

理解度チェック ➤ 度数分布表とヒストグラム

【問題】 次のデータは、女子大生 60 人の体重を調査した結果です。
度数分布表を作成し、そのヒストグラムを描いてください。

表 1.1.7 女子大生 60 人の体重

No.	体重	No.	体重	No.	体重
1	62	21	50	41	68
2	45	22	47	42	56
3	47	23	50	43	63
4	46	24	53	44	50
5	40	25	56	45	50
6	41	26	51	46	47
7	55	27	49	47	46
8	53	28	45	48	47
9	45	29	48	49	54
10	48	30	43	50	56
11	55	31	59	51	53
12	43	32	51	52	47
13	46	33	51	53	48
14	47	34	48	54	50
15	48	35	48	55	43
16	50	36	52	56	51
17	55	37	50	57	43
18	48	38	49	58	60
19	58	39	50	59	53
20	49	40	48	60	42

手順 1 最大値と最小値を探します。

最大値 = 68 最小値 = 40

手順 2 範囲を計算します。

範囲 = 70 - 40 = 30

手順 3 度数を求め、度数分布表を作成します。

表 1.1.8 度数分布表

階級	階級値	度数	相対度数	累積度数	累積相対度数
40～45	42.5	7	0.117	7	0.117
45～50	47.5	23	0.383	30	0.500
50～55	52.5	18	0.300	48	0.800
55～60	57.5	8	0.133	56	0.933
60～65	62.5	3	0.050	59	0.983
65～70	67.5	1	0.017	60	1.000
合計		60	1.000		

手順 4 ヒストグラムを描きます。

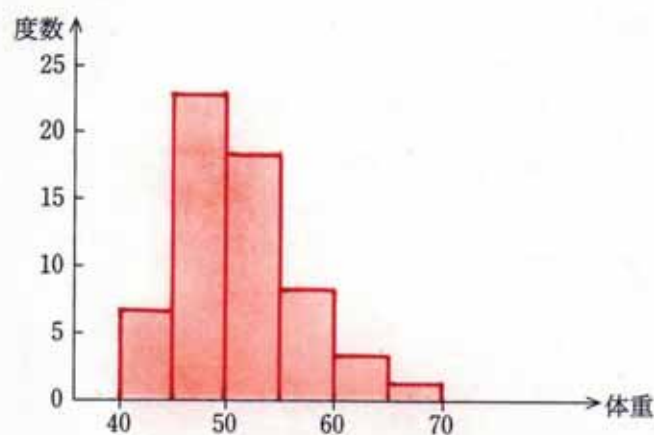


図 1.1.7 体重のヒストグラム

体重の3乗根の
ヒストグラムは
正規分布に近づくとい
われている
でござる！



理解度チェック ➤ 平均・分散・標準偏差

【問題 1】 次のデータは、9 本のカサの紫外線カット率を調査した結果です。
空欄を埋めて、平均・分散・標準偏差を求めてください。

表 1.2.10 カサの紫外線カット率

No.	カット率 x (%)	x^2
1	74	5476
2	65	4225
3	62	3844
4	72	5184
5	61	3721
6	58	3364
7	70	4900
8	64	4096
9	68	4624
合計	594	39434

$$\text{平均 } \bar{x} = \frac{594}{9} = 66$$

$$\text{分散 } s^2 = \frac{9 \times 39434 - 594^2}{9 \times (9 - 1)} = 28.75$$

$$\text{標準偏差 } s = \sqrt{28.75} = 5.36$$

【問題 2】 次のデータは、C 町の地下水に含まれているヒ素濃度を測定した結果です。

空欄を埋めて、平均・分散・標準偏差を求めてください。

表 1.2.11 地下水に含まれているヒ素濃度

No.	ヒ素濃度 x [ppm]	x^2
1	0.007	0.000049
2	0.012	0.000144
3	0.025	0.000625
4	0.009	0.000081
5	0.006	0.000036
6	0.018	0.000324
7	0.022	0.000484
8	0.019	0.000361
合計	0.118	0.002104

$$\text{平均 } \bar{x} = \frac{0.118}{8} = 0.0148$$

$$\text{分散 } s^2 = \frac{8 \times 0.002104 - 0.118^2}{8 \times (8 - 1)} = 5.19 \times 10^{-5}$$

$$\text{標準偏差 } s = \sqrt{5.19 \times 10^{-5}} = 0.0072$$

理解度チェック >>> 相関係数・共分散の求め方

【問題 1】 次のデータは、乗用車の価格と最大出力を調査した結果です。
次の空欄を埋めて、相関係数・分散・共分散を求めてください。

表 1.3.10 乗用車の価格と最大出力の統計量

No.	価格 x	最大出力 y	x^2	y^2	xy
1	140	120	19600	14400	16800
2	101	113	10201	12769	11413
3	149	160	22201	25600	23840
4	82	76	6724	5776	6232
5	118	105	13924	11025	12390
6	98	93	9604	8649	9114
7	143	140	20449	19600	20020
8	69	73	4761	5329	5037
9	137	140	18769	19600	19180
合計	1037	1020	126233	122748	124026

価格と出力の相関係数

$$= \frac{9 \times 124026 - 1037 \times 1020}{\sqrt{9 \times 126233 - 1037^2} \sqrt{9 \times 122748 - 1020^2}} = 0.9358$$

$$\text{価格の分散} = \frac{9 \times 126233 - 1037^2}{9 \times (9 - 1)} = 843.4$$

$$\text{出力の分散} = \frac{9 \times 122748 - 1020^2}{9 \times (9 - 1)} = 893.5$$

$$\text{価格と出力の共分散} = \frac{9 \times 124026 - 1037 \times 1020}{9 \times (9 - 1)} = 812.4$$

$$\text{相関係数} = \frac{\text{共分散}}{\sqrt{x \text{ の分散}} \sqrt{y \text{ の分散}}}$$



【問題 2】 次のデータは、発展途上国の妊産婦受診率と新生児死亡率を調査した結果です。

次の空欄を埋めて、相関係数・分散・共分散を求めてください。

表 1.3.11 受診率と死亡率の統計量

No.	受診率 x	死亡率 y	x^2	y^2	xy
1	1.54	4.26	2.3716	18.1476	6.5604
2	2.18	5.35	4.7524	28.6225	11.663
3	9.59	3.68	91.9681	13.5424	35.2912
4	5.16	4.72	26.6256	22.2784	24.3552
5	7.39	3.46	54.6121	11.9716	25.5694
6	2.08	3.91	4.3264	15.2881	8.1328
7	4.64	3.85	21.5296	14.8225	17.864
8	3.81	5.02	14.5161	25.2004	19.1262
9	2.38	4.36	5.6644	19.0096	10.3768
10	9.07	4.15	82.2649	17.2225	37.6405
11	3.74	5.79	13.9876	33.5241	21.6546
12	1.28	5.63	1.6384	31.6969	7.2064
合計	52.86	54.18	324.2572	251.3266	225.4405

受診率と死亡率の相関係数

$$= \frac{12 \times 225.4405 - 52.86 \times 54.18}{\sqrt{12 \times 324.2572 - 52.86^2} \sqrt{12 \times 251.3266 - 54.18^2}} = -0.5341$$

受診率の分散 = $\frac{12 \times 324.2572 - 52.86^2}{12 \times (12 - 1)} = 8.3099$

死亡率の分散 = $\frac{12 \times 251.3266 - 54.18^2}{12 \times (12 - 1)} = 0.6094$

受診率と死亡率の共分散 = $\frac{12 \times 225.4405 - 52.86 \times 54.18}{12 \times (12 - 1)} = -1.2020$

相関係数の関係式に
当てはめて
みるべし！



理解度チェック >> ケンドールの順位相関係数

【問題】 次の空欄を埋めて、ケンドールの順位相関係数を求めてください。

表 1.4.7 20 代女性と 40 代女性の海外旅行先の好み

観光地	韓国	中国	イギリス	フランス	イタリア	ロシア
20 代の順位	2	6	4	1	3	5
40 代の順位	1	5	2	3	4	6

観光地	韓国	中国
20 代	2	6
40 代	1	5
向き	同じ	

観光地	韓国	イギリス
20 代	2	4
40 代	1	2
向き	同じ	

観光地	韓国	フランス
20 代	2	1
40 代	1	3
向き	逆	

観光地	韓国	イタリア
20 代	2	3
40 代	1	4
向き	同じ	

観光地	韓国	ロシア
20 代	2	5
40 代	1	6
向き	同じ	

観光地	中国	イギリス
20 代	6	4
40 代	5	2
向き	同じ	

観光地	中国	フランス
20 代	6	1
40 代	5	3
向き	同じ	

観光地	中国	イタリア
20 代	6	3
40 代	5	4
向き	同じ	

観光地	中国	ロシア
20 代	6	5
40 代	5	6
向き	逆	

観光地	イギリス	フランス
20 代	4	1
40 代	2	3
向き	逆	

観光地	イギリス	イタリア
20 代	4	3
40 代	2	4
向き	逆	

観光地	イギリス	ロシア
20 代	4	5
40 代	2	6
向き	同じ	

観光地	フランス	イタリア
20 代	1	3
40 代	3	4
向き	同じ	

観光地	フランス	ロシア
20 代	1	5
40 代	3	6
向き	同じ	

観光地	イタリア	ロシア
20 代	3	5
40 代	4	6
向き	同じ	

$$P = \boxed{11}, \quad Q = \boxed{4}$$

$$\tau = \frac{2 \times (\boxed{11} - \boxed{4})}{\boxed{6} \times (\boxed{6} - 1)} = \boxed{0.47}$$

理解度チェック ➤ オッズ比

【問題】 自動車事故でシートベルト未着用の場合と着用の場合とでは、死亡のリスクはどの程度違うのでしょうか。

表 1.5.4 自動車事故の死傷者に関する調査結果

	シートベルト未着用	シートベルト着用
死亡した人	167 人	31 人
死亡しなかった人	8896 人	10533 人

次の空欄を埋めて、オッズとオッズ比を求めてください。

$$\begin{cases} \text{シートベルト未着用者が死亡した確率 } p = 0.018427 \\ \text{シートベルト着用者が死亡した確率 } q = 0.002934 \end{cases}$$

シートベルト未着用者のオッズ

$$\frac{p}{1-p} = \frac{0.018427}{1-0.018427} = 0.018772$$

シートベルト着用者のオッズ

$$\frac{q}{1-q} = \frac{0.002934}{1-0.002934} = 0.002943$$

$$\text{オッズ比} = \frac{0.018772}{0.002943} = 6.3784$$

このオッズ比は
次のようにしても
求められます



$$\text{オッズ比} = \frac{167 \times 10533}{31 \times 8896} = 6.3784$$

理解度チェック ➤ 確率分布

確率分布の数表を利用して、次の値を求めてください。

【問題 1】 $\chi^2(4; 0.95)$ 0.710723

$\chi^2(4; 0.05)$ 9.48773

$\chi^2(9; 0.95)$ 3.32511

$\chi^2(9; 0.05)$ 16.9190

【問題 2】 $t(7; 0.05)$ 1.895

$t(7; 0.025)$ 2.365

$t(12; 0.05)$ 1.782

$t(12; 0.025)$ 2.179

【問題 3】 $F(5, 7; 0.05)$ 3.9715

$F(5, 7; 0.95)$ 0.2051

理解度チェック

母平均の区間推定

【問題】 N社のサラダ油 800 g 入りのボトル 9 本を選んで、内容量を測定したところ、その値は

{807 g 811 g 801 g 798 g 798 g 795 g 803 g 805 g 804 g}

でした。このボトルの内容量の平均は何 g でしょうか。

次の空欄を埋めて、母平均 μ の 99% 信頼区間を求めてください。

t 分布の $100(1-\alpha)\%$ 点を求めます。

$$100(1-\alpha) = 99, \quad \frac{\alpha}{2} = 0.005$$

$$t(9-1; 0.005) = 3.355$$

この母集団は？



標本平均 $\bar{x}$ と標本分散 s^2 を計算します。

表 3.3.2 サラダ油の内容量

No.	x	x^2
1	807	651249
2	811	657721
3	801	641601
4	798	636804
5	798	636804
6	795	632025
7	803	644809
8	805	648025
9	804	646416
合計	7222	5795454

$$\text{標本平均 } \bar{x} = \frac{7222}{9} = 802.4$$

$$\text{標本分散 } s^2 = \frac{9 \times 5795454 - 7222^2}{9 \times (9-1)} = 25.03$$

最後に、信頼係数 99% の信頼区間を求めます。

$$802.4 - 3.355 \times \sqrt{\frac{25.03}{9}} \leq \mu \leq 802.4 + 3.355 \times \sqrt{\frac{25.03}{9}}$$

$$796.849 \leq \mu \leq 808.040$$

理解度チェック ➤ 母分散の区間推定

【問題】 N 自動車会社から発表された乗用車 C のガソリン 1 l 当たりの走行距離を測定したところ、次のようなデータを得ました。

{15.4 km 16.1 km 15.7 km 16.6 km 14.9 km 15.5 km 16.2 km}

この乗用車 C のガソリン 1 l 当たりの走行距離の分散は？

次の空欄を埋め、信頼係数 95% で母分散 σ^2 の区間推定をしてください。

カイ 2 乗分布の $100 \cdot \frac{\alpha}{2} \%$ 点と $100 \left(1 - \frac{\alpha}{2}\right) \%$ 点を求めます。

$$100(1 - \alpha) = 95, \quad \frac{\alpha}{2} = 0.025$$

$$\chi^2(7 - 1; 0.025) = 14.4494, \quad \chi^2(7 - 1; 0.975) = 1.2373$$

次に、標本分散 s^2 を計算します。

表 3.4.2 走行距離

No.	x	x^2
1	15.4	237.16
2	16.1	259.21
3	15.7	246.49
4	16.6	275.56
5	14.9	222.01
6	15.5	240.25
7	16.2	262.44
合計	110.4	1743.12

$$\begin{aligned} \text{標本分散 } s^2 &= \frac{1743.12}{7} - \frac{110.4^2}{7} \\ &= \frac{1743.12}{7} - \frac{12188.16}{7} \\ &= 0.3257 \end{aligned}$$

最後に、母分散 σ^2 の信頼係数 95% 信頼区間を求めます。

$$\frac{(7 - 1) \times 0.3257}{14.4494} \leq \sigma^2 \leq \frac{(7 - 1) \times 0.3257}{1.2373}$$

$$0.1352 \leq \sigma^2 \leq 1.5794$$

理解度チェック 母比率の区間推定

【問題】 人工降雨の実験をしたところ、54回のうち37回成功しました。
次の空欄を埋めて、この実験の成功率の区間推定をしてください。
ただし、信頼係数は90%とします。

はじめに、標準正規分布の $100 \cdot \frac{\alpha}{2} \%$ 点を求めます。

$$100(1-\alpha) = 90, \quad \frac{\alpha}{2} = 0.05$$

$$z(0.05) = 1.64$$

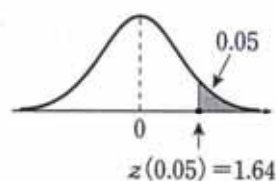


図 3.5.4

次に、標本比率 $\frac{m}{N}$ を計算します。

$$\text{標本比率 } \frac{m}{N} = \frac{37}{54} = 0.6852$$

最後に、母比率の区間推定の公式に代入します。

$$0.6852 - 1.64 \times \sqrt{\frac{0.6852 \times (1 - 0.6852)}{54}} \leq p \leq 0.6852 + 1.64 \times \sqrt{\frac{0.6852 \times (1 - 0.6852)}{54}}$$

$$0.5815 \leq p \leq 0.7888$$



データ数 N が少ないときは……

$$\frac{d_2}{d_1 F(d_1, d_2 : \frac{\alpha}{2}) + d_2} \leq p \leq \frac{e_1 F(e_1, e_2 : \frac{\alpha}{2})}{e_1 F(e_1, e_2 : \frac{\alpha}{2}) + e_2}$$

ただし、 $d_1 = 2(N - m + 1)$, $d_2 = 2m$

$e_1 = 2(m + 1)$, $e_2 = 2(N - m)$

理解度チェック ➤ 母平均の検定

【問題】 登山専門店I社のパンフレットによると、15 mm ザイルの破断強度は 4500 kg であると書いてあります。そこで、12本のザイルについて破断強度を調査したところ

$$\begin{Bmatrix} 4480 \text{ kg} & 4510 \text{ kg} & 4570 \text{ kg} & 4360 \text{ kg} & 4240 \text{ kg} & 4520 \text{ kg} \\ 4260 \text{ kg} & 4650 \text{ kg} & 4380 \text{ kg} & 4130 \text{ kg} & 4530 \text{ kg} & 4290 \text{ kg} \end{Bmatrix}$$

という測定結果になりました。

このパンフレットの内容は正しいとっていいのでしょうか？

次の空欄を埋めて、母平均 μ の検定をしてください。

検定の
手順1 仮説 H_0 と対立仮説 H_1 をたてます。

仮説 H_0 : 母平均 $\mu = 4500$

対立仮説 H_1 : 母平均 $\mu \neq 4500$

片側検定の方がいいのでは？

母平均: $\mu < 4500$

検定の
手順2 検定統計量 $T(\bar{x}, s^2, N)$ を計算します。

表 4.2.3 破断強度

No.	x	x^2
1	4480	20070400
2	4510	20340100
3	4570	20884900
4	4360	19009600
5	4240	17977600
6	4520	20430400
7	4260	18147600
8	4650	21622500
9	4380	19184400
10	4130	17056900
11	4530	20520900
12	4290	18404100
合計	52920	233649400

$$\text{標本平均 } \bar{x} = \frac{52920}{12} = 4410$$

$$\text{標本分散 } s^2 = \frac{12 \times \frac{233649400}{12} - \frac{52920^2}{12}}{12 \times (12 - 1)}$$

$$= 24745.45$$

したがって、検定統計量 $T(\bar{x}, s^2, N)$ は……

$$T(\bar{x}, s^2, N) = \frac{4410 - 4500}{\sqrt{\frac{24745.45}{12}}} = -1.9820$$

となります。

**検定の
手順 3** この検定統計量 $T(\bar{x}, s^2, N)$ は自由度 $(12 - 1)$ の t 分布に従うので、有意水準 $\alpha = 0.05$ の棄却域は、次のようになります。

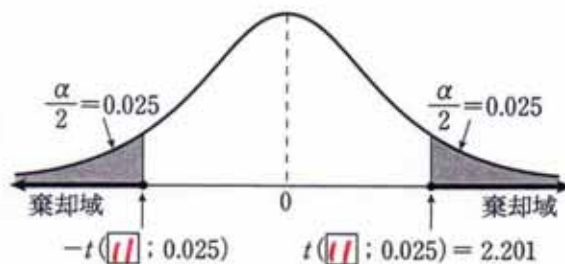


図 4.2.9 棄却域

検定統計量 $T(\bar{x}, s^2, N) = -1.9820$ は棄却域に 入らない ので、
仮説 H_0 は 棄却されない。

理解度チェック ➤ 2つの母平均の差の検定

【問題】 次のデータは、糖尿病の女性と男性の被験者 15 人ずつに対して、総コレステロール値を測定した結果です。

表 4.5.5 糖尿病患者の総コレステロール値

女性					男性				
292	351	284	278	322	265	272	248	276	284
295	282	317	305	296	258	289	307	284	273
267	272	343	298	275	301	268	293	284	318

次の空欄を埋めて、2つの母平均の差の検定をしてください。

検定の
手順 1 仮説 H_0 と対立仮説 H_1 をたてます。

仮説 $H_0: \mu_1 = \mu_2$ 対立仮説 $H_1: \mu_1 \neq \mu_2$

検定の
手順 2 検定統計量 $T(\bar{x}_1, \bar{x}_2, s^2, N_1, N_2)$ を計算します。

表 4.5.6 総コレステロール値

(a) 女性のグループ

(b) 男性のグループ

No.	x_1	x_1^2	No.	x_2	x_2^2
1	292	85264	1	265	70225
2	351	123201	2	272	73984
3	284	80656	3	248	61504
4	278	77284	4	276	76176
5	322	103684	5	284	80656
6	295	87025	6	258	66564
7	282	79524	7	289	83521
8	317	100489	8	307	94249
9	305	93025	9	284	80656
10	296	87616	10	273	74529
11	267	71289	11	301	90601
12	272	73984	12	268	71824
13	343	117649	13	293	85849
14	298	88804	14	284	80656
15	275	75625	15	318	101124
合計	4477	1345119	合計	4220	1192118

女性のグループ	男性のグループ
標本平均 $\bar{x}_1 = \frac{4477}{15} = 298.4667$ 標本分散 $s_1^2 = \frac{15 \times 134519 - 4477^2}{15 \times (15 - 1)} = 634.5524$	標本平均 $\bar{x}_2 = \frac{4420}{15} = 294.6667$ 標本分散 $s_2^2 = \frac{15 \times 101124 - 4420^2}{15 \times (15 - 1)} = 349.38095$

共通の分散 $s^2 = \frac{(15 - 1) \times 634.5524 + (15 - 1) \times 349.38095}{15 + 15 - 2} = 491.97$

検定統計量 $T(\bar{x}_1, \bar{x}_2, s^2, N_1, N_2)$

$$\begin{aligned}
 &= \frac{298.4667 - 294.6667}{\sqrt{\left(\frac{1}{15} + \frac{1}{15}\right) \times 491.9667}} = 2.115
 \end{aligned}$$

検定の

手順 3 この検定統計量 $T(\bar{x}_1, \bar{x}_2, s^2, N_1, N_2)$ は、

自由度 $(15 + 15 - 2)$ の t 分布に従うので、

有意水準 $\alpha = 0.05$ の棄却域は次のようになります。

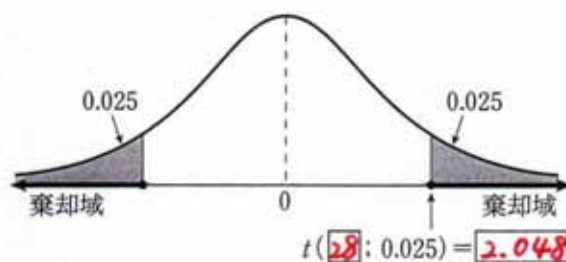


図 4.5.13 棄却域

検定統計量 2.115 は棄却域に 2.048 ので、

仮説 H_0 は ~~棄却される~~。

理解度チェック ➤ 対応のある2つの母平均の差の検定

【問題】 リンゴダイエットをする前と後の体重を測定した結果です。

表 4.6.3 ダイエット前後の体重

No.	ダイエット前	ダイエット後
1	53.0	51.2
2	50.2	48.7
3	59.4	53.5
4	61.9	56.1
5	58.5	52.4
6	56.4	52.9
7	53.4	53.3

次の空欄を埋めて、対応のある2つの母平均の差の検定をしてください。

検定の
手順1 仮説 H_0 と対立仮説 H_1 をたてます。

仮説 $H_0: \mu_1 - \mu_2 = 0$ 対立仮説 $H_1: \mu_1 - \mu_2 \neq 0$

検定の
手順2 検定統計量 $T(\bar{x}, s^2, N)$ を計算します。

表 4.6.4 体重の差

No.	x	x^2
1	1.8	3.24
2	1.5	2.25
3	5.9	34.81
4	5.8	33.64
5	6.1	37.21
6	3.5	12.25
7	0.1	0.01
合計	24.7	123.41

$$\text{標本平均 } \bar{x} = \frac{24.7}{7} = 3.529$$

$$\text{標本分散 } s^2 = \frac{7 \times 123.41 - (24.7)^2}{7 \times (7 - 1)} = 6.0424$$

$$\begin{aligned} \text{検定統計量 } T(\bar{x}, s^2, N) \\ = \frac{3.529 - 0}{\sqrt{\frac{6.0424}{7}}} = 3.7979 \end{aligned}$$

検定の
手順3 この検定統計量は、自由度 ($7 - 1$) の t 分布に従うので、

有意水準を $\alpha = 0.05$ とすると、 $t(6; 0.025) = 2.447$ となります。

よって、検定統計量 3.7979 は棄却域に $入る$ ので

仮説 H_0 は ~~棄却される~~。

検定の
手順2 検定統計量 $T(s_1^2, s_2^2)$ を計算します。

表 4.7.4 2つの証券の投資収益率

(a) 証券 A のグループ			(b) 証券 B のグループ		
No.	x_1	x_1^2	No.	x_2	x_2^2
1	-2.8	7.84	1	-6.3	39.69
2	-14.8	219.04	2	-8.5	72.25
3	-9.3	86.49	3	-2.4	5.76
4	2.7	7.29	4	4.9	24.01
5	-19.8	392.04	5	-11.6	134.56
6	-15.2	231.04	6	-2.7	7.29
7	21.4	457.96	7	10.3	106.09
8	14.2	201.64	8	6.9	47.61
9	-10.8	116.64	9	1.5	2.25
10	-4.5	20.25	10	-3.6	12.96
合計	-38.9	1740.23	合計	-11.5	452.47

$$\begin{aligned}
 \text{標本分散 } s_1^2 &= \frac{10 \times 1740.23 - (-38.9)^2}{10 \times (10 - 1)} = \frac{176.5454}{9} = 19.616167 \\
 \text{標本分散 } s_2^2 &= \frac{10 \times 452.47 - (-11.5)^2}{10 \times (10 - 1)} = \frac{48.805}{9} = 5.422778 \\
 \text{検定統計量 } T(s_1^2, s_2^2) &= \frac{s_1^2}{s_2^2} = \frac{19.616167}{5.422778} = 3.6174
 \end{aligned}$$

検定の
手順3 この検定統計量 $T(s_1^2, s_2^2)$ は自由度 $(10-1, 10-1)$ の F 分布に従うので、有意水準 $\alpha=0.05$ の棄却域は次のようになります。

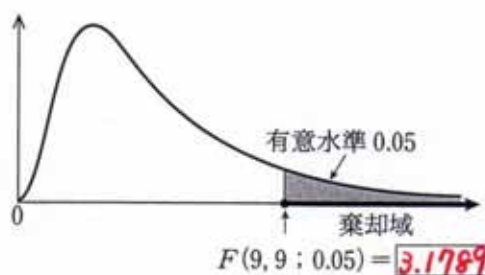


図 4.7.5 棄却域

検定統計量 $T(s_1^2, s_2^2) = 3.6174$ は棄却域に 入る ので、
仮説 H_0 は 棄却される。

理解度チェック >> 2つの母比率の差の検定

【問題】 次のデータは、ランダムに選ばれた人に対しておこなった内閣支持率の調査結果です。

女性と男性とでは、内閣支持率に差があるのでしょうか？

表 4.8.2 内閣支持率

性別	支持する	支持しない	合計
女性	175 人	225 人	400 人
男性	184 人	316 人	500 人

次の空欄を埋めて、2つの母比率の差の検定をしてください。

検定の

手順 1 仮説 H_0 と対立仮説 H_1 をたてます。

仮説 $H_0: p_1 = p_2$ 対立仮説 $H_1: p_1 \neq p_2$

検定の

手順 2 検定統計量 $T(m_1, m_2, N_1, N_2)$ を計算します。

$$\text{標本比率 } \frac{m_1}{N_1} = \frac{175}{400} = 0.4375 \quad \text{標本比率 } \frac{m_2}{N_2} = \frac{184}{500} = 0.368$$

$$\text{共通の比率 } p^* = \frac{175}{400} + \frac{184}{500} = 0.3989$$

$$\begin{aligned} \text{検定統計量 } T(m_1, m_2, N_1, N_2) &= \frac{0.4375 - 0.368}{\sqrt{0.3989 \times (1 - 0.3989) \times \left(\frac{1}{400} + \frac{1}{500}\right)}} = 2.1158 \end{aligned}$$

検定の

手順 3 この検定統計量 $T(m_1, m_2, N_1, N_2)$ は、標準正規分布に従うので、

$$\text{有意水準を } \alpha = 0.05 \text{ とすると、} z\left(\frac{\alpha}{2}\right) = 1.96$$

検定統計量 2.12 は棄却域に 1.96 ので、

仮説 H_0 は 棄却される。

理解度チェック ➤ 母相関係数の検定

【問題】 次のデータは、12 カ国でカロリー摂取量とエイズの患者数について調査した結果です。母相関係数 $\rho_0 = -0.3$ の検定をしてください。

表 4.9.3 12 カ国のカロリーとエイズ

No.	カロリー x	エイズ y	x^2	y^2	xy
1	2750	249	7562500	62001	684750
2	2956	713	8737936	508369	2107628
3	2675	1136	7155625	1290496	3038800
4	3198	575	10227204	330625	1838850
5	1816	5654	3297856	31967716	10267664
6	2233	2107	4986289	4439449	4704931
7	2375	915	5640625	837225	2173125
8	2288	4193	5234944	17581249	9593584
9	1932	7225	3732624	52200625	13958700
10	2036	3730	4145296	13912900	7594280
11	2183	472	4765489	222784	1030376
12	2882	291	8305924	84681	838662
合計	29324	27260	73792312	123438120	57831350

検定の

手順 1 仮説 H_0 と対立仮説 H_1 をたてます。仮説 $H_0: \rho = -0.3$ 対立仮説 $H_1: \rho \neq -0.3$

検定の

手順 2 検定統計量 $T(r)$ を計算します。

$$r = \frac{12 \times 57831350 - 29324 \times 27260}{\sqrt{12 \times 73792312 - 29324^2} \sqrt{12 \times 123438120 - 27260^2}} = -0.766550$$

$$T(r) = \sqrt{12} - 3 \left(\frac{1}{2} \log \frac{1+r}{1-r} - \frac{1}{2} \log \frac{1-0.3}{1+0.3} \right) = -0.915$$

-0.766 -0.3

検定の

手順 3 この検定統計量 $T(r)$ は、標準正規分布で近似されるので、

$$-z(0.025) = -1.96$$

検定統計量 -0.915 は棄却域に 入らない ので、仮説 H_0 は 棄却しない

理解度チェック ➤ 適合度検定

【問題】 サヤエンドウ豆の4種類の比は

黄色丸型：黄色角型：緑色丸型：緑色角型＝9：3：3：1

といわれています。

そこで、200本のサヤエンドウ豆をランダムに取り出し、
4種類の比を調査したところ、次のような結果を得ました。

表 4.10.4 4種類のサヤエンドウ豆

サヤエンドウ豆	黄色丸型	黄色角型	緑色丸型	緑色角形
度数	105本	38本	41本	16本

次の空欄を埋めて、適合度検定をしてください。

検定の
手順 1 仮説 H_0 をたてます。

$$\text{仮説 } H_0: p_1 = \frac{9}{16}, p_2 = \frac{3}{16}, p_3 = \frac{3}{16}, p_4 = \frac{1}{16}$$

検定の
手順 2 検定統計量 $T(f_i, N)$ を計算します。

表 4.10.5 4種類のサヤエンドウ豆の比率

エンドウ	比率 p_i	実測度数 f_i	期待度数 Np_i	$f_i - Np_i$	$(f_i - Np_i)^2$	$(f_i - Np_i)^2 / Np_i$
黄丸	0.5625	105	112.5	-7.5	56.25	0.5
黄角	0.1875	38	37.5	0.5	0.25	0.006667
緑丸	0.1875	41	37.5	3.5	12.25	0.326667
緑角	0.0625	16	12.5	3.5	12.25	0.98
合計 N	1	200	200			1.813333

$$\text{検定統計量 } T(f_i, N) = 1.81333$$

検定の
手順 3 この検定統計量 $T(f_i, N)$ は、自由度 ($4 - 1$) のカイ 2 乗分布に従うので、有意水準 $\alpha = 0.05$ の棄却域は次のようになります。

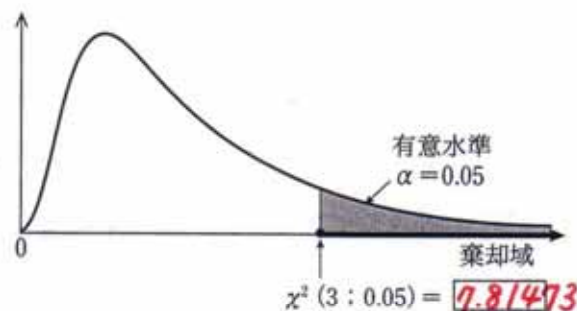


図 4.10.5 棄却域

検定統計量 $T(f_i, N) = 1.8133$ は棄却域に 入らない ので、
仮説 H_0 は 棄却されない。



.....
実測度数が 10 倍になったとすると.....

個数			
観測度数 N	期待度数 N	残差	
16.00	16	12.5	3.5
38.00	38	37.5	.5
41.00	41	37.5	3.5
105.00	105	112.5	-7.5
合計	200		

個数			
観測度数 N	期待度数 N	残差	
160.00	160	125.0	35.0
380.00	380	375.0	5.0
410.00	410	375.0	35.0
1050.00	1050	1125.0	-75.0
合計	2000		

↓

検定統計量

個数	
カイ 2 乗	1.813
自由度	3
漸近有意確率	.612



↓

検定統計量

個数	
カイ 2 乗	18.133
自由度	3
漸近有意確率	.000

理解度チェック ➤ 独立性の検定

【問題】 次のデータは、高段者剣道七段・八段の選手権大会と低段者剣道三段・四段の選手権大会において、打突部位の本数を調査した結果です。

表 4.11.7 剣道の段位と打突部位

		打突部位			
		面	小手	胴	突き
段位	七段・八段	17 本	19 本	1 本	2 本
	三段・四段	21 本	12 本	9 本	5 本

段位と打突部位という 2 つの属性の間に関連があるかどうか、独立性の検定をしてください。

検定の

手順 1 仮説 H_0 と対立仮説 H_1 をたてます。

仮説 H_0 : 段位と打突部位は 独立である

対立仮説 H_1 : 段位と打突部位の間には 関連がある

検定の

手順 2 検定統計量 $T(f_{ij}, N)$ を計算します。

計算その 1 セルの合計を計算します。

表 4.11.8 剣道の段位と打突部位

	面	小手	胴	突き	合計
七段・八段	17	19	1	2	<u>39</u>
三段・四段	21	12	9	5	<u>47</u>
合計	<u>38</u>	<u>31</u>	<u>10</u>	<u>7</u>	<u>86</u>

計算その2 各セルの $\frac{(Nf_{ij} - f_{i.}f_{.j})^2}{Nf_{i.}f_{.j}}$ を計算します。

表 4.11.9 剣道の段位と打突部位

	面	小手	胴	突き
七段・八段	0.003138	1.737213	2.755397	0.4344919
三段・四段	0.002604	1.441517	2.286393	0.3605358

計算その3 検定統計量 $T(f_{ij}, N)$ を計算します。

$$T(f_{ij}, N) = \boxed{} + \boxed{} + \boxed{} + \boxed{} \\ + \boxed{} + \boxed{} + \boxed{} + \boxed{} \\ = \boxed{9.02129}$$

すべて合計する

検定の
手順3 この検定統計量 $T(f_{ij}, N)$ は自由度 $(\boxed{2} - 1)(\boxed{4} - 1)$ の
カイ2乗分布に従うので、有意水準 $\alpha = 0.05$ の棄却域は
次のようになります。

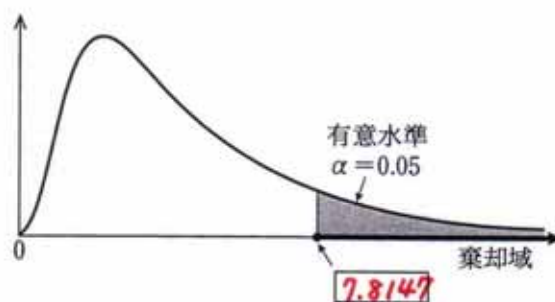


図 4.11.5 棄却域

よって、検定統計量 $\boxed{9.02129}$ は棄却域に $\boxed{入る}$ ので、
仮説 H_0 は $\boxed{棄却される}$ 。
したがって、段位と打突部位の間に関連が $\boxed{ある}$ 。

理解度チェック ➤ グラプス・スミルノフの外れ値の検定

【問題】 マンガンの融点を7回測定したところ、次の結果を得ました。

{1276 1265 1302 1283 1271 1269 1285}

この標本の中で、 $x_3=1302$ だけが大きく飛び離れたデータのように見えます。

次の空欄を埋めて、グラプス・スミルノフの外れ値の検定をしてください。

検定の
手順 1 仮説 H_0 と対立仮説 H_1 をたてます。

仮説 H_0 : $x_3=1302$ のデータは外れ値ではない

対立仮説 H_1 : $x_3=1302$ のデータは外れ値である

検定の
手順 2 検定統計量 $T(x_k)$ を計算します。

表 4.12.2 融点

No.	x	x^2
1	1276	1628176
2	1265	1600225
3	1302	1695204
4	1283	1646089
5	1271	1615441
6	1269	1610361
7	1285	1651225
合計	8951	11446721

$$\begin{aligned}
 \text{標本平均 } \bar{x} &= \frac{8951}{7} = 1278.714 \\
 \text{標本分散 } s^2 &= \frac{7 \times 11446721 - 8951^2}{7 \times (7 - 1)} \\
 &= 158.2381 \\
 \text{検定統計量 } T(x_k) &= \frac{1302 - 1278.714}{\sqrt{158.2381}} \\
 &= 1.851118
 \end{aligned}$$

検定の
手順 3 グラプス・スミルノフの数表から、 $G_7(0.05) = 1.938$ となります。

検定統計量 1.851 は棄却域に 入らない ので、

仮説 H_0 は 棄却されない。

理解度チェック >> ウィルコクソンの順位和検定

【問題】 次のデータは、コンパニオンとウェイトレスの時給について調査した結果です。

表 5.2.8 2つのグループの時給

コンパニオンのグループ ウェイトレスのグループ

No.	時給(円)
1	1400
2	1500
3	2200
4	2000
5	2500
6	2300
7	1600

No.	時給(円)
1	850
2	1000
3	1100
4	950
5	1200
6	900
7	1050
8	800

2つのグループの時給に差がありますか？

次の空欄を埋めて、ウィルコクソンの順位和検定をおこなってください。

検定の
手順1 母集団に、仮説 H_0 という対立仮説 H_1 をたてます。

仮説 H_0 : 2つのグループの時給は同じ

対立仮説 H_1 : 2つのグループの時給は異なる

検定の
手順2 検定統計量 W を計算します。

$$W = \boxed{9} + \boxed{10} + \boxed{13} + \boxed{12} + \boxed{15} + \boxed{14} + \boxed{11} = \boxed{84}$$

検定の
手順3 有意水準 $\alpha=0.05$ の棄却域は

$$W_{7,8} = \boxed{38} \quad \bar{W}_{7,8} = \boxed{74}$$

なので、検定統計量 $W = \boxed{84}$ は棄却域に 入る。

したがって、仮説 H_0 は 棄却される。

理解度チェック ➤ 回帰直線の求め方

【問題】 次のデータは、海水の温度とクルマエビの夜間の活動時間について調査した結果です。

表 6.1.6 海水温度とクルマエビの活動時間

No.	1	2	3	4	5	6	7	8	9	10
海水温度	30.4	27.2	30.9	22.5	19.0	16.4	12.1	12.7	13.7	23.6
活動時間	5.7	6.7	7.6	7.7	6.9	4.6	3.6	6.4	7.5	6.4

次の空欄を埋めて、回帰直線の式を求めてください。
海水の温度を x 、夜間の活動時間を y とします。

表 6.1.7 いろいろな統計量

No.	x	y	x^2	xy
1	30.4	5.7	924.16	173.28
2	27.2	6.7	739.84	182.24
3	30.9	7.6	954.81	234.84
4	22.5	7.7	506.25	173.25
5	19.0	6.9	361	131.1
6	16.4	4.6	268.96	75.44
7	12.1	3.6	146.41	43.56
8	12.7	6.4	161.29	81.28
9	13.7	7.5	187.69	102.75
10	23.6	6.4	556.96	151.04
合計	208.5	63.1	4807.37	1348.78

クルマエビが
1日中活動する
海水温度は
何度ですか？



$$\begin{aligned}
 \text{傾き } b &= \frac{10 \times \frac{1348.78}{208.5} - \frac{4807.37}{208.5}}{10 \times \frac{208.5}{208.5} - \frac{208.5^2}{208.5^2}} = 0.072 \\
 \text{切片 } a &= \frac{\frac{4807.37}{208.5} - \frac{1348.78}{208.5} \times \frac{208.5}{208.5}}{10 \times \frac{208.5}{208.5} - \frac{208.5^2}{208.5^2}} = 4.808 \\
 \text{回帰直線 } Y &= 4.808 + 0.072x
 \end{aligned}$$

理解度チェック ➤ 回帰分析

【問題】 次のデータは、生命保険会社 10 社の営業職員数と保険新契約高を調査した結果です。

表 6.3.3 営業職員数と保険新契約高

No.	営業職員	保険新契約高
1	43	157
2	42	158
3	38	154
4	37	148
5	36	143
6	34	135
7	33	124
8	31	116
9	30	109
10	27	105

営業職員数を x ，保険新契約高を y として

回帰直線， 決定係数， 分散分析表
を求めてください。

はじめに，次の統計量を計算します。

表 6.3.4 いろいろな統計量

No.	x	y	x^2	y^2	xy
1	43	157	1849	24649	6751
2	42	158	1764	24964	6636
3	38	154	1444	23716	5852
4	37	148	1369	21904	5476
5	36	143	1296	20449	5148
6	34	135	1156	18225	4590
7	33	124	1089	15376	4092
8	31	116	961	13456	3596
9	30	109	900	11881	3270
10	27	105	729	11025	2835
合計	351	1349	12557	185645	48246

次に、回帰直線の傾き b と切片 a を求めます。

$$\text{傾き } b = \frac{10 \times 48246 - 351 \times 1349}{10 \times 12557 - 351^2} = 3.7826$$

$$\text{切片 } a = \frac{12557 \times 1349 - 48246 \times 351}{10 \times 12557 - 351^2} = 2.1304$$

次に、決定係数 R^2 を求めます。

$$S_{yy} = 185645 - \frac{1349^2}{10} = 3664.9$$

$$S_{xy} = 48246 - \frac{351 \times 1349}{10} = 896.1$$

$$S_R = 3389.60 \times 3.7826 = 3389.60$$

$$R^2 = \frac{3389.60}{3664.9} = 0.9249$$

最後に、分散分析表を作ります。

表 6.3.5 分散分析表

変 動	平方和	自由度	平均平方	F 値
回帰による変動	3389.596	1	3389.596	98.497
残差による変動	275.304	8	34.413	
全 変 動	3664.900			

理解度チェック ➤ 指数平滑化

【問題】 次のデータは、ブナの木を伐採して丸太で出荷した量を調査した結果です。

指数平滑化を使って、一期先の予測値を求めてください。
ただし、 $\alpha = 0.7$ とします。

表 7.3.3 ブナの丸太の出荷量

時点 t	伐採量	予測値
1	2206	
2	2406	2206
3	2259	2346
4	2407	2285.1
5	2718	2370.43
6	2267	2613.729
7	2089	2371.019
8	1868	2173.606
9	1778	1959.682
10	1577	1832.505
11	1486	1653.651
12	1999	1536.295
13	1059	1860.189
14	1122	1299.357
15	1034	1175.207
16	960	1076.362
17	938	994.9086
18	854	955.0726
19	767	884.3218
20	805	802.1965
21		804.159