ゼミ試験（統計学）のサンプル問題

2006 年度の入ゼミ試験では１問だけ統計学の問題を出題します。出題範囲は「差の有意性検定」（秋学期に各クラスでやるはず）だけですので、しっかり勉強してきてください。
特に参考文献は示しませんが、鳥居先生の『よくわかる統計学』にも出ている基礎的な内容しか出題しないのでご安心ください。
代表的な２パターンの練習問題を示しておきますので勉強の参考にしてください。

A　平均値の差の有意性検定

(point)

実際に良く使うのは「分散は未知だが $\sigma=\sigma_1=\sigma_2$」のケースです。

$\bar{x}_1-\bar{x}_2$ ～$N(\mu_1-\mu_2, \frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2})$だから帰無仮説 $H_0: \mu_1=\mu_2$ が真の時、標本平均の差で定義される $t=(\bar{x}_1-\bar{x}_2)\Big/\left(s\sqrt{\frac{1}{n1}+\frac{1}{n2}}\right)$～$t(n_1+n_2-2)$

(例題)

１　ある学校でみかんを普段食べる子供 407 人と食べない子供 411 人に対してこの冬に風邪をひいた回数をたずねたところ、みかんを食べる子供は平均 1.38 回、標準偏差 1.23 回、みかんを食べない子供は平均 1.48 回、標準偏差 1.14 回であった。この結果から「みかんを食べることが風邪の予防に役立つ」という仮説を有意水準 5%で検定しなさい。

２　ある町の小売業で e-Business に進出した企業 18 社と進出していない企業 18 社について利益率を調査した結果、前者は標本平均 5%標本標準偏差 2%、後者は標本平均３％標本標準偏差１％という結果を得た。この結果から「e-Business への進出した企業の方が利益率は高くなる」という結論を導いても良いか？　有意水準５％で検定をしなさい。

B　比率の差の有意性検定

(point)

比率 $\bar{p}=x/n$ は n が大きい時、平均 np/n=p, 分散 $np(1-p)/n^2=p(1-p)/n$ の正規分布に従う

$H_0:P_1=P_2(=P)$の検定では $Z=(\bar{p}_1-\bar{p}_2)\Big/\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$～$N(0,1)$

(例題)

A,B という２人の候補者だけが立候補しているある選挙区で 2500 人の有権者に出口調査をしたところ 52%が A 候補に、48%が B 候補に投票したという。

1 この調査から A 候補は当選確実である（有意水準１％で仮説が認められる）と判断してよいか。

2 52%という割合で当選確実と判定するには、標本数を最低でいくつ集める必要があるか。

3 男性 50 人と女性 50 人に対して A 候補者の支持率を調べたところ男性は 56%、女性は 50%であった。この結果から A 候補者の支持率に男女差があると考えられるか。有意水準 5%で検定をしなさい。

解答

A　平均値の差の有意性検定

A-1

①仮説の設定

　みかんを食べる子を１、食べない子を２とすると風邪をひく回数の母平均μについて

　$H_0: \mu_1 = \mu_2$、$H_1: \mu_1 < \mu_2$

②統計量の設定と帰無分布の導出

　帰無仮説が正しいとき統計量 $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ は自由度 n_1+n_2-2 の t 分布に従う。

③標本の値と臨界値の関係

　得られた標本では $s^2=\{(407-1)1.23^2+(411-1)1.14^2\}/(407+411-2)=1.405727 \to s=1.185634$

　$t=\dfrac{1.38-1.48}{s\sqrt{\frac{1}{407}+\frac{1}{411}}}=-1.20612 > -t_{5\%}(816)$　$-Z_{5\%}=-1.645$

④判定

　これより帰無仮説は棄却できないので、みかんに有意な風邪予防の効果は認められない。

A-2

①仮説の設定

　e-Business 進出企業を１、非進出企業を２とすると収益率の母平均μについて

　$H_0: \mu_1 = \mu_2$、$H_1: \mu_1 > \mu_2$

②統計量の設定と帰無分布の導出

　帰無仮説が正しいとき統計量 $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ は自由度 n_1+n_2-2 の t 分布に従う。

③標本の値と臨界値の関係

　得られた標本では $s^2=\{(18-1)2^2+(18-1)1^2\}/(18+18-2)=2.5 \to s=1.581139$

　$t=\dfrac{5-3}{s\sqrt{\frac{1}{18}+\frac{1}{18}}}=3.7947 > t_{5\%}(34)=1.690$（　$Z_{5\%}=1.645$）

④判定

　これより帰無仮説は棄却されるので、e-Business 進出は利益率を有意に引き上げる。

B　比率の差の有意性検定

B-1

①仮説の設定

　A 候補の母集団における支持率を P_A とすると

　　$H_0: P_A=0.5$、$H_1: P_A > 0.5$

②統計量の設定と帰無分布の導出

　帰無仮説が正しいとき統計量 $z=\dfrac{p_A-0.5}{\sqrt{0.5(1-0.5)/n}}$ は標準正規分布に従う。

③標本の値と臨界値の関係

　得られた標本では $(0.52-0.5)/\sqrt{0.5\cdot 0.5/2500}=0.02/(0.5/50)=2<Z_{1\%}=2.326$

④判定

　これより帰無仮説が棄却できないので「当選確実であるとはいえない」

B-2

(0.52-0.5)/$\sqrt{0.5 \cdot 0.5 / n}$ >2.326→$\sqrt{n}$ >2.326*0.5/0.02=58.15→n>3381.423 より標本数は少なくとも 3382 人必要。

B-3

①仮説の設定

男性を１女子を２とすると A 候補の母集団における支持率Ｐについて

H_0: $P_1=P_2$、H_1: $P_1 \neq P_2$

②統計量の設定と帰無分布の導出

帰無仮説 H_0: $P_1=P_2=P$ が正しい時､統計量 z=$\dfrac{p_1 - p_2}{\sqrt{P(1-P)(\frac{1}{n_1}+\frac{1}{n_2})}}$ は標準正規分布に従う

③標本の値と臨界値の関係

Ｐを(0.56+0.50)/2=0.53 で評価すると得られた標本では

Ｚ=0.06/$\sqrt{0.53 \cdot 0.47 \cdot (1/50+1/50)}$ =0.6010829＜$z_{2.5\%}$=1.96 となるので、帰無仮説のもとで十分起こりうる現象が生じたことになる。

④判定

これより帰無仮説は採択され「A 候補者の支持率に男女差がない」といえる。