Some ANCOVA theory

Let X be a random variable

Let X_i be iid $\sim f_x$ (pdf or pmf: probability density or mass function) with mean μ and variance σ^2.

Think of X_i as repeated observations from the same population or the same statistic calculated for repeated experiments.

$$\mathrm{var}(X) \equiv \sigma_x^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Let Y be another random variable. Let f_{xy} be the joint pdf (or pmf) of X and Y. f_x and f_y are called marginal pdf's. We now have an additional characteristic of the joint pdf: the covariance of X and Y.

$$\mathrm{cov}(X,Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-2}$$

If X is independent of Y then $f_{xy} = f_x f_y$ and $\mathrm{cov}(X,Y) = 0$.

(We won't use it today, but $\mathrm{cor}(X,Y) = \mathrm{cov}(X,Y)/(\sigma_x \sigma_y.)$

Without any normality requirement, it is easy to show that

$$E(aX + bY) = aE(X) + bE(Y), \text{ but not } E(XY) = E(X)E(Y)$$

and

$$\mathrm{cov}(aX, bY) = ab\,\mathrm{cov}(X,Y) \text{ and } \mathrm{cov}(aX+bY, cZ) = ac\,\mathrm{cov}(X,Z) + bc\,\mathrm{cov}(Y,Z)$$

and

$$\mathrm{var}(aX + bY + c) = a^2\mathrm{var}(X) + b^2\mathrm{var}(Y) + 2ab\,\mathrm{cov}(X,Y).$$

What is $\mathrm{var}(aX + bY + cZ)$?

$$\begin{aligned}
\mathrm{var}(aX + bY + cZ) &= \mathrm{var}((aX + bY) + cZ) \\
&= \mathrm{var}(aX + bY) + c^2\mathrm{var}(Z) + 2c\ \mathrm{cov}(aX + bY, Z) \\
&= a^2\mathrm{var}(X) + b^2\mathrm{var}(Y) + 2ab\ \mathrm{cov}(X, Y) + c^2\mathrm{var}(Z) \\
&\quad +2c(a\mathrm{cov}(X, Z) + b\mathrm{cov}(Y, Z)) \\
&= a^x\mathrm{var}(X) + b^2\mathrm{var}(Y) + c^2\mathrm{var}(Z) + 2ab\ \mathrm{cov}(X, Y) \\
&\quad +2ac\ \mathrm{cov}(X, Z) + 2bc\ \mathrm{cov}(Y, Z)
\end{aligned}$$

In simple linear regression, where $\hat{\beta}$ is a vector of length 2

$$\mathrm{var}(\hat{\beta}) = \sigma^2[X'X]^{-1}$$

where X here is a matrix with the first column all 1's and the second column equal to the n explanatory variables, x_i.

Using standard matrix properties, $[X'X]$ has diagonal elements n and $\sum x_i^2$, and off-diagonal elements equal to $\sum x_i$, Also

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = g \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

where g is $\frac{1}{ad-bc}$, so $\mathrm{var}(\widehat{\beta_0})$ and $\mathrm{var}(\widehat{\beta_1})$ can be explicitly written out. In practice we need to use S.E.$(\widehat{\beta_j})$ where σ^2 is replaced by an estimator, $\widehat{\sigma^2} = SS_{res}/\mathrm{df}$. Note the effect of "centering" the explanatory variable. Also note that it is CI's and p-values that need normality.

F-test of two nested models:

The numerator is an estimator of σ^2 under the null hypothesis that the extra components of the "larger" model are useless, and the denominator is always an estimator of σ^2. The numerator is $(SS_{small} - SS_{big})/dfn$ with df equal to dfn, the difference in the number of parameters between the two models. The denominator is SS/dfd for the "larger" model where $dfd = n - pl$ and n is the number of subjects in the regression and pl is the number of parameters in the "larger" model. Under the null distribution, this F statistic has a central F distribution with dfn and dfd degrees of freedom.

Scheffe multiple (infinite) comparison procedure for contrasts:

$$C = \sum_{i=1}^{r} c_i Y_i$$

$$\mathrm{var}(C) \equiv \sigma_C^2 = \sum_{i=1}^{r} c_i^2 \mathrm{var}(Y_i) + 2 \sum_{i<j} c_i c_j \ \mathrm{cov}(Y_i Y_j)$$

We need a value of m such that the experiment-wise error rate of any number of confidence intervals of the form $C \pm m\ \sigma_C$ is bounded by α. Scheffe found m to be $\sqrt{(r-1)F_{(\alpha,r-1,dfd)}}$, where dfd is the df of σ_C.

Summarizing adjusted means in a model with a single covariate fixed at value x_0 and T treatments and different slopes and intercepts for each treatment:

Pick a few meaningful values of x_0 such as Q1, Q2, Q3.

Let $\widehat{\mu_j}$ represent $E(Y|X = x_0,\ \ T = t_j)$. The model says that adjusted mean $\widehat{\mu_j} = \widehat{\alpha_j} + x_0\widehat{\beta_j}$. So $\mathrm{var}(\widehat{\mu_j}) \equiv \sigma^2_{\widehat{\mu_j}} = \mathrm{var}(\widehat{\alpha_j}) + x_0^2\mathrm{var}(\widehat{\beta_j}) + 2x_0\ \mathrm{cov}(\widehat{\alpha_j}, \widehat{\beta_j})$.

A confidence region can be written as $\widehat{\mu_j} \pm t_{(1-\alpha/2,df)}\ \sigma_{\widehat{\mu_j}}$.

Multiple testing that $E(Y|X = x_0,\ \ T = t_1)$ differs from $E(Y|X = x,\ \ T = t_2)$ in a model with a single covariate fixed at value x_0 and T treatments and different slopes and intercepts for each treatment:

Let $\widehat{\mu_j}$ represent $E(Y|X = x_0,\ \ T = t_j)$. The model says $\widehat{\mu_j} = \widehat{\alpha_j} + x_0\widehat{\beta_j}$. So $\mathrm{var}(\widehat{\mu_j}) = \mathrm{var}(\widehat{\alpha_j}) + x_0^2\mathrm{var}(\widehat{\beta_j}) + 2x_0\ \mathrm{cov}(\widehat{\alpha_j}, \widehat{\beta_j})$. For any pair of levels of treatment, the β's are uncorrelated. So $\mathrm{var}(\widehat{\mu_1} - \widehat{\mu_2}) = \mathrm{var}(\widehat{\mu_1}) + \mathrm{var}(\widehat{\mu_2})$.

To find the region of x's where there is a "significant difference" in adjusted outcomes between a pair of treatments, make an infinite number of CIs (for all x's, or all in a reasonable range) using Scheffe's method, and see which ones exclude zero for the difference.