基礎統計（火１　廣松）頻出問題演習（２）

〔4〕正規分布

正規分布の問題はよく出ます。しっかり勉強しましょう。正規分布表の見方は、教科書の正規分布表のページの下にも書いてあります。しっかり使いこなせるようにしておくべきでしょう。

【例題１】

統計学の期末試験の得点がＮ（50、225）に従っているとする。得点の高いものからＡ，Ｂ，Ｃ，Ｄ，Ｅの評価を、それぞれ受験者総数の10%、20%、40%、20%、10%の割合でつけたい。どのように評価をつければよいか。ただし、試験得点は整数で表されるものとする。

【解答例】

まず、題意より、$\mu = 50 \quad \sigma = \sqrt{225} = 15$

正規分布表で、$Q(u) = 0.1$に対して、$u = 1.28\cdots \cong 1.28$ である。したがって、正規分布の上10%と下10%の境目はそれぞれ、$50 + 1.28 \times 15 = 69.2 \quad 50 - 1.28 \times 15 = 30.8$ となる。まったく同様に、正規分布表で、$Q(u) = 0.3$を調べると、$u = 0.52$となるから、上下30%、すなわちＣ評価が得られる上限・下限は、$50 + 0.52 \times 15 = 57.8 \quad 50 - 0.52 \times 0.52 = 42.2$

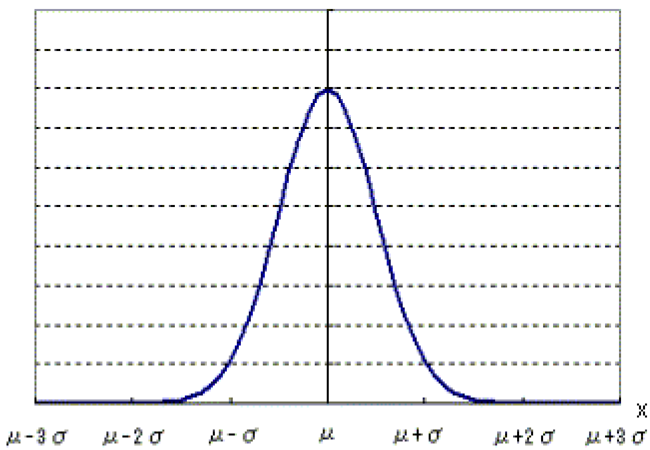
点数が整数値であることから、求める評価のつけ方は、

Ａ:70点以上　Ｂ：58点から69点　Ｃ：43点～58点　Ｄ：31点～42点　Ｅ：30点以下

【解説】

正規分布表を用いる問題です。正規分布表の見方がわかればあとは簡単です。

一般に、正規分布のグラフは下図のようになります。



Ａ評価は上位10%なので、まず、正規分布表で面積$Q(u)$が0.1になるuを求めます。このと

き、正規分布表の値は標準正規分布のときの値だから、実際はσをかけた値が、上位10%の下限と、平均μの差です。したがって【解答例】のような計算をすれば、答えは求まります。

本番の期末の分布はこの問題のようにならないで欲しいですね。

【例題２】正規分布の主な特徴を３つあげよ。その上で、以下の問いに答えよ。
ある試験の成績は、ほぼ正規分布内Ｎ（７０、１００）に従っている。このとき、不合格者の数を全体の１０％以内におさえるためには、何点以上を合格にすればよいか。

(97年夏)

【解答例・解説】

正規分布の特徴は、教科書120㌻の6.6をよく読めばわかると思います。しかし、どの程度書けばいいのかわかりません。左右対称とかのレベルでいいのか、それとも線形変換とか、標準化などのレベルまで書くのか。こればかりは各自の判断に任せたいと思います。何かいい意見があったら教えてください。

解き方は【例題１】とまったく同じです。下位10%の点を調べればいいので、正規分布表を見ましょう。面積$Q(u)=0.1$となるuを調べます。だいたい、1.28くらいでしょう。ここまでくればもう終わりです。σ＝10だから、

$$\mu - u\times\sigma = 70 - 1.28\times 10 = 56.2$$

よって、56点以下は不合格である。（竹村かよ）

練習問題として、優３割規定を守れるか、また、３割を超えるなら全員から何点引けばいいか求めてみると面白いかもしれません。

〔５〕推定・検定

推定・検定の問題は必ず出ると思います。結構ムズイと思いますが、練習あるのみです。

【例題】母平均＝μ、母分散＝25の正規母集団について、以下の問いに答えよ。

（１）この母集団から大きさnの標本を取り出して、母平均μを信頼水準99%で区間推定したい。信頼区間の幅を４以内にするためにはnの大きさをどのように取ればよいか。

（２）この母集団から取り出した大きさ10の標本の観測値が

32, 39, 34, 45, 39, 29, 36, 42, 37, 35

であったとする。仮説：μ＝35 を有意水準５％で検定せよ。

(2005冬)

【解答例・解説】

(1) この問題では、母平均を推定します。母平均の推定にはｔ分布を用います。正規母集団に対する推定の場合、**母平均の推定→ｔ分布を使うことを考える　母分散の推定χ²分布を使うことを考える**　です。

次に、信頼水準99％というのはどういうことか。教科書p281のｔ分布表を見ましょう。ｔ分布のグラフの端の面積αの部分を両側取って、1－2αが、99％になるということである。つまり、この表でα＝0.005 とすればよいでしょう。

ここで、この問題では、σが既知か未知かを確認します。（超重要）

σが既知→ｔ分布表の自由度が∞のところをみる（正規分布だから）

σが未知→自由度n－1のｔ分布を考える。

ここでは、σ＝5 とわかっているので、ｔ分布表自由度∞のところをチェックです。α(片側)は0.005のところをみて、ｔ＝2.576　このｔは、自由度∞の値なのでＺ＝2.576と表すことが多い。つまり、Ｚを求めることが大切なわけです。

ここまできたら、信頼区間は求まります。次の式に代入してください。

σが既知のときの信頼区間は、$(\overline{X}-Z\sqrt{\frac{\sigma^2}{n}},\overline{X}+Z\sqrt{\frac{\sigma^2}{n}})$ **である。**

σが未知のときの信頼区間は、$(\overline{X}-t_{n-1}\sqrt{\frac{s^2}{n}},\overline{X}+t_{n-1}\sqrt{\frac{s^2}{n}})$ **である。（自由度n－1）**

この問題では、信頼区間の幅を4以内におさえたいので、

$2Z\sqrt{\frac{\sigma^2}{n}}\leq 4$　σ=5とＺ=2.576を代入して、電卓使ってちゃっちゃと計算すると、

$n\geq 41.4736$とかいう結果が出ます。求める答えは42以上です。

⑵仮説検定の問題はまず、帰無仮説・対立仮説が何かをはっきりさせましょう。

この問題では、帰無仮説H_0：μ=35　対立仮説H_1：$\mu\neq 35$

となります。

では、仮説検定では何をすればいいのか。**母平均の検定→ｔ分布を使うことを考える　母分散の検定→χ^2分布を使うことを考える**　です。この問題はμの値を検定するので、当然ｔ分布です。$t=\dfrac{\overline{X}-\mu}{s/\sqrt{n}}$　に代入するため、$\overline{X}$ と s を求めましょう。

電卓使って$\overline{X}=36.8$はすぐ求まります。ｓは少し面倒です。まず、ｓは検定の場合、**不偏分散ｓ²（n－１で割ったやつ）を使います。**

$s^2=\dfrac{1}{n-1}\sum(\overline{X}-X)^2$ に与えられた標本を代入しましょう。関数電卓使ったほうが楽ですが、電卓でも十分できます。計算すると、$s=4.7093\cdots=4.709$

よって、$t=\dfrac{36.8-35}{4.709/\sqrt{10}}=1.208\cdots=1.21$

ここで、有意水準５％という条件を使います。有意水準５％というのは、ｔ分布のグラフで、両端をあわせて5%分切り取っても大丈夫ってことです。

要するに、ｔ分布表でα=0.025(2α=0.05)、自由度はn－1=9のところを見ると、t=2.262です。μ=35という仮説では、ｔ=1.21でしたから、有意水準５％では棄却されない範囲です。ｔが2.262を超えている場合には帰無仮説は棄却されます。結局この仮説は棄却されないということになります。

仮説検定のとき方は次の通りです。

１.何を検定するのか（母平均or母分散）→母平均の問題しか出ません

２.σ^2は既知or未知（だいたいは未知・既知の場合はp240参照）

以下はσ^2が未知のとき

３.標本平均、標本分散を求める。標本分散は不偏分散のほうを使う。（廣松は明記すればどっちでもいいって言ってたけどね）

４.ｔ分布のｔの式に代入してｔを求める。

５.有意水準をチェック。例えば5%なら、ｔ分布のグラフで両端をそれぞれ2.5%のあわせて5%棄却する。αの値は有意水準からわかるので、ｔ分布表で自由度n－1のところをみて、この場合のｔが求まる。

６.４と５を比較する。4で求めた値が5で求めたｔの値を超えなければ、帰無仮説は棄却されない。

慣れれば難しくないのでがんばりましょう。練習問題として、次の問題をやってみましょう。

【練習】ある機械で生産された１０個の製品の重量を測定したところ（単位はｇ）

１０１．１、１０３．２、１０２．１、９９．２、１００．５、１０１．３、９９．７、１００．５、９８．９、１０１．４

であった。母平均は１００ｇと考えてよいか。有意水準５％で検定せよ。

（97夏、99夏）

その他、指数分布、負の二項分布も要チェックかもしれないです。