Jack Todhunter
STAT 642
Assignment 06
2/24/22

Read H.O 16, Chp 17 Design & ANOVA books:

1.) • Factors: • Field (15 levels - Random)
• Section nested w/in fields , S(F), (30 levels - Random)
• Location nested w/in section , L(S,F) (2 levels - Random)

Q: ask if this is the correct way to specify.

(Our treatment is a field w/ 15 levels), EU is section , Subsample is location (either 1 or 2 w/in field)

(H.O. 6 pg 40) →

(a) Write a linear model for this experiment. Make sure to explain each term in your model along w/ any distributional requirements for the term.

• Model: $y_{ijk} = \mu + A_i + e_{ij} + d_{ijk}$

• μ is the overall mean of the porosity index on the experimental farm

Q: why doesn't this say random effect → • A_i is the effect of the i^{th} field $(i = 1, \dots, 15)$

• $A_i \overset{iid}{\sim} N(0, \sigma_A^2)$

• e_{ij} is the random effect of the j^{th} section of the i^{th} field $(j = 1, \dots, n_i = 1, 2)$ $\forall i \in [15]$ → $n_i = r = 2$

• $e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$

• d_{ijk} is the random effect of the k^{th} subsample in section j of field i $(k = 1, \dots m_{ij})$ and all other sources.

• $d_{ijk} \overset{iid}{\sim} N(0, \sigma_d^2)$

$m_{ij} = \begin{cases} 2, & (i,j) \in \{(1,1), (1,2), (6,1), (6,2), (11,1), (11,2)\} \\ 1, & o.w. \end{cases}$

• A_i's, e_{ij}'s & d_{ijk}'s are mutually independent.

(b)

Source	DF	Sum Squares	Mean Square	EMS	Error Term	Error df	F value	Pr > F
Field	14	14.43	1.03	var(residual) + 1.9075 var(section(Field)) + [illegible] var(Field)	0.9921 MS(section(Field)) + 0.0079 MS(residual)	15.452	1.33	0.2919
section(Field)	15	11.53	[illegible]	var(residual) + 1.2 (var(section(Field)))	MS(Residual)	6	0.52	0.8546
Residual	6	3.80	[illegible]	var(Residual)				

(c) Display the proportion of the variance in the porosity readings due to field, section & residuals.

$\hat{\sigma}_y = \hat{\sigma}_A^2 + \hat{\sigma}_e^2 + \hat{\sigma}_d^2 = 0.05944 + 0 + 0.93600 = 0.99544$

• Proportion of variance due to field: $\hat{\sigma}_A^2 / \hat{\sigma}_y^2 = \frac{0.05944}{0.99544} = 0.05971229$

• Proportion of variance due to section(field): $\hat{\sigma}_e^2 / \hat{\sigma}_y^2 = 0/0.99544 = 0$

• Proportion of variance due to residuals: $\hat{\sigma}_d^2 / \hat{\sigma}_y^2 = \frac{0.93600}{0.99544} = 0.94028771$

(d) $H_0: \sigma_A^2 \leq 0$ $H_a: \sigma_A^2 > 0$. Looking at the above table we see our $F_{stat} = 1.33$ w/ $P(F > 1.33) = 0.2919 \Rightarrow$ we fail to reject H_0.

(e) $H_0: \sigma_e^2 \leq 0$ $H_a: \sigma_e^2 > 0$. looking at the above table we see our $F_{stat} = 0.52$ w/ $P(F > 0.52) = 0.8546 \Rightarrow$ Fail to reject H_0.

2) A plant pathologist wants to evaluate cottonseed stored at various cotton gins for Aflatoxin, a toxin produced by organisms associated w/ cotton seed. The pathologist will randomly select eight cotton gins for analysis. Then at each cotton gin, she will take a 3 lbs sample from r randomly selected 50-ton lots of cotton seed.

(HO.6 pg7) → (a) Would a test of the difference in the mean aflatoxin levels over the eight cotton gins be a reasonable test to conduct? Justify your answer.

- No, a test of the difference in the means would not be a reasonable test to conduct because the levels of the cotton gin factor are randomly selected. We are interested in determining if there are differences in the population of treatments, not just the treatments observed in the study. Thus, our model is a random effects model $y_{ij} = \mu + A_i + e_{ij}$.
  In this model the treatment effects A_i are random variables w/ σ_A reflecting the difference in the treatment means across all treatments in the population of treatments. Thus what we want to test is not a difference in the mean aflatoxin levels, but instead we want to test $H_0: \sigma_A = 0$

Q: I'm still not sure why we test $\sigma_A = 0$ vs $\sigma_A > 0$? → I understand how this test works, if the variance of the treatment effects is $\neq 0$ then there is differences in the treatment effect population. However, how is this not the same as testing $\mu_1 = \mu_2 = \cdots = \mu_t$ or $\tau_1 = \tau_2 = \cdots = \tau_t$? Is it the case that in our sample of treatments we get $\tau_1 = \tau_2 = \cdots = \tau_t$ but that for some $\tau_i, \tau_j \in S$ $\tau_i \neq \tau_j$ where S is the population of treatments? Yes, that could happen,

① What is the probability of that happening?

② If we get $\sigma_A = 0$ why does the same problem not arise?

(b) What is the minimum number r that the plant pathologist would have to take from each of the eight cotton gins in order to have power of at least 90% to detect a ratio $\sigma_A^2 / \sigma_e^2 = 2$ at the 0.01 significance level, where σ_A^2 is the variance associated w/ differences in Aflatoxin level from cotton gin to cotton gin and σ_e^2 is the variance across 50-ton lots of cottonseed within each cotton gin.

$\Rightarrow \sigma_A^2 = 2\sigma_e^2$

$\sigma_A = \sqrt{2}\,\sigma_e$

- Using code from RepSizeRandEffectFixed_Fundt.R we get $\boxed{r = 5}$

3.) Using RepSizeRandEffectFixed_Fundt.R I get $\boxed{t = 10.}$

$\sigma_e = 2 \Rightarrow \sigma_e^2 = 4.$

$\sigma_A > 2.1 \Rightarrow \sigma_A^2 > 4.41 = 1.1025(4) = 1.1025\,\sigma_e^2$

$\Rightarrow \tau = 1.1025$

$r = 5, \ \alpha = 0.01, \ \gamma = 0.90$