Stats 8 winter 2019 (Baldi) – Final – SOLUTIONS

Problem 1 (11 points) < 15 min
Advertisement to children is tightly regulated, but social media content isn't. An Australia study published earlier this month examined whether social media influencers can influence children's food choices. The researchers created fake Instagram profiles for popular YouTubers, known in the media business as "influencers." At random, 176 Australian children (ages 9 to 11) were divided into three groups shown Instagram profiles that featured the influencer in different settings: with unhealthy snacks, with healthy snacks, or with non-food products. After viewing their assigned Instagram page, the children were offered a variety of snack options.

The researchers recorded how many calories of unhealthy food (in Cal) each child ate. Here are a summary of the findings and software output for the corresponding ANOVA *F* test:

Setting	N	Mean	St. Dev.
Unhealthy	59	384.83	141.21
Healthy	59	319.51	143.77
Non-food	58	292.24	146.85

```
Source of Variation           SS     DF          MS        F        P
Between Groups:         265532.1750    2   132766.0875   6.4076   0.0021
Within Groups:         3584585.0485  173    20720.1448
Total:                 3850117.2235
```

a) Write the corresponding null and alternative hypotheses.
H_0: **$\mu_1 = \mu_2 = \mu_3$** H_a: **H_0 is not true, at least one μ differs**

b) Assuming for now that the conditions for inference are met, interpret the findings in context.
The study found strong evidence ($P = 0.002$) that, on average, how many calories of unhealthy food Australian children consume for snack after watching Instagram pages of YouTube influencers is affected by the type of product featured with the influencers.

c) What are the conditions for this inference procedure?

A) Independent random samples AND normal data or large enough total sample size.
B) Normal data AND large enough expected counts.
C) Independent random samples AND normal data or large enough total sample size AND similar sample standard deviations.
D) No conditions are needed for this inference procedure.
E) One random sample AND normal data or large enough sample size.
F) Random selection without repeated measures AND large enough expected counts.

d) Are the conditions for inference met?
A) Yes, all conditions are met. **B)** No, only some are met. **C)** No, none of them are met.

Problem 2 (7 points) < 10 min
In a genetic experiment, chickens with white feathers and a large head crest (WL) were bred with chickens with dark feathers and a small crest (DS). It was theorized that the genetic inheritance of these phenotypes may be based on two independent genes, from which a 9:3:3:1 ratio in second generation would be expected (with the phenotype WS being the most common and the phenotype DL the least common). The breeding program actually produced in second generation 111 WS chicken, 37 WL, 34 DS, and 8 DL.

Here is the software output for the corresponding chi-square for goodness of fit test.

Category	Observed	Historical Counts	Test Proportion	Expected	Contribution to Chi-Sq
WS	111	9	0.5625	106.875	0.15921
WL	37	3	0.1875	35.625	0.05307
DS	34	3	0.1875	35.625	0.07412
DL	8	1	0.0625	11.875	1.26447

N	DF	Chi-Sq	P-Value
190	3	1.55088	0.671

a) Write the corresponding null and alternative hypotheses.

H_0: $\boldsymbol{p_{WS} = 9/16; p_{WL} = 3/16; p_{DS} = 3/16; p_{DL} = 1/16}$

H_a: **H_0 is not true** [at least one proportion differs]

b) Use the software output to conclude in context for the whole population of second generation chickens that would result from crossing WL and DS chickens.

A) The test assumptions are not met and, therefore, we should not conclude anything.
B) There is strong evidence (*P* = 0.671) that the 4 phenotypes are not equally represented.
C) There is weak evidence (*P* = 0.671) that the 4 phenotypes are not equally represented.
D) The findings are consistent with an equal representation of the 4 phenotypes (*P* = 0.671), which does not imply that they are truly equally represented.
E) The findings are consistent with an equal representation of the 4 phenotypes (*P* = 0.671), indicating that they are truly equally represented.
F) There is strong evidence (*P* = 0.671) that the 4 phenotypes are not represented according to the proportions expected under a genetic model with two independent genes.
G) There is weak evidence (*P* = 0.671) that the 4 phenotypes are not represented according to the proportions expected under a genetic model with two independent genes.
H) The findings are consistent with the proportions expected under a genetic model with two independent genes (*P* = 0.671), which does not imply that they are truly inherited according to this model.
I) The findings are consistent with the proportions expected under a genetic model with two independent genes (*P* = 0.671), indicating that are truly inherited according to this model.

Problem 3 (9 points) < 12 min

Fluoroquinolone is a common antibiotic, but fetal risk is an important consideration when treating pregnant women. An observational study examined the medical records of 783 women exposed to a fluoroquinolone during the first trimester of their pregnancy and 3433 women with no such exposure during pregnancy.

Following are further details on the study. For each one, select the appropriate inference procedure:

A) One sample or matched pairs *t* for a mean
B) Two sample *t* for two means
C) ANOVA for several means
D) Chi-square goodness of fit
E) Chi-square for two-way table of counts

a) The researchers compared the mean maternal age in the two groups to verify that the two groups were not significantly different in this respect.

A　　　**B**　　　C　　　D　　　E

b) The researchers compared the percent of newborns with a birth defect in the two groups to see if there was a significant difference between the groups.

A B C D **E**

c) The researchers examined the 783 women who had been exposed to a fluoroquinolone in their first trimester and sorted them according to the timing of the exposure (first month, second month, or third month). They compared the mean head circumference (in cm) of newborns in these three groups to see if it may be affected by the timing of fluoroquinolone exposure.

A B **C** D E

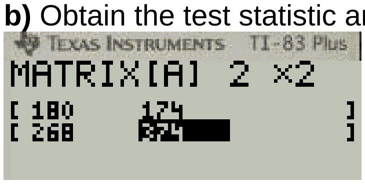
Problem 4 (15 points) < 20 min

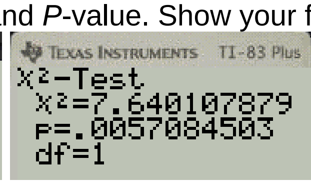
In the summer of 2016, the Cleveland Clinic started using delayed bathing for newborns (bathing no sooner than 12 hours after birth, “intervention”), hoping that the uninterrupted mother-infant time would help promote breastfeeding. But does it actually have any effect on breastfeeding? A study published earlier this month compared the rates of exclusive breastfeeding during the hospital stay in a random sample of 448 babies born at the hospital in the spring of 2016 who were bathed within 2 hours of birth (“preintervention”) and a random sample of 548 babies born in the fall of 2016 who got a delayed bath (“postintervention”). The reported findings are shown below.

Variable[a]	Total	Preintervention ($n = 448$)	Postintervention ($n = 548$)
Exclusive breastfeeding, *n* (%)			
No	354 (35.5)	180 (40.2)	174 (31.8)
Yes	642 (64.5)	268 (59.8)	374 (68.2)

a) Write the null and alternative hypotheses corresponding to the study’s stated objective.
H_0: **there is no association between the timing of the first bath and exclusive breastfeeding for all hospital births** [or hospitals similar to Cleveland Clinic]**; [these two variables are independent; the rate of exclusive breastfeeding is the same with both bath timings]**
H_a: **the null hypothesis is not true**

b) Obtain the test statistic and *P*-value. Show your full calculator work.





➔ **X^2 = 7.64, P = 0.0057**

c) What are the conditions for this inference procedure?

A) Independent random samples AND normal data or large enough total sample size.
B) Normal data AND large enough expected counts.
C) Independent random samples AND normal data or large enough total sample size AND similar sample standard deviations.
D) No conditions are needed for this inference procedure.
E) One random sample AND normal data or large enough sample size.
F) Random selection without repeated measures AND large enough expected counts.

d) Conclude in the context of this problem. **The study provides strong evidence (*P* = 0.006) of an association between the timing of the first bath and exclusive breastfeeding for all hospital births** [or hospitals similar to Cleveland Clinic]**,** [with a higher exclusive breastfeeding rate under the delayed bathing process (68% versus 60% in this study)].

e) Can we conclude from this study that delaying the first bath of newborns by at least 12 hours can be recommended as an effective way of increasing the rate of exclusive breastfeeding while the mother and infant are at the hospital? Briefly EXPLAIN your reasoning.
Whether such a conclusion is valid is debatable. Causality is a legitimate conclusion only for some study designs, typically experiments but not observational studies. This study is not a true experiment because the births were NOT randomly assigned to delayed or immediate bathing in similar settings. Yet, the conditions were imposed specifically to see how they affect the response variable. [*A common term for this type of study is "quasi experiment."*]
[What aspects of the design might factor in our ability to evaluate causation? Women did not choose to give birth in the spring or the fall, obviously, therefore their own life choices would likely not be confounded with the bathing procedure; however, perhaps breastfeeding choices vary somewhat across seasons (regardless of bathing type). On the hospital side, many things at the clinic may have been done differently once the delayed bathing was implemented, creating confounding variables; in particular, it is quite possible that the hospital staff would more strongly encourage women to breastfeed in the delayed bathing condition since it was introduced specifically in hope that it would lead to better breastfeeding outcomes.]
[Note: Points will be given for a sound discussion of the relevance of the design for a causal conclusion, regardless of whether the study is considered experimental or observational.]
[Lastly, the fact that the test is statistically significant is not directly relevant here, because a "recommendation" implies a causal effect; therefore, no points will be given for basing the conclusion on the significant P-value.]

Problem 5 (5 points each) < 25 min
For each following question, select the appropriate inference procedure. Then state the corresponding null and alternative hypotheses, using math symbols whenever possible.

a) Alzheimer's disease (AD) can be diagnosed using a detailed questionnaire resulting in a score on the Everday Cognition scale (Ecog, with higher scores indicating a worse cognitive impairment). This questionnaire is typically administered in person by a trained clinician, but a new online tool lets people self-administer the questionnaire. Out of concern that the two approaches may yield significantly different results, researchers compared a random sample of AD patients who took the questionnaire at a clinic and a random sample of patients who took the questionnaire online.

A) One sample or matched pairs *t* for a mean
B) Two sample *t* for two means
C) ANOVA for several means
D) Chi-square goodness of fit
E) Chi-square for two-way table of counts

H_0: $\mu_{online} = \mu_{clinic}$

H_a: $\mu_{online} \neq \mu_{clinic}$

b) We know that infants fall asleep easily when being gently rocked, but what would be the effect on adults? A 2019 study recruited a random sample of 18 healthy young adults to monitor their quality of sleep over two nights, measured by the time (in minutes) spent in REM sleep. In random order, on one night the participants slept in a stationary bed and on the other night they slept in a bed that gently rocks.

A) One sample or matched pairs *t* for a mean
B) Two sample *t* for two means
C) ANOVA for several means
D) Chi-square goodness of fit
E) Chi-square for two-way table of counts

H_0: $\mu_{diff\ rock\text{-}stationary} = 0$ [*matched pairs*]

H_a: $\mu_{diff\ rock\text{-}stationary} \neq 0$
H_a: $\mu_{diff\ rock\text{-}stationary} > 0$ [*also acceptable*]

c) French people commonly greet each other with a light kiss on each cheek ("la bise"). Is there a social preference for starting on the right versus the left cheek, or is it just a chance event? Researchers recorded a random sample of French individuals greeting each other with "la bise" and found the proportion who started with a kiss on the right cheek.

A) One sample or matched pairs *t* for a mean
B) Two sample *t* for two means
C) ANOVA for several means
<u>D) Chi-square goodness of fit</u>
E) Chi-square for two-way table of counts

H_0: $\boldsymbol{p_R = p_L = 1/2}$ [or 0.5]

H_a: $\boldsymbol{H_0}$ **is not true** [at least one proportion differs]

d) A 2019 study examined the effect of bone marrow transplants on the "physiological youth" of mice. The researchers compared random samples of young mice, old mice, old mice transplanted with bone marrow taken from young mice, and old mice transplanted with bone marrow taken from old mice. The mice were placed in transparent cages and videotaped to measure their activity level (in meters walked for the first 20 minutes after being placed in the cage).

A) One sample or matched pairs *t* for a mean
B) Two sample *t* for two means
<u>C) ANOVA for several means</u>
D) Chi-square goodness of fit
E) Chi-square for two-way table of counts

H_0: $\boldsymbol{\mu_1 = \mu_2 = \mu_3 = \mu_4}$

H_a: $\boldsymbol{H_0}$ **is not true, at least one μ differs**

e) The American Heart Association has formal recommendations for the maximum amount time that children should spend looking at screens, with a limit of 1 hour per day for children ages 2 to 5 years. A 2017 survey of parents in the United States found that the parents of children aged 2 to 5 reported that their children spend 2.6 hours per day in screen-based activities, on average.

<u>A) One sample or matched pairs *t* for a mean</u>
B) Two sample *t* for two means
C) ANOVA for several means
D) Chi-square goodness of fit
E) Chi-square for two-way table of counts

H_0: $\boldsymbol{\mu = 1}$ [hour/day]

H_a: $\boldsymbol{\mu > 1}$ [hour/day]
H_a: $\boldsymbol{\mu \neq 1}$ [*also acceptable*]

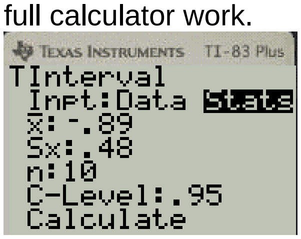
Problem 6 (22 points) < 30 min

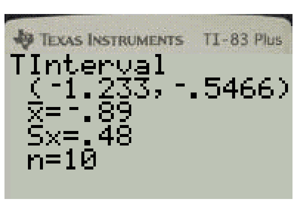
A 2019 study aimed to compare the health impact of different physical activities in women with type-2 diabetes. At random, 10 women with type-2 diabetes participated in a resistance training program and 8 participated in an endurance training program. Body weight issues are always a concern for patients with diabetes. The participants' body mass index (BMI, in kg/m^2) was recorded at the beginning of the study and again after 10 weeks of training. The findings are show below (with the format mean ± standard deviation) as they appear in the published report.

	Before	**After**	**Change**
Endurance (n = 8)	28.50 ± 2.38	27.70 ± 2.20	−0.80 ± 0.67
Resistance (n = 10)	28.45 ± 3.32	27.56 ± 3.15	−0.89 ± 0.48

<u>**Part 1**</u>: What is the impact of <u>resistance</u> training for 10 weeks on BMI in this population?

a) Obtain a 95% confidence interval for the mean BMI change with resistance training. Show your full calculator work.





➔ **−1.233 to −0.547** [kg/m^2]

b) The value of the margin of error for this confidence interval is: **0.34** kg/m^2 [*½ width of CI*]

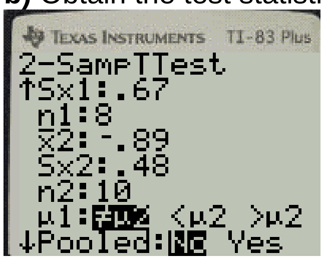
c) Interpret your confidence interval in the context of this problem.
We are 95% confident that, in the population of women with type-2 diabetes, the mean BMI change after 10 weeks of resistance training is somewhere between –1.2 to –0.5 kg/m^2 [BMI decreases by 0.5 to 1.2 kg/m^2, on average, after 10 weeks of resistance training]

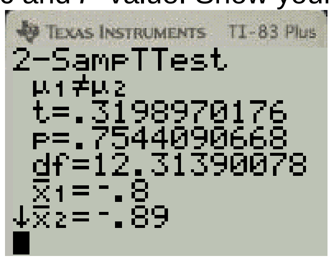
d) Based on this confidence interval, can we conclude that resistance training for 10 weeks can help significantly reduce BMI in this population? Briefly explain your reasoning.
Yes, because the confidence interval for the mean BMI change is entirely negative.

Part 2: Do the findings indicate that resistance and endurance training have significantly different effects on BMI change in the population of women with type-2 diabetes?

a) Write the corresponding null and alternative hypotheses.
H_0: $\mu_{endurance} = \mu_{resistance}$ H_a: $\mu_{endurance} \neq \mu_{resistance}$

b) Obtain the test statistic and *P*-value. Show your full calculator work.





➔ **t = 0.32, P = 0.754**

c) Conclude in the context of this problem.
This randomized experiment **failed to provide evidence (*P* = 0.75) that, in the population of women with type-2 diabetes, resistance training or endurance training for 10 weeks have a different impact on mean BMI change.**

Problem 7 (3 points) < 3 min
The margin of error of a confidence interval to estimate a population parameter accounts for the sample-to-sample variations in the value of the statistic

A) due only to measurement errors.
B) due only to the random sampling process.
C) due to nonresponse and response bias.
D) due to measurement errors and response bias.
E) due to anything that may influence the statistic as long as the data are a random sample.
F) due to anything that may influence the statistic as long as the population is Normal.
G) due to anything that may influence the statistic as long as the population is Normal and the data are a random sample.