STA 305 Midterm practice problems

1.

	df	SS	MS	F
Model (treatment)	5-1=4	300	300/4	~~$\frac{300/4}{50/67}$~~
Error	~~72-5=67~~ 60-5=55	~~50~~ 500	~~50/67~~ 500/55=9.1	F=8.24
Total	~~72-1=71~~ 60-1=59			

Compare F to ~~$F_{4,67,\alpha}$~~ F(4,55,alpha)

2. $$(y_{ij} - \bar{y}) = (y_{ij} - \bar{y}_{i.}) + (\bar{y}_{i.} - \bar{y})$$

add and subtract $\bar{y}_{i.}$ then square and sum over treatments and observations

$$(y_{ij} - \bar{y})^2 = \left[(y_{ij} - \bar{y}_{i.}) + (\bar{y}_{i.} - \bar{y})\right]^2$$

$$SSTotal = \sum_{i=1}^{K}\sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\left[(y_{ij} - \bar{y}_{i.}) + (\bar{y}_{i.} - \bar{y})\right]^2$$

$$= \sum_i\sum_j \left[(y_{ij} - \bar{y}_{i.})^2 + 2(y_{ij} - \bar{y}_{i.})(\bar{y}_{i.} - \bar{y}) + (\bar{y}_{i.} - \bar{y})^2\right]$$

Show that $\sum_i\sum_j 2(y_{ij} - \bar{y}_{i.})(\bar{y}_{i.} - \bar{y}) = 0$

$$= \sum_i\sum_j (y_{ij} - \bar{y}_{i.})^2 + \sum_i\sum_j (\bar{y}_{i.} - \bar{y})^2$$

$$= SSE + SST$$

3. $$SSE = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2$$

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

$$\Rightarrow y_{ij} \sim N(\mu + \tau_i, \sigma^2)$$

$$\therefore E(y_{ij}^2) = \sigma^2 + (\mu + \tau_i)^2 \quad \text{(why?)}$$

$$E(\bar{y}_{i\cdot}^2) = \frac{\sigma^2}{n_i} + (\mu + \tau_i)^2 \quad \text{(why?)}$$

Can write SSE as (why?)

$$\sum_{i=1}^{K} \left(\sum_{j=1}^{n_i} y_{ij}^2 - n_i \bar{y}_{i\cdot}^2 \right)$$

$$\Rightarrow E(SSE) = \sum_{i=1}^{K} \left(\sum_{j=1}^{n_i} E(y_{ij}^2) - n_i E(\bar{y}_{i\cdot}^2) \right)$$

$$= \sum_{i=1}^{K} \left(n_i (\sigma^2 + (\mu + \tau_i)^2) - \sigma^2 - n_i (\mu + \tau_i)^2 \right)$$

$$= \sum_{i=1}^{K} (n_i - 1)\sigma^2 = (n - K)\sigma^2, \quad n = \sum_{i=1}^{K} n_i$$

$$E(MSE) = E\left(\frac{SSE}{n-K} \right) = \sigma^2$$

4. (1) additive model $Y_{it} = \mu + \tau_i + \varepsilon_{it}$

(2) ε_{it} independent, constant variance for $i=1,\dots,v$, and normally distributed

5. (1) $H_0: \mu_A = \mu_B$. Under H_0 treatment A and treatment B have equal probability of being assigned to an experimental unit.

There are $\binom{6}{3} = 20$ possible ways of allocating 3 A's and 3 B's to 6 people.

(2) Called the randomization distribution

(3) Calculate empirical CDF (cumulative distribution function) $\hat{F}(x) = \#\{\text{values} \le x\}/20$

$$\text{p-value} = 1 - \hat{F}(\text{observed } \bar{x} - \bar{y})$$

$$= \#\{\text{values} \ge \text{observed } \bar{x} - \bar{y}\}/20$$

$$\text{values} = \{\text{all possible differences}\}$$

$$= \{\text{diff}_1, \text{diff}_2, \dots, \text{diff}_{20}\}$$

6. (i) Two Sample experiment compares

two <u>independent</u> Samples on different experimental units.

(ii) Paired experiment compares

two samples on the same experimental

unit

7. (i) increase Sample size

(ii) Increase type I error.

(iii) decrease variability through better design.

(iv) increase effect to be detected.

5/5

8. Multiple comparisons problem is the problem of testing more than one hypothesis simultaneously. This leads to a theoretical increase in the type I error rate.

The collection of comparisons is called a "family".

The familywise error rate is the probability that at least one comparison will include a type I error.

9. Conduct analysis using R.

10. Same as 9.