Ⅰ. Short Review

1. STATA
   - Command : mean 'var', ci 'var' , ttest 'var' == 'H_0 value'
   - For two sample t test : ttest 'var' , by('var')
   - Interpreting the table in STATA will be covered in the session.

2. Testing Hypotheses for Difference Between Two Means

   - Step 1: Determine Appropriate Test

     If N_1 and $N_2 \geq 20$, $\overline{x_1} - \overline{x_2} \sim N(\mu_{\overline{x_1}-\overline{x_2}}, \sigma_{\overline{x_1}-\overline{x_2}})$. $\sigma_{\overline{x_1}-\overline{x_2}} = \sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}$

     If N_1 or $N_2 < 20$, , $\overline{x_1} - \overline{x_2} \sim t(\mu_{\overline{x_1}-\overline{x_2}}, \sigma_{\overline{x_1}-\overline{x_2}}, N_1+N_2-2)$.

     $$\sigma_{\overline{x_1}-\overline{x_2}} = \hat{\sigma}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}, \quad \hat{\sigma}\text{(pooled variance)} = \sqrt{\frac{\left(S_1^2(N_1-1)\right) + (S_2^2(N_2-1))}{N_1+N_2-2}}$$

   - Step 2: Formulate the Null Hypothesis

     H_0: $\mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$; H_A: $\mu_1 \neq \mu_2$ or $\mu_1 >$ or $< \mu_2$

   - Step 3: Calculate the Test Statistic

     Z_{obs} or $t_{obs} = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_{\overline{x_1}-\overline{x_2}})}{\sigma_{\overline{x_1}-\overline{x_2}}}$

   - Step 4: Find Critical Value (95%)

     $Z_{crit} = 1.96$ (two tailed test. For one tail : 1.65, sign is important)

     t_{crit} : Find a value with DoF and $\alpha = 0.05$

   - Step 5: Compare Critical to Observed

   - Step 6: Decide on Null Hypothesis Reject H_0 , and Interpretation.

Ⅱ. Problems

1. Fill in the blank (?)

ttest realrinc == 19000	
One-sample t test	
Variable Obs Mean Std. Err. Std. Dev.	[95% Conf. Interval]
realrinc 689 21833.32 788.6 20699.81	(?)
mean = mean(realrinc)	t = 3.5929
Ho: mean = 19000	DoF = 688
Ha: mean < 19000 Ha: mean != 19000	Ha: mean > 19000
Pr(T < t) = 0.9998 Pr(T > t) = 0.0004	Pr(T > t) = 0.0002

2. Fill in the blank, and what is the result of the test

Two-sample t test with unequal variances	
Group Obs Mean Std. Err. Std. Dev.	[95% Conf. Interval]
0 351 16850.78 899.9806 16861.13	15080.73 18620.83
1 338 27007.51 1248.421 22951.94	24551.83 29463.18
combined 689 21833.32 788.6 20699.81	20284.97 23381.68
Diff: (? mean) (? Std.err)	
diff = mean(0) - mean(1)	t = (?)

3. Ten cigarettes of Brand A had an average nicotine content of 3.1mm with standard deviation of 0.5mm, while eight cigarettes of Brand B had an average nicotine content of 2.7mm with standard deviation of 0.7mm. Test the difference. (Assumption : two sets of data is independent)