

Express your answers neatly and show how to do each problem for credit. No work = no credit.

1. Determine whether the study uses independent samples (IS) or matched pairs (MP) (circle one for each):

- (IS) (MP) An agronomist compares the yields of two varieties of soybean by planting each variety in 10 separate plots of land (a total of 20 plots).
- (IS) (MP) A psychologist measures the response times of subjects under two stimuli; each subject is observed under both of the stimuli in a random order.
- (IS) (MP) A survey is conducted of teens from inner city schools and suburban schools to compare the proportion who have tried drugs.
- (IS) (MP) Two computing algorithms are compared in terms of the CPU times required to do the same six test problems.

2. A person's muscle mass is expected to decrease with age. To explore this relationship, a nutritionist randomly selected 40 subjects between the ages of 40 and 80 and recorded the age and muscle mass (in kg) for each. The fitted regression equation was $\hat{y} = 157.45 - 1.22x$. The standard error of the slope was reported to be 0.133 and the linear correlation coefficient was -0.775 .

- (a) Give brief statistical evidence to show that there is a statistically significant linear correlation between age and muscle mass.

$$H_0: \rho = 0, H_1: \rho \neq 0$$

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{-0.775}{\sqrt{\frac{1-(0.775)^2}{40-2}}} = -7.48 \quad df = 38, \quad P < .01$$

Reject H_0 at $\alpha = .05$ ($P < .01$). There is strong evidence to show evidence of a population linear correlation.

- (b) Give the coefficient of determination and interpret what it means for this model *in context*.

$$r^2 = (-0.775)^2 = .601 \quad \text{60.1\% of the total variation in muscle mass is explained by the linear relationship with age.}$$

- (c) Construct a 95% confidence interval for the population slope and interpret its meaning *in the context of this problem*.

$$b_1 \pm t_{\alpha/2} S_{b_1} = -1.22 \pm 2.024 (.133) = -1.22 \pm .27 \rightarrow (-1.49, -.95) \quad (df = 38)$$

We are 95% confident that, on average, muscle mass decreases .95 to 1.49 kg/year.

- (d) Given that $SE_{\hat{\mu}} = 1.68$ and $SE_{\hat{y}} = 9.06$, construct a 90% interval estimate of the muscle mass of an individual person who is 63 years of age. Interpret your interval.

$$\hat{y} = 157.45 - 1.22(63) = 80.59, \quad df = 38$$

$$80.59 \pm 1.686(9.06) = 80.59 \pm 15.28 \rightarrow (65.31, 95.87)$$

90% of individuals, at age 63, have muscle mass between 65.31 kg & 95.87 kg.

3. Suppose we suspected an unusual distribution of blood groups in patients undergoing one type of surgical procedure. We know that the expected distribution for the population served by the hospital which performs this surgery is 45% group O, 43% group A, 9% group B and 3% group AB. A random sample of 187 patients undergoing this surgical procedure have the following distribution of blood types:

Blood Group	O	A	B	AB
Count	67	83	29	8
	2 (84.15)	(80.41)	(16.83)	(5.61)

Is there statistically significant evidence ($\alpha = .05$) that patients undergoing this procedure have a different distribution of blood types than the general population served by this hospital?

- 3 \sum $H_0: p_O = .45, p_A = .43, p_B = .09, p_{AB} = .03$
 $H_1: \text{at least one of the proportions differs from its hypothesized value}$
 • **Test Statistic:** The correct value of the test statistic is $\chi^2 = 53.26$. Fill in the first expected cell count and show how to *start* computing the test statistic. 13.39

2 $\chi^2 = \frac{(84.15 - 67)^2}{84.15} + \dots$

2 • **P-value:** $df = 3 \rightarrow P < .005$

1 • **Reject H_0 | Fail to reject H_0**

• **Give a practical conclusion:** There is strong evidence to show that the distribution

2 of blood types is different for patients undergoing this procedure

4. A manufacturer of platinum tipped sparkplugs believes that they last longer than conventional sparkplugs. To show this, one platinum plug and one conventional plug were installed in each of six two-cylinder engines. The effective life in hours was measured for each plug:

Engine	1	2	3	4	5	6
Platinum Plug	640	570	530	410	600	580
Conventional Plug	480	360	460	480	380	410

| pairs

$\bar{x}_P = 538.3, s_P = 78.8, \bar{x}_C = 428.3, s_C = 52.3, d = P - C, \bar{x}_d = 110.0, s_d = 100.2$
 $\bar{x}_d = 126.7, s_d = 110.0$

- (a) Do the platinum plugs last longer on average? Conduct a complete hypothesis test at the 10% level of significance.

3 • $H_0: \mu_d = 0$ $H_1: \mu_d > 0$

• **Test Statistic:**

3 $t = \frac{\bar{x}_d - 0}{s_d / \sqrt{n}} = \frac{110.0 - 0}{100.2 / \sqrt{6}} = 2.69 \quad (2.82)$

2 • **P-value:** $df = 5, .01 < P < .025$

2 • **Reject H_0 | Fail to reject H_0**

2 • **Give a practical conclusion:** There is evidence that platinum plugs last longer on average.

- (b) In the conclusions to the test above, what type of error (I or II) might you have just made?

Type I (may have falsely rejected true H_0)

- (c) If the null hypothesis were true here, then what was the probability of a type I error?

$\alpha = .10$

5. A manufacturer of television sets is interested in the effect of the tube conductivity of four different types of coating for color picture tubes. Conductivity is obtained for independent, random samples of tubes using each coating type. Each sample is of size 4.

(a) If the sum of squares for groups is 865.71 and the sum of squares for errors is 148.93, complete the ANOVA table below:

Source	df	SS	MS	F
Groups	3	865.71	288.57	23.25
Error	12	148.93	12.41	
Total	15			

(b) Conduct a complete ANOVA test to determine if the type of coating used has an effect on conductivity. Use $\alpha = 0.05$.

- $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$
- $H_1: \text{at least one mean is different}$
- Test Statistic: (from ANOVA table)

$$F = 23.25$$

- critical value:

$$df_{num} = 3, df_{den} = 12, F_{crit} = 3.4903$$

- Reject H_0 | Fail to reject H_0

- Give a practical conclusion: There are statistically significant differences among the ~~four~~ 4 population mean tube conductivities

(c) Bonferroni 94% comparisons have been constructed from the data. Construct and fill in the missing interval and interpret your findings (which population means are larger).

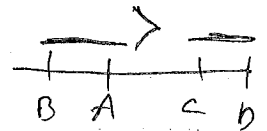
Coating	Sample Mean Conductivity
A	145.00
B	145.25
C	132.25
D	129.25

$$\alpha' = \frac{0.06}{6} = .01$$

$$df \text{ of CI's} = \frac{k(k-1)}{2} = \frac{4(4-1)}{2} = \frac{4 \cdot 3}{2} = 6$$

Comparison	94% CI's
$\mu_A - \mu_B$	(-6.92, 6.42)
$\mu_A - \mu_C$	(6.08, 12.75)
$\mu_A - \mu_D$	(9.08, 22.42)
$\mu_B - \mu_C$	(6.33, 19.67)
$\mu_B - \mu_D$	(8.39, 23.61)
$\mu_C - \mu_D$	(-3.67, 9.67)

$\mu_A \approx \mu_B$
 $\mu_A > \mu_C$
 $\mu_A > \mu_D$
 $\mu_B > \mu_C$
 $\mu_B > \mu_D$
 $\mu_C \approx \mu_D$



$$(\bar{x}_B - \bar{x}_D) \pm t_{\alpha'/2} \sqrt{MSE \left(\frac{1}{n_B} + \frac{1}{n_D} \right)} = (145.25 - 129.25) \pm 3.055 \sqrt{12.41 \left(\frac{1}{4} + \frac{1}{4} \right)}$$

$$= 16 \pm 7.61 \rightarrow (8.39, 23.61)$$

We are 94% confident that pop. mean conductivities A, B are larger than C and D and that there are no other significant differences.

6. Little Billy is asked to consult an agricultural research project. Researchers wish to determine if a kelp extract protects tomato plants from frost damage. Two similar plots are planted with the same variety of tomato plants. Plants in plot 1 are treated with the kelp extract and plants in plot 2 are not. After the first frost in autumn, 23 of 92 tomato plants in plot 1 exhibit damage ($\hat{p}_1 = .2500$). In plot 2, 36 of 104 tomato plants exhibit damage ($\hat{p}_2 = .3462$).

(a) Little Billy concludes that since the sample proportion of plants damaged in the kelp-treated plot is lower than in the plot that is not kelp-treated ($\hat{p}_1 < \hat{p}_2$) the kelp treatment must be effective at reducing frost damage. What is wrong with Little Billy's argument?

Little Billy's analysis doesn't account for random variation. It could be the case that the kelp treatment is ineffective & we are just observing a chance variation or fluke.

4

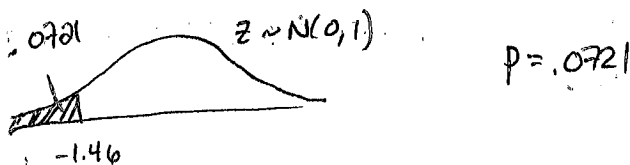
(b) Give a solid inferential statistics argument that either supports or refutes Little Billy's conclusion (at 95% or .05 if needed).

$$H_0: p_1 = p_2, H_1: p_1 < p_2, \alpha = .05$$

$$\bar{p} = \frac{23 + 36}{92 + 104} = .3010$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{.2500 - .3462}{\sqrt{.3010 \left(\frac{1}{92} + \frac{1}{104} \right)}} = -1.46$$

12



Do not reject H_0 at $\alpha = .05$ ($p = .0721$). There is not evidence to suggest that a lower proportion of all kelp-treated plants will be frost damaged than those that are not frost damaged.

(CI \rightarrow $(-.2234, .03108)$ 0 in CI, no difference)

14