

統計データ分析士 3 種検定試験問題

2010 年 9 月

統計科学研究所

統計データ分析士資格認定機構

解答は、解答用紙に記入しなさい。

問 1. シャンプーの購入について、次の 6 項目の重要度を調べるため、下記の質問文を考えた。

①価格 ②容量 ③CMの印象 ④メーカー ⑤効果 ⑥香り

- (1) シャンプーを購入するとき、最も重視することは何ですか。ひとつだけ選んでください。
- (2) シャンプーを購入するとき、重視することを、2つ以内で選んでください。
- (3) シャンプーを購入するときに重視することは何ですか。6つの項目に重視する方から順位を付けてください。
- (4) シャンプーを購入するとき、6つの項目をどの程度重視しますか。それぞれの項目に対して、
1. 重視する 2. やや重視する 3. どちらとも言えない
4. あまり重視しない 5. 重視しない
の 5 段階で評価してください。
- (5) シャンプーを購入するときに重視することは何ですか。いくつでも選んでください。

以上の (1) ~ (5) の質問はそれぞれ何法と呼ばれていますか。以下の中から選びなさい。

- ① 単一選択法 ② 多項選択法 ③ 制限選択法 ④ 一対比較法 ⑤ 評定法
⑥ 順位法 ⑦ 自由回答法 ⑧ 数量回答法

解答例

(1)	(2)	(3)	(4)	(5)
1	3	6	5	2

問 2. 「シャンプーを購入するとき、『① 価格』から『⑥ 香り』までの 6 項目の中で重視するものを、2つ以内で選んでください」という質問を 500 人に対し行った結果、次のような単純集計の結果が得られた。ただし、下段の数字はサンプル計 (500 人) に対する比率をパーセントで示している。

サンプル計	価格	容量	CM	メーカー	効果	香り	回答数計
500	114	90	110	150	345	145	954
100 (%)	22.8	18.0	22.0	30.0	69.0	29.0	190.8

以下の各問に答えなさい。

- (1) 比率の合計が100%を超えるのはなぜか。
 (2) 1人当たり平均して何個選んだことになるか。
 (3) この表をグラフ化する場合に相応しくないものはどれか。
 ①円グラフ ②横棒グラフ ③縦棒グラフ ④折れ線グラフ

解答例

(1) 2項以内の選択数をサンプル数で単純に割ったため

(2) 1.9 個

(3) 4

問3. 性別に「シャンプーを購入するとき、『① 価格』から『⑥ 香り』までの6項目の中で重視するものを、2つ以内で選んでください」という質問のクロス集計を行った結果、次のようになった。ただし、それぞれの枠内の下段の数字は性別及び全体のサンプル計に対する比率をパーセントで示している。

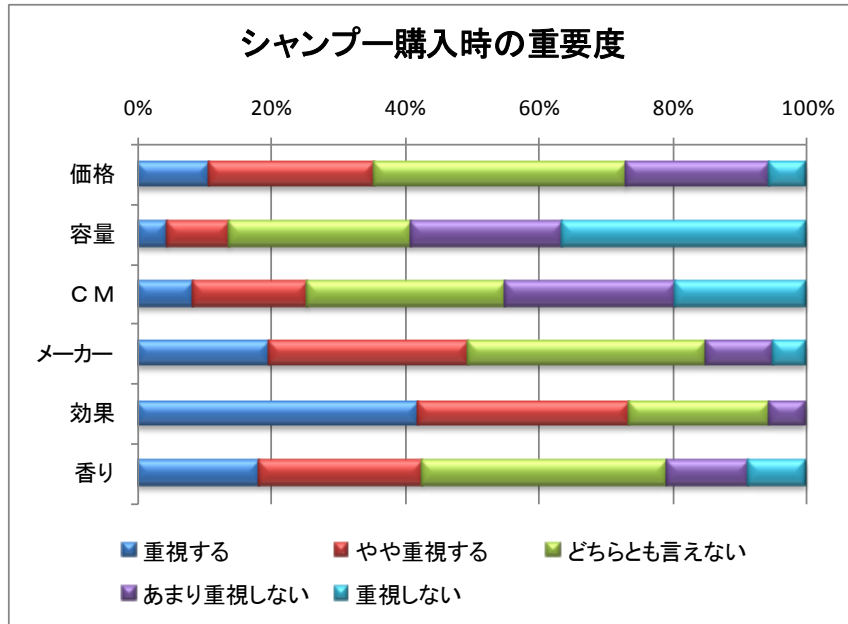
	サンプル計	価格	容量	CM	メーカー	効果	香り	回答数計
計	500	114	90	110	150	345	145	954
	100 (%)	22.8	18.0	22.0	30.0	69.0	29.0	190.8
男	200	70	60	50	30	120	40	370
	100 (%)	35.0	30.0	25.0	15.0	60.0	20.0	185.0
女	300	44	30	60	120	225	105	584
	100 (%)	14.7	10.0	20.0	40.0	75.0	35.0	194.7

このクロス集計の結果から、性別にどのような違いがあることが分かるか、述べなさい。

解答例

男女とも、効果を一番に重視する一方で、次に重視するものが男性では、価格や容量であるのに対し、女性はメーカーや香りを重視していることが分かる。

問4. 以下の図は、シャンプーを購入する際に重視する事柄に関して、『① 価格』から『⑥ 香り』までの6項目を、「1. 重視する」から「5. 重視しない」までの5段階で評価してもらった結果をグラフにしたものである。



以下の各問に答えなさい。

- (1) 上の図からどのようなことが分かるか、述べなさい。
- (2) 特徴をより明確にするために、上のグラフを改善するとしたら、上記6項目をどのような順に並べればよいか。答えは、6項目を上から順番に書きなさい。

解答例

(1)

日用品なので実際の使用結果を重視、そのうえでブランドにこだわっている。また、価格は重視する(やや重視するを含む)と重視しない(あまり重視しない)が同じ程度になっている。

(2)

(1)	(2)	(3)	(4)	(5)	(6)
効果	メーカー	香り	価格	CM	容量

問5. 文章や数表などで事柄を表現するだけではその内容が伝わりにくいとき、視覚的な表現としてグラフ表現を活用することによって、次のような内容のポイントを容易に把握することができる。

- (1) データの時間的な変化を視覚的に表現するのに適している。例えば、年度別、月別、日別の売上高やシェアの変化を時系列的に表現する場合など。
- (2) 同一の種類的项目同士を比較するのに適している。例えば、ある時期における支店ごとの売上高の比較やセールスマンごとの比較など。
- (3) 全体の中に占める構成比率を表現するのに適している。例えば、支店別売上高の全社に対する構成比率を表現する場合など。

以上の(1)～(3)は、次のどのようなグラフが相応しいか、番号を記入しなさい。

- ①レーダーチャート ②棒グラフ ③線グラフ ④散布図 ⑤円グラフ

解答例

(1)	(2)	(3)
3	2	5

問6. 次の表は新生児の生後1日目のヘモグロビンレベル (g/dl) のデータである。平均値、中央値、分散、標準偏差、変動係数を計算しなさい。

No.	1	2	3	4	5	6	7	8	9	10	11	12	13
生後1日目	16.2	17.2	29.8	22.0	23.6	26.4	29.2	21.4	25.5	18.6	22.4	16.2	27.1
No.	14	15	16	17	18	19	20	21	22	23	24	25	26
生後1日目	15.4	23.2	25.0	17.6	22.2	25.6	26.0	21.2	20.2	16.6	25.0	28.4	20.5
No.	27	28	29	30	31	32	33	34	35	36	37	38	39
生後1日目	22.3	23.8	20.4	16.0	20.2	24.7	22.0	21.0	19.7	22.0	20.0	20.0	24.4
No.	40	41	42	43	44	45	46	47	48	49	50		
生後1日目	19.2	21.2	25.0	25.0	23.2	25.2	23.2	21.2	24.4	21.2	21.2		

問7. 次の表は新生児の生後24週目のヘモグロビンレベルのデータ (g/dl) である。平均値、中央値、分散、標準偏差、変動係数を計算しなさい。

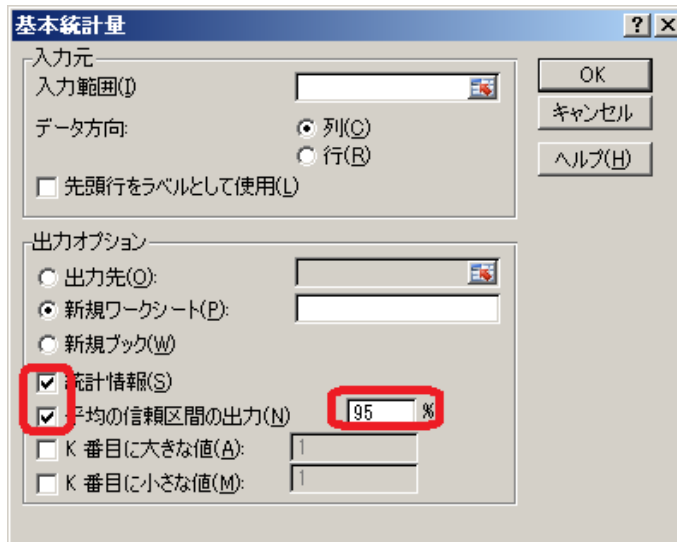
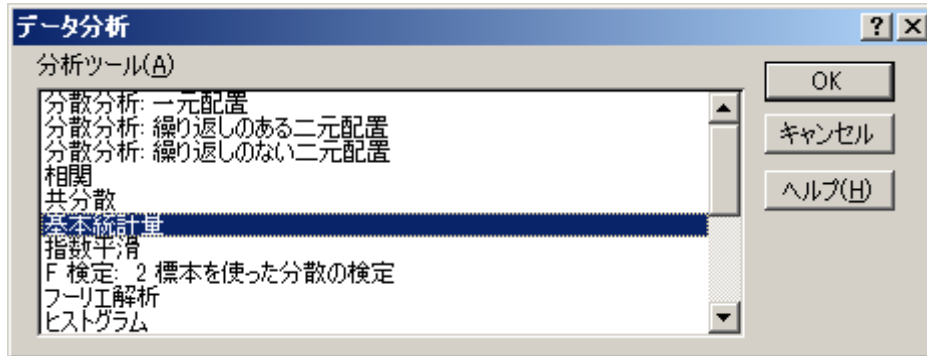
No.	1	2	3	4	5	6	7	8	9	10	11	12	13
生後24週目	11.6	9.6	12.1	13.8	12.2	12.8	15.0	11.2	11.2	11.6	13.1	12.8	11.9
No.	14	15	16	17	18	19	20	21	22	23	24	25	26
生後24週目	13.3	11.0	10.2	14.2	12.5	13.8	12.6	12.9	10.4	14.4	11.7	13.6	11.1
No.	27	28	29	30	31	32	33	34	35	36	37	38	39
生後24週目	10.6	10.2	13.8	12.8	11.0	14.5	10.8	11.9	11.5	10.7	13.2	14.7	12.5
No.	40	41	42	43	44	45	46	47	48	49	50	51	52
生後24週目	13.0	13.4	12.6	12.8	13.8	12.2	12.6	10.8	12.3	11.7	13.5	10.8	13.6
No.	53	54	55	56	57	58	59	60	61	62			
生後24週目	13.0	11.5	11.0	12.0	12.4	12.6	13.7	11.8	13.2	12.9			

注意) 分散を計算する場合の分母は、(データの個数-1)で計算してください。以下も同じです。

解答例

	問6	問7
平均値	22.18	12.35
中央値	22.00	12.50

分散	11.86	1.54
標準偏差	3.44	1.24
変動係数	15.5%	10.0%



問 8. 次の表は新生児の生後 1 日目 (x) と生後 24 週目 (y) の対になったヘモグロビンレベルのデータである。

No.	1	2	3	4	5	6	7	8
生後 1 日目	16.2	17.2	26.4	25.5	16.2	27.1	15.4	17.6
生後 24 週目	11.6	9.6	11.2	11.9	12.5	12.9	14.4	11.1
No.	9	10	11	12	13	14	15	16
生後 1 日目	25.6	16.6	25.0	16.0	21.0	23.2	21.2	21.2
生後 24 週目	13.8	11.0	14.5	12.5	12.6	11.5	12.4	12.6

- (1) 相関係数を計算しなさい。
- (2) 相関図を描いてみて、1 日目 (x) と 24 週目 (y) との間に、直線的な関連性が認められるか否かを考察し、①と②どちらかを記号で答えなさい。

①直線的な関連性は認められると思う。

②直線的な関連性は認められないと思う。

(3) 問6、問7、問8の数値結果から、気の付いたことを自由に述べなさい。

注：成人の参考基準値 男 14-18 g/dl 女 11.5-15.5 g/dl と言われている。

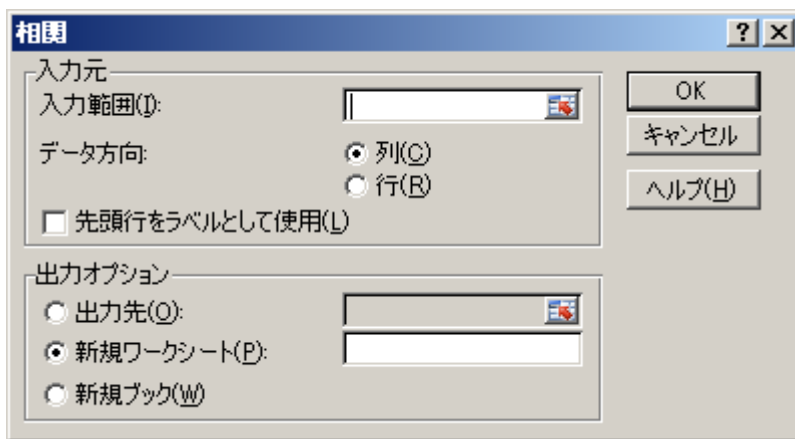
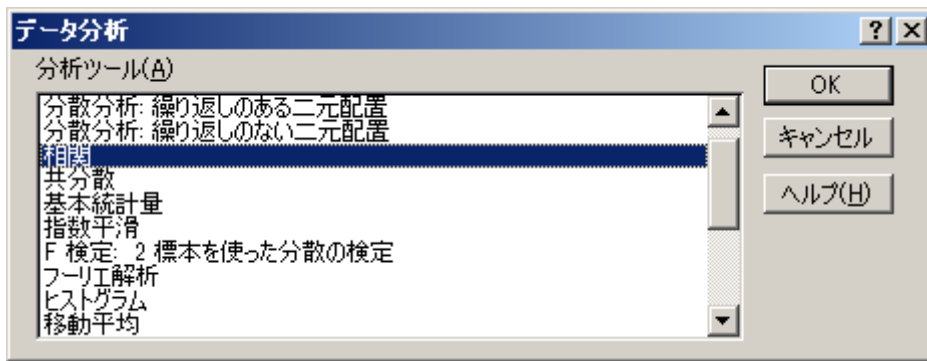
赤血球の約 2/3 は水で、残り 1/3 はヘモグロビンである。ヘモグロビンは鉄を含み、酸素を運搬している。ある程度以上ヘモグロビンが減少すると動悸・めまい・息切れなどの貧血症状が出現する。一般に、男性では 13 g/dl 以下、女性では 11 g/dl 以下を貧血とされている。

解答例

(1) 相関係数 $r = 0.234274$

(2) 2

(3) 生後1日目の方がばらつきが大きい。生後24週目に入るとばらつきが小さくなる。相関係数からは、相関は弱いあるいはほとんどないように思われる。



問9. 関西人と関東人で味の好みに違いがあるか、20人に以下のような設問のアンケートを実施した。

設問1 どちらの味付けが好みか。

1. 濃い味
2. うす味

設問2 あなたの出身エリアはどこか。

1. 関東 2. 関西

以下は上記質問に対する回答結果である。

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
設問1	2	1	2	1	1	2	1	2	1	2	1	2	2	1	2	1	2	1	2	2
設問2	2	1	2	2	1	2	2	1	1	2	1	1	2	1	1	1	2	1	2	2

- (1) 上記のデータから、設問1と設問2の回答をクロス集計しなさい。
 (2) 味の好みが出身エリアによって違いが見られるか否かを考察しなさい。

解答例

(1)

件数の集計表

		全体	味の好み	
			濃い味	うす味
合計		20	9	11
出身	関東	10	7	3
	関西	10	2	8

%の集計表

		全体	味の好み	
			濃い味	うす味
合計		100%	45	55
出身	関東	100%	70	30
	関西	100%	20	80

(2)

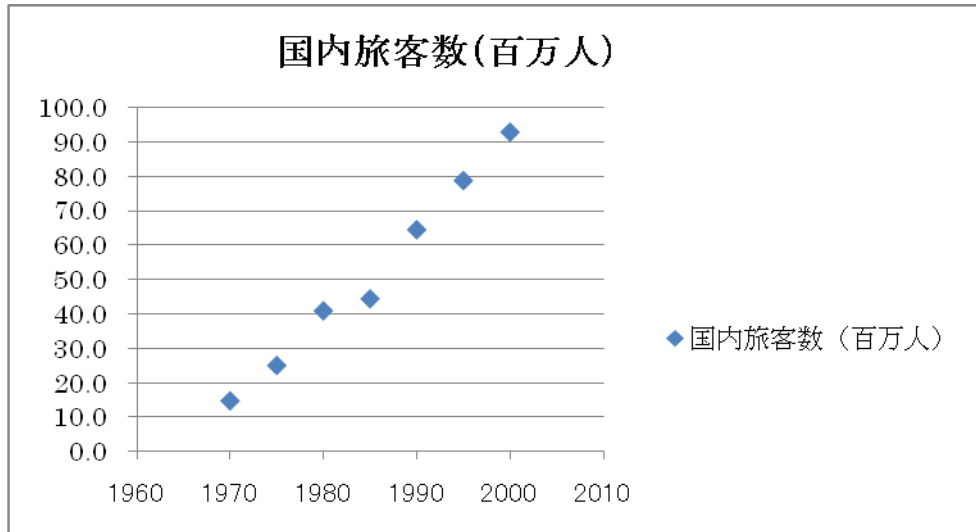
考察

濃い味の好みは全体では半数弱だが出身地別で見ると関東が関西よりも濃い味を好むことが示されている。

問10. 次の表は日本の航空旅客数のデータである。回帰直線を表す式を求めなさい。

年度	1970	1975	1980	1985	1990	1995	2000
国内旅客数(百万人)	14.7	25.0	40.9	44.4	64.5	78.8	92.9

(国土交通省「航空輸送統計年報」より)



解答例

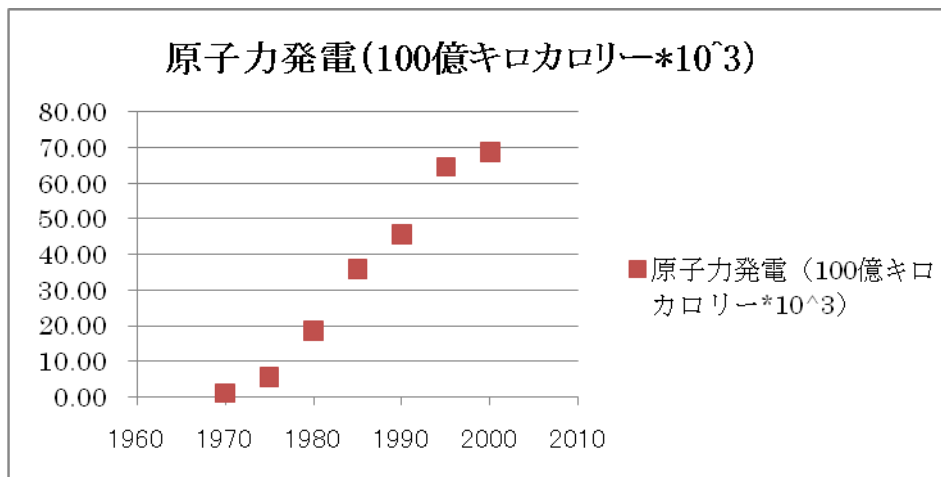
回帰直線 $y = 2.61x - 5134.9$

問 1 1. 次の表は日本の原子力発電のデータである。回帰直線を表す式を求めなさい。

(単位：100 億キロカロリー×10³)

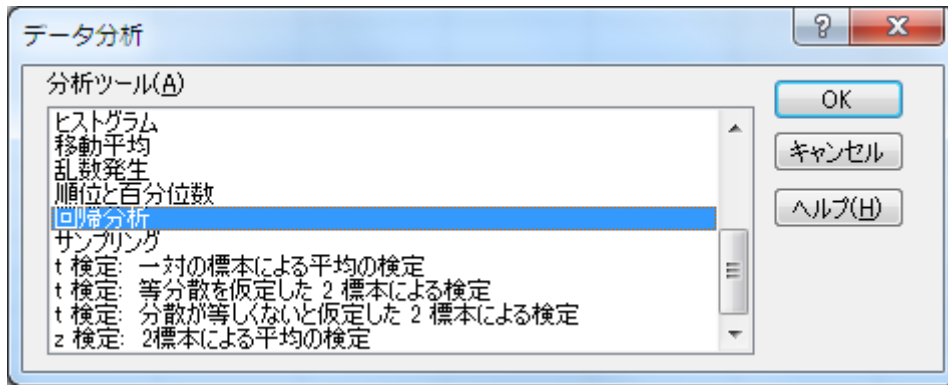
年度	1970	1975	1980	1985	1990	1995	2000
原子力発電	1.05	5.65	18.58	35.91	45.51	64.50	68.64

(資源エネルギー庁「総合エネルギー統計」より)



解答例

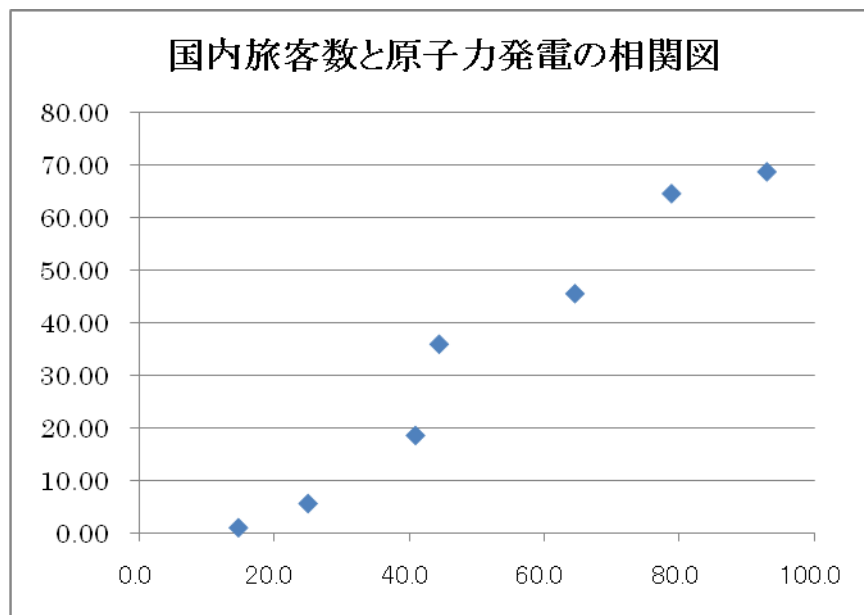
回帰直線 $y = 2.48x - 4891.4$



問 1 2. 次の表は日本の航空旅客数と原子力発電のデータである。相関係数を求めなさい。

(原子力発電単位：100 億キロカロリー×10³)

年度	1970	1975	1980	1985	1990	1995	2000
国内旅客数 (百万人)	14.7	25.0	40.9	44.4	64.5	78.8	92.9
原子力発電	1.05	5.65	18.58	35.91	45.51	64.50	68.64



解答例

相関係数 $r=0.981631$

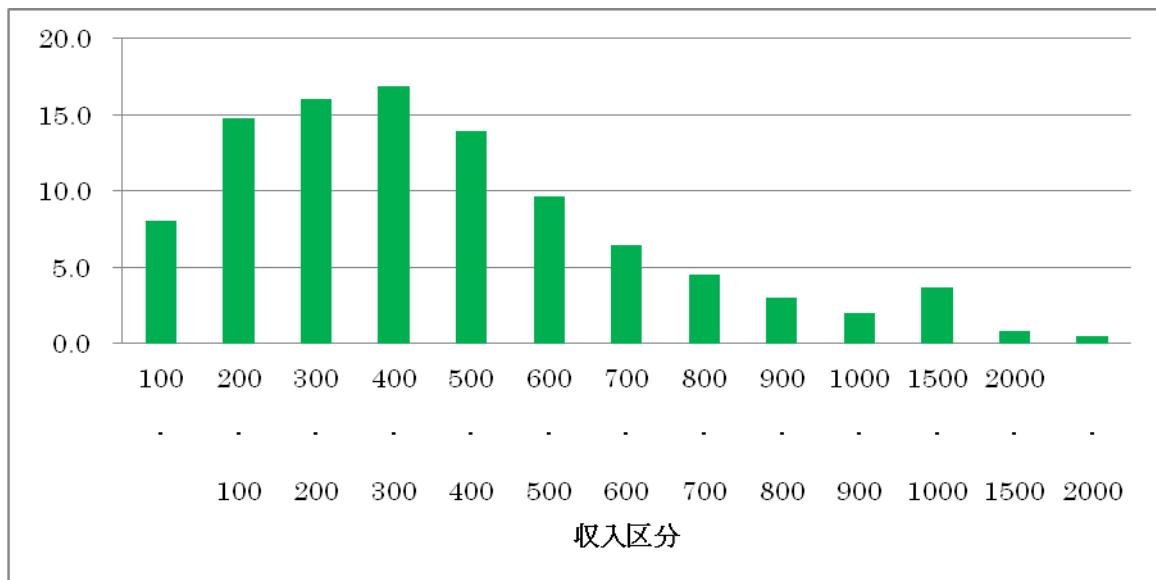
問 1 3. 次のデータは年収別給与所得者の分布である。中央値を求めなさい。また、第 1 四分位数、第 3 四分位数を求めなさい。

年収別給与所得者の分布 (2006 年)

収入区分 (万円)	人数 (千人)	割合 (%)
～ 100	3,605	8.0

100 ～ 200	6,623	14.8
200 ～ 300	7,180	16.0
300 ～ 400	7,562	16.9
400 ～ 500	6,250	13.9
500 ～ 600	4,313	9.6
600 ～ 700	2,859	6.4
700 ～ 800	2,002	4.5
800 ～ 900	1,329	3.0
900 ～ 1000	881	2.0
1000 ～ 1500	1,655	3.7
1500 ～ 2000	364	0.8
2000 ～	223	0.5

(国税庁「民間給与の実態調査結果」より)



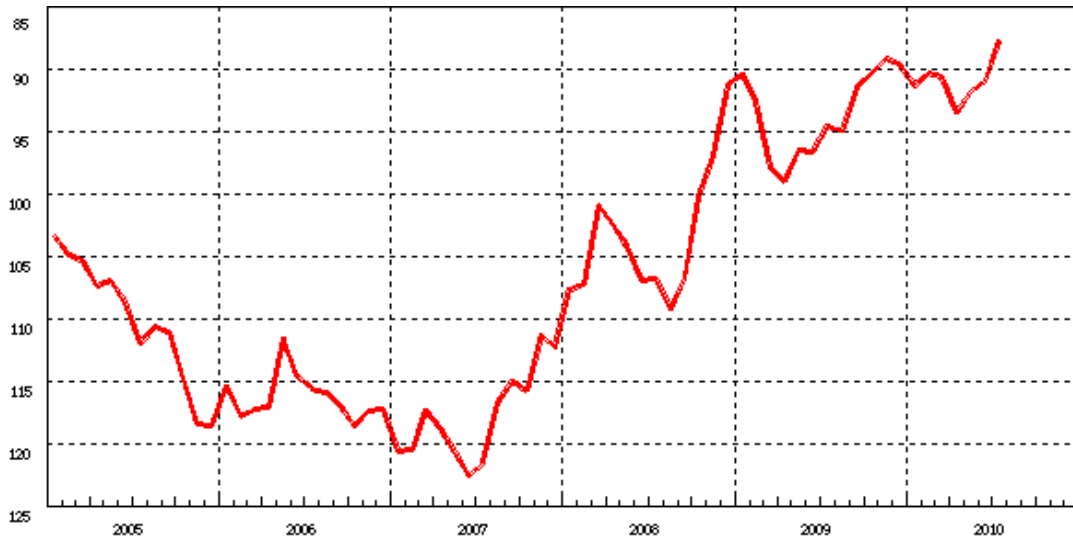
中央値	366.3 万円
第1四分位数	213.7 万円
第3四分位数	556.0 万円

問14. 下記は1ドル当たりの円価額です。2009年1月から12月までの5項移動平均値を算出し、その結果、グラフがどう変わったかを述べなさい。

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
2005年	103.27	104.84	105.3	107.35	106.94	108.62	111.94	110.65	111.03	114.84	118.45	118.6
2006年	115.33	117.81	117.31	117.13	111.53	114.57	115.59	115.86	117.02	118.59	117.33	117.26
2007年	120.59	120.49	117.29	118.81	120.77	122.64	121.56	116.74	115.01	115.77	111.24	112.28
2008年	107.6	107.18	100.83	102.41	104.11	106.86	106.76	109.24	106.71	100.2	96.89	91.21

2009年	90.35	92.53	97.83	98.92	96.43	96.58	94.49	94.9	91.4	90.28	89.11	89.52
2010年	91.26	90.28	90.56	93.43	91.79	90.89	87.67					

(日本銀行「金融市場に関する統計」より)



問 1 5. 以下は近年の日本の広告費の推移である。

(1) 下記の表から媒体別広告費の変化について考察しなさい。

年度	金額(億円)	前年比	新聞	雑誌	ラジオ	テレビ	衛星メディア	インターネット
2005	68235	102.9	98.3	99.3	99.1	99.9	111.7	148.0
2006	69399	101.7	96.2	98.7	98.1	98.8	111.7	127.8
2007	70191	101.1	94.8	96.0	95.8	99.1	110.8	124.4
2008	66926	95.3	87.5	88.9	92.7	95.6	112.1	116.3
2009	59222	88.5	81.4	74.4	88.4	89.8	104.9	101.2

(電通「日本の広告費」より)

(2) 2007年度と2009年度の円グラフを描きなさい。

媒体別広告費 (%)

年度	テレビ	インターネット	新聞	雑誌	ラジオ	衛星メディア	S P 広告
2007	28.5	8.6	13.5	6.5	2.4	0.8	39.7
2008	28.5	10.4	12.4	6.1	2.3	1.0	39.3
2009	29.0	11.9	11.4	5.1	2.3	1.2	39.1

新聞	全国日刊紙、業界紙の広告料および新聞広告制作費
雑誌	全国月刊誌、週刊誌、専門誌の広告料および雑誌広告制作費
ラジオ	全国民間放送の電波料および番組制作費とラジオCM制作費 [注、事業費は含まない]
テレビ	全国民間放送の電波料および番組制作費とテレビCM制作費 [注、事業費は含まない]

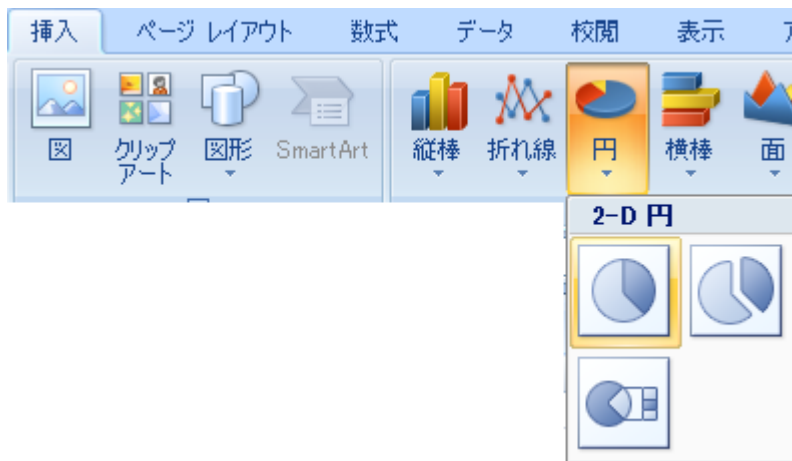
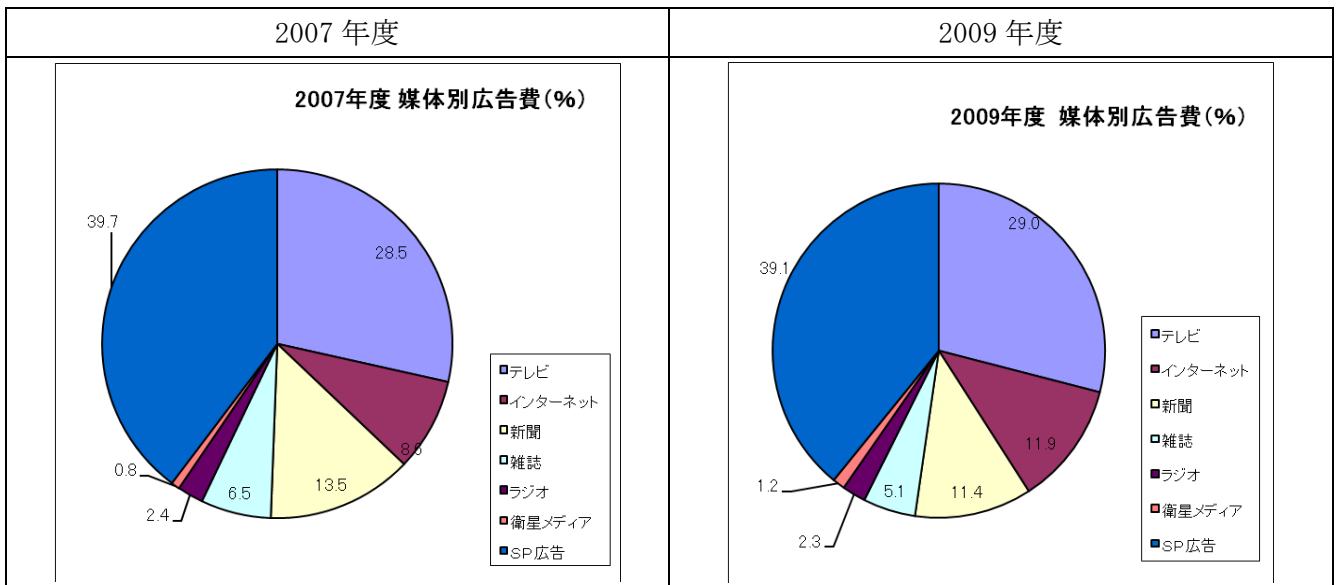
衛星メディア関連広告費	衛星放送、CATV、文字放送などに投下された広告費（媒体費および番組制作費）
インターネット広告費	インターネットサイト上の広告掲載費（モバイル広告を含む）および広告制作費
S P 広告	DM、折り込み、屋外、交通、POP、電話帳、展示・映像などの広告費

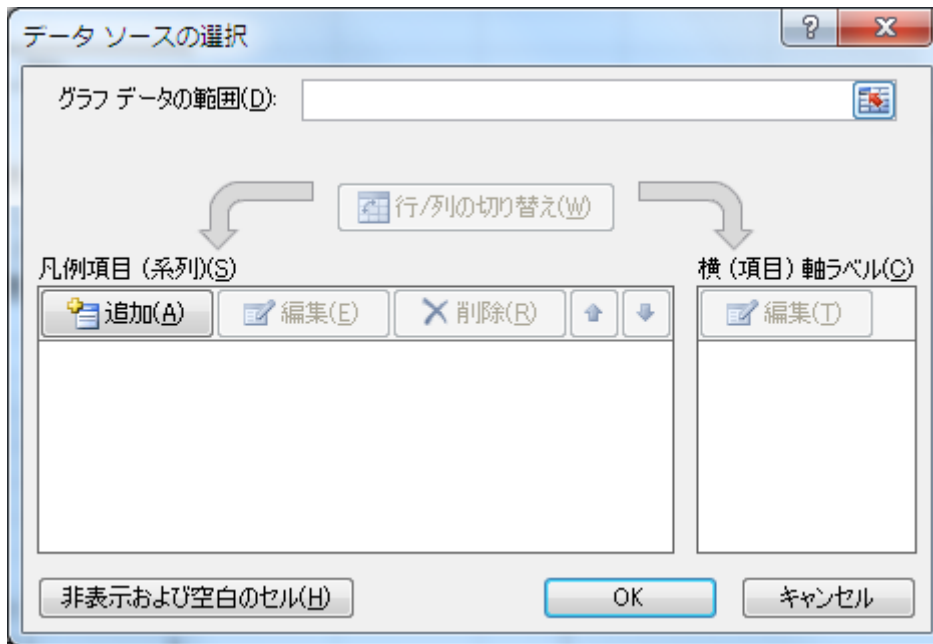
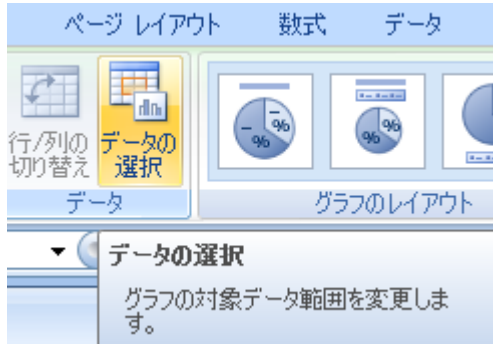
解答例

(1)

考察

テレビ、ラジオ、衛星の広告費の伸び率の落ち込みが全体の伸び率に大きな影響を与えている。
新聞、雑誌、インターネットの伸び率の落ち込みは緩やかであった。





問 16. 下記の数値は、東京都心における 2009 年、2010 年の 7 月、8 月の最高気温のデータである。各月の平均気温を求めなさい。また、各月の第 1 四分位数、第 3 四分位数を求めなさい。

東京地区最高気温 (2009 年 7 月 1 日～7 月 31 日)										
25.7	22.6	26.3	27.6	26.3	24.7	31.5	27.8	30.5	30.3	27.4
29.2	33.6	31.3	33.6	34.2	28.4	29.8	33.9	30.0	23.7	27.7
29.7	27.9	31.9	33.8	30.9	28.4	31.2	33.2	26.5		

東京地区最高気温 (2009 年 8 月 1 日～8 月 31 日)										
29.0	26.4	30.3	29.6	30.0	30.5	33.2	30.4	32.1	28.6	29.1
30.6	31.7	30.5	31.7	33.2	30.8	30.1	31.3	31.6	32.5	31.6
30.7	30.2	27.7	27.6	29.6	31.4	32.3	27.1	22.2		

東京地区最高気温 (2010 年 7 月 1 日～7 月 31 日)										
29.5	31.0	28.8	31.6	30.3	30.6	27.8	30.1	27.8	31.1	28.9

29.1	27.2	31.3	31.4	31.9	32.1	31.7	34.5	34.5	36.3	36.1
35.7	35.8	34.4	33.3	33.8	34.2	27.9	29.2	32.2		

東京地区最高気温 (2010年8月1日～8月31日)										
34.1	32.2	33.3	33.8	33.4	33.6	33.2	32.5	28.9	30.5	31.9
32.3	30.3	32.6	35.5	36.3	37.2	35.1	33.1	31.8	32.8	34.6
35.1	33.9	33.8	34.7	33.8	34.7	34.4	34.8	35.8		

(気象庁より)

解答例

	2009年		2010年	
	7月	8月	7月	8月
平均気温	29.34	30.12	31.62	33.54
第1四分位数	27.50	29.35	29.35	32.55
第3四分位数	31.40	31.60	34.00	34.70

問17. 以下は2009年の月平均小売価格のデータである。それぞれの平均、分散と標準偏差を求めなさい。また、標準偏差のそれぞれの値からわかることを述べなさい。

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
干しのり(10枚)	448	450	453	454	454	456	458	452	457	460	455	458
レタス(1kg)	629	449	452	460	384	324	310	523	449	279	283	426
トマト(1kg)	693	658	748	739	674	559	532	579	658	579	540	605

(総務省統計局「小売物価統計調査」より)

解答例

	干しのり(10枚)	レタス(1kg)	トマト(1kg)
平均	454.58	414.00	630.33
分散	12.27	10860.18	5622.61
標準偏差	3.50	104.21	74.98

標準偏差のそれぞれの値からわかることを述べよ。

干しのりは、レタスやトマトに比べてはるかに平均小売価格のばらつきが小さい。

レタスやトマトなどの価格は気候などの変化により変動しやすいが、干しのりはそうでないことが分かる。

以上で問題は終了です