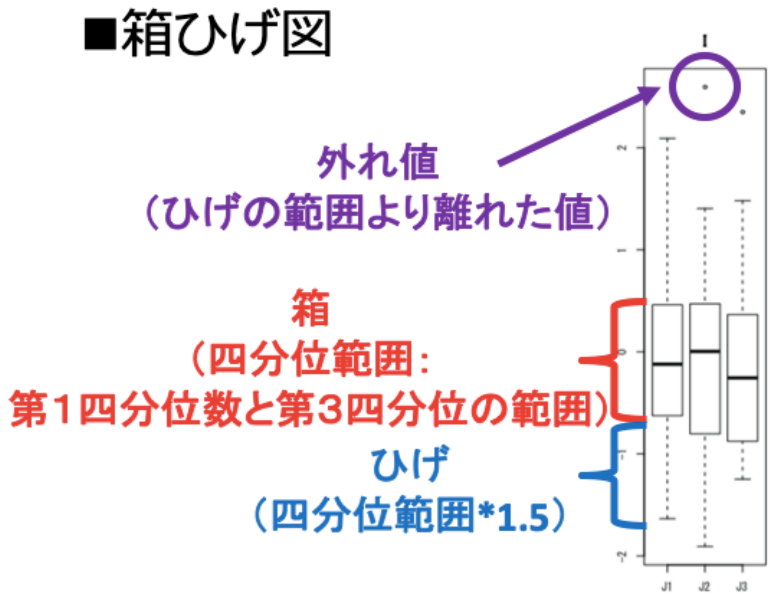
一番優しい、医薬品開発に必要な統計学の教本

統計検定２級の問題解説

2018 年 6 月の問題

2018年6月に実施された問題の解説です。
問題を含めた解説は、無料メルマガに登録していただくと入手できます。
https://best-biostatistics.com/mail

問１（１）
３つの箱ひげ図から適切なものを選ぶ問題。
箱ひげ図に関しては、以下の図を参照ください。



また、偏差の定義は「個々のデータ-平均値」です。
標準化とは、「偏差/平均値」で示されます。（標準化後のデータを Z で示すことが多いです。）

$$Z = \frac{\text{データ} - \text{平均値}}{\text{標準偏差}}$$

標準化することによるメリットは、標準化することで平均値 0、標準偏差 1 の正規分布データになることです。
そのため、-2～2 の間に、95%のデータが入ることになります。（正規分布の性質です。）

I：標準化得点のグラフ
縦軸が-2～2 のグラフです。すなわち、標準化得点のグラフということです。
II：偏差のグラフ
縦軸が 0 を中心に、約-25～30 の間に散らばっています。

すなわち、偏差のグラフを示しています。
III：総得点のグラフ
縦軸が 20～80 の間に散らばっています。
すなわち、総得点のグラフです。

問１（２）
この問題では、標準化の性質を使います。
標準化することで、平均 0、標準偏差 1 の正規分布のデータになります。
つまり、「標準偏差の 2 倍より離れた値があるかどうか」は、「標準化データのグラフで、-2 より小さいデータまたは 2 より大きいデータ、があるかどうか」と言い換えることができます。
I が標準化データのグラフですので、そこから読み取ると、1 つだけ 2 より大きい外れ値があります。
そのため、１つ、という答えになります。

問２（１）

I：◯

横軸（人口）が大きくなればなるほど、縦軸（常設映画関数）が大きくなる傾向があります。

また、その傾向は直線に近いため、正の相関があると言えます。

II：◯

相関係数は外れ値の影響を受けやすいため、散布図を一緒に見る必要があります。

III：×

正の相関は見ることができました。

相関に関しての詳細はこちらの記事をご参照ください。

https://best-biostatistics.com/correlation_regression/correlation.html

問２（２）

I：×

北海道は人口が 500 万人の付近にあります。

500 万人の付近のデータだけを見ると、北海道は一番上にありますので、一般病院病床数は多いです。

II：◯

変動係数は標準偏差を平均で割ったものです。

人口 1 人当たりの一般病院病床数の変動係数は、一般病院病床数の変動係数よりも小さくなることが予想されます。

例えば、以下のようなデータがあったとします。（数値は適当です）

病院数	人口	1 人あたりの病院数
50.0	5.0	10.0
40.0	3.0	13.3
30.0	7.0	4.3
50.0	5.0	10.0
40.0	8.0	5.0

70.0	5.0	14.0

このとき、変動係数（CV）は以下のようになります。

	Mean	SD	CV
病院数	46.7	13.7	3.4
1 人あたりの病院数	9.4	4.1	2.3

1 人あたりの病院数のほうが、CV が小さくなることがわかります。

変動係数に関しての詳細は、こちらの記事をご参照ください。
https://best-biostatistics.com/summary/cv.html

III：×
人口が多い都道府県に限っても、正の相関が見られます。

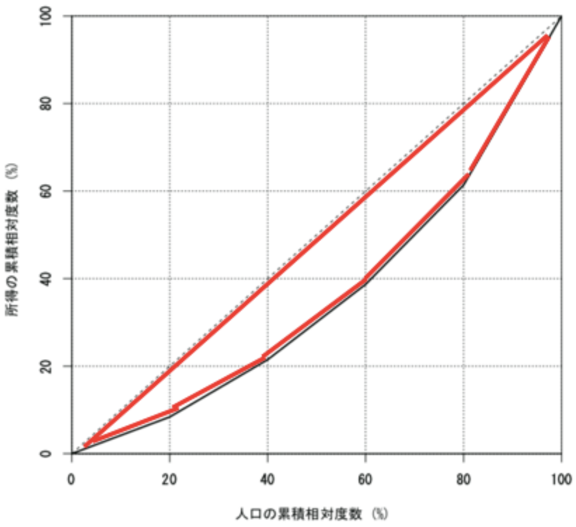
問２（３）
I：◯
偏相関係数とは、第 3 の因子の影響を除いた相関係数のことです。
今回の場合、常設映画関数と一般病院病床数に関して、人口の影響を取り除いた相関係数を指しています。
II：◯
残差 e1 と残差 e2 の散布図を見ると、相関があるようには見えません。
つまり、人口の影響を除くと相関が見られなくなることから、常設映画関数と一般病院病床数の相関は、擬相関であったことがわかります。
III：×
病院と併設している映画館のデータはどこにもないため、ここからは読み取れません。

問３（１）
ローレンツ曲線とは、横軸に世帯（人口）の累積度数を、縦軸に所得の累積度数を示した曲線である。
ローレンツ曲線の特徴は、所得格差が存在しないなら 45 度線と一致し、所得に偏りがあると、曲線は下に膨らむ、ということ。
与えられた表は各階級の度数であるため、この表から累積度数分布表を作成する。
すると、以下のような表ができる。

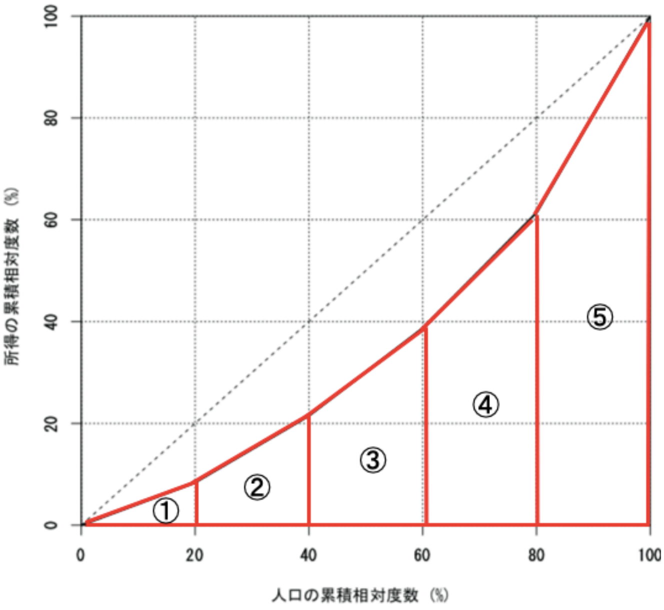
	~20%	**~40%**	**~60%**	**~80%**	**~100%**
日本	5.4%	16.1%	32.4%	56.5%	100%
アメリカ	5.1%	15.4%	30.8%	53.5	100%
スウェーデン	8.7%	23.0%	40.8%	63.8%	100%
中国	5.2%	15.0%	29.9%	52.2%	100%
ドイツ	8.4%	21.5%	38.7%	61.4%	100%

この累積度数分布表と、曲線を見比べると、ドイツが当てはまっていることがわかる。

問３（２）
ジニ係数とは、「45 度線とローレンツ曲線の面積の２倍」で計算できる。
すなわち、以下の図の枠で囲んだ部分の面積の２倍である。



赤い枠の面積を直接求めることができないため、以下のように 0.5 - (① + ② + ③ + ④ + ⑤)を計算する。



三角形と台形の面積の計算を 5 つ実施し、ジニ係数を求めると、0.28 になる。

問３（３）

I：◯

どの国も 45 度線より下に曲線が描かれます。

II：×

ジニ係数が大きいと、格差が大きいことを表しています。
問題文は、説明が逆です。
III：◯
スウェーデンと中国を比べると、中国のローレンツ曲線が下に描かれます。
よって、中国の方が、格差が大きいということです。

問４（１）

2011 年の輸出物価指数の、前年からの変化率は以下の式で算出できます。

$$\text{変化率} = \frac{2011\text{年の値} - 2010\text{年の値}}{2010\text{年の値}} \times 100$$

よって、以下のようになります。

$$\text{変化率} = \frac{87.5 - 89.5}{89.5} \times 100 = -2.23$$

問４（２）

問題文を解釈できたかどうかが鍵となる。

問題文を図示すると、以下の条件を満たす r を計算するということ。

	2010	2011	2012	2013	2014	2015
輸出物価指数(%)	89.5	X1	X2	X3	X4	100

ではまず、X1 を計算します。

$$\text{変化率}(r) = \frac{2011\text{年の値} - 2010\text{年の値}}{2010\text{年の値}} \times 100$$

$$\mathrm{r} = \frac{X1 - 89.5}{89.5} \times 100$$

$$\mathrm{X1} = 89.5 \times (1 + \frac{r}{100})$$

次に、X２を計算します。

$$\mathrm{r} = \frac{X2 - X1}{X1} \times 100$$

$$\frac{rX1}{100} = X2 - X1$$

$$X2 = \frac{rX1}{100} + X1 = X1\left(1 + \frac{r}{100}\right) = 89.5 \times (1 + \frac{r}{100})^2$$

と計算することができます。
ということは、あとは 2015 年までの変化率を計算していけば良いということです。

$$100 = 89.5 \times (1 + \frac{r}{100})^5$$

$$\frac{100}{89.5} = (1 + \frac{r}{100})^5$$

$$1 + \frac{r}{100} = (\frac{100}{89.5})^{\frac{1}{5}}$$

$$r = 100 \times \{(\frac{100}{89.5})^{\frac{1}{5}} - 1\}$$

問５

フィッシャーの３原則とは、「繰り返し」「局所管理」「無作為化」のことです。

問６

抽出の問題は毎回必ずでてくるので、理解したいです。

今回の問題では、男女という「性別」というカテゴリごとに無作為抽出をしたい場合です。

この時カテゴリのことを「層」とよびます。

「層」は一般的にも使われていて、例えば「年齢層」という使い方をしますよね。

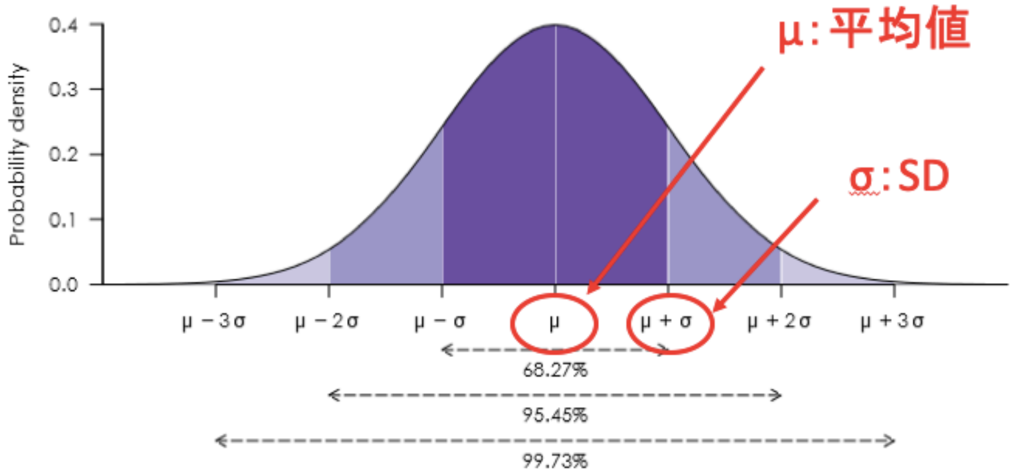
そのため、層化（層別）抽出といいます。

問７（１）
まず、お菓子をもらえる条件は「２回連続で勝つ」ということ。
つまり、お菓子がもらうためには、以下のうちいずれかの場合。
1. １回目と２回目に勝つ：確率は pq（独立なので、そのまま掛けて良い）
2. １回目負けて、２回目と３回目に勝つ：確率は(1-p)pq

そのため、お菓子がもらえるのは上記の２つを足したもの。
pq+(1-p)pq が正解。

問７（２）
T-U-T でお菓子をもらえる確率は、（１）で計算できました。
では、U-T-U でお菓子をもらえる確率を計算してみます。
（１）と同様に計算すると、qp+(1-q)qp となります。
では、「pq+(1-p)pq」と「qp+(1-q)qp」ではどちらが大きいでしょうか。
２つの違いは、1-p と 1-q のみです。
問題文より $p<q$ ですので、$1-p>1-q$ となります。
つまり、T-U-T の順番の方がお菓子をもらいやすいということがわかります。

問８（１）
まず必ず知っておきたいのが、正規分布の性質。
正規分布は「平均値（μ）」と「標準偏差（SD）」の２つで形が決まります。
そして、$\mu \pm$SD の間には、約 68%のデータが含まれています。
$\mu \pm$2SD の間には、約 95%のデータが含まれています。



そして、もう一つ知っておきたいのが、標準化。
問１でもさらっと出てきましたが、ここでもおさらいしましょう。

$$Z = \frac{\text{データ} - \text{平均値}}{\text{標準偏差}}$$

標準化をすることで、平均 0、標準偏差 1 の正規分布（標準正規分布）のデータにすることができます。

以上の２点を知ったところで、問題です。
4,800 円を標準化のデータに直します。
すると（4800-4000）/500=1.6 となります。
つまり、標準正規分布で 1.6 以上となる確率はどれぐらいか？と問われているのと一緒です。
標準正規分布表は問題の最後の方についておりますので、1.6 以上がどのぐらい

の確率かを確認します。

付表 1. 標準正規分布の上側確率

u	.00	.01	.02
0.0	0.5000	0.4960	0.4920
0.1	0.4602	0.4562	0.4522
0.2	0.4207	0.4168	0.4129
0.3	0.3821	0.3783	0.3745
0.4	0.3446	0.3409	0.3372
0.5	0.3085	0.3050	0.3015
0.6	0.2743	0.2709	0.2676
0.7	0.2420	0.2389	0.2358
0.8	0.2119	0.2090	0.2061
0.9	0.1841	0.1814	0.1788
1.0	0.1587	0.1562	0.1539
1.1	0.1357	0.1335	0.1314
1.2	0.1151	0.1131	0.1112
1.3	0.0968	0.0951	0.0934
1.4	0.0808	0.0793	0.0778
1.5	0.0668	0.0655	0.0643
1.6	0.0548	0.0537	0.0526
1.7	0.0446	0.0436	0.0427

0.548 なので、0.55 が正解です。

問 8（2）
この問題も、与えられた数字を Z に直して、標準正規分布表と見比べる、ということをします。
つまり、最終的には以下の式に代入します。

$$Z = \frac{\text{データ} - \text{平均値}}{\text{標準偏差}}$$

ここで問題なのが、「今年 6 月と昨年 6 月の差の標準偏差」はどう求めるのか、ということ。
これは、分散の定義から計算します。

$$V[A-B] = V[A] + V[B] = 500^2 + 500^2$$

このように求めることができるので、最終的には以下の式になります。

$$Z = \frac{\text{データ} - \text{平均値}}{\text{標準偏差}} = \frac{800 \ - 0}{\sqrt{500^2 + 500^2}} = 1.13$$

ということで、また標準正規分布と見比べて、答えは 0.129 となります。

問８（３）

難しそうに見えて、簡単な問題です。

3 年分の電気料金は独立なので、その大小関係が出る確率はそれぞれ等しいです。

つまり、３年分のデータの並び順は 3!=3*2*1 で 6 通り。

そのうち、今年が一番大きい組み合わせは「前々年＜前年＜今年」と「前年＜前々年＜今年」の 2 通り。

つまり、2/6=1/3 が正解。

問９（１）
平均、分散、共分散、相関の４つの定義をおさらいしましょう。
こちらの４つの定義は、暗記したほうが良いです。

$$E[aX+b]=aE[X]+b$$
$$V[aX+b]=a^2V[X]$$
$$V[X]=E[X^2]-\{E[X]\}^2$$
$$Cov[X,Y]=E[XY]-E[X]E[Y]$$
$$r=\frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

これを知っていれば、あとは当てはめるだけです。

問９（２）
少々複雑に思いますが、こちらの問題も上記の定義がわかっていれば解ける問題です。
計算間違いさえしなければ解くことができます。

$$Cov[U,V]=E[UV]-E[U]E[V]$$

$$\begin{aligned}E[UV]&=E[(3X-2)(-2Y-4)]\\&=E[-6XY-14X+4Y+8]\\&=-6E[XY]-14E[X]+4E[Y]+8\\&=-6\times6.3-14\times2+4\times3+8\\&=-41.8\end{aligned}$$

$$Cov[U,V] = E[UV] - E[U]E[V]$$

$$E[U] = E[(3X-2)] \qquad E[V] = E[(-2Y-4)]$$

$$= 3E[X] - 2 \qquad = -2E[Y] - 4$$

$$= 3\times 2 - 2 \qquad = -2\times 3 - 4$$

$$= 4 \qquad = -10$$

$$Cov[U,V] = -41.8 - (40) = -1.8$$

$$r = \frac{Cov(U,V)}{\sqrt{V(U)V(V)}}$$

$$V[U] = V[3X-2] = 9V[X] = 9$$

$$V[V] = V[-2Y-4] = 4V[Y] = 4$$

$$r = \frac{Cov(U,V)}{\sqrt{V(U)V(V)}} = \frac{-1.8}{\sqrt{9\times 4}} = -0.3$$

問 10（1）
95%信頼区間の知識と、ちょっとした応用力が必要な問題でした。
まず、95%信頼区間は以下の式です。

$$P(\bar{X}-\underbrace{\boxed{Z_{\alpha/2}}}_{1.96}\times\sqrt{\frac{s^2}{n}}\leq\mu\leq\bar{X}+\underbrace{\boxed{Z_{\alpha/2}}}_{1.96}\times\sqrt{\frac{s^2}{n}})\geq 0.95$$

95%信頼区間の定義がわかっていれば、あとは与えられた問題文を上記の式に近づけます。

$$P(|\bar{X}-\mu|\leq 0.5)\geq 0.95$$

絶対値を外す

$$P(-0.5\leq\bar{X}-\mu\leq 0.5)\geq 0.95$$

真ん中をμにする

$$P(\bar{X}-0.5\leq\mu\leq\bar{X}+0.5)\geq 0.95$$

かなり、95%信頼区間の式に近づきました。
そして、上記の式と 95%信頼区間の式と見比べます。

$$P(\bar{X}-0.5\leq\mu\leq\bar{X}+\boxed{0.5})\geq 0.95$$

$1.96\times\sqrt{\frac{s^2}{n}}\leq 0.5$ を満たせば良い！

$$P(\bar{X}-1.96\times\sqrt{\frac{s^2}{n}}\leq\mu\leq\bar{X}+\boxed{1.96\times\sqrt{\frac{s^2}{n}}})\geq 0.95$$

よって、あとは計算していきます。

$$1.96 \times \sqrt{\frac{s^2}{n}} \leq 0.5$$

$$\sqrt{n} \geq \frac{1.96}{0.5}$$

$$n \geq 15.37$$

問 10（2）

n=20 が与えられたとき、T 分布に従います。

つまり、95%信頼区間は以下のような式になります。

T 分布表より、自由度 19 の $t_{0.025}$ は 2.093 です。

$$P(\bar{X} - \underbrace{\boxed{t_{\alpha/2}}}_{2.093} \times \sqrt{\frac{s^2}{n}} \leq \mu \leq \bar{X} + \underbrace{\boxed{t_{\alpha/2}}}_{2.093} \times \sqrt{\frac{s^2}{n}}) \geq 0.95$$

あとは、与えられた数字を代入します。

$$10.5 - 2.093 \times \sqrt{\frac{5.41}{20}} \leq \mu \leq 10.5 + 2.093 \times \sqrt{\frac{5.41}{20}}$$

問 11（1）

比率の信頼区間。

統計検定２級では、かなりの頻度で出題されます。

割合をpとして、信頼区間は以下の式で計算できます。(暗記しておいていいレベルです)

$$P \pm 1.96 \times \sqrt{\frac{p(1-p)}{n}}$$

今回のデータを当てはめると、以下のようになります。

$$0.071 \pm 1.96 \times \sqrt{\frac{0.071(1-0.071)}{4633}} = 0.071 \pm 0.007$$

問 11（2）

母比率の推定値は以下の式で計算することができます。

$$\textbf{母比率の推定値} = \frac{\textbf{北海道と沖縄の全体の野球行動者}}{\textbf{北海道と沖縄全体の人口}}$$

$$= \frac{\hat{p}_1 N_1 + \hat{p}_2 N_2}{N_1 + N_2}$$

そして、標準誤差の算出には、問９で出てきた分散の定義を使います。

$$E[aX+b] = aE[X] + b$$

$$V[aX+b] = a^2 V[X]$$

$$V[X] = E[X^2] - \{E[X]\}^2$$

$$Cov[X,Y] = E[XY] - E[X]E[Y]$$

$$r = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

つまり、以下のようになります。

$$V\left[\frac{\hat{p}_1N_1+\hat{p}_2N_2}{N_1+N_2}\right]=V\left[\frac{\hat{p}_1N_1}{N_1+N_2}\right]+V[\frac{\hat{p}_2N_2}{N_1+N_2}]$$

$$=(\frac{N_1}{N_1+N_2})^2\times V[\hat{p}_1]+(\frac{N_2}{N_1+N_2})^2\times V[\hat{p}_2]$$

$$=(\frac{N_1}{N_1+N_2})^2\times V[\hat{p}_1]+(\frac{N_2}{N_1+N_2})^2\times V[\hat{p}_2]$$

ここまでできたらあとは、標準誤差（Standard Error, SE）とは何か？がわかっていれば解くことができます。
データの標準誤差とは、推定値の標準偏差のことです。
ここはぜひ理解しておきましょう。
つまり、上記の分散をルートすれば、そのまま標準誤差になります。

$$V=(\frac{N_1}{N_1+N_2})^2\times V[\hat{p}_1]+(\frac{N_2}{N_1+N_2})^2\times V[\hat{p}_2]$$

$$=(\frac{N_1}{N_1+N_2})^2\times\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+(\frac{N_2}{N_1+N_2})^2\times\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}$$

$$SE=\sqrt{(\frac{N_1}{N_1+N_2})^2\times\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+(\frac{N_2}{N_1+N_2})^2\times\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

問 12（１）

T検定の基礎知識を問う問題です。

2群のT検定の際のT統計量は以下の式で計算できます。

これは暗記していいレベルです。

$$t = \frac{\text{群1の平均値} - \text{群2の平均値}}{\sqrt{\text{2群のプールした分散}\left(\frac{1}{\text{群1の例数}} - \frac{1}{\text{群2の例数}}\right)}}$$

そして、プールした分散とは、以下の通りです。

$$\text{2群のプールした分散} = \frac{(\text{群1の例数} - 1)\times\text{群1の不偏分散} + (\text{群2の例数} - 1)\times\text{群2の不偏分散}}{\text{群1の例数} + \text{群2の例数} - 2}$$

これがわかれば、あとは計算ミスしないようにするだけ。

計算式は割愛しますが、セリーグの不偏分散は以下の通り。

$$\frac{13549}{6-1} = 2709.8$$

パリーグの不偏分散は以下の通り。

$$\frac{7763}{6-1} = 1552.6$$

よって、t統計量は以下の通りになります。

$$t = \frac{\text{群1の平均値} - \text{群2の平均値}}{\sqrt{\text{2群のプールした分散}\left(\frac{1}{\text{群1の例数}} - \frac{1}{\text{群2の例数}}\right)}}$$

$$= \frac{233.7 - 185.3}{\sqrt{2131.2\left(\frac{1}{6} - \frac{1}{6}\right)}} = 1.82$$

問 12（2）
分散分析のF値を問う問題です。
分散分析表を思い浮かべられるかどうかが鍵です。

分散分析、および分散分析表に関しての詳細は、こちらの記事をご参照ください。
https://best-biostatistics.com/stat-test/anova.html

分散分析表は以下の通りです。

要因	平方和S	自由度df	不偏分散V	F値
群	S(群)	df(群) (群の数-1)	V(群) (=S(群)/df(群))	V(群)/V(残)
残差	S(残)	df(残) (全データ-群の数-1)	V(残) (=S(残)/df(残))	
全体	S(全)	df(全) (全データ-1)		

つまり、F値を算出するには群の分散と、残差の分散が必要です。
そのために、まずはそれぞれの平方和を算出します。

	セリーグ	パリーグ	全体
	218	209	
	303	177	
	198	167	
	296	145	
	201	161	
	186	253	
データ数	6	6	12
平均	233.7	185.3	209.5
標準偏差	52.1	39.4	50.7
分散	2709.9	1552.7	2574.6
平方和	13549.3	7763.3	28321

ここまでくれば、あとは分散分析表の通りに当てはめるだけです。

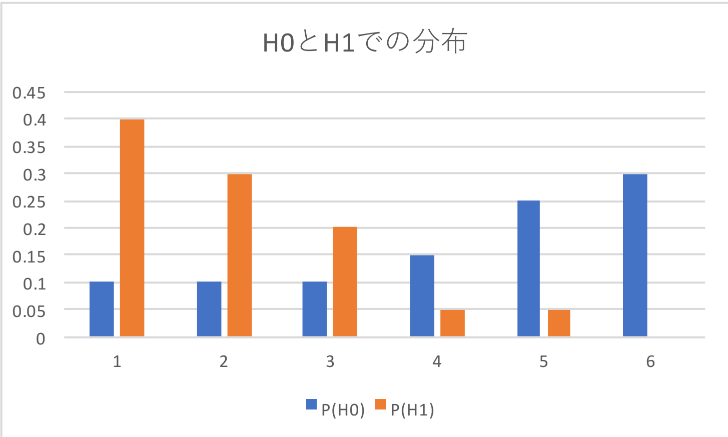
自由度に関しては、こちらの記事をご確認ください。
https://best-biostatistics.com/contingency/degree-freedom.html

要因	平方和S	自由度df	不偏分散V	F値
群	7009	1	7009	3.29
残差	21312	10	2131.2	
全体	28321	11		

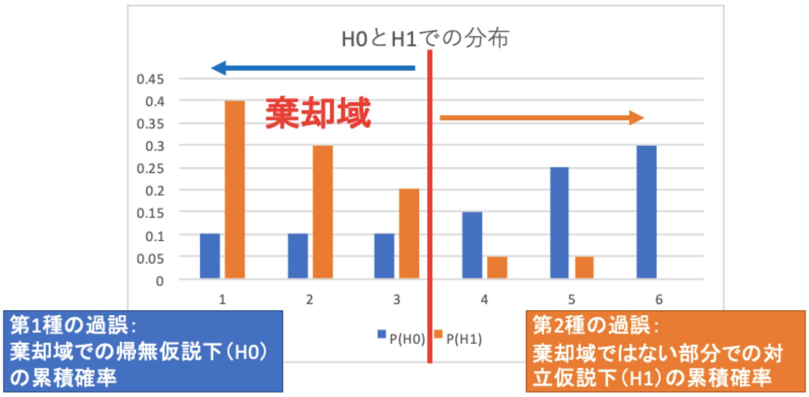
問 13（1）

まず、問題文を可視化してみます。

Ｈ０とＨ１の数字が与えられていますので、それをそのままグラフにします。



（1）ではX≦3としていますので、以下の通りの棄却域になります。



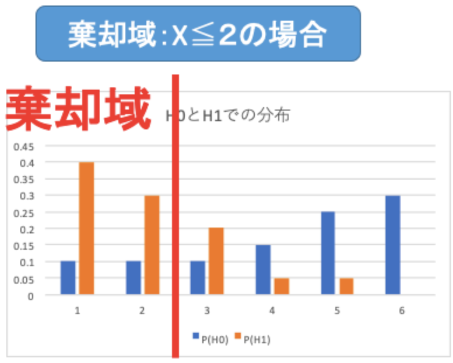
そして、第１種の過誤、第２種の過誤、検出力の用語を整理します。

- 第１種の過誤：棄却域での帰無仮説下（Ｈ０）の累積確率
- 第２種の過誤：棄却域ではない部分での対立仮説下（Ｈ１）の累積確率
- 検出力：1-第２種の過誤

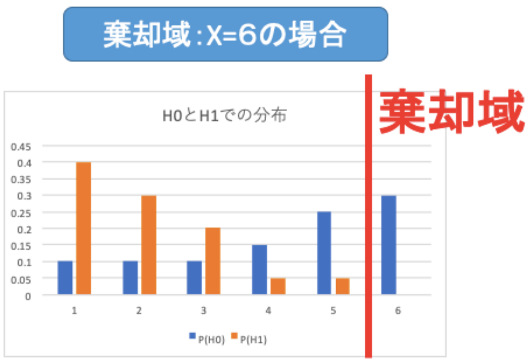
よって、第１種の過誤は0.1+0.1+0.1=0.3となります。

第２種の過誤は0.05+0.05=0.1となります。
検出力は1-0.1=0.9となります。

問 13（２）
X≦２が棄却域の場合と、X＝６が棄却域の場合、第１種の過誤、第２種の過誤、検出力はそれぞれ以下の通り。
有意水準とは、第１種の過誤とほぼ同じ意味です。



第1種の過誤:0.1+0.1=0.2
第2種の過誤: 0.2+0.05+0.05=0.3
検出力: 1-0.3=0.7



第1種の過誤:0.3
第2種の過誤: 1
検出力: 1-1=0

問 14（1）
回帰分析の読み取り方です。
問題文で与えられた回帰式に、出力結果を当てはめると以下の式になります。

$$log(\textbf{犯罪発生率}) = -7.1 + 0.09\times\text{失業率} + 2.4\times\log(\text{賃金}) - 0.065\times\log(\text{警察官数}) + \text{誤差}$$

ここに、問題文の数字を当てはめます。

$$\begin{aligned} log(\textbf{犯罪発生率}) &= -7.1 + 0.09\times\text{失業率} + 2.4\times\log(\text{賃金}) - 0.065\times\log(\text{警察官数}) + \text{誤差} \\ &= -7.1 + 0.09\times2.8 + 2.4\times5.6 - 0.065\times5.3 \\ &= -7.1 + 0.252 + 13.44 - 0.3445 \\ &\cong 6.4 \end{aligned}$$

問 14（2）
回帰分析の結果から、t 統計量を求める式は以下の通り。

$$t - value = \frac{estimate\ -\ \textbf{帰無仮説となる数字}}{\mathrm{SE}}$$

よって、値を代入するとこのようになる。

$$t - value(\beta_3\textbf{の帰無仮説}: -0.5) = \frac{-0.06498\ -(-0.5)}{0.22718} = 1.915$$

計算結果の t 統計量を、t 分布表と見比べる。
自由度は回帰分析結果の Residual Standard Error を見る。
43degree of freedom とあるため、自由度 40 の t 分布表を確認する。
すると、得られた t=1.915 は両側 10%と両側 5%の間の t 値に相当する。

そのため、両側 10%であれば棄却される。

問 14（3）

I：◯

Intercept と log（賃金）の 2 つが有意水準 1%で 0 ではない。

II：×

Log（賃金）の推定値は正の値であり、有意水準 1%で 0 ではない。

そのため、賃金が上がると犯罪発生率は上がる傾向にある。

III：◯

自由度調整済み決定係数は、回帰分析結果の Adjusted R-squared の値である。

問 15（1）
分割表の基本的な知識を問う問題。
与えられた表から期待度数の表を作成すると、以下のようになります。

期待度数の表	北	それ以外	合計
冬季	$365\times\frac{120}{365}\times\frac{207}{365}=68.05$	$365\times\frac{120}{365}\times\frac{158}{365}=51.95$	120
それ以外	$365\times\frac{245}{365}\times\frac{207}{365}=138.95$	$365\times\frac{245}{365}\times\frac{158}{365}=106.05$	245
合計	207	158	365

分割表に関する詳細は、こちらの記事をご参照ください。
https://best-biostatistics.com/contingency/contingency-kiso.html

問 15（2）
カイ二乗値は、以下の式で求めることができる。

$$\chi^2=\frac{(\text{得られたデータ}-\text{期待度数})^2}{\text{期待度数}}$$

よって、数字を代入すると以下の通り。

$$\chi^2=\frac{(105-68.05)^2}{68.05}+\frac{(15-51.95)^2}{51.95}+\frac{(102-138.95)^2}{138.95}+\frac{(143-106.05)^2}{106.05}$$

カイ二乗検定に関しての詳細は、こちらをご参照ください。
https://best-biostatistics.com/contingency/chi-square.html

問 15（３）

分割表の自由度は、(行の数-1)*(列の数-1)で計算できる。

よって、2*2 分割表の場合の自由度は 1 である。

カイ二乗分布表より、自由度 1 のときに 0.05 となるカイ二乗値は 3.84

	α							
ν	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
$\boxed{1}$	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63
2	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21
3	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09
6	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81

また、（２）より計算されたカイ二乗値は 69.04 である。

69.04>3.84 であるため、風向と季節には関連があると言える。

問 16

F 値の求め方は、以下の通り。

$$F値 = \frac{大きい値の分散}{小さい値の分散} = \frac{19.5^2}{14.5^2} = 1.81$$

また、自由度（20, 40）の両側 5%の F 統計量は下記の通り 2.068 となる。

$\alpha = 0.025$

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	15	$\boxed{20}$	40
5	10.007	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619	6.428	6.[illegible]29	6.175
10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717	3.522	3.[illegible]19	3.255
15	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060	2.862	2.[illegible]56	2.585
20	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837	2.774	2.573	2.[illegible]64	2.287
25	5.686	4.291	3.694	3.353	3.129	2.969	2.848	2.753	2.677	2.613	2.411	2.[illegible]00	2.118
30	5.568	4.182	3.589	3.250	3.026	2.867	2.746	2.651	2.575	2.511	2.307	2.[illegible]95	2.009
$\boxed{40}$	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.452	2.388	2.182	2.068	1.875
60	5.286	3.925	3.343	3.008	2.786	2.627	2.507	2.412	2.334	2.270	2.061	1.944	1.744
120	5.152	3.805	3.227	2.894	2.674	2.515	2.395	2.299	2.222	2.157	1.945	1.825	1.614

1.81<2.068 であるため、棄却されない。