DISCUSSION 12

1 Review

1.1 The Paired t test

1. Assumption
   - The data is of n independent pairs $(X_1, Y_1), \cdots, (X_n, Y_n)$ with $E(X_i) = \mu_1, E(Y_i) = \mu_2$.
   - Difference $D_1 = X_1 - Y_1, \cdots D_n = X_n - Y_n$.
   - $D_1, \cdots, D_n \overset{iid}{\sim} N(\mu_D, \sigma_D^2)$, where $\mu_D = \mu_1 - \mu_2$.
   - Then $\bar{D} = \frac{1}{n}\sum_{i=1}^n D_i \sim N(\mu_D, \frac{\sigma_D^2}{n})$
2. The Paired t procedure
   - A CI for μ_D : $\bar{d} \pm t_{\alpha/2,n-1} s_D \frac{1}{\sqrt{n}}$
   - The paired t test
     (a) Null hypothesis : $H_0 : \mu_D = \Delta_0$
     (b) Test statistic : $t = \frac{\bar{d} - \Delta_0}{s_d/\sqrt{n}}$
     (c) Decision rule: Reject H_0
       - $H_a : \mu_D > \Delta_0$: $t \geq t_{\alpha,n-1}$, P-value: $P(t > t_{n-1})$
       - $H_a : \mu_D < \Delta_0$: $t \leq -t_{\alpha,n-1}$, P-value: $P(t < t_{n-1})$
       - $H_a : \mu_D \neq \Delta_0$: $|t| \geq t_{\alpha/2,n-1}$, P-value: $2P(|t| > t_{n-1})$

1.2 Inference Concerning a Difference Between Population Proportions

- $X \sim Bin(m, p_1)$ and $Y \sim Bin(n, p_2)$, where X and Y are independent.
- A large-sample CI for $p_1 - p_2$: $\widehat{p_1} - \widehat{p_2} \pm \sqrt{\frac{\widehat{p_1}\widehat{q_1}}{m} + \frac{\widehat{p_2}\widehat{q_2}}{n}}$
- A large-sample test procedure for $p_1 - p_2$
  1. Null hypothesis: $H_0 : p_1 - p_2 = 0$
  2. Test statistic : $z = \frac{\widehat{p_1} - \widehat{p_2}}{\sqrt{\widehat{p}\widehat{q}(\frac{1}{m} + \frac{1}{n})}}$, where $\widehat{p} = \frac{X+Y}{m+n} = \frac{m}{m+n}\widehat{p_1} + \frac{n}{m+n}\widehat{p_2}$,
  3. Reject H_0 if
     - $H_a : p_1 - p_2 > 0$: $z \geq z_\alpha$, P-value: $P(Z > z)$
     - $H_a : p_1 - p_2 < 0$: $z \leq -z_\alpha$, P-value: $P(Z < z)$
     - $H_a : p_1 - p_2 \neq 0$: $|z| \geq z_{\alpha/2}$, P-value: $2P(Z > |z|)$

1.3 Inference Concerning Two Population Variances

1. The F distribution has two parameters, ν_1 and ν_2.
2. $F_{1-\alpha,\nu_1,\nu_2} = 1/F_{\alpha,\nu_2,\nu_1}$.
3. The F test for equality of variances, $H_0 : \sigma_1^1 = \sigma_2^2$.
   - Assumption: $X_1, \ldots, X_m \overset{iid}{\sim} N(\mu_1, \sigma_1^2)$ and $Y_1, \ldots, Y_n \overset{iid}{\sim} N(\mu_2, \sigma_2^2)$, where X and Y are independent.
   - Test statistic: $f = \frac{s_1^2}{s_2^2}$
   - Decision Rule: Reject H_0 if
     - $H_1 : \sigma_1^2 > \sigma_2^2 : f \geq F_{\alpha,m-1,n-1}$
     - $H_1 : \sigma_1^2 < \sigma_2^2 : f \leq F_{1-\alpha,m-1,n-1}$
     - $H_1 : \sigma_1^2 \neq \sigma_2^2 : f \geq F_{\alpha/2,m-1,n-1}$ or $f \leq F_{1-\alpha/2,m-1,n-1}$

2 Example

1. Cushing's disease is characterized by muscular weakness due to adrenal or pituitary dysfunction. To provide effective treatment, it is important to detect childhood Cushing's disease as early as possible. Age at onset of symptoms and age at diagnosis for 10 children suffering from the disease were given in an article. Here are the values of differences between age at onset of symptoms and age at diagnosis (in months).

Patient	1	2	3	4	5	6	7	8
Onset	7	11	15	9	8	11	5	20
Diagnosis	31	23	70	24	38	71	19	41
Difference	24	12	55	15	30	60	14	21

   (a) Does the data suggest it takes more than 17 years to detect Cushing's disease since after its first onset? State and test appropriate hypotheses using α=0.05.

   (b) Calculate a lower 95 % confidence bound for the population mean differences, and interpret the resulting bound.

   (c) Does the (incorrect) use of two-sample t test to test the hypotheses suggested in (a) lead to the same conclusion that you obtained there?

2. Some survey was done in Wisconsin and Texas to see if these two states differ in the preference in animals. In Wisconsin, 204 out of 323 said they preferred donkeys to elephants, while 114 out of 312 said so in Texas.

   (a) Test $H_0 : p_1 = p_2$ versus $H_a : p_1 \neq p_2$ using $\alpha = 0.05$, where p_1 and p_2 refers the proportion of people prefer donkeys in Wisconsin and Texas, respectively.

(b) If the true proportion favoring the donkeys are actually $p_1 = 0.6$(WI) and $p_2 = 0.4$(TX), what is the probability that H_0 will be rejected using $\alpha = 0.05$ test with $m = 300, n = 180$?

(c) Estimate the difference between the proportion of people favoring donkeys by calculating a CI.

3. Obtain or compute the following quantities.

(a) $F_{0.05,5,8}$

(b) $F_{0.95,8,5}$

(c) The 99th percentile of the F distribution with $\nu_1 = 10, \nu_2 = 12$.

(d) The 99th percentile of the F distribution with $\nu_1 = 10, \nu_2 = 12$.

4. Suppose we have Two types of anchor bolts: 3/8" and 1/2" diameter.

Diameter	Sample Size	Sample Mean (of the weight)	Sample SD
3/8	13	4.25	3.70
1/2	12	7.14	1.68

Does the data suggest that there is more variability in 3/8" bolts? Assuming normality, carry out a test of hypotheses at $\alpha = 0.05$.