Statistische Inferenz | 03. Einfache Tests für Mittelwerte Musterlösung

Prof. Dr. Roland Schäfer | Germanistische Linguistik FSU Jena

1. Dezember 2024

Hinweis: Wo nicht anders angegeben, runden Sie die Ergebnisse auf zwei Nachkommastellen.

1 z-Test bzw. t-Test für ein Stichprobenmittel

Gegeben sei eine in vorherigen Split-100-Experimenten abgesicherte mittlere Bewertung von ob-wohl-Sätzen mit Verbzweitstellung bei Komma vor obwohl und fehlendem Interpunktionszeichen nach obwohl von $\mu=22.3$ (Standardabweichung $\sigma=8.4$) gegenüber identisch markierten obwohl-Sätzen mit Verb-Letzt-Stellung (Mittel $\mu'=77.7$). (In Split-100-Experimenten bewerten Probanden zwei gegebene Varianten eines Satzes, indem Sie 100 Punkte zwischen den beiden Möglichkeiten verteilen.) In einem Folge-Experiment soll getestet werden, ob sich die Bewertung ändert, wenn vor obwohl ein Punkt gesetzt wird. Die 17 Probanden vergeben die folgenden Punktzahlen für die Verb-Zweit-Sätze.

$$x = [7, 23, 20, 10, 37, 21, 40, 53, 28, 30, 44, 35, 46, 27, 51, 23, 38]$$

(a) Rechnen Sie zunächst von Hand Schritt für Schritt einen z-Test mit der gegebenen Standardabweichung $\sigma=8.4$ bei sig=0.01. (b) Vergleichen Sie das Ergebnis mit einem t-Test, ebenfalls bei sig=0.01. (c) Falls Sie Abweichungen feststellen, wie kommen diese Abweichungen zustande? (d) Quantifizieren Sie die Effektstärke für den t-Test mit d und r^2 . Interpretieren Sie die Testergebnisse und die Effektstärke.

(a)

$$\bar{x} = 31.35 \tag{1}$$

$$SF_z = \frac{8.4}{\sqrt{17}} = 2.04 \tag{2}$$

$$z = \frac{31.35 - 22.3}{2.04} = 4.44\tag{3}$$

$$4.44 > 2.58$$
 (sig-Niveau erreicht.) (4)

(b)

$$sd_x = 13.35 \tag{5}$$

$$SF_t = \frac{13.35}{\sqrt{17}} = 3.24 \tag{6}$$

$$t = \frac{31.35 - 22.3}{3.24} = 2.8\tag{7}$$

$$2.8 \ge 2.92$$
 (sig-Niveau nicht erreicht.) (8)

(c) Die Abweichung zwischen den Ergebnissen hat zwei Ursachen: (i) Die Varianz in der Stichprobe ist deutlich größer als die in der Approximation der Grundgesamtheit. Dadurch wird der Standardfehler größer, und der t-Wert ist kleiner als der z-Wert. (ii) Durch die Abhängigkeit der t-Verteilung von den Freiheitsgraden (hier 16) ist der kritische Wert größer (2.92 gegenüber 2.58), und der t-Wert liegt knapp darunter.

(d)

$$d = \frac{31.35 - 22.3}{13.35} = 0.68 \tag{9}$$

$$r^2 = \frac{2.8^2}{2.8^2 + 16} = 0.33\tag{10}$$

Gemäß z-Test ist die mittlere Bewertung der Sätze durch Probanden nicht dieselbe, wenn die besondere Interpunktion hinzugefügt wird, oder es ist ein seltenes Ereignis eingetreten. Selten heißt, dass nur in einem von hundert Fällen ein solches oder ein extremeres Ereignis zu erwarten ist, wenn die Mittellwerte in der Grundgesamtheit gleich sind und man dasselbe Experiment immer wieder macht. Der z-Test liefert also vorläufige Evidenz dafür, dass es einen Unterschied geben könnte. Das Ergebnis des t-Tests liefert keine Evidenz für irgendetwas und darf nicht inferenziell interpretiert werden. Die Werte für d und r^2 dürfen daher auch nicht interpretiert werden. Insgesamt sollte auch das Ergebnis des z-Tests mit Vorsicht behandelt werden, da die Varianz in der Stichprobe erheblich vom bekannten Wert σ abweicht. Es muss überprüft werden, ob das aktuelle Experiment im Design oder in der Durchführung von den Referenz-Experimenten abweicht.

2 t-Test für zwei Stichprobenmittel

Nehmen wir an, ein Experiment wie das aus Aufgabe 1 hätte nicht auf bekannten bzw. etablierten Vorergebnissen beruht. Stattdessen wurden zwei Gruppen (*O* für *ohne besondere Interpunktion* und *M* für *mit besonderer Interpunktion*) in einem Experiment verglichen. In den Gruppen wurden für die Verbzweitstellung folgende Werte vergeben:

$$x_O = [15, 26, 17, 37, 36, 18, 29, 32, 42, 27, 32, 47, 5, 20, 22]$$
 (11)

$$x_M = [25, 38, 45, 32, 43, 32, 27, 37, 58, 33, 26, 32, 43, 19, 47]$$
 (12)

(a) Rechnen Sie einen t-Test für zwei Stichproben bei sig=0.05 und sig=0.01. (b) Quantifizieren Sie die Effektstärke mit d und r^2 . Interpretieren Sie die Testergebnisse und die Effektstärke. (c) (Bonus) Es stellt sich heraus, dass die Person, die Gruppe M betreut hat, den letzten vier Teilnehmenden aus

Versehen verraten hat, um welches Phänomen es im Experiment geht. Die letzten vier Werte aus x_M müssen also entfernt werden. Berechnen Sie den neuen korrekten t-Test und interpretieren Sie das Ergebnis.

(a)

$$\bar{x}_O = 27 \tag{13}$$

$$\bar{x}_M = 35.8$$
 (14)

$$var_{x_0} = 124.57$$
 (15)

$$var_{x_M} = 102.6$$
 (16)

$$SF_{\bar{x}_O - \bar{x}_M} = \sqrt{\frac{124.57}{15} + \frac{102.6}{15}} = 3.89$$
 (17)

$$t = \frac{27 - 35.8}{3.89} = -2.26 \tag{18}$$

$$-2.26 < -2.05$$
 (sig-Niveau bei sig=0.05 erreicht.) (19)

$$-2.26 < -2.76$$
 (sig-Niveau bei sig=0.01 nicht erreicht.) (20)

(b)

$$var_{X_O, X_M} = \frac{1744 + 1436.4}{14 + 14} = 113.59 \tag{21}$$

$$d = \frac{27 - 35.8}{\sqrt{113.59}} = -0.83\tag{22}$$

$$r^2 = \frac{5.11}{5.11 + 28} = 0.15\tag{23}$$

(c)

$$\bar{x}_{M_{\bullet}} = 36 \tag{24}$$

$$var_{X_O,X_M} = \frac{1744 + 962}{14 + 10} = 112.75$$
 (25)

$$SF_{\bar{x}_O - \bar{x}_M}$$
, = $\sqrt{\frac{112.75}{15} + \frac{112.75}{11}} = 4.22$ (26)

$$t = \frac{27 - 36}{422} = -2.14\tag{27}$$

Die Interpretation bleibt im Wesentlichen dieselbe, auch wenn vier Datenpunkte gelöscht werden und die Varianz damit als zusammengelegte Varianz berechnet wird.

3 Transfer: Tests und Konfidenzintervalle

Berechnen Sie die 95%- und 99%-Konfidenzintervalle für Aufgabe 2 (mit den vollen Stichproben, nicht mit dem um vier Datenpunkte reduzierten Datensatz). Versuchen Sie zu verstehen, in welcher Relation die Tests und die Konfidenzintervalle stehen.

$$SF_O = \sqrt{\frac{124.57}{15}} = 2.88 \tag{28}$$

$$KI_{95,O} = 27 \pm 1.96 \cdot 2.88 = [21.35, 32.65]$$
 (29)

$$KI_{99,O} = 27 \pm 2.85 \cdot 2.88 = [18.79, 35.21]$$
 (30)

$$SF_M = \sqrt{\frac{102.6}{15}} = 2.62 \tag{31}$$

$$KI_{95,M} = 35.8 \pm 1.96 \cdot 2.62 = [30.67, 40.93]$$
 (32)

$$KI_{99,M} = 35.8 \pm 2.85 \cdot 2.62 = [28.35, 43.25]$$
 (33)

Die KIs fokussieren die Genauigkeit einer Schätzung, während die p-Werte bzw. der Signifikanztest eine Ja/Nein-Entscheidung über den Wert der Evidenz für eine wissenschaftliche Inferenz zum Ziel haben. Im Prinzip gilt dabei: Die äußeren Grenzen der Konfindenzintervalle korrespondieren mit dem Erreichen oder Verfehlen der kritischen Werte des Tests. Es gilt ein Zusammenhang der folgenden Art: Überlappen die KIs bei einem bestimmten Konfidenzniveau, so wird der Unterschiedstest beim zugehörigen sig-Niveau nicht signifikant. Überlappen die KIs nicht, wird der Test signifikant. Das ist wünschenswert, da die KIs mit steigendem Konfidenzniveau breiter werden, weil die Schätzung bei einem hohen Konfidenzniveau mehr Werte einschließt. Bei einem niedrigeren sig-Niveau ist es dementsprechend "schwerer", ein signifikantes Testergebnis (also Nicht-Überlappung der Intervalle) zu erreichen. Im gegebenen Fall überschneiden die 95%-Konfidenzintervalle einander allerdings knapp um 32.65 - 30.67 = 1.98, obwohl der t-Test sig=0.05 erreicht hat. Das liegt in diesem Fall daran, dass die asymptotisch normalen KIs zwei einzelne Schätzungen darstellen, der t-Test aber ein einzelner asymptotischer Test der Mittelwertdifferenz auf $H0 = \bar{x}_O - \bar{x}_M = 0$ ist. Ohne auf die Details einzugehen, ist es plausibel, dass hier nicht dasselbe herauskommt, da es sich nicht um die genau zum gerechneten Test korrespondierenden KIs handelt. Berechnet man hingegen das KI für $\bar{x}_O - \bar{x}_M$, so konvergiert dies sowohl in der normalen als auch der t-Approximation mit dem Testergebnis: Das 95%-KI schließt in beiden Fällen o ein, das 99%-KI nicht.

$$KI_{z,95,\bar{x}_M-\bar{x}_O} = (27 - 35.8) \cdot 1.96 \cdot 3.89 = [-16.43, -1.17]$$
 (34)

$$KI_{z,99,\bar{x}_M-\bar{x}_O} = (27 - 35.8) \cdot 2.85 \cdot 3.89 = [-18.84, 1.24]$$
 (35)

$$KI_{t_{28},95,\bar{x}_M-\bar{x}_O} = (27-35.8) \cdot 2.05 \cdot 3.89 = [-16.78, -0.82]$$
 (36)

$$KI_{t_{28},99,\bar{x}_M-\bar{x}_O} = (27-35.8) \cdot 2.76 \cdot 3.89 = [-19.54, 1.94]$$
 (37)

4 Transfer: Einseitige und zweiseitige Tests

Wir haben sogenannte zweiseitige Tests gerechnet. Der Unterschied zu einem einseitigen Test besteht in der Art der Ho und der zugehörigen theoretischen Erwartung an das Testergebnis. Überlegen Sie anhand der untenstehenden Hos für den Zwei-Stichproben-t-Test, wie sich die Tests unterscheiden könnten, und was sich mathematisch ändern müsste.

Ho_{einseitig,kleiner}: Der wahre Mittelwert für die Grundgesamtheit zu Gruppe O ist nicht kleiner als der wahre Mittelwert für die Grundgesamtheit zu Gruppe M.

 ${
m Ho_{einseitig,g\"oßer}}$: Der wahre Mittelwert für die Grundgesamtheit zu Gruppe O ist nicht gr\"oßer als der wahre Mittelwert für die Grundgesamtheit zu Gruppe M.

Ho_{zweiseitig}: Die wahren Mittelwerte der Grundgesamtheiten zu Gruppe O und Gruppe M sind gleich groß.

Kritische Werte für zweiseitige z- und t-Tests

Alle Werte sind als \pm zu lesen, da es sich um die Werte für einen zweiseitigen Test handelt.

	0.75	0.76	0.77	0.78	0.79	0.8	0.81	0.82	0.83	0.84	0.85	0.86	0.87	0.88	0.89	0.9	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99
z	1.15	1.17		1.23	1.25		1.31			1.41			1.51			1.64		1.75	1.81	1.88	1.96	2.05	2.17	2.33	2.58
t(49)	1.16	1.19	1.22	1.24	1.27		1.33	1.36		1.43			1.54		1.63	1.68	1.73	1.79	1.85	1.93	2.01	2.11	2.24	2.40	2.68
t(48)	1.16	1.19	1.22	1.24	1.27	1.30	1.33	1.36	1.39	1.43	1.46	1.50	1.54	1.58	1.63	1.68	1.73	1.79	1.85	1.93	2.01	2.11	2.24	2.41	2.68
t(47)	1.16	1.19	1.22	1.24	1.27	1.30	1.33	1.36	1.39	1.43	1.46	1.50	1.54	1.58	1.63	1.68	1.73	1.79	1.85	1.93	2.01	2.11	2.24	2.41	2.68
t(46)	1.17	1.19	1.22	1.24	1.27	1.30	1.33	1.36	1.39	1.43	1.46	1.50	1.54	1.58	1.63	1.68	1.73	1.79	1.86	1.93	2.01	2.11	2.24	2.41	2.69
t(45)	1.17	1.19	1.22	1.24	1.27	1.30	1.33	1.36	1.39	1.43	1.46	1.50	1.54	1.58	1.63	1.68	1.73	1.79	1.86	1.93	2.01	2.12	2.24	2.41	2.69
t(44)	1.17	1.19	1.22	1.24	1.27	1.30	1.33	1.36	1.40	1.43	1.47	1.50	1.54	1.59	1.63	1.68	1.73	1.79	1.86	1.93	2.02	2.12	2.24	2.41	2.69
t(43)	1.17	1.19	1.22	1.24	1.27	1.30	1.33	1.36	1.40	1.43	1.47	1.50	1.54	1.59	1.63	1.68	1.73	1.79	1.86	1.93	2.02	2.12	2.24	2.42	2.70
t(42)	1.17	1.19	1.22	1.25	1.27	1.30	1.33	1.36	1.40	1.43	1.47	1.50	1.54	1.59	1.63	1.68	1.74	1.79	1.86	1.93	2.02	2.12	2.25	2.42	2.70
	,	1.19		1.25	1.27	1.30	1.33	1.36	1.40	1.43	1.47	1.50	1.55	1.59	1.63	1.68	1.74	1.80	1.86	1.93	2.02	2.12	2.25	2.42	2.70
	,	1.19		1.25	1.27	-	1.33	1.36	1.40	1.43	1.47	1.51		1.59	1.63	1.68	1.74	1.80	1.86	1.94	2.02	2.12	2.25	2.42	2.70
	,	1.19		1.25	1.27	-	1.33	1.37	-	1.43		-	1.55	1.59	1.64		1.74	1.80	1.86	1.94	2.02	2.12	2.25	2.43	2.71
	,	1.19		1.25	1.28	-		1.37				-		1.59	1.64	-	1.74	1.80	1.86	1.94	2.02	2.13	2.25	2.43	2.71
	1.17			1.25		-		1.37	-	1.43		1.51		1.59	1.64	-	1.74	1.80	1.87	1.94	2.03	2.13	2.26	2.43	2.72
	1.17		1.22	-		1.31		1.37	-	1.43		1.51		1.59	1.64	-	, -	1.80	1.87	1.94	2.03	2.13	2.26	2.43	2.72
()	1.17				1.28					1.44				1.59	1.64	-	1.74	1.80	1.87	1.94	2.03	2.13	2.26	2.44	2.72
()	1.17			1.25		1.31				1.44		1.51		1.59	1.64	-	1.75	1.80	1.87	1.95	2.03	2.14	2.27	2.44	2.73
t(33) t(32)	1.17	1.20		1.25	1.28	1.31	1.34			1.44			1.55	1.60	1.64	-	1.75	1.81	1.87	1.95	2.03	2.14	2.27	2.44	2.73
	1.17			1.25		1.31			1.40	1.44		-	1.56		1.65	-	1.75	1.81	1.88	1.95 1.95	2.04	2.14	2.27	2.45 2.45	2.74 2.74
	1.17			1.25		1.31			-		-	-	1.56		1.65		1.75	1.81	1.88	1.95	2.04	2.14	2.28	2.46	2.75
t(29)	,		1.23	1.25		1.31		1.37				-	1.56	1.60	1.65	1.70	1.75	1.81	1.88	1.95	2.04	2.15	2.28	2.46	2.76
	,		-	1.25		1.31			-		-	-	1.56		-	1.70	, .	1.82	1.88	1.96	2.05	2.15	2.29	2.47	2.76
` ′	1.18		1.23	1.26		1.31		1.38					1.56		-	1.70	,	1.82	1.89	1.96	2.05	2.16	2.29	2.47	2.77
t(26)		1.20	1.23	1.26	1.29			1.38					1.56		-	1.71	,	1.82	1.89	1.97	2.06	2.16	2.30	2.48	2.78
t(25)		1.20	1.23	1.26	1.29	-	1.35		1.41		1.49			1.61	1.66		1.76	1.82	1.89	1.97	2.06	2.17	2.30	2.49	2.79
t(24)	1.18	1.20	1.23	1.26		1.32		-	-			-	1.57			1.71	,	1.83	1.90	1.97	2.06	2.17	2.31	2.49	2.80
t(23)	1.18	1.21	1.23	1.26	1.29	1.32	1.35	1.38	1.42	1.45	1.49	1.53	1.57	1.61	1.66	1.71	1.77	1.83	1.90	1.98	2.07	2.18	2.31	2.50	2.81
t(22)	1.18	1.21	1.23	1.26	1.29	1.32	1.35	1.38	1.42	1.45	1.49	1.53	1.57	1.62	1.67	1.72	1.77	1.84	1.90	1.98	2.07	2.18	2.32	2.51	2.82
t(21)	1.18	1.21	1.24	1.26	1.29	1.32	1.35	1.39	1.42	1.46	1.49	1.53	1.58	1.62	1.67	1.72	1.78	1.84	1.91	1.99	2.08	2.19	2.33	2.52	2.83
t(20)	1.18	1.21	1.24	1.27	1.30	1.33	1.36	1.39	1.42	1.46	1.50	1.54	1.58	1.62	1.67	1.72	1.78	1.84	1.91	1.99	2.09	2.20	2.34	2.53	2.85
t(19)	1.19	1.21	1.24	1.27	1.30	1.33	1.36	1.39	1.43	1.46	1.50	1.54	1.58	1.63	1.68	1.73	1.79	1.85	1.92	2.00	2.09	2.20	2.35	2.54	2.86
t(18)	1.19	1.22	1.24	1.27	1.30	1.33	1.36	1.39	1.43	1.47	1.50	1.54	1.59	1.63	1.68	1.73	1.79	1.86	1.93	2.01	2.10	2.21	2.36	2.55	2.88
t(17)	-	1.22	-	1.27	-	1.33		-		1.47	-			1.64	1.69	, -	1.80	1.86	1.93	2.02	2.11	2.22	2.37	2.57	2.90
t(16)	-	1.22	-	1.28		1.34								1.64	1.69	, ,	1.80	1.87	1.94	2.02	2.12	2.24	2.38	2.58	2.92
t(15)	1.20	1.22	-	1.28		1.34								1.65	,	1.75		1.88	1.95	2.03	2.13	2.25	2.40	2.60	2.95
t(14)	1.20	1.23	1.26	1.28	-		1.38							1.66	1.71	1.76	1.82	1.89	1.96	2.05	2.14	2.26	2.41	2.62	2.98
t(13)	1.20	1.23		1.29	1.32		1.38					1.57		1.66	1.72		1.83	1.90	1.97	2.06	2.16	2.28	2.44	2.65	3.01
` ′	1.21	-	1.26	-		1.36							1.63		1.73		1.84	1.91	1.99	2.08	2.18	2.30	2.46	2.68	3.05
1 1	1.21	-	1.27	1.30		1.36						1.59	-	1.69	1.74	1.80	1.86	1.93	2.01	2.10	2.20	2.33	2.49	2.72	3.11
t(10) t(9)	1.22	1.25	1.28	1.31	1.34	1.37				1.52		1.60	-	1.70	1.75		1.88	1.95	2.03	2.12	2.23	2.36	2.53	2.76 2.82	3.17
t(8)	-	1.26	1.29	1.32	00	-	1.42		1.49	1.53	- ,	1.64	,	1.72 1.74	1.77 1.80	-	1.90	1.97 2.00	2.06	2.15	2.26	2.40	2.57 2.63	2.02	3.25 3.36
t(7)	1.24	1.28	1.31	1.35		1.40 1.41				1.55		1.66	1.72	, -	1.83	1.89	1.93	2.05	2.14	2.19	2.36	2.45 2.52	2.71	3.00	3.50
t(6)	1.27		1.34	1.37		1.44		1.52		1.60			,	1.81	1.87	1.94	2.02	2.10	2.14	2.31	2.45	2.61	2.83	3.14	3.71
t(5)	1.30	1.33		1.40	-	1.48	-	-	-		-		1.81		1.94	2.02	2.10	2.10	2.30	2.42	2.57	2.76	3.00	3.36	4.03
t(4)	1.34			1.45		1.53		1.62				1.84		1.97	2.05		2.23	2.33	2.46	2.60	2.78	3.00	3.30	3.75	4.60
t(3)	1.42	-	1.50		1.59	1.64	1.69		,	1.86	,	2.00	-	2.16	2.25	2.35	2.47	2.61	2.76	2.95	3.18	3.48	3.90	4.54	5.84
t(2)	1.60	1.65	1.71			1.89	-	2.03			2.28	2.38	,		2.76	2.92	3.10	3.32	3.58	3.90	4.30	4.85	5.64	6.96	9.92
1.1	2.41	-	2.65			-							4.83			-	7.03			10.58	12.71	15.89	21.20	31.82	63.66
/						-					- /	/													