

Übungen zur Vorlesung
Datenanalyse - Dr. Terveer, Vogt, Pohl
Sommersemester 2022 Blatt 5 17.05.2022

Aufgabe 12 (Mittelwerttests) Das Testaufkommen in der täglich geöffneten Covid-Teststation auf dem Galileo-Campus lag bis zur weitgehenden Aufhebung der Testpflicht im April 2022 bei 300 Tests pro Tag. Nach Aufhebung der Testpflicht geht man davon aus, dass sich diese Zahl halbieren wird. Sollte die Zahl allerdings noch weiter sinken, so soll die Teststation aus Rentabilitätsgründen geschlossen werden. Zur Unterstützung einer entsprechenden Entscheidung soll im Mai 2022 täglich die Anzahl der Tests festgehalten und dann anhand eines geeigneten Mittelwerttests entschieden werden, ob sich der Betrieb ab Juni weiter lohnt.

- Formulieren Sie das zugehörige Mittelwert-Testproblem auf Grundlage der im Mai zu beobachtenden Testzahlen X_1, \dots, X_{31}
- Stellen Sie einen geeigneten Mittelwert-Test zum Signifikanzniveau 5% auf.
- Ermitteln Sie die Testentscheidung an für die folgenden Testzahlen im Mai: 154, 136, 130, 144, 121, 145, 148, 146, 136, 130, 132, 136, 139, 139, 151, 149, 147, 150, 149, 140, 116, 145, 152, 131, 134, 137, 138, 144, 139, 123, 151.

Aufgabe 13 (Einkommen im Praktikum) Für das Einkommen studentischer Praktikanten werde eine Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ mit $\sigma = 100$ zugrunde gelegt. Eine Stichprobe von $n = 90$ studentischen Praktikanten ergebe das Durchschnittseinkommen $\bar{X} = 480$. Kann die Hypothese $H_0 : \mu \geq 550$ bei einem Signifikanzniveau von $\alpha = 0.01$ verworfen werden? Bestimmen Sie auch den p^* -Wert des Testes.

Aufgabe 14 (Einstichproben tests, Klausur WS1415) Eine Druckerei verwendet Papier mit einer Soll-Dicke von $\mu_0 = 300$ Mikrometern. Die an $n = 11$ aufeinanderfolgenden Zeitpunkten gemessenen Papierdicken lauten: 305, 273, 307, 305, 288, 249, 261, 274, 280, 328, 295. Es darf davon ausgegangen werden, dass diesen Werten stochastisch unabhängige Zufallsvariablen X_1, \dots, X_{11} mit jeweils derselben Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ zugrundeliegen, dabei sind $\sigma > 0$, $\mu \in \mathbb{R}$ unbekannt.

- Berechnen Sie die t -Statistik $V = \sqrt{n} \cdot \frac{\bar{X} - \mu}{\hat{\sigma}}$ für den vorliegenden Datensatz.
- Betrachten Sie die Hypothesen

- (1) $H_0^* : \mu = \mu_0$ gegen $H_1^* : \mu \neq \mu_0$
- (2) $H_0^{**} : \mu \geq \mu_0$ gegen $H_1^{**} : \mu < \mu_0$
- (3) $H_0^{***} : \mu \leq \mu_0$ gegen $H_1^{***} : \mu > \mu_0$

Berechnen Sie zum vorliegenden Datensatz auf Basis der Teststatistik in a) für jede der Hypothesen den p -Wert und untersuchen Sie, welche der Hypothesen zum Niveau 5% abgelehnt wird bzw. werden. Nutzen Sie die nachfolgende Tabelle der Werte $F_{t(k)}(x)$ (Verteilungsfunktion der t -Verteilung mit k Freiheitsgraden).

x	$k=2$	3	4	5	6	7	8	9	10	11	12	13	14	15	x	$k=2$	3	4	5	6	7	8	9	10	11	12	13	14	15	x	$k=2$	3	4	5	6	7	8	9	10	11	12	13	14	15	
0.0	500	500	500	500	500	500	500	500	500	500	500	500	500	500	1.0	789	804	813	818	822	825	827	828	830	831	831	832	833	833	2.0	908	930	942	949	954	957	960	962	963	965	966	967	967	968	
0.1	535	537	537	538	538	538	539	539	539	539	539	539	539	539	1.1	807	824	833	839	843	846	848	850	851	853	854	854	855	856	2.1	915	937	948	955	960	963	966	967	969	970	971	972	973	973	
0.2	570	573	574	575	576	577	577	577	577	577	578	578	578	578	1.2	823	842	852	858	862	865	868	870	871	872	873	874	875	876	2.2	921	942	954	960	965	968	971	972	974	975	976	977	977	978	
0.3	604	608	610	612	613	614	614	615	615	615	615	616	616	616	1.3	838	858	868	875	879	883	885	887	889	890	891	892	893	893	2.3	926	948	959	965	969	973	975	977	978	979	980	981	981	982	
0.4	636	642	645	647	648	649	650	651	651	652	652	652	652	653	1.4	852	872	883	890	894	898	900	902	904	905	907	908	908	909	2.4	931	952	963	969	973	976	978	980	981	982	983	984	985	985	
0.5	667	674	678	681	683	684	685	685	686	687	687	687	688	688	1.5	864	885	896	903	908	911	914	916	918	919	920	921	922	923	2.5	935	956	967	973	977	980	982	983	984	985	986	987	987	988	
0.6	695	705	710	713	715	716	717	718	719	720	720	721	721	721	1.6	875	896	908	915	920	923	926	928	930	931	932	933	934	935	2.6	939	960	970	976	980	982	984	986	987	988	988	989	990	990	
0.7	722	733	739	742	745	747	748	749	750	751	751	752	752	753	1.7	884	906	918	925	930	934	936	938	940	941	943	944	944	945	2.7	943	963	973	979	982	985	986	988	989	990	990	991	991	992	
0.8	746	759	766	770	773	775	777	778	779	780	780	781	781	782	1.8	893	915	927	934	939	943	945	947	949	950	951	952	953	954	2.8	946	966	976	981	984	987	988	989	990	991	991	992	992	993	993
0.9	768	783	790	795	799	801	803	804	805	806	807	808	808	809	1.9	901	923	935	942	947	950	953	955	957	958	959	960	961	962	2.9	949	969	978	983	986	989	990	991	992	992	993	993	994	994	995

Die angegebenen Werte sind noch durch 1000 zu teilen. Nicht in der Tabelle stehende Werte von x runden Sie bitte auf den nächstgelegenen Wert \tilde{x} in der Tabelle.

- c) Nehmen Sie an, dass die Daten in einem R-Skript in einem Vektor $\mathbf{x} = \text{round}(100 * \text{rnorm}(n=11, \text{mean}=2.8, \text{sd}=.2), 0)$ gegeben sind. Nutzen Sie die im **stats**-Paket bereitgestellten Funktion zum t-Test um die t -Statistik $V = \sqrt{n} \cdot \frac{\bar{X} - \mu}{\hat{\sigma}}$ und den p-Wert eines der in b) genannten Hypothesenpaare zu berechnen.