[8.4] The t, F, and Beta Distributions.

Let's introduce the "Student's t distribution".

–Named after William Gosset, Guiness Brewrey in Dublin, Ireland

–He could not publish under his own name (prob for competition)

–Why "Student"? Silly name isn't it!

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

By STUDENT.

Why use it?

–$\overline{X}$ can be used to make inferences about μ

–But, $\overline{X}$ depends on σ^2, which is most likely not known.

–We'd like to substitute S^2 in for σ^2, but what effect does it have on the dist. of $\overline{X}$

Thm: If $Z \sim N(0,1)$, and $C \sim \chi^2(\nu)$, and if $Z \perp\!\!\!\perp C$, then

$$T = \frac{Z}{\sqrt{\frac{C}{\nu}}}$$

is referred to as Student's t Distribution with ν degrees of freedom, denoted by

$$T \sim t(\nu)$$

Proof: The joint density of Z and C is

$$f_{Z,C}(z,c) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \cdot \frac{1}{2^{\nu/2}\Gamma(\nu/2)} c^{\nu/2-1} e^{-c/2} = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})\sqrt{2\pi}} c^{\frac{\nu}{2}-1} e^{-\frac{z^2}{2}-\frac{c}{2}}$$

Make the transformation $T = \frac{Z}{\sqrt{\frac{C}{\nu}}}$ and $W = C$. Then $C = W$ & $Z = T\sqrt{\frac{W}{\nu}}$.

The Jacobian is: $|J| = \begin{vmatrix} \frac{\partial C}{\partial W} & \frac{\partial C}{\partial T} \\ \frac{\partial Z}{\partial W} & \frac{\partial Z}{\partial T} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ \sim & \sqrt{\frac{W}{\nu}} \end{vmatrix} = \sqrt{\frac{W}{\nu}}$

So the joint dist. of W, T is $f_{W,T}(w,t) = f_{Z,C}(z,c)|J| = f_{Z,C}\left(t\sqrt{\frac{w}{\nu}}, w\right)\sqrt{\frac{w}{\nu}}$

$$= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})\sqrt{2\pi}} w^{\frac{\nu}{2}-1} e^{-\frac{1}{2}t^2\frac{w}{\nu}+w} \sqrt{\frac{w}{\nu}} = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})\sqrt{2\pi}\sqrt{\nu}} w^{\frac{\nu-1}{2}} e^{-\frac{1}{2}(\frac{t^2}{\nu}+1)w}$$

$$f_T(t) = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})\sqrt{2\pi}\sqrt{\nu}}\int_0^\infty w^{\frac{\nu-1}{2}} e^{-\frac{1}{2}(\frac{t^2}{\nu}+1)w} dw$$

Let $u = \frac{1}{2}\left(1+\frac{t^2}{\nu}\right)w$

Then $w = \dfrac{u}{\frac{1}{2}\left(1+\frac{t^2}{\nu}\right)}$ and $dw = \dfrac{du}{\frac{1}{2}\left(1+\frac{t^2}{\nu}\right)}$

$$f_T(t) = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})\sqrt{2\pi}\sqrt{\nu}}\int_0^\infty \left(\frac{u}{\frac{1}{2}(1+\frac{t^2}{\nu})}\right)^{\frac{\nu-1}{2}} e^{-u}\left(\frac{du}{\frac{1}{2}\left(1+\frac{t^2}{\nu}\right)}\right)$$

Factor out all constants in the integral and you are left with:

$$f_T(t) = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})\sqrt{2\pi}\sqrt{\nu}}\left(\frac{1}{\frac{1}{2}(1+\frac{t^2}{\nu})}\right)^{\frac{\nu-1}{2}} \underbrace{\int_0^\infty u^{\frac{\nu+1}{2}-1}e^{-u}du}_{\Gamma(\frac{\nu+1}{2})}$$

Which leaves the pdf of T as:

$$f_T(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}}\left(1+\frac{t^2}{\nu}\right)^{-(\frac{\nu+1}{2})}$$

Thm: If $T \sim t(\nu)$, then for $\nu > 2r$,

$$E[T^{2r}] = \frac{\Gamma(r+\frac{1}{2})\Gamma(\frac{\nu}{2}-r)}{\sqrt{\pi}\Gamma(\frac{\nu}{2})}$$

$$E[T^{2r-1}] = 0, r = 1, 2, \ldots$$

$$\text{Var}(T) = \frac{\nu}{\nu-2}, \nu > 2$$

Thm: If $X_1, X_2, \ldots, X_n$ denotes a random sample from $N(\mu, \sigma^2)$, then

$$\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$$

Snedecor's F Distribution

ν_1 = numerator degrees of freedom

ν_2 = denominator degrees of freedom

Thm: If $V_1 \sim \chi^2(\nu_1)$ and $V_2 \sim \chi^2(\nu_2)$ are independent, then the random variable

$X = \frac{V_1/\nu_1}{V_2/\nu_2} = \frac{\nu_2}{\nu_1}\frac{V_1}{V_2} \sim F(\nu_1, \nu_2)$ is known as Snedecor's F Distribution

Its pdf is:

$$f_F(x) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} x^{\frac{\nu_1}{2}-1}\left(1+\frac{\nu_1}{\nu_2}x\right)^{-\frac{\nu_1+\nu_2}{2}}$$

Thm: If $X \sim F(\nu_1, \nu_2)$, then

$$E(X^r) = \frac{\left(\frac{\nu_2}{\nu_1}\right)^r \Gamma\left(\frac{\nu_1}{2} + r\right)\Gamma\left(\frac{\nu_2}{2} - r\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}, \nu_2 > 2r$$

$$E(X) = \frac{\nu_2}{\nu_2 - 2}, \nu_2 > 2$$

Note: when $\nu_2 < 2$, then F has NO MEAN!

$$\text{Var}(X) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}, \nu_2 > 4$$

Note: when $\nu_2 < 4$, then F has NO VARIANCE!

And even stranger: when $2 < \nu_2 < 4$, then

F has a mean, but no variance!

Example: Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ be independent random samples from $X_i \sim N(\mu_1, \sigma_1^2)$ and $Y_i \sim N(\mu_2, \sigma_2^2)$.

If $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$, then

$$\frac{\nu_1 S_1^2}{\sigma_1^2} \sim \chi^2(\nu_1) \text{ and } \frac{\nu_2 S_2^2}{\sigma_2^2} \sim \chi^2(\nu_2) \quad \text{and} \quad \frac{\frac{\nu_1 S_1^2}{\sigma_1^2}/\nu_1}{\frac{\nu_2 S_2^2}{\sigma_2^2}/\nu_2} = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} \sim F(\nu_1, \nu_2)$$

Percentiles are provided in Table 7 (Appendix C)

If $X \sim F(\nu_1, \nu_2)$, then $Y = \dfrac{1}{X} = F(\nu_2, \nu_1)$.

The γth percentile, $f_\gamma(\nu_1, \nu_2)$, is defined by $P[X \leqslant f_\gamma(\nu_1, \nu_2)] = \gamma$

So $\gamma = P[X \leqslant f_\gamma(\nu_1, \nu_2)] = P[1/Y \leqslant f_\gamma(\nu_1, \nu_2)] = P\left[Y \geqslant \dfrac{1}{f_\gamma(\nu_1, \nu_2)}\right]$

So $1 - \gamma = 1 - P\left[Y \geqslant \dfrac{1}{f_\gamma(\nu_1, \nu_2)}\right] = P\left[Y \leqslant \dfrac{1}{f_\gamma(\nu_1, \nu_2)}\right]$

Since $1 - \gamma = P[Y \leqslant f_{1-\gamma}(\nu_2, \nu_1)]$, then

$$f_{1-\gamma}(\nu_2, \nu_1) = \frac{1}{f_\gamma(\nu_1, \nu_2)}$$

The Beta Distribution

If $X \sim F(\nu_1, \nu_2)$, then

$$Y = \frac{(\nu_1/\nu_2)X}{1 + (\nu_1/\nu_2)X} \qquad Y \sim BETA\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)$$

Let $a = \frac{\nu_1}{2}, b = \frac{\nu_2}{2}$, The pdf is:

$$f(y; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1} I_{(0,1)}(y)$$

$$\mu = \frac{a}{a+b} \qquad \sigma^2 = \frac{ab}{(a+b+1)(a+b)^2}$$