10. Übung am 02. Juni 2020

UV Angewandte Statistik (405.170), Assoz.Prof. Dr. Wolfgang Trutschnig
Link Ankreuzliste: siehe www.trutschnig.net/courses
Mit 'F' versehene Aufgaben sind freiwillig, mit * versehene Aufgaben haben einen erhöhten Schwierigkeitsgrad.
Alle Verweise beziehen sich auf das Statistik-Skriptum.

Übungsaufgabe 50. Beweisen Sie die in der VO getätigte Aussage, dass im Falle des zweiseitigen t-tests die zwei Definitionen des p-Werts† übereinstimmen.

Übungsaufgabe 51. Finden Sie heraus, wofür die R-Funktion *power.t.test* verwendet werden kann.

Übungsaufgabe 52. Erika Musterfrau schlägt den folgenden Test auf Normalverteilung vor: Ausgehend von einer Stichprobe $X_1, \ldots X_n$ von $X \sim F$ setzen wir $\mu_0 = \overline{X}_n$ und $\sigma_0 = \sqrt{S_n^2}$ und testen dann (mittels Aufgabe 49) $H_0 : F = F_{\mu_0, \sigma_0^2}$, wobei F_{μ_0, σ_0^2} die Verteilungsfunktion von $\mathcal{N}(\mu_0, \sigma_0^2)$ bezeichnet. Überprüfen Sie mittels Simulationen, dass der resultierende Test für kleine Sample Sizes ($n = 10, 20, 30$) sehr geringe Power hat und überlegen Sie sich, warum dem so ist.

Definition 5.15. Wir betrachten wie bisher $\emptyset \neq \Theta_0 \subset \Theta$ und eine Stichprobe $\mathbf{X} = (X_1, \ldots, X_n)$ von $X \sim (P_\theta)_{\theta \in \Theta}$. Mit $\mathcal{X} := Rg(X)^n$ bezeichnen wir den Stichprobenraum. Weiters sei $\mathcal{G}$ eine endliche Gruppe von Transformationen $g : \mathcal{X} \to \mathcal{X}$. Wir sagen, dass *unter H_0 die Randomisierungseigenschaft* gilt, genau dann, wenn für jedes $X \sim P_\theta$ mit $\theta \in \Theta_0$, jede Stichprobe $\mathbf{X} = (X_1, \ldots, X_n)$ von X, und jedes $g \in \mathcal{G}$ gilt: $\mathbf{X}$ und $g \circ \mathbf{X}$ haben die selbe Verteilung.

Übungsaufgabe 53. Sei $(X_1, \ldots, X_m)$ eine Stichprobe von $X \sim \mathcal{N}(\mu_1, \sigma^2)$, $(Y_1, \ldots, Y_n)$ eine Stichprobe von $Y \sim \mathcal{N}(\mu_2, \sigma^2)$, und gelte $H_0 : \mu_1 = \mu_2$. Wir setzen $\mathcal{X} = \mathbb{R}^{m+n}$ und $\mathbf{X} = (X_1, \ldots, X_m, Y_1, \ldots, Y_n)$. Für jede Permutation π von $\{1, \ldots, m+n\}$ sei $g_\pi : \mathbb{R}^{m+n} \to \mathbb{R}^{m+n}$ definiert durch

$$g_\pi(z_1, \ldots, z_{m+n}) = (z_{\pi(1)}, \ldots, z_{\pi(m+n)}),$$

$\mathcal{G}$ bezeichne die Gruppe all dieser Transformationen. Zeigen Sie, dass in diesem Setting die Randomisierungseigenschaft gilt.

Übungsaufgabe 54. Wir betrachten das Setting von Definition 5.15 und schreiben $M := \#G$ für die Kardinalität von G. Zusätzlich sei T eine beliebige Teststatistik, für jede Stichprobe $\mathbf{x} = (x_1, \ldots, x_n)$ bezeichne $T_{(1)}(\mathbf{x}) \leq T_{(2)}(\mathbf{x}) \leq \cdots \leq T_{(n)}(\mathbf{x})$ die Ordnungsstatistik der Werte $\{T \circ g(\mathbf{x}) : g \in \mathcal{G}\}$. Für jedes $\alpha \in (0,1)$ setzen wir $k := M - \lfloor M\alpha \rfloor$ sowie

$$\begin{aligned} M^0(\mathbf{x}) &= \#\Big\{ j \in \{1, \ldots, M\} : T_{(j)}(\mathbf{x}) = T_{(k)}(\mathbf{x}) \Big\} \\ M^+(\mathbf{x}) &= \#\Big\{ j \in \{1, \ldots, M\} : T_{(j)}(\mathbf{x}) > T_{(k)}(\mathbf{x}) \Big\}, \end{aligned}$$

und definieren einen 'drei'-wertigen Test φ durch

† Definition 5.5 und die implizite Definition des p-Werts als 'die Wahrscheinlichkeit, unter H_0 etwas mindestens so Extremes wie den aktuellen Wert (der Teststatistik) zu beobachten.'

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{if } T(\mathbf{x}) > T_{(k)}(\mathbf{x}), \\ \frac{M\alpha - M^{+}(\mathbf{x})}{M^{0}(\mathbf{x})} & \text{if } T(\mathbf{x}) = T_{(k)}(\mathbf{x}), \\ 0 & \text{if } T(\mathbf{x}) < T_{(k)}(\mathbf{x}). \end{cases}$$

Beweisen Sie, dass dann für jedes $\theta \in \Theta_0$ die folgende Gleichheit gilt:

$$\mathbb{E}_\theta(\varphi \circ \mathbf{X}) = \alpha.$$