JOE SEIDEL

# STAT 245 HOMEWORK 6

## 1. *Sampling Distribution*

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

We have $y_1 = \sigma z_i$ where $z_i \sim N(0,1)$ and i.d.d. Therefore $\overline{y} = \sigma \overline{z}$, $(y_i - \overline{y})^2 = \sigma^2 (z_i - \overline{z})^2$.

Let $A = I_n - \frac{1}{n} \mathbb{1}\mathbb{1}^T$ where $\mathbb{1} \in \mathbb{R}^{n \times 1}$. Then

$$\begin{pmatrix} z_1 - \overline{z} \\ \vdots \\ z_n - \overline{z} \end{pmatrix} = A \cdot Z = (I - \frac{1}{n} \mathbb{1}\mathbb{1}^T) \cdot Z.$$

By eigen decomposition,

$$A = Q \begin{pmatrix} I_{n-1} & \\ & 0 \end{pmatrix} Q^T.$$

Therefore

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

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

Therefore

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

it follows

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

## 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.

(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 $i = 1, ..., n$.

Since $\alpha_1 = \alpha_2 = ... = \alpha_n$ then we have $\sum_{i=1}^n \alpha_i = n\alpha = 0$ which implies $\alpha_i = 0$.

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

Under the null $\alpha_i = 0$ for $i = 1, ..., n$. Therefore $y_{ij} \sim N(\mu, \sigma^2)$, i.i.d. for $i = 1, ..., n$, $j = 1, ..., m$ and $\overline{y}_i \sim N(\mu, \frac{\sigma}{n})$. Since $\overline{y}_i - \mu = \frac{1}{m}\sum_{j=1}^{m}(y_{ij} - \mu)$ then $(y_{ij} - \mu) \sim N(0, \sigma^2)$. Applying the sampling distribution from question 1 we have

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

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

   Observe

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - \overline{y}_{i.})^2}{\sigma^2} = \sum_{i=1}^{n}\frac{\sum_{j=1}^{m}(y_{ij} - \overline{y}_{i.})^2}{\sigma^2}$$

   where we've found in part (b) that

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

   For each $i, j$, $y_{ij}$ are independent. Therefore sum of $\chi^2_{m-1}$ is $\chi^2_{n(m-1)}$ e.g.

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

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

$$\begin{aligned}
m\sum_{i=1}^{n}(\overline{y}_{i.} - \overline{y})^2 &= \sum_{i=1}^{n}(\sqrt{m}\overline{y}_{i.} - \sqrt{m}\overline{y})^2 \\
&= \sum_{i=1}^{n}(x_i - \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n}\sqrt{m}y_{ij})^2 \\
&= \sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{i=1}^{n}(\sqrt{m}\cdot\frac{1}{m}\sum_{j=1}^{m}y_{ij}))^2 \\
&= \sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{i=1}^{n}\sqrt{m}\overline{y}_{i.})^2 \\
&= \sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{i=1}^{n}x_i)^2 \\
&= \sum_{i=1}^{n}(x_i - \overline{x})^2
\end{aligned}$$