

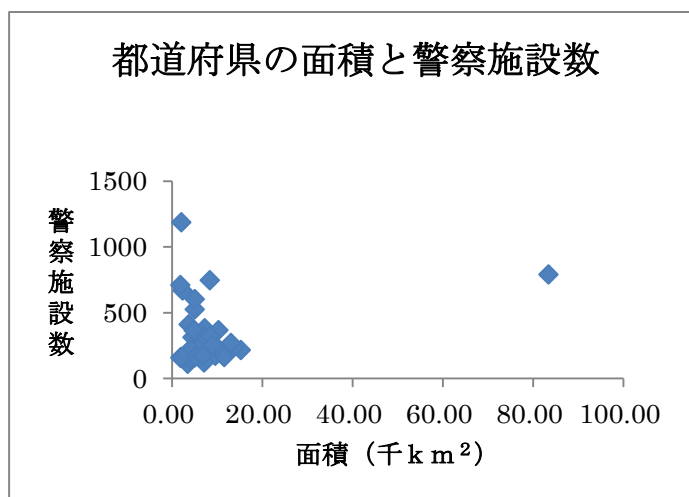
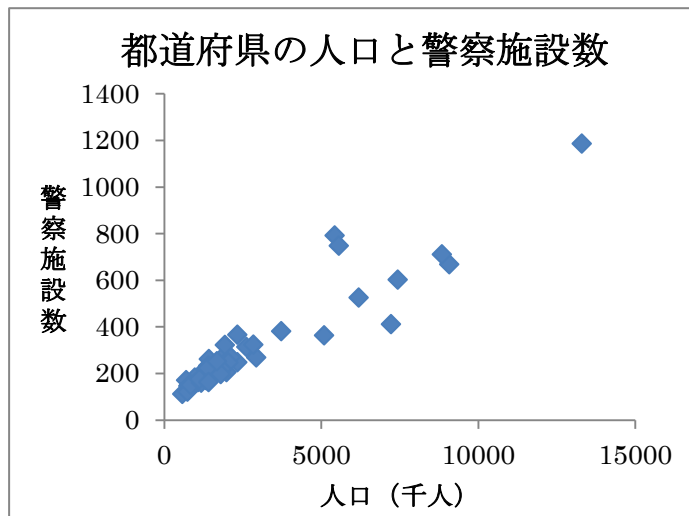
テーマ：クロス集計(その2), 冬休みの宿題の解答例

1 はじめに

テスト 第15回目(1/29)にテスト(総合演習)を行います。PCを使った問題と筆記型の試験問題の併用型です。資料(教科書、配布資料等の紙類)の持込みは可。ケータイの持込みは不可。冒頭に座席の並び替えを行います。持ち込み可です。冒頭に、座席を指定しますのでご協力下さい。

前回の演習 4 の解答例

(1) 散布図



(2) 相関係数

- ・ 人口と警察施設数の相関係数： 0.94
- ・ 面積と警察施設数の相関係数： 0.29
- ・ 面積と警察施設数の相関係数（北海道を除く）： -0.16

(注) 北海道のように「他の値から大きく外れた値」を「外れ値」といいます。相関係数は外れ値の影響を受けやすいので、「外れ値」を除外して計算すべきとされています。

(3) 相関はあると判断しますか？ ないと判断しますか？

- ・ 都道府県の人口と警察施設数： 相関あり
- ・ 都道府県の面積と警察施設数： 相関なし

2 冬休みの宿題の解答例

問1 90kmの道を、はじめの30kmは時速60kmで、次の60kmは時速40kmで走りました。平均時速を求めてください。

$$\text{最初にかかった時間は、} 30\text{km} \div 60\text{km/時} = 0.5 \text{ 時間}$$

$$\text{次にかかった時間} = 60\text{km} \div 40\text{km/時} = 1.5 \text{ 時間}$$

$$\text{よって全体でかかった時間は } 0.5 + 1.5 = 2 \text{ 時間}$$

$$\text{よって、平均速度} = 90\text{km} \div 2 = 45\text{km/時} \dots (\text{答})$$

問2 {14cm, 2cm, 13cm, 20cm, 16cm} について、以下の値を求めて下さい。

(1) 平均値

$$(14+2+13+20+16) \div 5 = 13\text{cm} \dots (\text{答})$$

(2) 分散

データ	14	2	13	20	16	合計
偏差	-1	-11	0	7	3	0
偏差の二乗	1	121	0	49	9	180

$$\text{よって分散} = 180 \div 5 = 36 \dots (\text{答})$$

(3) 標準偏差

$$\text{標準偏差} = \sqrt{36} = 6 \quad \therefore 6\text{cm} \dots (\text{答})$$

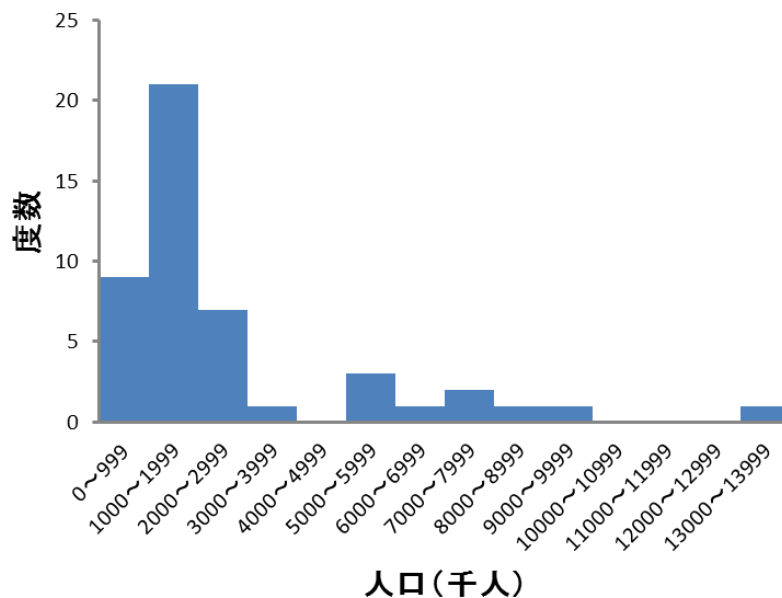
問3 以下の表は、各都道府県の人口（2016.10 現在）です。

- (1) 手書きの度数分布図を作ってください。
- (2) 分布の様子について気が付いたことを書いて下さい。

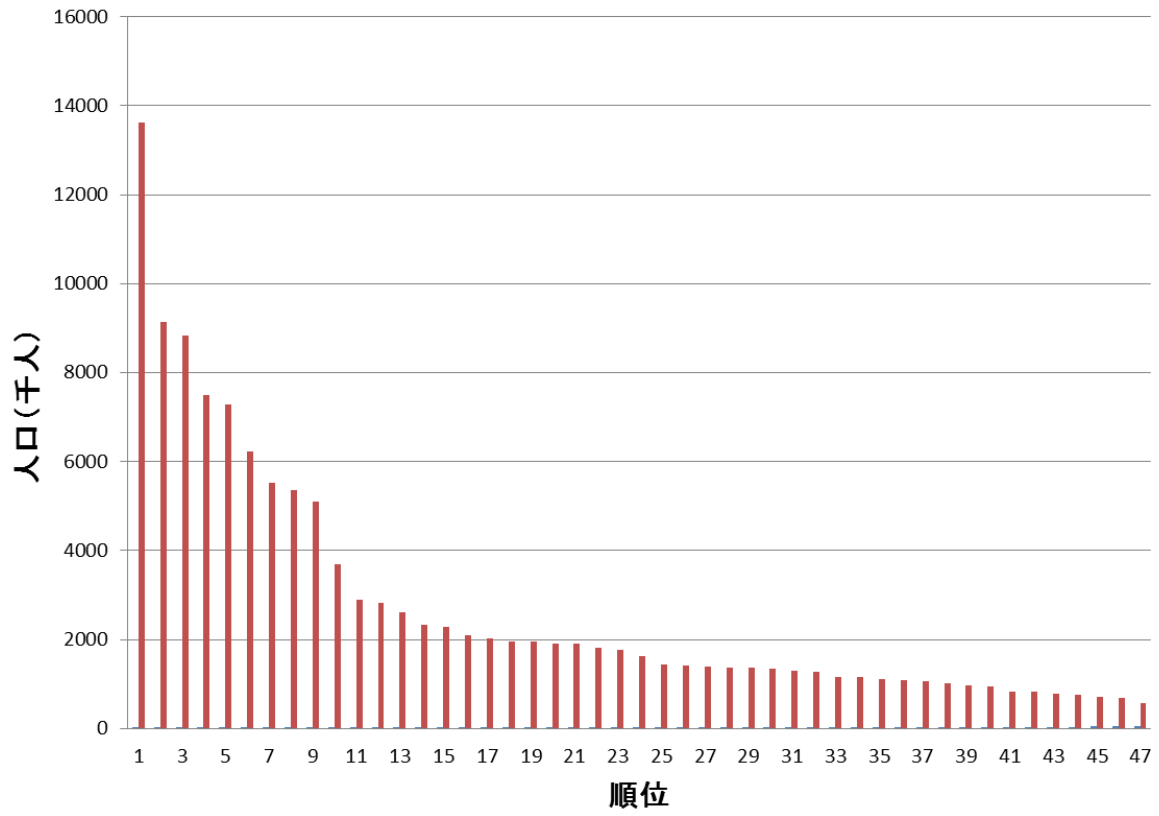
順位	都道府県	人口 (千人)
1	東京都	13624
2	神奈川県	9145
3	大阪府	8833
4	愛知県	7507
5	埼玉県	7289
6	千葉県	6236
7	兵庫県	5520
8	北海道	5352
9	福岡県	5104
10	静岡県	3688
11	茨城県	2905
12	広島県	2837
13	京都府	2605
14	宮城県	2330
15	新潟県	2286
16	長野県	2088
17	岐阜県	2022
18	群馬県	1967
19	栃木県	1966
20	岡山県	1915
21	福島県	1901
22	三重県	1808
23	熊本県	1774
24	鹿児島県	1637
25	沖縄県	1439
26	滋賀県	1413
27	山口県	1394
28	愛媛県	1375
29	長崎県	1367
30	奈良県	1356
31	青森県	1293
32	岩手県	1268
33	大分県	1160
34	石川県	1151
35	山形県	1113
36	宮崎県	1096
37	富山県	1061
38	秋田県	1010
39	香川県	972
40	和歌山県	954
41	山梨県	830
42	佐賀県	828
43	福井県	782
44	徳島県	750
45	高知県	721
46	島根県	690
47	鳥取県	570

都道府県の人口の度数分布表	
人口(千人)	度数
0～999	9
1000～1999	21
2000～2999	7
3000～3999	1
4000～4999	0
5000～5999	3
6000～6999	1
7000～7999	2
8000～8999	1
9000～9999	1
10000～10999	0
11000～11999	0
12000～12999	0
13000～13999	1

都道府県の人口の度数分布図



各都道府県の人口順位と人口の関係



問4 2014年度の体力運動能力調査結果(文部科学省)によると、19歳大学女子の握力は、平均26.6kg、標準偏差4.9kgです。握力が正規分布に従っているとすると、95パーセントの19歳大学女子の握力は、何kgから何kgの間に分布していると言えますか？

$$26.6\text{kg} - 4.9\text{kg} * 1.96 = 69.96 = 16.996 = 17.0\text{kg}$$

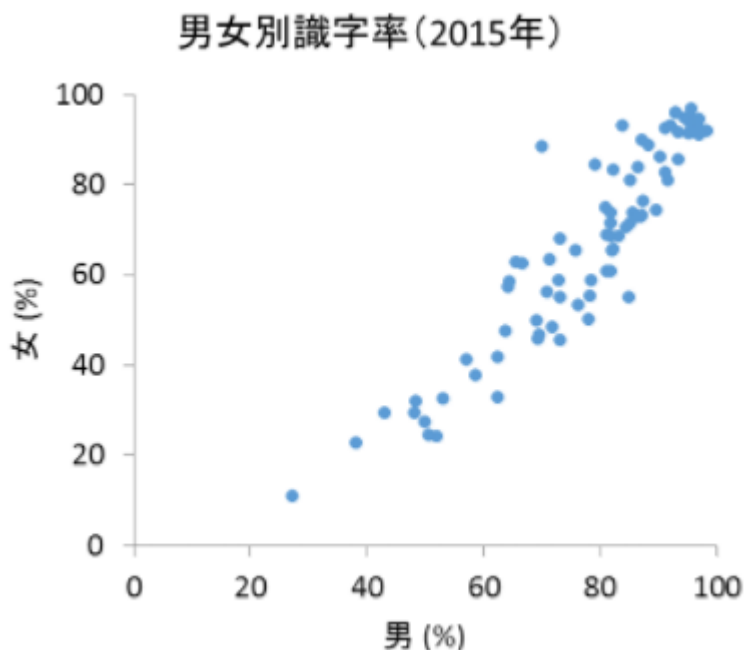
$$26.6\text{kg} + 4.9\text{kg} * 1.96 = 36.2 = 36.204 = 36.2\text{kg}$$

17.0kg から 36.2kg の間 . . . (答)

問5 A君の学校で数学の試験をしたところ、平均が54点、標準偏差が12点でした。A君の得点は69点です。A君の偏差値を求めて下さい。

$$\text{偏差値} = \frac{69-54}{12} * 10 + 50 = 62.5 \quad \dots \text{(答)}$$

問6 次の図は2015年における「様々な発展途上国における男女別識字率」(総務省「世界の統計2017」)を散布図で表したものです。識字率とは、簡単な文章を理解して読み書きできる能力がある人の割合です。この図から気づくことを書いて下さい。



男性の識字率が高い国は、女性の識字率も高い。

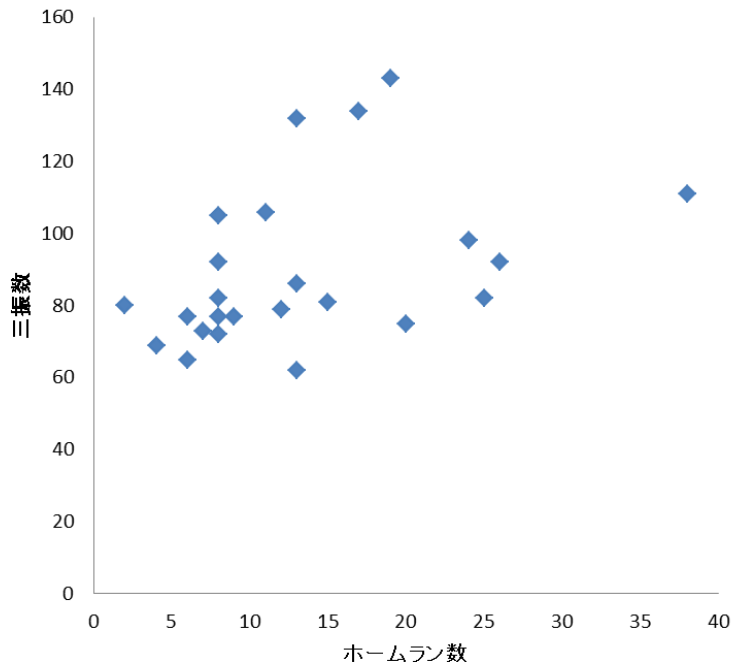
男性に比べ女性の識字率が低い国が多い

(参考) 男性の識字率と女性の識字率の相関係数は0.94である

問7 次の表は、2015年のプロ野球セリーグの打率上位24名のデータです。「ホームラン数が多い打者は三振数も多い」と言えるでしょうか？ 手書きの散布図を作って分析して下さい。

打率順位	選手	チーム名	ホームラン数	三振数
1	川端 慎吾	(ヤ)	8	72
2	山田 哲人	(ヤ)	38	111
3	筒香 嘉智	(デ)	24	98
4	ルナ	(中)	8	77
5	ロペス	(デ)	25	82
6	平田 良介	(中)	13	86
7	鳥谷 敬	(神)	6	77
8	福留 孝介	(神)	20	75
9	マートン	(神)	9	77
10	梶谷 隆幸	(デ)	13	132
11	新井 貴浩	(広)	7	73
12	田中 広輔	(広)	8	105
13	ゴメス	(神)	17	134
14	エルナンデス	(中)	11	106
15	雄平	(ヤ)	8	82
16	坂本 勇人	(巨)	12	79
17	畠山 和洋	(ヤ)	26	92
18	大島 洋平	(中)	6	65
19	バルディリス	(デ)	13	62
20	菊池 涼介	(広)	8	92
21	上本 博紀	(神)	4	69
22	長野 久義	(巨)	15	81
23	丸 佳浩	(広)	19	143
24	中村 悠平	(ヤ)	2	80

ホームラン数と三振数の関係



「ホームラン数が多いと三振数も多い」と言えるが、両者の関係は弱い正の相関である。
(参考) 相関係数は0.4程度である。

3 Excel の操作方法【ピボットテーブル】

(1) クロス集計を自動的に行ってくれる機能に「ピボットテーブル」がある。

- クロス集計表とは、たくさんの項目からなるデータのうち2つの項目に着目し、1つの項目を行に、一つの項目を列した表に集計したもの。
- ピボット (pivot) とは、回転する、要約するという意味。
(バスケットで「片方の足を軸にクルクルと方向を変える技」をピボットという。)

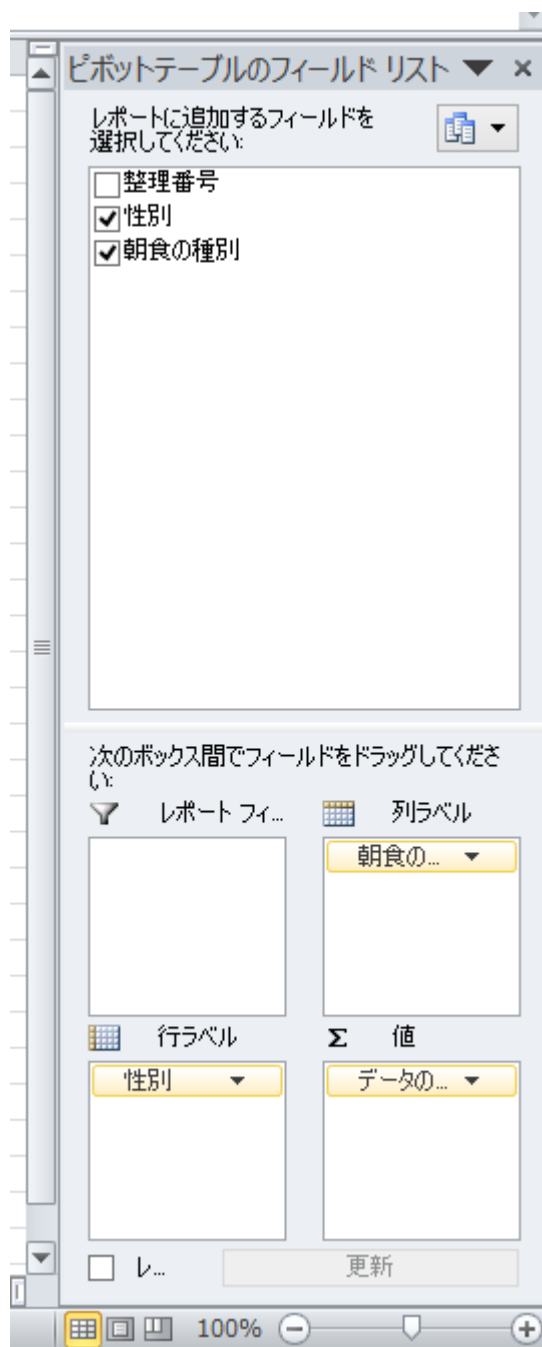
演習 1 20人の性別（男、女）と朝食の種別（パン、ご飯）に関するアンケート結果（架空のデータ）です。クロス集計表（度数）を完成させてください。

朝食アンケートの結果		
整理番号	性別	朝食の種別
1	男	パン
2	男	ご飯
3	女	パン
4	女	パン
5	男	パン
6	女	ご飯
7	男	ご飯
8	女	パン
9	男	ご飯
10	女	ご飯
11	女	パン
12	男	パン
13	女	パン
14	女	ご飯
15	男	ご飯
16	男	パン
17	女	パン
18	男	ご飯
19	男	ご飯
20	男	ご飯

(1) ピボットテーブルを挿入する

- 表中の任意のセルをクリック→「挿入」タブ→「ピボットテーブル」
 - ◇ 「テーブル／範囲」に表の範囲が表示されていることを確認
 - ◇ 「ピボットテーブルを配置する場所」で「新規ワークシート」を選択
(注)「既存のワークシート」を選択する場合は、シート名とセルを指定

(2) ピボットテーブルで集計表を作成する



- 「性別」フィールドを「行ラベル」にドラッグ
- 「朝食の種別」フィールドを「列ラベル」にドラッグ
- 「性別」フィールドを「値」にドラッグ

◇ （注）表示は、「データの個数／性別」となる。

※ 「データの個数」を集計する場合は、
このように文字形式を持つフィールド名を「値」にドラッグする。

データの個数 / 朝食の種別 行ラベル	列ラベル ご飯	パン	総計
女	3	6	9
男	7	4	11
総計	10	10	20

(3) ピボットテーブルのフィールド名（又はレコード名）の表示順を変える

- (方法1) 「行ラベル」セル（又は、「列ラベル」セル）の横の矢印ボタンをクリックし、表示順を変える（例：昇順、降順）
- (方法2) 順番を変える列名（又は行名）をクリックし、そのセルの端までカーソルを移動して、十字の形（十字の4つの先に矢印がついた十字）になったらドラッグして表示順を入れ替える。

データの個数 / 朝食の種別 行ラベル	列ラベル パン	ご飯	総計
男	4	7	11
女	6	3	9
総計	10	10	20

(4) 作成したクロス集計表（度数）から、男性と女性の差異に関するクロス集計表（横方向の割合）を作成する

- ピボットテーブルを選択→「コピー」
→貼り付け先のセルを選択し右クリック
→「貼り付けのオプション」で「値」を選択
→（「行ラベル」などの項目名を削除）

	パン	ご飯	総計
男	4	7	11
女	6	3	9
総計	10	10	20

- 別の場所に、この表をコピーを作成し、数値部分を空白にする
- クロス集計表（横方向の割合）を作成する

	パン	ご飯	総計
男	36%	64%	100%
女	67%	33%	100%
総計	50%	50%	100%

4 クロス集計（応用問題）

演習 2

- ・ 1912年4月10日に英国を出発し米国のニューヨークに向かっていた豪華客船タイタニック号が、4月14日深夜、氷山に衝突、2時間半後に沈没しました。
- ・ タイタニック号のExcelシート（別途紹介）には、
 - ・ 船内にいた人々の識別番号（1～2201）、
 - ・ 船室等級（1等、2等、3等、乗務員）、
 - ・ 性別（男、女）、大人／子供（大人／子供）、
 - ・ 生還の可否（可／否）が文字型データとして格納されています。
- ・ 「どのような種類の人々の生還の割合が大きかったのか」を明らかにしたいと思います。
- ・ 適切なピボットテーブルを、必ず「**新規ワークシート**」に作成して下さい。
そのシートに、学籍番号、氏名を入力し、印刷の上、提出して下さい。

（ヒント）

- ピボットテーブルで以下の①～③のうち1つを作成
 - ① 「行ラベル」に性別、「列ラベル」に生還の可否、「値」に生還の可否
 - ② 「行ラベル」に大人／子供、「列ラベル」に生還の可否、「値」に性別
 - ③ 「行ラベル」に船室等級、「列ラベル」に生還の可否、「値」に性別
- 2つ目のピボットテーブルを作成する場合は、
1つ目のピボットテーブルと同じシートに作成。
 - ◇ そのためには以下のとおり操作する。
 - 「ピボットテーブルを配置する場所」で「既存のワークシート」を選択
 - 場所の欄の右にあるアイコンをクリック
 - 1つ目のピボットテーブルのあるシートのタブをクリック
 - 任意のセル1つをクリック
 - 場所の欄の右にあるアイコンをクリック
 - OK
- 印刷を1枚の紙に収めるためには、「ページレイアウト」タブ
 - 「拡大縮小印刷」の右下をクリック
 - 「次のページ数に合わせて印刷」で「横1×縦1」と指定

演習 2 の解答例

性別	生還の可否 (人)		
	可	不可	総計
男	367	1,364	367
女	344	126	344
総計	711	1,490	711

性別	生還の可否 (%)		
	可	不可	総計
男	21%	79%	100%
女	73%	27%	100%
総計	32%	68%	100%

性別	生還の可否 (人)		
	可	不可	総計
大人	654	1,438	654
子供	57	52	57
総計	711	1,490	711

性別	生還の可否 (%)		
	可	不可	総計
大人	31%	69%	100%
子供	52%	48%	100%
総計	32%	68%	100%

性別	生還の可否 (人)		
	可	不可	総計
1 等	203	122	203
2 等	118	167	118
3 等	178	528	178
乗務員	212	673	212
総計	711	1,490	711

性別	生還の可否 (%)		
	可	不可	総計
1 等	62%	38%	100%
2 等	41%	59%	100%
3 等	25%	75%	100%
乗務員	24%	76%	100%
総計	32%	68%	100%