STA 410/2102, Spring 2003 — Assignment #1

Due at **start** of class on February 14. Worth 17% of the final mark.

Note that this assignment is to be done by each student individually. You may discuss it in general terms with other students, but the work you hand in should be your own.

In this assignment, you will implement a variation on the standard t test for the overall mean being zero in a one-way random effects ANOVA model, and investigate the statistical properties of this variation using simulation.

The one-way random effects ANOVA model is appropriate in situations such as the following: Suppose we are interested in whether treating poplar wood with a certain chemical affects how rapidly it rots. To determine this, we take boards produced by a number of sawmills, from locally-harvested poplar trees, and treat half of each board with the chemical, leaving the other half untreated. We expose these boards to the weather for a year, after which we measure the ability of each half of each board to resist compression, as a way of measuring how much the wood has decayed. For each board, we compute the difference in this measurement for the treated and untreated halves. If the treatment has no effect, we would expect that the mean of this difference will be zero, except for random sampling variation. We therefore wish to test the hypothesis that the true population mean is zero to determine whether the treatment has an effect. A one-sided test might be appropriate here, but we will look at a two-sided test (as would be appropriate if we think the treatment might actually make things worse). In doing this test, we need to account for the possibility that the effect is different (on average) for boards produced by different sawmills (either because of the way the sawmill works, or because the sawmills get their trees from different regions). We intend to apply our conclusions not just to the sawmills in the study, but to other similar sawmills for which we did not gather data.

We can formalize models for such problems as follows. For item j in group i (eg, board j from sawmill i), we have a measurement, x_{ij} (eg, the difference in compressive strength between treated and untreated halves). We have n_g groups in all, for each of which we have measurements on n_s items (for simplicity, the number of measurements is here assumed to be the same for all groups). We model the measurements as follows:

$$x_{ij} \;=\; \mu + a_i + b_{ij}$$

where μ is the unknown overall mean, a_i is a "random effect" for group i, and b_{ij} is the random variation for a particular item (the jth in group i). The random effects for the groups, a_i, are assumed to be independent and normally distributed with mean zero and variance σ_g^2. The item effects, b_{ij}, are assumed to be independent and normally distributed with mean zero and variance σ_s^2 (the same for all groups). The values of the parameters μ, σ_g^2, and σ_s^2 are not known.

A standard approach to testing the null hypothesis that $\mu = 0$ versus the alternative that $\mu \neq 0$ is to compute the group means:

$$\bar{x}_i \;=\; \frac{1}{n_s}\sum_{j=1}^{n_s} x_{ij}$$

and to then do a standard t test of the null hypothesis that the mean of the $\bar{x}_i$ is zero. This test is based on the following t statistic:

$$t \;=\; \frac{\bar{x}}{\sqrt{s_g^2/n_g}}$$

where $\bar{x} = \dfrac{1}{n_g}\displaystyle\sum_{i=1}^{n_g}\bar{x}_i$ and $s_g^2 = \dfrac{1}{n_g-1}\displaystyle\sum_{i=1}^{n_g}(\bar{x}_i - \bar{x})^2$.

Note that the mean of the $\bar{x}_i$ is μ, so this null hypothesis is the same as what we wish to test. Note also that the $\bar{x}_i$ are independent, and their distribution is normal. These are the assumptions needed for concluding that the t statistic above has the t distribution with n_g-1 degrees of freedom, if the null hypothesis is true. We can therefore compute the p-value of a two-sided test as $P(|t| > t_{obs})$, where t_{obs} is the value of t actually observed, and t is assumed to have the $t(n_g-1)$ distribution.

There's something disturbing about this test, however. It is based on estimating the variance of the $\bar{x}_i$ by s_g^2 (which appears in the denominator of the t statistic). The true variance of the $\bar{x}_i$ is $\sigma_g^2 + \sigma_s^2/n_s$ — the sum of the variance in the true group means (a_i) and the variance in the corresponding sample means ($\bar{x}_i$) that is due to sampling variability within each group. We could estimate σ_s^2 by the following (which averages all the estimates from within each group):

$$s_s^2 \;=\; \frac{1}{n_g}\sum_{i=1}^{n_g}\frac{1}{n_s-1}\sum_{j=1}^{n_s}(x_{ij} - \bar{x}_i)^2$$

Now, what if s_g^2 is less than s_s^2/n_s? We know that the true variance of the $\bar{x}_i$ can't be less than σ_s^2/n_s, so if we think that s_s^2 is a good estimate of σ_s^2, we shouldn't believe that s_g^2 is a good estimate of the variance to use in the t test if it is less than s_s^2/n_s. But in using the standard t test, we are ignoring this possibility.

We might therefore consider using a modified hypothesis test, based on the following modified t statistic:

$$t' = \frac{\bar{x}}{\sqrt{\max(s_g^2, s_s^2/n_s)/n_g}}$$

Unfortunately, there is no reason to think that this modified test statistic has the $t(n_g-1)$ distribution under the null hypothesis, so a p-value computed from t' on this assumption will not be uniformly distributed between 0 and 1 if the null hypothesis is true (as p-values should be).

Nevertheless, we might investigate the properties of this modified test, to see whether it may be useful. In particular, we might focus on the Type I and Type II error probabilities if we decide to reject or not reject the null hypothesis according to whether the p-value obtained is less than or greater than a specified level, α, such as 0.05 or 0.01.

The Type I error probability is the probability of (incorrectly) rejecting if the null hypothesis is actually true, and should be equal to α, if p-values are uniformly distributed under the null hypothesis. This should be the case for the standard t test. For the modified test, it may not be true, and the actual Type I error rate may depend on the values of σ_g^2 and σ_s^2 (or more precisely, on their ratio).

The Type II error probability is the probability of not rejecting the null hypothesis when it is actually false. It will depend on the actual value of μ, since if μ is far from zero, we would expect to reject more often than if μ is close to zero. It will also depend on the values of σ_g^2 and σ_s^2.

To investigate the performance of the standard and modified tests, you should write the following R functions (giving them names and arguments as shown):

```
gen.data (ng, ns, mean, varg, vars)
```

> This function should generate a simulated data set with `ng` groups, each with `ns` items. The value returned should be a data frame with columns called `group` and `value`. The group id should be an integer from 1 to `ng`. The value should be randomly generated from the model with μ, σ_g^2, and σ_s^2 set to `mean`, `varg`, and `vars`.

```
standard.t (d)
```

> This function should perform a standard t test of $\mu = 0$ on the data in `d`, returning a list with an element called `p.value`. You may put other elements in the list returned as well (such as the values of the $\bar{x}_i$), which may be helpful in debugging, and which might be of use to someone using this procedure in practice.

`modified.t (d)`

> This function should perform the modified t test on the data in `d`, returning a list with an element called `p.value`. Again, you may put other elements in the list returned as well.

`simulate.test (k, ng, ns, mean, varg, vars, test)`

> This function should simulate `k` data sets using `gen.data` (with the indicated arguments), and apply the test given by `test` (which will be either `standard.t` or `modified.t`) to each data set. The value returned should be the vector of `k` p-values obtained.

`estimate.rejection.prob (k, level, means, ng, ns, varg, vars, test)`

> This function should estimate the probabilities of the indicated test rejecting the null hypothesis at the given level when the data is simulated using the given arguments to `gen.data`, but with the `mean` argument set to each element in the vector `means` in turn. The estimates should be based on `k` simulated data sets for each mean. The results should be a vector of estimated rejection probabilities for each element of `means`.

You should use these functions to investigate the performance of the standard and modified tests when $n_g = 5$ and $n_s = 4$, for a variety of values for μ, σ_g^2, and σ_s^2. To start, you should produce histograms of the distribution of p-values under the null hypothesis for the standard and modified tests, in order to see when and in what way the modified test behaves substantially differently. You should then investigate the Type I and Type II error probabilities of both tests, using a level of 0.05. You should summarize your results in this respect by producing plots of the estimated probability of rejection as a function of μ.

You should base your estimates on at least 2000 data sets; more would be better if it doesn't take too long. You should estimate roughly how accurate your estimates are.

Finally, you should discuss the implications of your results. Does it appear that the modified test produces a practical improvement over the standard test?

Here are some R functions that you may find useful: `rnorm`, `data.frame`, `unique`, `mean`, `var`, `pt`, `max`, `hist`, `plot`.