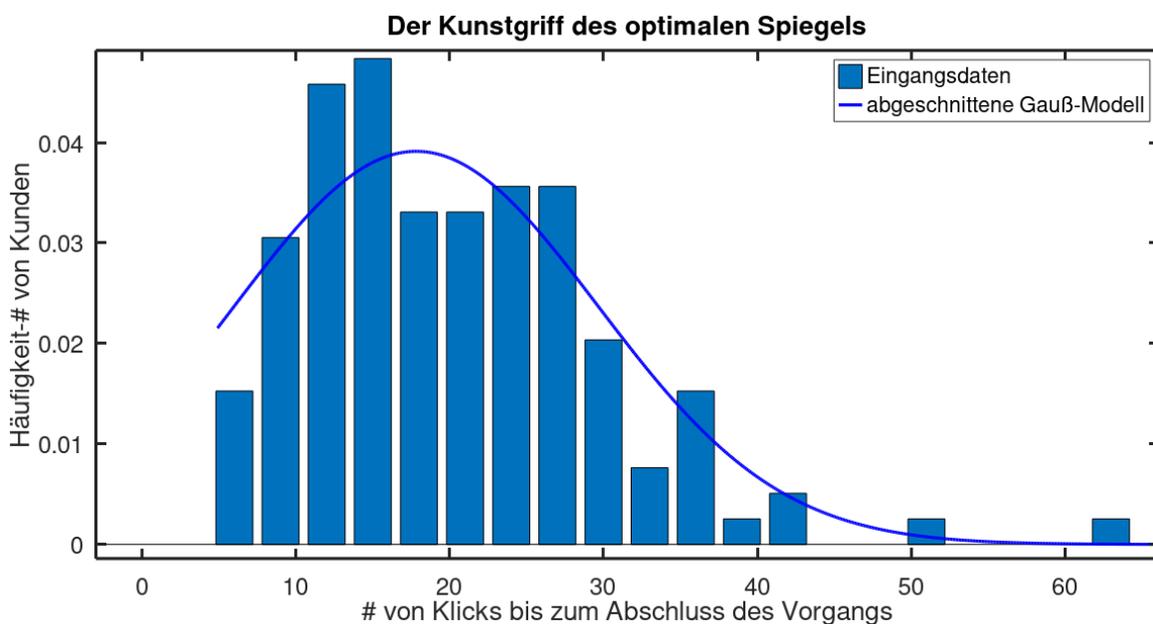
Der Kunstgriff des optimalen Spiegels [1] kann angewendet werden, um die abgeschnittene Kurve (die Daten) zu vervollständigen:



Diese können berechnet werden:

$\mu = 17.82$ (Mittelwert) $\sigma = 11.82$ (Standardabweichung)

Das abgeschnittene Gauß-Modell kann gezeichnet werden:

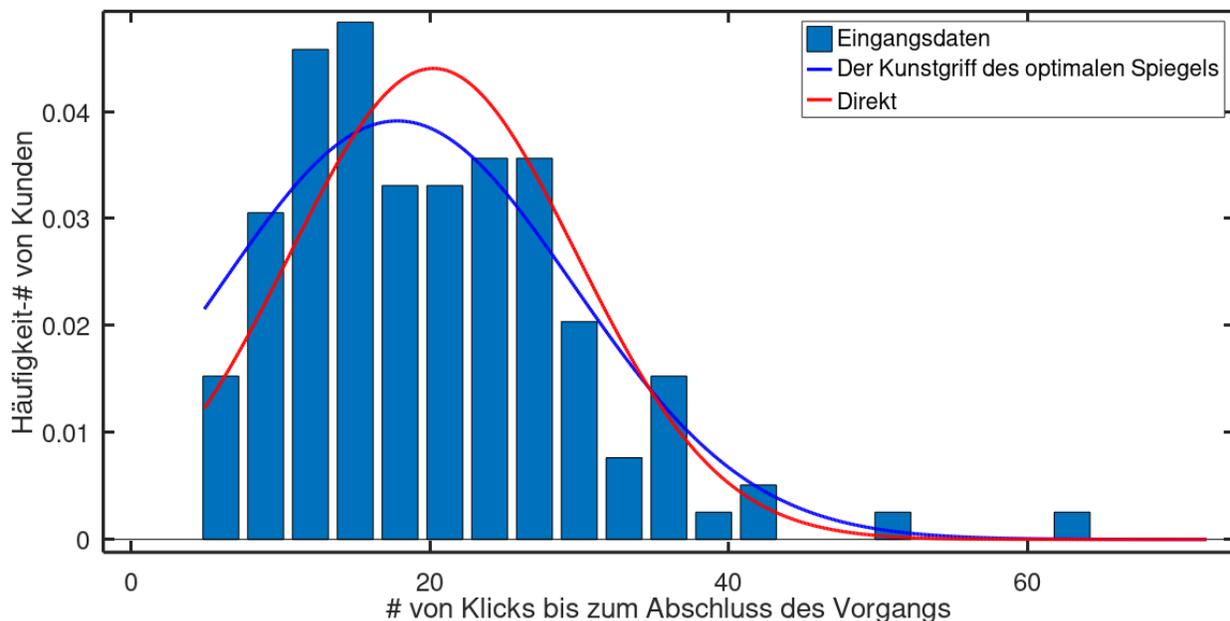


Bei dem bereits definierten Modell können viele Kriterien gewählt werden. Ein Kriterium kann sein, die Liste der Kunden auszuwählen, die weniger Klicks als der Durchschnitt (17,82) haben, und ihnen eine Nachricht zu schicken, um sie aufzufordern, den Prozess bald abzuschließen.

Was würde passieren, wenn der Kunstgriff des optimalen Spiegels nicht verwendet wird und stattdessen der direkte Wert der Parameter der Eingangsdaten verwendet wird?

Die direkten Parameter der Eingangsdaten sind:

$\mu=20.23$ (Mittelwert) $\sigma=9.59$ (Standardabweichung)



Der Fehler für den Mittelwert = $\frac{20.23-17.82}{17.82} * 100\% = 13.52\%$

Der Fehler für die Standardabweichung = $\frac{11.82-9.59}{11.82} * 100\% = 18.87\%$

Die Fehler sind signifikant. Die direkte Standardabweichung ist kleiner als die Standardabweichung des Kunstgriffs des optimalen Spiegels. Andererseits ist der direkte Mittelwert höher als der Mittelwert aus dem Kunstgriff des optimalen Spiegels.

Hinweis: Die Berechnungen wurden in Excel durchgeführt, die Grafiken wurden in Octave erstellt.

Referenzen

- [1] C. San Roman. Das praktische Handbuch der abgeschnittenen Verteilung. 2021.
https://sudamericaciencia.org/Handbuch_abgesch.html

