

Statistik 5

Übung 2

Einleitung und Ziele der Übung

Herzlich Willkommen zur Übung Statistik 5. Diese zweite Übung wendet die Funktionen aus der ersten Übung an, um Stichproben zu simulieren und deren Eigenschaften zu untersuchen. Sie werden `replicate()` benötigen aber keine der Schleifen oder Bedingungen

Agenda

- Befehle letzte Woche mündlich Wiederholen
- Zentraler Grenzwertsatz
- Populationsschätzungen und Standardfehler

1. Zentraler Grenzwertsatz - Einführung

In der Vorlesung haben wir behandelt, dass wenn wir Zufallszahlen addieren, die Verteilung der aufsummierten Zahlen mit steigender Anzahl eine Normalverteilung annimmt, selbst wenn die Variablen ursprünglich nicht normalverteilt sind, solange die Zufallsvariablen ähnliche Eigenschaften haben und unabhängig sind. Wir wollen das am Beispiel des Würfels zeigen.

1. Zentraler Grenzwertsatz - Übung 1

Führen Sie folgende Aufgaben der Reihe nach durch. Hinweis: Die erwähnten Barplots finden Sie auf Übungsseite 4. Überprüfen Sie Ihr Ergebnis aber erst am Ende.

- Überlegen Sie sich welche Ereignisse beim einmaligen Werfen eines (idealen) Würfels auftreten können. Wie gross ist die Wahrscheinlichkeit für jedes Ereignis?
- Simulieren Sie 10000 Stichproben eines einzelnen Wurfs eines idealen Würfels.
 - Speichern Sie das Ergebnis in einer Variable.
 - Ermitteln Sie die Häufigkeiten der Einzelereignisse mittels `table()` und der Variable als Argument.
 - Speichern Sie das Ergebnis in einer neuen Variable.
 - Ermitteln Sie die relativen Häufigkeiten, indem Sie die neue Variable als Argument des Befehls `prop.table()` nutzen.
 - Entsprechen die beobachteten relativen Häufigkeiten Ihren Erwartungen zur ersten Frage?
 - Speichern Sie auch dieses Ergebnis ab und visualisieren Sie es mittels `barplot()`.
- Fahren Sie auf der nächsten Seite fort.

1. Zentraler Grenzwertsatz - Übung 2

Führen Sie folgende Aufgaben der Reihe nach durch.

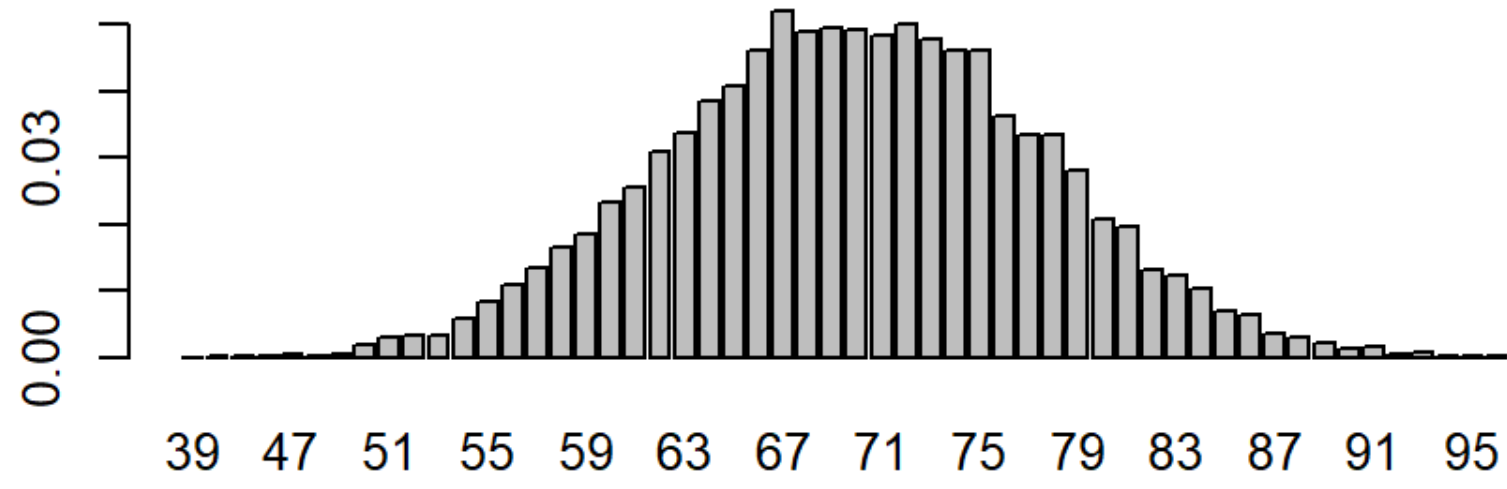
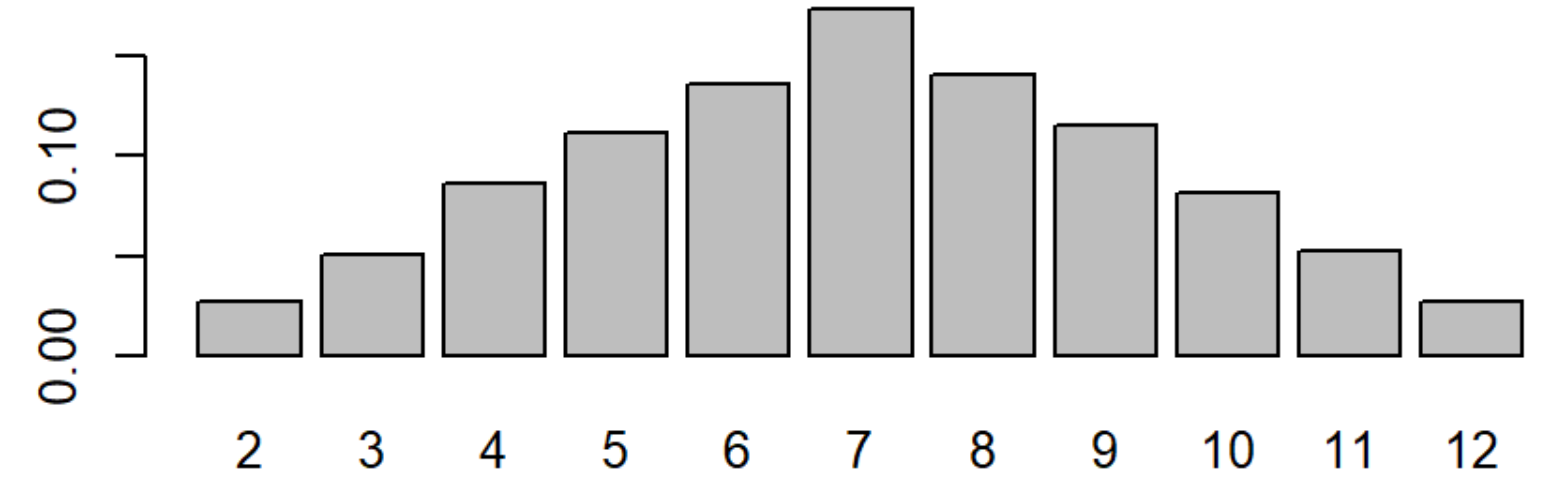
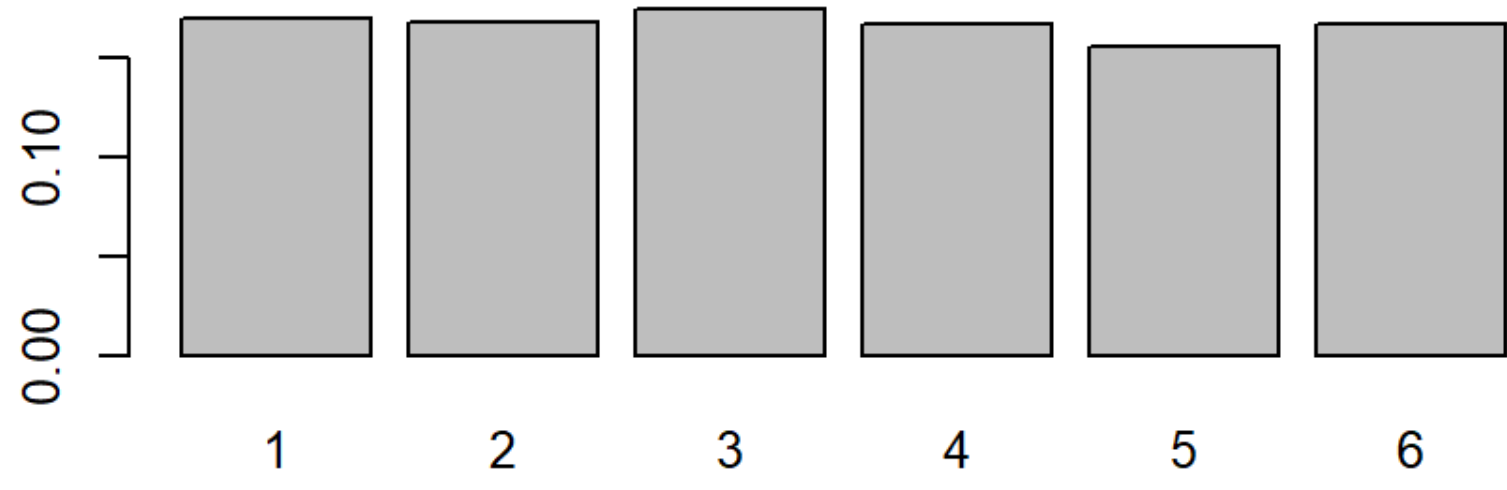
- Überlegen Sie sich welche Ereignisse beim zweimaligen Werfen eines (idealen) Würfels und dem Addieren der Werte auftreten können. Wie gross ist die Wahrscheinlichkeit für jedes Ereignis?
- Simulieren Sie 10000 Stichproben des Werfens 2er idealer Würfel
 - Bilden Sie für jede Stichprobe die Summe der beiden Würfel und speichern Sie das Ergebnis in einer neuen Variable.(siehe Vorlesung)
 - Ab jetzt gehen Sie exakt so vor, wie bei der vorherigen Übung
 - Ermitteln Sie die Häufigkeiten der Einzelereignisse mittels `table()` und der Variable als Argument.
 - Speichern Sie das Ergebnis in einer neuen Variable.
 - Ermitteln Sie die relativen Häufigkeiten, indem Sie die neue Variable als Argument des Befehls `prop.table()` nutzen.
 - Entsprechen die beobachteten relativen Häufigkeiten Ihren Erwartungen zur ersten Frage?
 - Speichern Sie auch dieses Ergebnis ab und visualisieren Sie es mittels `barplot()`.
 - Vergleichen Sie Ihre beiden Diagramme. Was stellen Sie fest?
- Fahren Sie auf der nächsten Seite fort.

1. Zentraler Grenzwertsatz - Übung 3

Führen Sie folgende Aufgaben der Reihe nach durch.

- Jetzt wollen wir das Ganze ein letztes Mal wiederholen, allerdings mit 20 Würfeln je Stichprobe.
- Simulieren Sie 10000 Stichproben des Werfens von 20 idealen Würfeln
 - Bilden Sie für jede Stichprobe die Summe der beiden Würfel und speichern Sie das Ergebnis in einer neuen Variable.(siehe Vorlesung)
 - Ab jetzt gehen Sie exakt so vor, wie bei der vorherigen Übung
 - Ermitteln Sie die Häufigkeiten der Einzelereignisse mittels `table()` und der Variable als Argument.
 - Speichern Sie das Ergebnis in einer neuen Variable.
 - Ermitteln Sie die relativen Häufigkeiten, indem Sie die neue Variable als Argument des Befehls `prop.table()` nutzen.
 - Entsprechen die beobachteten relativen Häufigkeiten Ihren Erwartungen zur ersten Frage?
 - Speichern Sie auch dieses Ergebnis ab und visualisieren Sie es mittels `barplot()`.
 - Vergleichen Sie Ihre drei Diagramme. Was stellen Sie fest?
- Fahren Sie auf der nächsten Seite fort.

1. Zentraler Grenzwertsatz - Übung 4



Populationsschätzungen und Standardfehler - Einführung

Für die Normalverteilung gelten folgende Gleichungen:

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$$\hat{\sigma}_{\tilde{x}} = \sqrt{\frac{\pi}{2}} * \frac{s}{\sqrt{n}}$$

$$\hat{\sigma}_{s^2} = s^2 * \sqrt{\frac{2}{n-1}}$$

Diese wollen wir uns in den nächsten Übungsaufgaben genauer anschauen.

Populationsschätzungen und Standardfehler - Übung 1

Ziehen Sie für jeden der folgenden Populationsparameter 5000 Stichproben aus einer Normalverteilung. Beachten Sie den Stichprobenumfang.

Tipp: `rnorm()` hat als Argument die Standardabweichung, nicht die Varianz!

- $n = 10, \mu = 100, \sigma^2 = 100$
- $n = 6, \mu = 100, \sigma^2 = 100$
- $n = 30, \mu = 100, \sigma^2 = 225$

Ermitteln (und speichern) Sie für jede der jeweils 5000 Stichproben

- den Mittelwert
- die Varianz
- den Median
- das Maximum

Fahren Sie mit der nächsten Seite fort

Populationsschätzungen und Standardfehler - Übung 2

Wir wollen jetzt zunächst die simulierten Stichprobenkennwerteverteilungen ausgeben. Dafür benutzen wir `hist()` mit drei Argumenten

- die Kennwerte der jeweiligen Stichprobe die wir zuvor berechnet haben
- `breaks = 50` (Damit wir 51 Balken haben)
- `xlim = c(Start, Stop)` (damit die Diagramme für den gleichen Parameter zwischen den simulierten Verteilungen und Stichproben besser vergleichbar sind)
 - für die Mittelwerte: `xlim = c(80, 120)`
 - für die Varianzen: `xlim = c(0, 700)`
 - für die Mediane: `xlim = c(80, 120)`
 - für die Maxima: `xlim = c(80, 160)`
- z.B. für die Mittelwerte der ersten Normalverteilung `hist(M1,breaks=50,xlim= c(80, 120))`

Sobald Sie alle Histogramme erfolgreich dargestellt haben, können Sie auf der nächsten Seite fortfahren.

Populationsschätzungen und Standardfehler - Übung 3

Schauen Sie die Stichprobenkennwerteverteilungen für die Mittelwerte an.

Welche Verteilungsform haben die Stichprobenmittelwerte?

Für welche Verteilung streuen die Mittelwerte am stärksten, für welche am wenigsten stark?

Prüfen Sie Ihre Aussage, indem Sie für alle drei Verteilungen den Standardfehler des Mittelwerts berechnen. Nehmen Sie dazu die “wahren” Werte der Population an.

- $\hat{\sigma}_{\bar{x}_{Verteilung1}} =$
- $\hat{\sigma}_{\bar{x}_{Verteilung2}} =$
- $\hat{\sigma}_{\bar{x}_{Verteilung3}} =$

Wenn Sie auf die Gleichungen schauen, wovon ist der Standardfehler des Mittelwerts abhängig? Wodurch nimmt er ab, wodurch zu?

Bestimmen Sie nun die für jede der Stichprobenmittelwertverteilungen die Standardabweichung. Was stellen Sie fest?

Fahren Sie auf der nächsten Seite fort!

Populationsschätzungen und Standardfehler - Übung 4

Schauen Sie die Stichprobenkennwerteverteilungen für die Varianzen an.

Welche Verteilungsform haben die Varianzen am ehesten?

Für welche Verteilung streuen die Varianzen am stärksten, für welche am wenigsten stark?

Prüfen Sie Ihre Aussage, indem Sie für alle drei Verteilungen den Standardfehler der Varianz berechnen.

Nehmen Sie dazu die “wahren” Werte der Population an.

Bestimmen Sie nun die für jede der Stichprobenvarianzverteilungen die Standardabweichung. Was stellen Sie fest?

- $\hat{\sigma}_{s^2_{Verteilung1}} =$
- $\hat{\sigma}_{s^2_{Verteilung2}} =$
- $\hat{\sigma}_{s^2_{Verteilung3}} =$

Wenn Sie auf die Gleichungen schauen, wovon ist der Standardfehler der Varianz abhängig? Wodurch nimmt er ab, wodurch zu?

Fahren Sie auf der nächsten Seite fort!

Populationsschätzungen und Standardfehler - Übung 5

Schauen Sie die Stichprobenkennwerteverteilungen für den Median an.

Welche Verteilungsform haben die Stichprobenmediane?

Für welche Verteilung streuen die Mediane am stärksten, für welche am wenigsten stark?

Prüfen Sie Ihre Aussage, indem Sie für alle drei Verteilungen den Standardfehler des Medians berechnen. Nehmen Sie dazu die “wahren” Werte der Population an.

- $\hat{\sigma}_{\tilde{x}_{Verteilung1}} =$
- $\hat{\sigma}_{\tilde{x}_{Verteilung2}} =$
- $\hat{\sigma}_{\tilde{x}_{Verteilung3}} =$

Wenn Sie auf die Gleichungen schauen, wovon ist der Standardfehler des Medians abhängig? Wodurch nimmt er ab, wodurch zu?

Bestimmen Sie nun die für jede der Stichprobemedianverteilungen die Standardabweichung. Was stellen Sie fest?

Vergleichen Sie die Standardfehler des Medians mit den zugehörigen Standardfehlern des Mittelwerts. Was stellen Sie fest?

Fahren Sie auf der nächsten Seite fort!

Populationsschätzungen und Standardfehler - Übung 6

Schauen Sie die Stichprobenkennwerteverteilungen für die Maxima an.

Die Verteilung der Maxima ist etwas kompliziert und entspricht keiner uns bekannten Verteilung. Können Sie jedoch beurteilen, ob die Verteilung schief ist? Wenn ja, in welche Richtung und woher kommt dies?

Für welche Verteilung streuen die Maxima am stärksten, für welche am wenigsten stark?

Bestimmen Sie nun die für jede der Stichprobenmaximaverteilungen die Standardabweichung. Entspricht dies Ihrer obigen Aussage?

Ohne die Gleichung des Standardfehler des Medians zu kennen, wovon ist der Standardfehler des Maximums abhängig? Wodurch nimmt er ab, wodurch zu?

Sie haben die Übung nun abgeschlossen.