2010年度後期　統計データ解析Ｂ　第９回

総合演習　問題および解答例

［問題１］　次の「平均」に関する記述について，正しい（あるいは適切である）かどうかを答えよ．正しくない（適切でない）ときは，どういう点がどのように正しくない（適切でない）かを説明せよ．

1. 気象用語でいう「雲量」とは，快晴を０，本曇りを１０として空のうち雲の占める割合を表現する．平均は５程度である．したがって，一年間のうち多少雲がある曇りの日が一番多い．
2. 力士の能力は「心・技・体」の３項目で計られる．力士「甲の海」の評価は，１０点満点で心６点，技７点，体７点で平均 6.7 点である．力士「乙の花」の評価は，心０点，技１０点，体１０点で平均 6.7 点である．したがって，両力士の能力は同じである．
3. 「人生５０年」という言葉があったように，１００年前の日本人の平均寿命は５０歳程度であった．当時の女性は５人６人の子供を生むことは当たり前であったので，子育てが終わった後の人生はたった数年しかなかった．
4. A 社の電球の平均寿命は，B 社の電球の平均寿命よりも長い．したがって，価格，消費電力，光量など，寿命以外の条件が全く同じならば，A 社の電球を使うほうがよい．
5. 母集団の平均の区間推定を行なう問題で，取り出した標本を使って計算した結果，「母平均の 95 パーセント信頼区間は 20 から 30 である」という結論を得た．このとき，母平均が 20 以上 30 以下である確率は 95 パーセントである．
6. 東京証券取引市場の日経平均株価が 10,000 円程度であるとき，市場には 1 株あたり 10,000 円程度で取引されている銘柄がたくさんある．
（講義では扱っていませんし，試験にも出しません．参考に）

［問題２］　X薬品の「Y」という薬は，１つ１グラムの錠剤となっている．いま，10 個の錠剤を無作為抽出し，各々の錠剤に含まれる物質 P の量を調べた．その結果，各錠剤の物質 P の量（ミリグラム）は次の通りであった．

1.2　0.8　0.9　0.9　1.0　1.3　1.2　1.0　0.8　0.9

1. 講義で説明した知識を使って，「薬Ｙに含まれる物質Ｐの割合」を区間推定するには，この測定がどのようなものであると仮定できる必要があるか．
2. 1. で答えた仮定が正しいとして，信頼係数 95% で 1. の区間推定を行なえ．

［問題３］　ある金属材料は，鉄分を 10%含むと称している．この材料の塊から６個からなる標本を切り出して，鉄分の割合を計ると以下の通りであった（単位%）．

10　8　11　8　7　7

どうも，塊全体の鉄分の割合は 10%ではないように感じる．はたして，「鉄分の割合は 10%ではない」と言い切れるだろうか？ 講義で説明した知識を用いて，この問題に答えよ．なお，問題文中にない仮定が必要な場合は，それを明示せよ．

解答例

[問題１]

1. 雲量５は平均であって,「もっとも頻繁に現れる雲量」(モードという）と混同してはいけない．実は，雲量５の日数は一番少ない．つまり，晴れまたは雨（当然雲量１０）の日は多いが，中途半端な曇りの日は少ない．このように，ヒストグラムのピークがひとつの分布（単峰性分布という）ではない場合，平均値とモードとは大きく異なることもある．

2. 「総合能力」を，いつも算術平均で表してよいとは限らない．伝統的な相撲観では「心・技・体」はいずれも欠けてはいけないもので，その場合乙の花はまったくだめということになる．このような時に用いる平均として「幾何平均」というものがある．これは，項目が３つなら３つの項目の点数を掛け合わせて３乗根を計算したものである．

   あるいは,「心などどうでもよくて，技，体がよければ強いはず」という価値観もあるかもしれない．この場合,「心に１倍，技，体に５倍」という「重み」を定めて，$(1\times$心$+5\times$技$+5\times$体$)/(1+5+5)$ という平均を求める方法もある．これを重みつき平均という．

3. 平均寿命とは,「現在ある年齢の人が来年まで生きられる確率」がどの年齢層でも将来にわたって不変であるとしたときの，現在０歳の人が将来生きる人生の長さの平均を意味する．したがって，乳児死亡率が高いと平均寿命は短くなる．子育てが終わって現在４０歳になっている人が，平均してあと何年生きるかを表す「４０歳平均余命」は，乳児死亡率は関係ないので，(平均寿命 − ４０) とはだいぶ異なった値になる．

4. 家庭用の照明ならば，切れてから取り替えればよいのだから，平均寿命が長い電球のほうが，長期的に見れば家計の節約になる．しかし，信号機の電球などの「絶対に故障してはいけない機器」では，故障しないうちに一定の間隔で電球を交換することになるから，絶対に故障しないことが保証される期間，すなわち「最低保証寿命」が長い電球のほうが，平均寿命が長い電球よりも望ましい．

5. 母平均は，観測者が知らないだけで，実際にはひとつに決まっている．だから,「20 以上 30 以下」という具体的区間を求めた段階で，母平均が「20 以上 30 以下」であるかどうかはすでに決まっている．「95% 信頼区間」とは，この方法で信頼区間を何度も求めると，そのうち 95% が母平均を本当に含んでいる，という意味である．

6. 東証日経平均株価は，東京証券取引所に上場されている銘柄のうち代表的な 225 銘柄の株価から算出される．しかし，その値は「225 銘柄（額面 50 円換算の）株価の合計を 225 で割ったものではなく，24.869 で割ったものである（2010 年 10 月１日より)．したがって，10,000 円前後の株価の銘柄はほとんどなく，ずっと安い株価の銘柄がほとんどである．

   除数が 225 でないのは,「株式分割」があるからである．株式分割とは，企業が株主に対して現在の保有株数に応じて無償で追加の株を割り当てることである．このとき，各株主の保有する株数は増えるが，資産総額は増えないので，見かけ上株価が安くなる．例えば，株価 500 円の企業が，１株に対して１株を追加して２株を割り当てる株式分割をおこなった場合，分割後の株価は半分の 250 円となる．しかし，このように株価が安くなったのは見かけ上のもので，会社の価値が下がったわけではないので，この分割によって平均株価が変化しないようにするため，除数を小さくして対応する．
   （参考：日経平均プロフィル http://www.nikkei.co.jp/nkave/ ）

[問題２]

1. 薬Yの各錠に含まれる物質Pの割合が，正規分布にしたがっていること．

2. 薬Y全体での1グラムあたりの物質Pの量をμ，各錠剤での物質Pの量の平均を$\bar{X}$，不偏分散をs^2，標本サイズをnとする．このとき，問題文から，各錠剤が含む物質Pの量は，平均μの正規分布にしたがうと考えられる．したがって，

$$t = \frac{\bar{X} - \mu}{\frac{s^2}{n}} \tag{1}$$

という値（t統計量）は自由度$n-1$のt分布$t(n-1)$にしたがうので，$t_{0.025}(n-1)$を「t統計量がその値以上になる確率が0.025であるような値」とし，$-t_{0.025}(n-1)$を「t統計量がその値以下になる確率が0.025であるような値」とすると

$$P\left(\bar{X} - t_{0.025}(n-1)\sqrt{\frac{s^2}{n}} \le \mu \le \bar{X} + t_{0.025}(n-1)\sqrt{\frac{s^2}{n}}\right) = 0.95 \tag{2}$$

となるので，(2)式のかっこ内の不等式が，μの95%信頼区間の下限と上限を表している．

$X = (1.2 + 0.8 + \cdots + 0.8 + 0.9)/10 = 1.0$，$s^2 = \{(1.2 - 1.0)^2 + \cdots + (0.9 - 1.0)^2\}/(10 - 1) = 0.031$，$n = 10$，$t_{0.025}(9) = 2.262$を(2)式に代入すると，$\mu$の95%信頼区間は$[0.87, 1.13]$（ミリグラム）となる．したがって，物質Pの割合は$[0.087, 0.113]$(%)となる．

[問題３]

講義で学んだ知識によって問題に答えるために，標本が材料の塊からまんべんなく抽出されており，十分多くの回数測定を行ったときの各測定での鉄分の割合が，正規分布にしたがっていると仮定する．

そこで，有意水準5%のt検定を行うことにする．この場合，「10%ではないかどうか」検査するので，「10%より大きくても小さくても帰無仮説棄却する」必要があり，両側検定を行う．よって，材料全体での鉄分の割合をμ(%)とし，帰無仮説を「$\mu = 10$である」，対立仮説を「$\mu = 10$でない」とする．標本サイズを$n = 6$とし，標本から標本平均$\bar{X}$と不偏分散s^2を求めると，$\bar{X} = 8.5$，$s^2 = 2.7$となる．

$$t = \frac{\bar{X} - \mu}{\sqrt{s^2/n}} \tag{3}$$

は自由度$n-1$のt分布にしたがうので，$t \le -t_{0.025}(n-1)$または$t \ge t_{0.025}(n-1)$である確率は5%である．帰無仮説が正しいとするとき，上の数値を代入すると$t = -2.24$である．一方，$t_{0.025}(5) = 2.571$であるから，「$t \le -t_{0.025}(n-1)$または$t \ge t_{0.025}(n-1)$」は成り立たない．よって，「5%の確率でしか起きないことが起きている」とは言えず，帰無仮説は棄却されない．したがって，有意水準5%で，「材料全体での鉄分の割合は10%ではないとは，このデータからは言えない」という結論になる．

(有意水準を答えるのを忘れないようにしてください．有意水準によってはこの解答例とは異なる結果となります．)