Homework 6 (due May 30)

In your solution include your name and the homework number. Please staple your pages together. When solving the problems below, give detailed derivations in order to receive credit.

10 1. *(Sampling Distribution.)* Before the midterm, we learned that for i.i.d. observations $y_1, ..., y_n \sim N(0, \sigma^2)$, $\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sigma^2} \sim \chi^2_{n-1}$. Review the derivation of this result in your notes.

2. *(ANOVA.)* Today, we learned one-way ANOVA. In this setting, we have independent observations $y_{ij} \sim N(\mu + \alpha_i, \sigma^2)$ for $i = 1, ..., n$ and $j = 1, ..., m$. We also imposed the constraint $\sum_{i=1}^n \alpha_i = 0$ for the sake of identifiability.

   10 (a) Under the null hypothesis that $\alpha_1 = \alpha_2 = ... = \alpha_n$, together with the constraint $\sum_{i=1}^n \alpha_i = 0$, show $\alpha_i = 0$ for each $i = 1, ..., n$.

   10 (b) Show $\frac{\sum_{j=1}^m (y_{ij} - \bar{y}_{i\cdot})^2}{\sigma^2} \sim \chi^2_{m-1}$ under the null. You can use the sampling distribution theorem.

   10 (c) Show $\frac{\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_{i\cdot})^2}{\sigma^2} \sim \chi^2_{n(m-1)}$ under the null. This is the first conclusion in the theorem you learned today.

   10 (d) Define $x_i = \sqrt{m}\bar{y}_{i\cdot}$, and show $m \sum_{i=1}^n (\bar{y}_{i\cdot} - \bar{y})^2 = \sum_{i=1}^n (x_i - \bar{x})^2$.

   10 (e) What is the distribution of x_i under the null?

   10 (f) Show $\frac{m \sum_{i=1}^n (\bar{y}_{i\cdot} - \bar{y})^2}{\sigma^2} \sim \chi^2_{n-1}$ under the null using the sampling distribution theorem. This proves the second conclusion.

   10 (g) Calculate $\mathsf{Cov}(y_{ij} - \bar{y}_{i\cdot}, \bar{y}_{i\cdot} - \bar{y})$. Argue that $\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_{i\cdot})^2$ is independent of $m \sum_{i=1}^n (\bar{y}_{i\cdot} - \bar{y})^2$. This proves the third conclusion.

   10 (h) Prove the fourth conclusion.

   10 (i) Prove the fifth conclusion.

STAT 245 HW6 Solution.

1. Review: to show iid $y_1, \cdots, y_n \sim N(0, \sigma^2)$ we have $\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{\sigma^2} \sim \chi^2_{n-1}$

$y_i = \sigma z_i$ where $z_i \overset{iid}{\sim} N(0,1)$

$\therefore \bar{y} = \sigma \bar{z}$, $(y_i - \bar{y})^2 = \sigma^2 (z_i - \bar{z})^2$

let $A = I_n - \frac{1}{n} 1 1^T$ where $1 = \text{ones}(n,1)$

then $\begin{pmatrix} z_1 - \bar{z} \\ \vdots \\ z_n - \bar{z} \end{pmatrix} = A \cdot z = (I - \frac{1}{n} 1 1^T) \cdot z$

by eigendecomposition: $A = Q \begin{pmatrix} I_{n-1} & \\ & 0 \end{pmatrix} Q^T$, $QQ' = Q'Q = I$

$\therefore \sum_{i=1}^{n} (z_i - \bar{z})^2 = \|Az\|^2 = (Az)^T (A \cdot z) = z^T Q \begin{pmatrix} I_{n-1} & \\ & 0 \end{pmatrix} Q^T z$

let $\tilde{z} = Q^T z$, then $\tilde{z} \sim N(0, Q^T \cdot I \cdot Q) \sim N(0, I)$

$$\therefore \|Az\|^2 = \tilde{z}^T \begin{pmatrix} I_{n-1} & \\ & 0 \end{pmatrix} \tilde{z} = \sum_{i=1}^{n-1} \tilde{z}_i^2 \sim \chi^2_{n-1}$$

$$\therefore \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{\sigma^2} \sim \chi^2_{n-1}$$

2.

(a) Under the null: $\alpha_1 = \cdots = \alpha_n$

$\sum_{i=1}^{n} \alpha_i = 0 \quad \therefore n \alpha_i = 0. \ \forall i \in \{1, \cdots, n\}$

$\therefore \alpha_i = 0 \quad \forall i \in \{1, \cdots, n\}.$

2.

(b). Under the null $\alpha_i = 0$ for $i = 1, \cdots, n$

$\therefore y_{ij} \overset{iid}{\sim} N(\mu, \sigma^2)$ for $i = 1, \cdots, n$; $j = 1, \cdots, m$.

and $\bar{y}_{i\cdot} \sim N(\mu, \frac{\sigma^2}{n})$, $\bar{y}_{i\cdot} - \mu = \frac{1}{m}\sum_{j=1}^{m}(y_{ij} - \mu)$

$\therefore (y_{ij} - \mu) \overset{iid}{\sim} N(0, \sigma^2)$ we can apply the

Sampling distribution Thm in 1:

$$\frac{\sum_{j=1}^{m}(y_{ij} - \bar{y}_{i\cdot})^2}{\sigma^2} = \frac{\sum_{j=1}^{m}((y_{ij} - \mu) - (\bar{y}_{i\cdot} - \mu))^2}{\sigma^2} \sim \chi^2_{m-1}$$

(OK to use the sampling distribution theorem for not centered y's).

(c) $$\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - \bar{y}_{i\cdot})^2}{\sigma^2} = \sum_{i=1}^{n}\left(\sum_{j=1}^{m}\frac{(y_{ij} - \bar{y}_{i\cdot})^2}{\sigma^2}\right)$$

$\because \sum_{j=1}^{m}\frac{(y_{ij} - \bar{y}_{i\cdot})^2}{\sigma^2} \sim \chi^2_{m-1}$ from (b)

and for each i, j. y_{ij}'s are independent,

$\therefore$ for each i, $\sum_{j=1}^{m}\frac{(y_{ij} - \bar{y}_{i\cdot})^2}{\sigma^2} \overset{iid}{\sim} \chi^2_{m-1}$.

$$\therefore \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(y_{ij} - \bar{y}_{i\cdot})^2}{\sigma^2} \sim \chi^2_{n(m-1)}$$

(d) Define $x_i = \sqrt{m}\,\bar{y}_i$,

$$m \sum_{i=1}^{n} (\bar{y}_{i\cdot} - \bar{y})^2$$

$$= \sum_{i=1}^{n} (\sqrt{m}\,\bar{y}_{i\cdot} - \sqrt{m}\,\bar{y})^2$$

$$= \sum_{i=1}^{n} (x_i - \frac{1}{m\cdot n} \sum_{j=1}^{m} \sum_{i=1}^{n} \sqrt{m}\, y_{ij})^2$$

$$= \sum_{i=1}^{n} (x_i - \frac{1}{n} \sum_{i=1}^{n} (\sqrt{m} \cdot \frac{1}{m} \sum_{j=1}^{m} y_{ij}))^2$$

$$= \sum_{i=1}^{n} (x_i - \frac{1}{n} \sum_{i=1}^{n} \sqrt{m}\, \bar{y}_{i\cdot})^2$$

$$= \sum_{i=1}^{n} (x_i - \frac{1}{n} \sum_{i=1}^{n} x_i)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 .$$

(e). Under the null: $y_{ij} \overset{iid}{\sim} N(\mu, \sigma^2)$

$$\bar{y}_{i\cdot} = \frac{1}{m} \sum_{j=1}^{m} y_{ij} = \frac{1}{m} \cdot 1_m^T \cdot \vec{y}_{i\cdot} \sim N(\frac{1}{m} \cdot 1_m^T \cdot \mu \cdot 1_m , \sigma^2 \cdot \frac{1}{m} \cdot 1_m^T \cdot 1_m$$

$$\sim N(\mu , \frac{\sigma^2}{m})$$

$x_i = \sqrt{m}\,\bar{y}_{i\cdot}$ $\therefore x_i \sim N(\sqrt{m}\mu , \sigma^2)$

(f). $$\frac{m\sum_{i=1}^{n}(\bar{y}_{i\cdot}-\bar{y})^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2}$$

where $x_i \sim N(\sqrt{m}\mu, \sigma^2)$

By sampling distribution thm, $\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2} \sim \chi^2_{n-1}$.

(g). $Cov(y_{ij}-\bar{y}_{i\cdot}, \bar{y}_{i\cdot}-\bar{y})$

$= Cov(y_{ij}, \bar{y}_{i\cdot}) - Cov(y_{ij}, \bar{y}) - Cov(\bar{y}_{i\cdot}, \bar{y}_{i\cdot}) + Cov(\bar{y}_{i\cdot}, \bar{y})$

$= \frac{1}{m} Cov(y_{ij}, \sum_{j=1}^{m} y_{ij}) - \frac{1}{mn}\sum_{k=1}^{n}\sum_{\ell=1}^{m} Cov(y_{ij}, y_{k\ell}) - Cov(\bar{y}_{i\cdot}, \bar{y}_{i\cdot})$

$+ Cov(\bar{y}_{i\cdot}, \frac{1}{n}\cdot\sum_{k=1}^{n}\bar{y}_{k\cdot})$

$$= \frac{\sigma^2}{m} - \frac{1}{mn}\cdot\sigma^2 - \frac{\sigma^2}{m} + \frac{1}{n}\cdot\frac{\sigma^2}{m} = 0.$$

$\because$ $(y_{ij}-\bar{y}_{i\cdot})$ and $(\bar{y}_{i\cdot}-\bar{y})$ can both be written as

some $M_1\cdot\begin{pmatrix}\vdots\\ y_{ij}\\ \vdots\end{pmatrix}$ and $M_2\cdot\begin{pmatrix}\vdots\\ y_{ij}\\ \vdots\end{pmatrix}$ $\therefore$ they are both normally

distributed, $\therefore$ $Cov(y_{ij}-\bar{y}_{i\cdot}, \bar{y}_{i\cdot}-\bar{y}) \Rightarrow$ they are independant.

$\therefore$ $\underset{\uparrow\ \text{a function of } (y_{ij}-\bar{y}_{i\cdot})\text{'s}}{\sum_{i=1}^{n}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\cdot})^2}$ is indep of $\underset{\uparrow\ \text{function of } (\bar{y}_{i\cdot}-\bar{y})\text{'s}}{m\sum_{i=1}^{n}(\bar{y}_{i\cdot}-\bar{y})^2}$.

Also need to show $Cov(y_{ij}-\bar{y}_{i\cdot}, \bar{y}_{k\cdot}-\bar{y})=0$ for $i\neq k$.

(h). To show $\dfrac{\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\bar{y})^2}{\sigma^2} \sim \chi^2_{nm-1}$.

$\because$ under the null, $y_{ij} \overset{iid}{\sim} N(\mu, \sigma^2)$

$\therefore$ We can just reindex all the $n \cdot m$ terms of y_{ij} and use the sampling distribution theorem:

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\bar{y})^2}{\sigma^2} = \sum_{(i,j)\in\{1,\cdots,n\}\otimes\{1,\cdots m\}} \frac{(y_{ij}-\bar{y})^2}{\sigma^2} \sim \chi^2_{nm-1}.$$

To show:

(i). $$F = \frac{m\sum_{i=1}^{n}(\bar{y}_{i\cdot}-\bar{y})^2/(n-1)}{\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\bar{y}_{i\cdot})^2/n(m-1)} \sim F_{n-1,\, n(m-1)}.$$

From previous steps:

$$\frac{m\sum_{i=1}^{n}(\bar{y}_{i\cdot}-\bar{y})^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\bar{y}_{i\cdot})^2}{\sigma^2} \sim \chi^2_{n(m-1)}$$

and they are independent.

$\therefore$ by definition: $F \sim F_{n-1,\, n(m-1)}$.