ANOVA

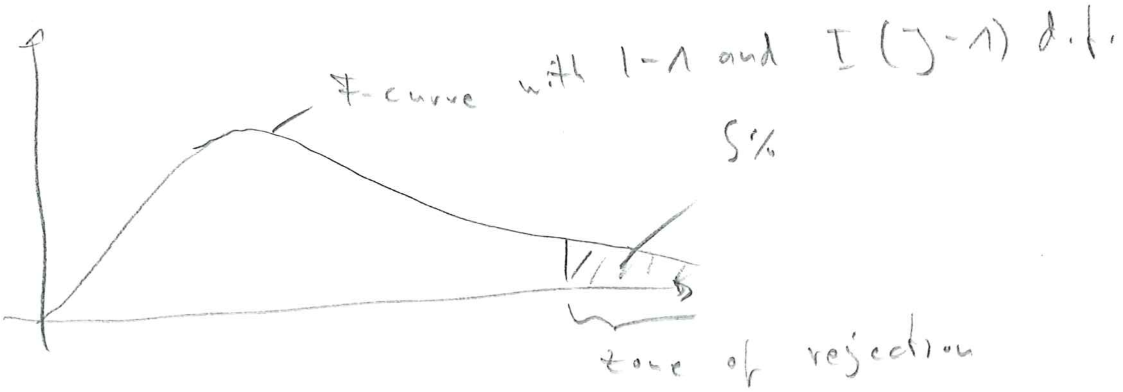
2. Recap
3. Proof fundamental ANOVA identity
4. Exercise: why/how does Levene's test work?

Recap

Treatment sum of squares (between-sample variation)	Error sum of squares (within-sample variation)
$SST = J \sum_{i=1}^{I} (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$	$SSE = \sum_{i,j=1}^{I,J} (X_{ij} - \bar{X}_{i\cdot})^2$
$\Downarrow$	$\Downarrow$
$SST/\sigma^2 \sim \chi^2$ distribution with $I-1$ d.f. when H_0 is true, thus	$SSE/\sigma^2 \sim \chi^2$-distr with $I(J-1)$ d.f. (even when H_0 is not true), thus
H_0 true $\Rightarrow E\Big(\underbrace{\frac{SST}{I-1}}_{MSTr}\Big) = \sigma^2$, MSTr (mean square for treatments)	$E\Big(\underbrace{\frac{SSE}{I(J-1)}}_{MSE}\Big) = \sigma^2$, MSE (mean square for error)
i.e. MSTr is an unbiased estimator of σ^2 if H_0 is true	i.e. MSE is an unbiased estimator of σ^2

$$F = \frac{MSTr}{MSE},$$ and

F takes $\begin{cases} \text{values around } 1 \text{ if } H_0 \text{ is true} \\ \text{values larger than } 1 \text{ if } H_0 \text{ is not true} \end{cases}$



Fundamental ANOVA identity

SST = SSTr + SSE

Proof: Let $x_{ij} - \bar{x}_{..} = (x_{ij} - \bar{x}_{i.}) + (\bar{x}_{i.} - \bar{x}_{..})$

$$\xRightarrow{(\dots)^2} (x_{ij}-\bar{x}_{..})^2 = (x_{ij}-\bar{x}_{i.})^2 + (\bar{x}_{i.}-\bar{x}_{..})^2 + 2(x_{ij}-\bar{x}_{i.})(\bar{x}_{i.}-\bar{x}_{..})$$

$$\xRightarrow{\sum_i \sum_j} \underbrace{\sum_i \sum_j (x_{ij}-\bar{x}_{..})^2}_{SST} = \underbrace{\sum_i \sum_j (x_{ij}-\bar{x}_{i.})^2}_{SSE} + \underbrace{\overset{\sum_j}{J} \sum_i (\bar{x}_{i.}-\bar{x}_{..})}_{SSTr}$$

$$+ \underbrace{2 \sum_i \sum_j (x_{ij}-\bar{x}_{i.})(\bar{x}_{i.}-\bar{x}_{..})}_{(*)}$$

$$(*) = 2 \sum_i \left[(\bar{x}_{i.} - x_{..}) \underbrace{\sum_j (x_{ij} - \bar{x}_{i.})}_{} \right]$$

deviations of all x_{ij} from their mean $\bar{x}_{i.}$ $= 0$, because:

$$-J \cdot \bar{x}_{i.} = -\cancel{J} \cdot \frac{1}{\cancel{J}} \sum_{j=1}^{J} x_{ij} = -\sum_j x_{ij}$$

Visualization of Levene's test

One-way ANOVA on $z_{ij} = |X_{ij} - \bar{X}_{i\cdot}|$ means: