Math 4720: Statistical Methods

12^{th} Week Summary (04/10/25)

- **Multiple Comparisons** ....
- The problem of **multiplicity** is serious when we are testing many hypotheses
- Instead of using Type-I error rate, we use familywise error rate (FEW):

  $\alpha_F = P($**Falsely reject at least one hypotheses**$)$
- Bonferroni Method:

  If there are m hypotheses, then Reject $H_0^{ij}: \mu_i = \mu_j$ if $|\bar{y}_i - \bar{y}_j| > t_{\alpha/2m}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$

  Pros: This will guarantee that $\alpha_F \leq 0.05$

  Cons: The chance of rejecting H_0 is small. In other word, the power of Bonferroni is very poor

  R: `with(exmp9.3, pairwise.t.test(yield, agent, p.adj = "bonf"))`
- Fisher's Least Significant Difference (LSD):

  Reject $H_0^{ij}: \mu_i = \mu_j$ if $|\bar{y}_i - \bar{y}_j| > t_{\alpha/2}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$

  Pros: Easy to discover differences (reject H_0^{ij}). In other word, high power

  Cons: This DOES NOT control the familywise error

  R: `with(exmp9.3, pairwise.t.test(yield, agent, p.adj = "none"))`
- Tukey's Method:

  Reject $H_0^{ij}: \mu_i = \mu_j$ if $|\bar{y}_i - \bar{y}_j| > \dfrac{q_\alpha(t, df_E)}{\sqrt{2}}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$

  Pros: Control the familywise error rate

  Cons: Lower power comparing to Fisher LSD and Dunnett methods

  R: `model = aov(yield agent, data= exmp9.3); TukeyHSD(model)`
- Dunnett's Method:

  Reject $H_0^{ij}: \mu_i = \mu_j$ if $|\bar{y}_i - \bar{y}_j| > d_\alpha(t-1, df_E)\sqrt{\dfrac{2MSE}{n}}$

  Pros: Control the familywise error rate. Higher power comparing to Bonferroni and Tukey's

  Cons: It's just for comparing with a control group.

  R: `library("DescTools"); with(exmp9.3, DunnettTest(yield, agent, control = "None"))`
- **Population proportion** ....
- Draw a large random sample of size n from a large population having unknown proportion p of successes. To test the hypothesis $H_0: \pi = \pi_0$, compute the Z statistic: $z = \dfrac{\hat{\pi} - \pi_0}{\sqrt{\dfrac{\pi_0(1-\pi_0)}{n}}}$
- Use this test when the sample size n is so large that both $n\pi_0$ and $n(1-\pi_0)$ are 5 or more.
- The p_value for a test of $H_0: \pi = \pi_0$ against

  $H_a: \pi > \pi_0$ is p-value $= P(Z \geq z)$

  $H_a: \pi < \pi_0$ is p-value $= P(Z \leq z)$

  $H_a: \pi \neq \pi_0$ is p-value $= 2P(|Z| \geq |z|)$
- Confidence interval for π can be obtained by $\hat{\pi} \pm z_{\alpha/2v}\sqrt{\dfrac{\hat{\pi}(1-\hat{\pi})}{n}}$. This interval should not be used unless $n\hat{\pi} \geq 5$ and $n(1-\hat{\pi}) \geq 5$.