Solution to Assignment #14

1. *(ANOVA Theory)*

   (a) Show that $\mathsf{E}(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2) = (n-a)\sigma^2$.

      i. Show that $y_{ij}-\bar{y}_i = (y_{ij}-\beta_i) - \frac{1}{n}\sum_{k=1}^{n_i}(y_{ik}-\beta_i)$.

      ii. Show that $\mathsf{E}((y_{ij}-\beta_i)(y_{ik}-\beta_i))$ is 0 when $j \neq k$ and σ^2 when $j = k$.

      iii. Evaluate $\mathsf{E}(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2)$ by making the substitution to represent the data in terms of $y_{ij}-\beta_i$, square the expression (keeping $y_{ij}-\beta_i$ together!), exchange the order of summation and expectation, and simplify.

Solution: Here is a complete derivation. Begin by writing everything in terms of the independent variables y_{ij} and centering these random variables by subtracting their means.

$$
\begin{aligned}
y_{ij}-\bar{y}_i &= y_{ij} - \frac{\sum_{k=1}^{n_i} y_{ik}}{n_i} \\
&= y_{ij} - \beta_i + \beta_i - \frac{\sum_{k=1}^{n_i} y_{ik} - \beta_i + \beta_i}{n_i} \\
&= (y_{ij}-\beta_i) - \frac{\sum_{k=1}^{n_i} y_{ik} - \beta_i}{n_i}
\end{aligned}
$$

Next, square this expression and simplify.

$$
\begin{aligned}
(y_{ij}-\bar{y}_i)^2 &= \left((y_{ij}-\beta_i) - \frac{\sum_{k=1}^{n_i} y_{ij} - \beta_i}{n_i}\right)^2 \\
&= (y_{ij}-\beta_i)^2 - 2\left((y_{ij}-\beta_i)\frac{\sum_{k=1}^{n_i} y_{ik} - \beta_i}{n_i}\right) + \left(\frac{\sum_{k=1}^{n_i} y_{ik} - \beta_i}{n_i}\right)^2 \\
&= (y_{ij}-\beta_i)^2 - \frac{2}{n_i}\left((y_{ij}-\beta_i)^2 + \sum_{k\neq j}(y_{ij}-\beta_i)(y_{ik}-\beta_i)\right) + \frac{1}{n_i^2}\sum_{k=1}^{n_i}\sum_{m=1}^{n_i}(y_{ik}-\beta_i)(y_{im}-\beta_i)
\end{aligned}
$$

Note that

$$\mathsf{E}(y_{ij}-\beta_i)^2 = \mathsf{Var}(y_{ij}) = \sigma^2$$

and that when $j \neq k$

$$\mathsf{E}(y_{ij}-\beta_i)(y_{ik}-\beta_i) = \mathsf{E}(y_{ij}-\beta_i)\mathsf{E}(y_{ik}-\beta_i) = 0$$

since the expected value of a product of independent random variables is the product of the expectations and $\mathsf{E}(y_{ij}-\beta_i) = 0$. It follows that

$$\mathsf{E}(y_{ij}-\bar{y}_i)^2 = \sigma^2 - \frac{2}{n_i}\sigma^2 + \frac{1}{n_i^2}(n_i\sigma^2) = \sigma^2 - \frac{1}{n_i}\sigma^2 \ .$$

To finish the derivation,

$$\begin{aligned}
\mathsf{E}\left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2\right) &= \sum_{i=1}^{a}\sum_{j=1}^{n_i}\mathsf{E}\left((y_{ij}-\bar{y}_i)^2\right)\\
&= \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(\sigma^2-\frac{1}{n_i}\sigma^2\right)\\
&= \sum_{i=1}^{a}\left(n_i\sigma^2-\sigma^2\right)\\
&= n\sigma^2-a\sigma^2\\
&= (n-a)\sigma^2
\end{aligned}$$

(b) Show that $\mathsf{E}(\sum_{i=1}^a n_i(\bar{y}_i-\bar{y})^2) = (a-1)\sigma^2+\sum_{i=1}^a n_i(\beta_i-\beta)^2$ where $\beta=\sum_{i=1}^a(n_i/n)\beta_i$ is the weighted mean of the means from each normal distribution.

i. Show that $\bar{y}_i-\bar{y} = (\bar{y}_i-\beta_i)+(\beta_i-\beta)-\frac{1}{n}\sum_{j=1}^a n_j(\bar{y}_j-\beta_j)$.

ii. Make the above replacement and square out the trinomial expression $(A+B-C)^2 = A^2+B^2+C^2+2AB-2AC-2BC$.

iii. Multiply by n_i and take expectations, recalling that $\mathsf{E}(\bar{y}_i-\beta_i)=0$, $\mathsf{E}(\bar{y}_i-\beta_i)^2=\sigma^2/n_i$, and $\mathsf{E}(y_i-\beta_i)(y_j-\beta_j)=0$ when $i\neq j$ since $\bar{y}_i$ and $\bar{y}_j$ are then independent. This shows that $\mathsf{E}(n_i(\bar{y}_i-\bar{y})^2)=n_i(\beta_i-\beta)^2+\sigma^2(1-n_i/n)$.

iv. Take the final sum, and simplify to complete the problem.

(c) Conclude $(\sum_{i=1}^a\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2)/(n-a)$ is an unbiased estimator of σ^2 for any $\{\beta_i\}$, but that $\sum_{i=1}^a n_i(\bar{y}_i-\bar{y})^2$ is an unbiased estimator of σ^2 only if the $\{\beta_i\}$ are all equal.

Solution: The solution begins again by writing the expression in terms of independent random variables (the sample means) and centering them all.

$$\begin{aligned}
(\bar{y}_i-\bar{y})^2 &= \left(\bar{y}_i-\frac{\sum_{j=1}^a n_j\bar{y}_j}{n}\right)^2\\
&= \left(\bar{y}_i-\beta_i+\beta_i-\frac{\sum_{j=1}^a n_j(\bar{y}_j-\beta_j+\beta_j)}{n}\right)^2\\
&= \left((\bar{y}_i-\beta_i)+(\beta_i-\beta)-\frac{\sum_{j=1}^a n_j(\bar{y}_j-\beta_j)}{n}\right)^2
\end{aligned}$$

where

$$\beta=\frac{\sum_{j=1}^a n_j\beta_j}{n}$$

is the weighted average of the (possibly different) means from each group. Next square out the terms and rearrange.

$$
\begin{aligned}
(\bar{y}_i - \bar{y})^2 &= (\bar{y}_i - \beta_i)^2 + (\beta_i - \beta)^2 + \left(\frac{\sum_{j=1}^a n_j(\bar{y}_j - \beta_j)}{n} \right)^2 \\
&\quad +2(\bar{y}_i - \beta_i)(\beta_i - \beta) - 2(\bar{y}_i - \beta_i)\left(\frac{\sum_{j=1}^a n_j(\bar{y}_j - \beta_j)}{n} \right) - 2(\beta_i - \beta)\left(\frac{\sum_{j=1}^a n_j(\bar{y}_j - \beta_j)}{n} \right) \\
&= (\bar{y}_i - \beta_i)^2 + (\beta_i - \beta)^2 + \frac{1}{n^2}\sum_{j=1}^a \sum_{k=1}^a n_j n_k (\bar{y}_j - \beta_j)(\bar{y}_k - \beta_k) \\
&\quad +2(\bar{y}_i - \beta_i)(\beta_i - \beta) - 2(\bar{y}_i - \beta_i)\left(\frac{\sum_{j=1}^a n_j(\bar{y}_j - \beta_j)}{n} \right) - 2(\beta_i - \beta)\left(\frac{\sum_{j=1}^a n_j(\bar{y}_j - \beta_j)}{n} \right)
\end{aligned}
$$

Taking expectations and simplifying using $\mathsf{E}((\bar{y}_i - \beta_i)(\bar{y}_j - \beta_j)) = \mathsf{Var}(\bar{y}_i) = \sigma^2/n_i$ when $i = j$ and 0 otherwise results in this expression.

$$
\begin{aligned}
\mathsf{E}(\bar{y}_i - \bar{y})^2 &= \frac{\sigma^2}{n_i} + (\beta_i - \beta)^2 + \frac{1}{n^2}\sum_{j=1}^a n_j^2(\sigma^2/n_j) + 0 - \frac{2n_i\sigma^2}{nn_i} - 0 \\
&= (\beta_i - \beta)^2 + \frac{\sigma^2}{n_i} - \frac{\sigma^2}{n}
\end{aligned}
$$

Thus, the expected value of the numerator mean square is

$$
\begin{aligned}
\mathsf{E}\left(\sum_{i=1}^a n_i(\bar{y}_i - \bar{y})^2 \right) &= \sum_{i=1}^a n_i \mathsf{E}(\bar{y}_i - \bar{y})^2) \\
&= \sum_{i=1}^a n_i \left((\beta_i - \beta)^2 + \frac{\sigma^2}{n_i} - \frac{\sigma^2}{n} \right) \\
&= \sum_{i=1}^a \left(n_i(\beta_i - \beta)^2 + \sigma^2 - \frac{n_i\sigma^2}{n} \right) \\
&= a\sigma^2 - \sigma^2 + \sum_{i=1}^a n_i(\beta_i - \beta)^2 \\
&= (a-1)\sigma^2 + \sum_{i=1}^a n_i(\beta_i - \beta)^2
\end{aligned}
$$

Thus,

$$
\mathsf{E}\left(\frac{\sum_{i=1}^a n_i(\bar{y}_i - \bar{y})^2}{a-1} \right) = \sigma^2 + \frac{\sum_{i=1}^a n_i(\beta_i - \beta)^2}{a-1}
$$

and is an unbiased estimate of σ^2 when $\beta_i = \beta$ for all i.