Worksheet 08 (Solutions)

1. Our goal is, broadly, to produce two independent random variabels that will have chi-squared distributions under the null hypothesis. One will depend only on the sample means and the other only on the sample variances. The latter is the easier of the two, so let's start there. Consider the following quantity:

$$\sum_{j=1}^{k} \frac{(n_j - 1)S_j^2}{\sigma^2} = \frac{(n_1 - 1)S_1^2}{\sigma^2} + \cdots + \frac{(n_j - 1)S_j^2}{\sigma^2}$$

What is its distribution? Note that this result does not depend on the null hypothesis

Solution: The jth term is, from our previous results, a $\chi^2(n_j - 1)$. Since all of the terms are independent, the sum is just a $\chi^2(N - K)$, a chi-squared with the sum of the degrees of freedom. This will be N (the total number of observations) minus K (the number of blocks).

2. Now, let's work out another chi-squared distribution based only on the means. What is the distribution of the following quantity, where μ is the hypothesized mean of all the blocks?

$$\left[\frac{\bar{X}_j - \mu}{\sqrt{\sigma^2 / n_j}} \right]^2$$

Solution: The part inside of the square is just a standardized sample mean. Squaring in results in a $\chi^2(1)$.

3. Next, what is this quantity?

$$\left[\frac{\bar{X} - \mu}{\sqrt{\sigma^2 / N}} \right]^2$$

Solution: By the same logic, this is also a $\chi^2(1)$.

4. Using a similar derivation from worksheet 2, we can show that the following is true:

$$\sum_{j=1}^{K} \left[\frac{\bar{X}_j - \mu}{\sqrt{\sigma^2 / n_j}} \right]^2 = \sum_{j=1}^{K} \left[\frac{\bar{X}_j - \bar{X}}{\sqrt{\sigma^2 / n_j}} \right]^2 + \left[\frac{\bar{X} - \mu}{\sqrt{\sigma^2 / N}} \right]^2$$

Based on this and your previous results, what is the distribution of the second sum in the equation above?

Solution: The left-hand side is a sum of K independent $\chi^2(1)$'s, which is a $\chi^2(K)$. The last term is a single independent $\chi^2(1)$. So, because chi-squared add together, the middle term must be a $\chi^2(K-1)$.

5. Put all of the previous results together to construct an F-statistic for the hypothesis. Simplify the cancel the unknown value σ^2. While technically either way is valid, put the chi-squared based on the means in the numerator and the chi-squared based on the variances in the denominator. This will follow the typical convention. Make sure to write down the distribution of the test statistic under H_0.

Solution: We have the following:

$$F = \frac{\frac{1}{K-1}\sum_{j=1}^{K}\left[\frac{\bar{X}_j-\bar{X}}{\sqrt{\sigma^2/n_j}}\right]^2}{\frac{1}{N-K}\sum_{j=1}^{k}\frac{(n_j-1)S_j^2}{\sigma^2}} = \frac{\frac{1}{K-1}\sum_{j=1}^{K} n_j \cdot \left[\bar{X}_j - \bar{X}\right]^2}{\frac{1}{N-K}\sum_{j=1}^{k}(n_j-1)S_j^2} \sim F(K-1, N-K)$$