Statistics 5.200, L01
First Term 2004/2005

1. In a study of iron deficiency among infants, random samples of infants following different feeding programs were compared. One group contained breast-fed infants, while the children in another group were fed by a standard baby formula without any iron supplements. Here are summary results of blood hemoglobin levels at 12 months of age.

Group	Sample Size	Mean	Std. Dev.
Breast-fed	8	13.3	1.7
Formula-fed	10	12.4	1.8

(a) Is there good evidence in the data that the mean hemoglobin level of the breast-fed infants is higher than that of the formula-fed infants? Test at 5% significance level.

(b) Find the 95% confidence interval for the difference of the mean hemoglobin levels between the two groups of infants.

(c) What are the appropriate assumptions for this problem?

Soln: First, we check the criterion to decide which t procedure we should use. We refer to the rule of thumb. Since $(larger \quad s)/(smaller \quad s) = 1.8/1.7 = 1.06 < 2$, we use the pooled t procedure.

(a) Let μ_B and μ_F be the mean hemoglobin levels of Breast-fed and Formula-fed infants, respectively. Then the hypotheses are Ho: $\mu_B - \mu_F = 0$ vs. Ha: $\mu_B - \mu_F > 0$.
Level of significance is α=0.05.
The pooled sample standard deviation is

$$s_p = \sqrt{\frac{7(1.7^2) + 9(1.8^2)}{16}} = 1.757$$

and the corresponding t statistic is evaluated as

$$t = \frac{13.3 - 12.4}{1.757\sqrt{1/8 + 1/10}} = \frac{0.9}{0.8334} = 1.0799\,.$$

The degrees of freedom are df = 8+10-2 = 16.

Rejection Region Method:
The 5% rejection region: we reject Ho if $t \geq 1.746$.
The observed t (1.0799) is less than 1.746. So we fail to reject Ho - same conclusion.
Alternatively, by P-value Method:
The P-value is found to be $P(t(16) > 1.0799) > 0.1$.
This is a relatively big P-value. So we fail to reject Ho.

Therefore we conclude that there is **no** strong evidence in the data that the mean hemoglobin level of the breast-fed infants is higher than that of the formula-fed infants.

(b) Using the pooled t procedure, the 95% confidence interval for μ_1 - μ_2 is
$0.9 \pm 2.120(0.8334) = 0.9 \pm 1.7668 = (-0.8668,\ 2.6668)$.

(c) The samples are independent - both SRSs.
The populations are both normal.
The populations have a common variance

2.Samples of hamburger were selected from two different outlets of a large supermarket to measure the percentage of fat present in the meat, with the following summary data.

	Outlet 1	Outlet 2
n	5	10
mean	10.3	12.7 (percent)
std.dev	1.1	2.4 (percent)

(a) Do these data support the claim that the mean fat in the hamburgers from these two outlets are the same? Test at 5% significance level.

(b) Calculate the 97% confidence interval for the difference of the mean fat in the hamburgers.

(c) What are the appropriate assumptions for this problem?

Soln: First, we check the criterion to decide which t procedure we should use. We refer to the rule of thumb. Since $(larger\quad s)/(smaller\quad s) = 2.4/1.1 = 2.18 > 2$, we use the conservative t procedure.

(a) Let μ_1 be the mean percentage of fat in the meat of all hamburgers from Outlet 1; and μ_2 be the mean percentage of fat in the meat of all hamburgers from Outlet 2. Then the hypotheses are Ho: $\mu_1 - \mu_2 = 0$ vs. Ha: $\mu_1 - \mu_2 \neq 0$. Level of significance is α=0.05.
The conservative t statistic is evaluated as

$$t = \frac{10.3 - 12.7}{\sqrt{1.1^2/5 + 2.4^2/10}} = \frac{-2.4}{0.9044} = -2.6537\,.$$

The degrees of freedom are the smaller of 5-1 and 10-1, so that df = 4.

Rejection Region Method:
The 5% rejection region: we reject Ho if $|t| \geq 2.777$.
The absolute value (2.6537) of observed t is less than 2.777. So we fail to reject Ho - same conclusion.
Alternatively, by P-value Method:
Using the t-Table, we found $0.025 < P(t(4) > 2.6537) < 0.05$. Hence the P-value satisfies $0.05 < 2P(t(4) > 2.6537) < 0.1$. Therefore we fail to reject Ho at $\alpha = 5\%$ significance level.

Therefore, we conclude that the data **do not** support the claim that the mean percentages of fat in the meat of all hamburgers from Outlet 1 and 2 are different.

(b) Using the conservative t procedure, the 97% confidence interval for $\mu_1 - \mu_2$ is
$-2.4 \pm 3.2976(0.9044) = -2.4 \pm 2.9823 = (-5.3823, 0.5823)$.

(c) The samples are independent - both SRSs.
The populations are both normal.
The populations do not have a common variance