Math 145 Test 3A Name: Solutions Row:

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**) Two computing algorithms are compared in terms of the CPU times requried to do the same six test problems.
   - (**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**) 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**) 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). 8

2. A person's muscle mass is expected to decrease with age. To explore this relationship, a nutritionist randomly selected 30 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 $-0.707$.

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

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

   $t = \dfrac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \dfrac{-.707}{\sqrt{\frac{1-.5}{28}}} = -5.29$

   $df = 28$, $P < \underline{.01}$ $(P = 6.3 \times 10^{-6})$

   or since $|r| = .707 > r_{crit} = \underline{.396}$

   either way —
   Reject $H_0$ at $\alpha = .05$ $(p < .01)$.
   There is strong evidence to show evidence of a population linear correlation. 6

   (b) Give the coefficent of determination and interpret what is means for this model *in context.*

   $r^2 = (-.707)^2 = .500$, 50% of the total variation in muscle mass is explained by the linear relationship with age. 4

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

   $b_1 \pm t_{\alpha/2} S_{b_1} = -1.22 \pm \underset{(df=28)}{2.048}(.133) \approx -1.22 \pm .27 \rightarrow (-1.49, -.95)$

   We are 95% confident that, on average, muscle mass decreases 1.49 to .95 Kg/year. 6

   (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.

   $\hat{y} = 157.45 - 1.22(63) = 80.59$, $df = 28$

   $80.59 \pm 1.701(9.06) = 80.59 \pm 15.41 \rightarrow (65.18, 96.00)$ 6

   90% of individuals, at age 63, have muscle mass between 65.18 Kg & 96.00 Kg.

32

30

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 44% group O, 45% group A, 8% 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 (82.28) | 83 (84.15) | 29 (14.96) | 8 (5.61) |

$exp_O = 187 * .44$

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?

- $H_0$: $p_O = .44,\ p_A = .45,\ p_B = .08,\ p_{AB} = .03$
- $H_1$: at least one of the proportions differs from its hypothesized value
- **Test Statistic:** The correct value of the test statistic is $\chi^2 =$ ~~57.27~~ 17.05. Fill in the first expected cell count and show how to *start* computing the test statistic.

$$\chi^2 = \frac{(82.28 - 67)^2}{82.28} + \ldots$$

- $P$-value: $df = 3$ (exact $P = 6.9 \times 10^{-4}$) $P < .005$
- **Reject $H_0$** | **Fail to reject $H_0$**
- **Give a practical conclusion:** There is strong evidence to show that the distribution of blood types is different for patients undergoing this procedure.

12

4. A manufacturer of platinum tipped sparkplugs believes that they last longer than conventional sparkplugs. To show this, one platinum plug and one convential 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 | 470 | 370 | 460 | 490 | 380 | 410 |

$\bar{x}_P = 538.3,\quad s_P = 78.8,\quad \bar{x}_C = 430.0,\quad s_C = 50.2,\quad d = P - C,\quad \bar{x}_d = 108.3,\quad s_d = 103.0$

$d = P - C$

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

- $H_0$: $\mu_d = 0$ $\qquad H_1$: $\mu_d > 0$
- **Test Statistic:**

$$t = \frac{\bar{x}_d - 0}{s_d/\sqrt{n}} = \frac{108.3 - 0}{103/\sqrt{6}} = 2.58,\quad df = 5$$

- $P$-value: $.01 < P < .025$
- **Reject $H_0$** | **Fail to reject $H_0$**
- **Give a practical conclusion:** There is evidence that platinum plugs last longer on average.

12

(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 $H_0$) 4

(c) If the null hypothesis were true here, then what was the probability of a type I error? 4

$\alpha = .10$

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 843.97 and the sum of squares for errors is 148.93, complete the ANOVA table below:

| Source | df | SS | MS | F |
|---|---|---|---|---|
| Groups | 3 | 843.97 | 281.32 | 22.67 |
| Error | 12 | 148.93 | 12.41 | |
| Total | 15 | | | |

5

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

- $H_0$ : $\mu_1 = \mu_2 = \mu_3 = \mu_4$ 2
- $H_1$ : at least one mean is different
- **Test Statistic:** (from ANOVA table)
  $F = 22.67$

12

- **critcal value:**
  $df_{num} = 3$, $df_{den} = 12$. $F_{crit} = \underline{3.4903}$ 2
- **Reject $H_0$** | **Fail to reject $H_0$** 1
- **Give a practical conclusion:** There are statistically significant differences among the four population mean tube conductivities. 2

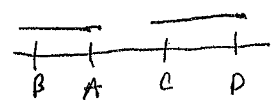
(c) Bonferroni 94% comparisons have been constructed from the data. Construct and fill in the missing interval and interpret your findings (which population means are larger).

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

$\alpha' = \frac{.06}{6} = .01$

$\#$ of CI's $= \frac{4 \cdot 3}{2} = 6$

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




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

7 10

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

We are 94% confident that pop. mean conductivities A, B are larger than C and D and that there are no other significant differences. 3

22

6. Little Billy is asked to consult an agricultural resarch 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

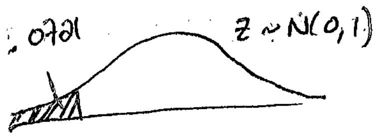
(b) Give a solid inferential statistics argument that either supports or refutes Little Billy's conclusion (at 95% or .05 if needed).

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

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

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

12




$p = .0721$

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 → (−.2234, .03108) 0 in CI, no difference)

14