Stat 20: Discussion at section

B. M. Bolstad, bolstad@stat.berkeley.edu

Nov 5, 2003

More on using the t-table to get P-values

We will continue our exploration of the t-distribution. In particular, we shall use it to compute P-values when testing hypotheses about the two sample normal model.

Refresher on the Two sample normal model

We concentrate on the homoskedastic two sample normal model (assume $\sigma_1 = \sigma_2 = \sigma$).

Estimate σ using the pooled sample standard deviation $s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$.

Test statistic: $t = \frac{\bar{Y}_1 - \bar{Y}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$. Under the assumptions of the two sample normal model this will have the t-distribution with $n_1 + n_2 - 2$ df.

Testing against the two sided alternative

We will use the two sided test in the following examples

$$H_0 : \mu_1 - \mu_2 = 0 \text{ vs } H_A : \mu_1 - \mu_2 \neq 0$$

the P-value will be given by $2P(T > |t|)$

Example 1.1

Let df = 21 and suppose the value of the t statistic is $t = 2.05$.

Method 1 - Test at a fixed level of significance

Consider both the 5% and 1% levels of significance. From the t-distribution table we find that

$$2P(T > 2.08) = .05$$

$$2P(T > 2.831) = .01$$

For the 5% significance level: if t is larger than 2.08 or smaller than -2.08 then we would reject the null hypothesis. If $-2.08 < t < 2.08$ we cannot reject the null hypothesis.

For the 1% significance level: if t is larger than 2.831 or smaller than -2.831 we would reject the null hypothesis. If $-2.831 < t < 2.831$ we cannot reject the null hypothesis.

Since $t = 2.05$ we cannot reject the null hypothesis at either the 1% or 5% levels.

Method 2 - Bracket/Bound the P-value

From the t-table we find that

$$1.721 < 2.05 < 2.080$$

converting to P-values

$$0.05 > P(T > 2.05) > 0.025$$

multiplying through by 2 yields

$$0.1 > 2P(T > 2.05) > 0.05$$

and so we can reasonably say that $0.05 <$ P-value < 0.1

Method 3 - Approximate the P-value using linear interpolation

$$2P\left(T > 2.05\right) \approx 2\left(.05 + \frac{2.05 - 1.721}{2.08 - 1.721}\left(0.05 - 0.1\right)\right) = 0.0542$$

for comparison the computer give the exact P-value as 0.053

Example 1.2

Let df $= 16$ and suppose the value of the t statistic is $t = 0.321$.

Method 1 - Test at a fixed level of significance

Consider both the 5% and 1% levels of significance. From the t-distribution table we find that

$$2P\left(T > 2.120\right) = .05$$

$$2P\left(T > 2.921\right) = .01$$

For the 5% significance level: if t is larger than 2.120 or smaller than -2.120 then we would reject the null hypothesis. If $-2.120 < t < 2.120$ we cannot reject the null hypothesis.

For the 1% significance level: if t is larger than 2.921 or smaller than -2.921 we would reject the null hypothesis. If $-2.921 < t < 2.921$ we cannot reject the null hypothesis.

Since $t = 0.321$ we cannot reject the null hypothesis at either the 1% or 5% levels.

Method 2 - Bracket/Bound the P-value

One thing to note is that because the T-distribution is symmetric about 0, 50% of the area under the curve is above 0. Therefore from the t-table we find that

$$0 < 0.321 < 0.690$$

converting to P-values

$$0.5 > P(T > 0.321) > 0.25$$

multiplying through by 2 yields

$$1 > 2P(T > 2.05) > 0.5$$

and so we can reasonably say that $0.5 <$ P-value < 1.

Method 3 - Approximate the P-value using linear interpolation

$$2P\left(T > 2.05\right) \approx 2\left(.5 + \frac{.321 - 0}{0.690 - 0}\left(0.25 - 0.5\right)\right) = 0.767$$

for comparison the computer give the exact P-value as 0.752

Example 1.3

Let df = 25 and suppose the value of the t statistic is $t = -4.01$.

Method 1 - Test at a fixed level of significance

Consider both the 5% and 1% levels of significance. From the t-distribution table we find that

$$2P\left(T > 2.06\right) = .05$$

$$2P\left(T > 2.787\right) = .01$$

For the 5% significance level: if t is larger than 2.06 or smaller than -2.06 then we would reject the null hypothesis. If $-2.06 < t < 2.06$ we cannot reject the null hypothesis.

For the 1% significance level: if t is larger than 2.787 or smaller than -2.787 we would reject the null hypothesis. If $-2.787 < t < 2.787$ we cannot reject the null hypothesis.

Since $t = -4.01$ we reject the null hypothesis at both the 1% and 5% levels.

Method 2 - Bracket/Bound the P-value

Now we know that

$$-4.01 < -3.725$$

multiplying both sides by -1 leads to

$$3.725 < 4.01$$

converting to P-values

$$0.0005 > P(T > 4.01)$$

multiplying through by 2 yields

$$0.001 > 2P(T > 4.01)$$

and so we can reasonably say that 0.001 > P-value.

Method 3 - Approximate the P-value using linear interpolation

Because we have no value for an upper bound on t we cannot use linear interpolation to approximate the P-value. For reference using a computer we compute the P-value is 0.00048

Testing against the greater than alternative

We will use the following one sided test in the following examples

$$H_0 : \mu_1 - \mu_2 \leq 0 \text{ vs } H_A : \mu_1 - \mu_2 > 0$$

so any P value we compute will be given by $P(T > t)$

Example 2.1

Let df = 31 and suppose the value of the t statistic is $t = 2.56$. Firstly we must choose a row of the t-table to use since $df = 31$ is not in the table. To be conservative pick the largest df that is in the table but smaller than df you want (in this case we pick the row $df = 30$). This approach will give us slightly wider confidence intervals and we will be slightly less likely to reject the null hypothesis.

Method 1 - Test at a fixed level of significance

Consider both the 5% and 1% levels of significance. From the t-distribution table we find that

$$P(T > 2.457) = .01$$

$$P(T > 1.697) = .05$$

For the 5% significance level: if t is larger than 1.697 then we would reject the null hypothesis. If $t < 1.697$ we cannot reject the null hypothesis.

For the 1% significance level: if t is larger than 2.457 we would reject the null hypothesis. If $t < 2.457$ we cannot reject the null hypothesis.

Since $t = 2.56$ we reject the null hypothesis at both the 1% and 5% significance levels.

Method 2 - Bracket/Bound the P-value

From the t-table we find that

$$2.457 < 2.56 < 2.750$$

converting to P-values

$$0.01 > P(T > 2.56) > 0.005$$

and so we can reasonably say that 0.005 < P-value < 0.01.

Method 3 - Approximate the P-value using linear interpolation

$$P\left(T > 2.56\right) \approx .01 + \frac{2.56 - 2.457}{2.750 - 2.457}\left(0.005 - 0.01\right) = 0.0082$$

for comparison the computer give the exact P-value as 0.0078 (note we use $df = 31$ for the computer)

Example 2.2

Let df = 11 and suppose the value of the t statistic is $t = -0.75$.

Method 1 - Test at a fixed level of significance

Consider both the 5% and 1% levels of significance. From the t-distribution table we find that

$$P\left(T > 1.796\right) = .01$$

$$P\left(T > 2.718\right) = .05$$

For the 5% significance level: if t is larger than 1.796 then we would reject the null hypothesis. If $t < 1.796$ we cannot reject the null hypothesis.

For the 1% significance level: if t is larger than 2.718 we would reject the null hypothesis. If $t < 2.718$ we cannot reject the null hypothesis.

Since $t = -0.75$ we cannot reject the null hypothesis at either the 1% or 5% significance levels.

Method 2 - Bracket/Bound the P-value

Note that because the t-distribution is symmetric about 0, the following holds

$$P\left(T > -0.75\right) = 1 - P(T < -0.75) = 1 - P(T > 0.75)$$

From the t-table we find that

$$0.697 < 0.75 < 0.876$$

converting to P-values

$$0.25 > P(T > 0.75) > 0.20$$

multiply through by -1

$$-0.25 < -P(T > 0.75) < -0.20$$

then add 1 to each side

$$0.75 < 1 - P(T > 0.75) < 0.8$$

and so

$$0.75 < P(T > -0.75) < 0.8$$

and so we may state that $0.75 <$ P-value < 0.8

Method 3 - Approximate the P-value using linear interpolation

$$P\left(T > 0.75\right) \approx .25 + \frac{0.75 - 0.697}{0.876 - 0.697}\left(0.2 - 0.25\right) = 0.237$$

and so

$$P\left(T > -0.75\right) \approx 1 - 0.237 = 0.763$$

Testing against the less than alternative

We will use the following one sided test in the following examples

$$H_0 : \mu_1 - \mu_2 \geq 0 \text{ vs } H_A : \mu_1 - \mu_2 < 0$$

and so the P-value will be given by $P(T < t)$

Example 3.1

Let df = 8 and suppose the value of the t statistic is $t = -1.97$.

Method 1 - Test at a fixed level of significance

Consider both the 5% and 1% levels of significance. From the t-distribution table (and by symmetry about 0) we find that

$$P\left(T < -1.86\right) = P\left(T > 1.86\right) = .05$$

$$P\left(T < -2.896\right) = P\left(T > 2.896\right) = .01$$

For the 5% significance level: if t is less than -1.86 then we would reject the null hypothesis. If $t > -1.86$ we cannot reject the null hypothesis.

For the 1% significance level: if t is less than -2.896 we would reject the null hypothesis. If $t > -2.896$ we cannot reject the null hypothesis.

Since $t = -1.97$ we cannot reject the null hypothesis the 1% significance level, but we can reject the null hypothesis at the 5% significance level.

Method 2 - Bracket/Bound the P-value

Observe that

$$-2.306 < -1.97 < -1.86$$

multiply through by -1 to give

$$1.86 < 1.97 < 2.306$$

converting to p-values from the t-table

$$0.05 > P(T > 1.97) < 0.025$$

and so we may reasonably conclude that 0.025 < P-value < .05

Method 3 - Approximate the P-value using linear interpolation

$$P\,(T < -1.97) = P(T > -1.97) \approx .05 + \frac{1.97 - 1.86}{2.306 - 1.86}\,(.025 - .05) = 0.0438$$

for comparative purposes a computer gives 0.0421.