Chs. 10 & 12 Proofs

Re: ANOVA, Simple Linear Regression

Question 10.1.10, Page 420

In single-factor ANOVA with I treatments and J observations per treatment, let $\overline{\mu} = \frac{1}{I}\sum_{i=1}^{I}\mu_i$.

Before beginning, we recall two of the numerous assumptions of ANOVA: i) all observations are independent of one another, ii) each $X_{ij} \sim N(\mu_i, \sigma^2)$.

(a) Express $\mathbf{E}\left(\overline{X}_{\cdot\cdot}\right)$ in terms of $\overline{\mu}$. [Hint: $\overline{X}_{\cdot\cdot} = \frac{1}{I}\sum_{i=1}^{I}\overline{X}_{i\cdot}$]

From the above note, it follows that:

$$\mathbf{E}\left(\overline{X}_{i\cdot}\right) = \mathbf{E}\left(\frac{1}{J}\sum_{j=1}^{J}X_{ij}\right) = \frac{1}{J}\sum_{j=1}^{J}\mathbf{E}\left(X_{ij}\right) = \frac{J\mu_i}{J} = \mu_i \quad \forall i = 1, 2, ..., I$$

Then:

$$\mathbf{E}\left(\overline{X}_{\cdot\cdot}\right) = \mathbf{E}\left(\frac{1}{I}\sum_{i=1}^{I}\overline{X}_{i\cdot}\right) = \frac{1}{I}\sum_{i=1}^{I}\mathbf{E}\left(\overline{X}_{i\cdot}\right) = \frac{1}{I}\sum_{i=1}^{I}\mu_i \equiv \overline{\mu}$$

(b) Determine $\mathbf{E}\left(\overline{X}_{i\cdot}{}^2\right)$. [Hint: Use the rearrangement of the variance formula]

$$\mathbf{Var}\left(\overline{X}_{i\cdot}\right) = \mathbf{Var}\left(\frac{1}{J}\sum_{j=1}^{J}X_{ij}\right) = \frac{1}{J^2}\sum_{j=1}^{J}\mathbf{Var}\left(X_{ij}\right) = \frac{J\sigma^2}{J^2} = \frac{\sigma^2}{J} \quad \forall i = 1, 2, ..., I$$

$$\mathbf{E}\left(\overline{X}_{i\cdot}\right) = \mu_i \implies \left(\mathbf{E}\left(\overline{X}_{i\cdot}\right)\right)^2 = {\mu_i}^2 \quad \forall i = 1, 2, ..., I$$

$$\mathbf{E}\left(\overline{X}_{i\cdot}^2\right) = \mathbf{Var}\left(\overline{X}_{i\cdot}\right) + \left(\mathbf{E}\left(\overline{X}_{i\cdot}\right)\right)^2 = \frac{\sigma^2}{J} + {\mu_i}^2 \quad \forall i = 1, 2, ..., I$$

(c) Determine $\mathbf{E}\left(\overline{X}_{\cdot\cdot}{}^2\right)$.

$$\mathbf{Var}\left(\overline{X}_{\cdot\cdot}\right) = \mathbf{Var}\left(\frac{1}{I}\sum_{i=1}^{I}\overline{X}_{i\cdot}\right) = \frac{1}{I^2}\sum_{i=1}^{I}\mathbf{Var}\left(\overline{X}_{i\cdot}\right) = \frac{I\sigma^2}{I^2 J} = \frac{\sigma^2}{IJ}$$

$$\mathbf{E}\left(\overline{X}_{\cdot\cdot}\right) = \overline{\mu} \implies \left(\mathbf{E}\left(\overline{X}_{\cdot\cdot}\right)\right)^2 = \overline{\mu}^2$$

$$\mathbf{E}\left(\overline{X}_{\cdot\cdot}^2\right) = \mathbf{Var}\left(\overline{X}_{\cdot\cdot}\right) + \left(\mathbf{E}\left(\overline{X}_{\cdot\cdot}\right)\right)^2 = \frac{\sigma^2}{IJ} + \overline{\mu}^2$$

(d) Determine $\mathbf{E}(\mathrm{SSTr})$ and then show that $\mathbf{E}(\mathrm{MSTr}) = \sigma^2 + \frac{J}{I-1}\sum_{i=1}^{I}(\mu_i - \overline{\mu})^2$.

$$\begin{aligned}
\mathrm{SSTr} &= \frac{1}{J}\sum_{i=1}^{I}\left(\sum_{j=1}^{J} X_{ij}\right)^2 - \frac{1}{IJ}\left(\sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij}\right)^2 \\
&= \frac{1}{J}\sum_{i=1}^{I}\left(\frac{J}{J}\sum_{j=1}^{J} X_{ij}\right)^2 - \frac{1}{IJ}\left(\frac{IJ}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij}\right)^2 \\
&= \frac{J^2}{J}\sum_{i=1}^{I}\left(\frac{1}{J}\sum_{j=1}^{J} X_{ij}\right)^2 - \frac{IJ^2}{IJ}\left(\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij}\right)^2 \\
&= J\sum_{i=1}^{I}\overline{X}_{i\cdot}^2 - IJ\overline{X}_{\cdot\cdot}^2
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}(\mathrm{SSTr}) &= \mathbf{E}\left(J\sum_{i=1}^{I}\overline{X}_{i\cdot}^2 - IJ\overline{X}_{\cdot\cdot}^2\right) \\
&= J\sum_{i=1}^{I}\mathbf{E}\left(\overline{X}_{i\cdot}^2\right) - IJ\mathbf{E}\left(\overline{X}_{\cdot\cdot}^2\right) \\
&= J\sum_{i=1}^{I}\left(\frac{\sigma^2}{J} + {\mu_i}^2\right) - IJ\left(\frac{\sigma^2}{IJ} + \overline{\mu}^2\right) \\
&= \frac{IJ\sigma^2}{J} + J\sum_{i=1}^{I}{\mu_i}^2 - \sigma^2 - IJ\overline{\mu}^2 \\
&= (I-1)\sigma^2 + J\left(\sum_{i=1}^{I}{\mu_i}^2 - I\overline{\mu}^2\right) \\
&= (I-1)\sigma^2 + J\left(\sum_{i=1}^{I}{\mu_i}^2 - 2I\overline{\mu}^2 + I\overline{\mu}^2\right) \\
&= (I-1)\sigma^2 + J\left(\sum_{i=1}^{I}{\mu_i}^2 - 2\overline{\mu}\sum_{i=1}^{I}\mu_i + \sum_{i=1}^{I}\overline{\mu}^2\right) \\
&= (I-1)\sigma^2 + J\sum_{i=1}^{I}({\mu_i}^2 - 2\overline{\mu}\mu_i + \overline{\mu}^2) \\
&= (I-1)\sigma^2 + J\sum_{i=1}^{I}(\mu_i - \overline{\mu})^2
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}(\mathrm{MSTr}) &= \mathbf{E}\left(\frac{1}{I-1}\mathrm{SSTr}\right) = \frac{1}{I-1}\mathbf{E}(\mathrm{SSTr}) \\
&= \sigma^2 + \frac{J}{I-1}\sum_{i=1}^{I}(\mu_i - \overline{\mu})^2
\end{aligned}$$

(e) Using the result of (d), what is $\mathbf{E}(\text{MSTr})$ when H_0 is true? When H_0 is false, how does $\mathbf{E}(\text{MSTr})$ compare to σ^2?

If H_0 is true, each μ_i is identical so $\overline{\mu} = \mu_i = \mu$. Then:

$$\sum_{i=1}^{I} (\mu_i - \overline{\mu})^2 = 0 \Longrightarrow \mathbf{E}(\text{MSTr}) = \sigma^2$$

If H_0 is not true, then:

$$\sum_{i=1}^{I} (\mu_i - \overline{\mu})^2 > 0 \Longrightarrow \mathbf{E}(\text{MSTr}) > \sigma^2$$

This proves the result on **Lecture 7 Slide 8** that $\mathbf{E}(\text{MSTr}) > \sigma^2$ when the treatment means are not all identical.

Question 12.S.79, Page 540

Show that $\text{SSE} = S_{yy} - \widehat{\beta}_1 S_{xy}$. (I used this formula in my solution for 12.3.31a in the Tutorial 11 Solutions file, but we should always prove formulas before using them.)

$$\begin{aligned}
\text{SSE} &= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\
&= \sum_{i=1}^{n} (y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i))^2 \\
&= \sum_{i=1}^{n} (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)^2 \\
&= \sum_{i=1}^{n} (y_i - (\bar{y} - \widehat{\beta}_1 \bar{x}) - \widehat{\beta}_1 x_i)^2 \\
&= \sum_{i=1}^{n} (y_i - \bar{y} + \widehat{\beta}_1 \bar{x} - \widehat{\beta}_1 x_i)^2 \\
&= \sum_{i=1}^{n} (y_i - \bar{y} - \widehat{\beta}_1 (x_i - \bar{x}))^2 \\
&= \sum_{i=1}^{n} (y_i - \bar{y})^2 - 2\widehat{\beta}_1 \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) + \widehat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 \\
&= S_{yy} - 2\frac{S_{xy}}{S_{xx}} S_{xy} + \frac{{S_{xy}}^2}{{S_{xx}}^2} S_{xx} \\
&= S_{yy} - 2\frac{{S_{xy}}^2}{S_{xx}} + \frac{{S_{xy}}^2}{S_{xx}} \\
&= S_{yy} - \frac{{S_{xy}}^2}{S_{xx}} \\
&= S_{yy} - \widehat{\beta}_1 S_{xy}
\end{aligned}$$

Question 12.S.83, Page 540

Show that $R^2 = 1 - \frac{\text{SSE}}{\text{SST}}$, where R is the sample correlation coefficient.

We know that SST $= S_{yy}$. From the previous proof, we have:

$$\text{SSE} = S_{yy} - \widehat{\beta}_1 S_{xy} = S_{yy} - \frac{{S_{xy}}^2}{S_{xx}}$$

$$\implies \frac{{S_{xy}}^2}{S_{xx}} = S_{yy} - \text{SSE} = \text{SST} - \text{SSE}$$

We also know that

$$R = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

Then:

$$\begin{aligned} R^2 &= \left(\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}\right)^2 \\ &= \frac{{S_{xy}}^2}{S_{xx}S_{yy}} \\ &= \frac{{S_{xy}}^2}{S_{xx}} \cdot \frac{1}{S_{yy}} \\ &= \frac{S_{yy} - \text{SSE}}{S_{yy}} \\ &= 1 - \frac{\text{SSE}}{\text{SST}} \end{aligned}$$