

Statistik 5

Übung 3

Einleitung und Ziele der Übung

Herzlich Willkommen zur Übung Statistik 5. Diese Übung wendet die Funktionen aus den ersten zwei Übungen an um Konfidenzintervalle anzuschauen. Auch die for-Schleife werden wir wieder nutzen. Wir bleiben dabei zunächst bei der Normalverteilung und χ^2 -Verteilung, auch wenn Konfidenzintervalle basierend auf t -Werten bessere Schätzungen liefern könnten.

Agenda

- Konfidenzintervall Mittelwert
- Konfidenzintervall Varianz

1. Konfidenzintervall Mittelwert - Einführung

In der Vorlesung haben wir uns bereits die Idee des Konfidenzintervalls angeschaut, jetzt wollen wir uns diese auf Basis von Simulationen nochmals anschauen.

Grundsätzlich sind Konfidenzintervalle definiert über den zugehörigen Standardfehler. Für den Mittelwert ist dies (falls die Populationsvarianz bekannt ist):

$$\bar{x} \pm z_{1-\frac{\alpha}{2}} * \sigma_{\bar{x}}$$

Das Konfidenzintervall hat dann die Grenzen

$$\left[\bar{x} - z_{1-\frac{\alpha}{2}} * \sigma_{\bar{x}}, \bar{x} + z_{1-\frac{\alpha}{2}} * \sigma_{\bar{x}} \right]$$

Wenn nicht, sollten wir die t -Verteilung verwenden (nächste/übernächste Woche).

$z_{1-\frac{\alpha}{2}}$ ist der Wert der Normalverteilung (z) für eine bestimmte Konfidenz ($1 - \frac{\text{Signifikanz oder } \alpha\text{-Fehler}}{2}$). Wir müssen α wieder halbieren, da wir ein zweiseitiges Konfidenzintervall berechnen wollen. Bei einem einseitigen Konfidenzintervall sind die Grenzen dagegen: $[-\infty, \bar{x} + z_{1-\alpha} * \sigma_{\bar{x}}]$ oder $[\bar{x} - z_{1-\alpha} * \sigma_{\bar{x}}, +\infty]$

1. Konfidenzintervall Mittelwert - Übung 1

Ziehen Sie für jeden der folgenden Populationsparameter 20 Stichproben aus einer Normalverteilung.

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

Bestimmen Sie mittel jeweils die 5000 Mittelwert je Verteilung und weisen Sie das Ergebnis wie folgt einem Datensatz zu (ersetzen Sie die Angaben in BLOCKSCHRIFT durch Ihre eigenen Variablen) und die Nummer des Datensatz durch die Verteilung, für den Sie den Datensatz erstellen.

```
1 Stichproben <- 30
2 Datensatz1  <- data.frame(Stichprobennummer = 1:Stichproben,
3                           M_SP = MITTELWERTE_1._VERTEILUNG)
```

Wie unterscheiden sich die Standardfehler und warum?

1. Konfidenzintervall Mittelwert - Übung 2

Bestimmen Sie nun die Standardfehler des Mittelwerts $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ für die drei Verteilungen:

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

Bestimmen Sie jetzt jeweils $z_{1-\frac{\alpha}{2}} * \sigma_{\bar{x}}$ für ein 90%-Konfidenzintervall, ein 95%Konfidenzintervall, ein 99%-Konfidenzintervall und ein 100% Konfidenzintervall. Vergessen Sie nicht, dass Sie den zugehörigen z-Wert mittels qnorm() bestimmen können.

Konfidenz	Verteilung 1	Verteilung 2	Verteilung 3
90%			
95%			
99%			

Wann ist das Konfidenzintervall am grössten, wann am kleinsten?

1. Konfidenzintervall Mittelwert - Übung 3

Sie werden jetzt insgesamt 2 neue Spalten für jeden Datensatz (basierend auf dem 90%-Konfidenzintervall) berechnen:

- uG_90
- oG_90

wobei uG die untere Grenze und oG die obere Grenze des Konfidenzintervalls sein soll.

Berechnen Sie jetzt beide Spalten für alle drei Verteilungen. Denken Sie daran, dass Sie die Mittelwerte bereits in den Datensätzen haben.

1. Konfidenzintervall Mittelwert - Übung 4

Jetzt soll das Ergebnis visualisiert werden. Eventuell müssen Sie zuvor das Package ggplot2 installieren. Danach können Sie den folgenden Code direkt ausführen. Falls Sie die Datensätze oder die Variablen im Datensatz angepasst haben, müssen Sie natürlich diese Anpassungen auch hier vornehmen. Für die beiden anderen Datensätze müssen Sie nur die zweite Zeile anpassen (Datensatz2 und Datensatz3).

```
1 library(ggplot2)
2 ggplot(Datensatz1, aes(y = Stichprobennummer, x = M_SP)) +
3 geom_vline(xintercept=100) +
4 geom_errorbarh(aes(xmin = uG_90, xmax = oG_90), height = 0.2) +
5 geom_point(size = 2) +
6 labs(x = "Mittelwert", y = "Stichprobe") +
7 xlim(80, 120) +
8 theme_minimal()
```

Jeder Graph stellt die Konfidenzintervalle (horizontale Linien), Stichprobenmittelwerte (Punkte) und den wahren Mittelwert in der Population (senkrechte Linie) dar. Vergleichen Sie die Graphen, was fällt Ihnen auf? Was verraten Ihnen diese Graphen. Was würde sich ändern, wenn Sie das 95%-Konfidenzintervall verwendet hätten?

2. Konfidenzintervall Varianz - Einleitung

Auch für die Varianz von Stichproben können Konfidenzintervalle berechnet werden, genau genommen für jede Statistik (zumindest theoretisch).

Wir wissen, dass die Varianz einer normalverteilten Population χ^2 -verteilt ist.

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

Das Konfidenzintervall der Varianz hat dann die Grenzen:

$$\left[\frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}, n-1}^2}, \frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}, n-1}^2} \right]$$

Die χ^2 -Verteilung ist nicht symmetrisch, daher benötigen wir diesmal zwei Werte aus der Verteilung. Beachten Sie, dass der grössere χ^2 -Wert mit der unteren Grenze assoziiert ist, da durch diesen dividiert wird. Für das Konfidenzintervall der Standardabweichung wäre die Rechnung identisch, es muss nur jeweils noch die Wurzel gezogen werden.

2. Konfidenzintervall Varianz - Übung 1

Fügen Sie für jeden Datensatz eine Spalte Var_SP hinzu, die sich aus den Varianzen der simulierten Stichproben ergibt.

Bestimmen Sie anschliessend die benötigten Kennwerte der χ^2 -Verteilung für ein 90%-Konfidenzintervall für $n = 10$ und $n = 30$.

Stichprobe mit $n = 10$

$$\chi^2_{1-\frac{\alpha}{2}, n-1} = \text{[]}$$

$$\chi^2_{\frac{\alpha}{2}, n-1} = \text{[]}$$

Stichprobe mit $n = 30$

$$\chi^2_{1-\frac{\alpha}{2}, n-1} = \text{[]}$$

$$\chi^2_{\frac{\alpha}{2}, n-1} = \text{[]}$$

Berechnen Sie jetzt jeweils eine Spalte Var_uG_90 und Var_oG_90 für jeden Datensatz für die untere und obere Grenze des Konfidenzintervalls jeder Stichprobe.

2. Konfidenzintervall Varianz - Übung 2

Kopieren Sie Ihren R-Code für die Konfidenzintervalle der Mittelwerte. Passen Sie ihn an, dass er stattdessen die Konfidenzintervalle der Varianzen visualisiert. Achten Sie auf den Namen des Datensatzes, die verwendeten Variablen/Spalten und das σ^2 statt μ eingezeichnet werden muss.

Stellen Sie die Grenzen der x-Achse mit `xlim()` auf 0 und 1500 ein. Damit können Sie die Graphen gut vergleichen.

Schauen Sie sich die Graphen an, was fällt Ihnen auf, wenn Sie die Graphen miteinander vergleichen? Was fällt Ihnen auf im Vergleich zu den Graphen der Mittelwerte?