## Name (Print):

You may use notes and a calculator on this exam. In order to receive full credit, either show all work or write the calculator function and inputs used to find a confidence interval or perform a hypothesis test.

1. (20 points) Write True(T) or False(F) for each of the following.
(a) If on average $y$ increases as $x$ increases, the correlation coefficient is positive.
(b) If $r$ is close to 1 , then the points lie close to a straight line with a positive slope. $T$
(c) Two events are said to be disjoint or mutually exclusive when they have no outcomes in common.
(d) Two events are independent if they cannot occur simultaneously. $F$
(e) A type 2 error occurs if you reject the null hypothesis when it is actually true. $F$
(f) The closer a value's z score is to 0 , the more extreme the value is. $F$
(g) In a negatively skewed distribution, the value of the mean is greater than the value of the median.
(h) The width of the one-sample confidence interval for $\mu$ decreases as the sample size grows $\boldsymbol{T}$ larger.
(i) The best estimate of the population proportion p is $\hat{p} \quad F$
(j) In order to decide whether the observed data is compatible with the null hypothesis, the observed cell counts are compared to the cell counts that would be expected when the alternative hypothesis is true.
2. (2 points) In order to determine the average salary of MLB players, you obtain a list of the rosters of all MLB teams. You randomly select 5 players from each of the teams and ask them to report their salary. What sampling technique did you use?
(a) Simple Random Sample
(b) Cluster
(c) Stratified
(d) Systematic
3. (2 points) In a hypothesis test, determine the only possible conclusions:

Select all that apply
(a) Fail to reject the alternative hypothesis
(b) Fail to reject the null hypothesis
(c) Reject the null hypothesis
(d) Reject the alternative hypothesis
4. (2 points) The area under a density curve
(a) Is always 1
(b) Is always 0
(c) Can be a negative number
(d) Can be any positive number
5. (2 points) The best graph/plot to use to determine if there is a linear relationship between 2 quantitative variables is a
(a) box plot
(b) histogram
(c) scatterplot
(d) stem and leaf plot
6. (2 points) Each person in a random sample of 200 students and a random sample of 150 faculty at a large university was asked "Do you take public transportation to campus?" $58 \%$ of the students and $46 \%$ of the faculty responded "Yes". This data was used to carry out a test of the hypothesis $H_{0}: P_{\text {student }}-P_{\text {faculty }}=0$ versus $H_{a}: P_{\text {student }}-P_{\text {faculty }} \neq 0$. The value of the test statistic was $\mathrm{z}=2.226$. For a significance level of 0.01 , which of the following is correct?
(a) $H_{0}$ should be rejected; there is convincing evidence that the proportion of people who take public transportation to campus is different for students and faculty.
(b) $H_{0}$ should be rejected; there is convincing evidence that the proportion of people who take public transportation to campus is the same for students and faculty.
(c) $H_{0}$ should not be rejected; there is not convincing evidence that the proportion of people who take public transportation to campus is different for students and faculty.
(d) $H_{0}$ should not be rejected; there is not convincing evidence that the proportion of people who take public transportation to campus is the same for students and faculty.
7. The following are waiting times (in seconds) between ordering and receiving coffee for 18 customers at Starbucks.

| 50 | 80 | 80 | 90 | 100 | 100 |
| :--- | :--- | :--- | :---: | :---: | :---: |
| 120 | 120 | 130 | 130 | 130 | 150 |
| 160 | 190 | 200 | 210 | 250 | 380 |

(a) (2 points) Compute the values of the mean and standard deviation. Round answer to two decimal places. No work necessary. Label solutions with correct notation.

$$
\bar{x}=148.33 \quad s=77.48
$$

(b) (2 points) List the 5 - number summary. Label each number in the 5 number summary.

8. The mean number of text messages sent per month by customers of a cell phone service provider is 2,050 and the standard deviation is 850 .
(a) (2 points) Find the z score associated with 3,500 messages sent. Round to 2 decimal places. Show work.

$$
z=\frac{3500-2050}{850}=1.70588 \ldots \approx 1.71
$$

(b) (2 points) What is the approximate value of the 95th percentile? Round to the nearest whole number. Show work.

$$
\begin{aligned}
& z=i u v \operatorname{Norm}(0.95,0,1)=1.645 \\
& x=2050+1.645(850)=3448 \text { text messages }
\end{aligned}
$$

kilograms
9. The following data on mother's age and birth weight (in grams) are consistent with data published by the National Center for Health Statistics. This data describes 10 mothers and their baby's birth weight.

| Mother's <br> Age | 15 | 15 | 16 | 16 | 17 | 17 | 18 | 18 | 19 | 19 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Baby's <br> weight | 5.1 | 5.7 | 5.5 | 6.5 | 6.6 | 7.5 | 6.9 | 7.2 | 7.4 | 7.8 |

(a) (1 point) Find the correlation coefficient. Round to 4 places.

$$
v=0.8832
$$

(b) (1 point) Give the equation for the regression line.

$$
\text { Baby weight }=-2.645+0.545 \text { (mother age) }
$$

(c) (1 point) What would you predict for the birth weight of a baby born to a 15 year old mother?

$$
\text { Baby weight }=-2.645+0.545(15)=5.53
$$

kilograms (can't be grams)
(d) (1 point) Interpret the slope of the regression line.

As the mother's age increases by 1 year, the baby's weight increases by 0.545 kilograms, on average.
0.25
10. If you randomly select a number in a legitimate financial report the probability that the first digit is 1 is 02 , that the first digit is 2 is 0.21 , and in general, the probabilities for each of the possible first digits is given by the following table

| X | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathrm{P}(\mathrm{X})$ | 0.25 | 0.21 | 0.15 | 0.10 | 0.09 | 0.07 | 0.06 | 0.03 | 0.04 |

(a) (1 point) Is this a continuous or a discrete probability distribution?

Discrete
(b) (1 point) Determine $\mathrm{P}(\mathrm{X}>3)$.

$$
1-(0.25+0.21+0.15)=0.39
$$

(c) (1 point) Find the expected value (mean). Round to two decimal places. No work necesspry.

$$
\text { mean }=3.41
$$

(d) (1 point) Find the standard deviation. Round to two decimal places. No work necessary.

$$
\text { St. del }=2.30
$$

11. What is the appropriate critical value for each of the following confidence levels? Round to two decimals.
(a) (1 point) Confidence level $90 \%$ for proportions

$$
z=\operatorname{invNorm}\left(\begin{array}{c}
(0.95,0,1) \\
(1+0.90) / 2
\end{array}\right.
$$

(b) (1 point) Confidence level $95 \%$ for means if $\sigma$ is known

$$
z=\operatorname{iuvNovm}\left(\begin{array}{c}
0.975,0,1)=1.96 \\
(1+0.95) / 2
\end{array}\right.
$$

(c) (1 point) Confidence level $90 \%$ for means if $\sigma$ is unknown and degrees of freedom is 10

$$
\begin{array}{r}
t=\operatorname{iuvT}(0.95,10)=1.81 \\
(1+0.90) / 2
\end{array}
$$

(d) (1 point) Confidence level $85 \%$ for two proportions and $\mathrm{n}=50$ for both samples

$$
\begin{gathered}
z=\operatorname{inv} \operatorname{Norm}(0.925,0,1)=1.44 \\
(1+0.85) / 2
\end{gathered}
$$

12. In a survey of 900 college students in the United States, 625 indicated that they believe that a student or faculty member on campus who uses language considered racist, sexist, homophobic, or offensive should be subject to disciplinary action. Assuming that the sample is representative of college students in the United States, construct and interpret a $95 \%$ confidence interval for the proportion of college students who have this belief.
(a) (2 points) Define the parameter we are estimating
$p=$ population proportion of college students in US who have this belief
(b) (2 points) Answer the 4 key questions

Question type Estimation or hypothesis test
Study type Sample r experiment data
Type of Dato 10 variables Categorical or Numen
(c) (2 points) Check the conditions that need to be met
Number of samples:

$$
n=900
$$

- assumed random sample
- students gave inde pendent responses

$$
-u p(1-p)=900 . \frac{625}{900}\left(1-\frac{625}{900}\right)=190.97>10
$$

(d) (2 points) Find the confidence interval. Show work or write the calculator function used with inputs. Round to 4 places.

$$
\begin{array}{ll}
\text { 1- Prop ZIut } & \\
\text { x: } 625 & (0.6643,0.7245) \\
n: 900 & \\
\text { c-level : } 0.95 &
\end{array}
$$

(e) (2 points) Interpret the confidence interval.

We are $95 \%$ confident thar the true proportion of US college students who have this belief is between 0.664 and 0.725 .
13. Acrylic bone cement is sometimes used in hip and knee replacements to secure an artificial joint in place. The force required to break an acrylic bone cement bond was measured for six specimens, and the resulting mean and standard deviation were 315.25 Newtons and 45.50 Newtons, respectively. Assuming that is is reasonable to believe that breaking force has a distribution that is approximately normal, use a confidence interval with $95 \%$ confidence to estimate the mean breaking force for acrylic bone cement.
(a) (2 points) Define the parameter we are estimating
$\mu=$ population mean breaking force for
acrylic bone cement.
(b) (2 points) Answer the 4 key questions

Question type: Estimation or hypothesis test Study type: Sample experiment data

Type of Dato 1 gr 2 variables? Categorical o Numerical?
Number of samples:

$$
n=6
$$

(c) (2 points) Check the conditions that need to be met

- approximately normal distribution
- random sample
- measurements are independent
(d) (2 points) Find the confidence interval. Show work or write the calculator function used with inputs. Round to 4 places.

$$
\begin{aligned}
& \text { TIuterval, Stats } \\
& \bar{x}: 315.25 \\
& S_{x}: 45.50 \\
& u: 6 \\
& \text { c-cevel }: 0.95
\end{aligned} \quad(267.5007,362.9993)
$$

(e) (2 points) Interpret the confidence interval. We are $95 \%$ confident shat the mean breaking force for acrylic bone cement is between 267.5 and 363.0 Newtons.
14. To determine if chocolate milk is as effective as other carbohydate replacement drinks, nine male cyclists performed an intense workout followed by a drink and a rest period. At the end of the rest period, each cyclist performed an endurance trial in which he exercised until exhausted, and the time to exhaustion was measured. Each cyclist completed the entire regimen on two different days. On one day, the drink provided was chocolate milk, and on the other day the drink provided was a carbohydate replacement drink. Is there evidence that the mean time to exhaustion is greater after chocolate milk than after a carbohydrate replacement drink? Use significance level of $\alpha=0.01$.

|  | Time to Exhaustion (minutes) |  |  |  |  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Cyclist | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Chocolate Milk | 25.00 | 50.25 | 38.30 | 26.10 | 36.55 | 27.30 | 36.45 | 48.35 | 35.80 |
| Carbohydrate Replacement | 10.00 | 29.10 | 37.50 | 16.75 | 10.11 | 21.60 | 31.20 | 22.04 | 17.20 |

(a) (2 points) Determine the hypotheses

$$
\begin{aligned}
& \text { (2 points) Determine the hypotheses } \\
& d=\text { chocolate milk - carbohydrate replace men }
\end{aligned}
$$

$$
H_{0}: \mu_{d}=0, H_{a}: \mu_{d}>0
$$

(b) (2 points) Answer the 4 key questions

Question type: Estimation or hypothesis test
Study type: Sample or experiment data

Type of Data: 1 or 2 variables? Categorical or Numerical?
Number of samples:

$$
n=9
$$

(c) (2 points) Check the conditions that need to be met

- assuming differences have normal distr.
- results for each cyclist are independent
- assuming a random sample
(d) (2 points) Calculate the following. Show work or write the calculator function used with inputs. Round test statistic to 2 places and the p value to 4 places.

$$
\begin{aligned}
& \bar{x}_{d}=14.289 \text { Test statistic } t=4.52 \text { p-value }=0.0010 \\
& T \text {-Test, Data } \\
& \mu_{0}: 0 \\
& L_{i s t}: L_{1} \text { (contains differences) } \\
& \mu:>\mu_{0}
\end{aligned}
$$

(e) (2 points) Make a conclusion in the context of the problem.

Reject $H_{0}: ~ p<\alpha=0.01$
There is convincing evidence that the mean time to exhaustion is greater after chocolate milk than after a carbon kyd rare replacement drink.
15. The Knight Foundation asked each person in a representative sample of high school students and in a representative sample of high school teachers which of the rights guaranteed by the First Amendment they thought was the most important. Suppose that the sample size for each sample was 1000. Data consistent with summary values given in the paper are summarized below. Carry out a hypothesis test to determine if there is convincing evidence that the proportions falling into the five First Amendment rights categories are not the same for teachers and students. Use a significance level of $\alpha=0.05$.

|  | Most Important First Amendment Right |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
|  | Freedom of Speech | Freedom of the Press | Freedom of Religion | Freedom to Peacefully <br> Assemble | Freedom to Petition the <br> Government |
| Students | 635 | 40 | 235 | 35 | 55 |
| Teachers | 410 | 50 | 430 | 40 | 70 |

(a) (2 points) Determine the hypotheses

Ho: proportions are the same for teachers and Ha: proportions are not the same students
(b) (2 points) Which of the $\chi^{2}$ tests should be used?
$x^{2}$ test of independence
(c) (2 points) Check the conditions that need to be met

- assumed random samples of student ts and teachers
- expected frequencies are $\geqslant 5$.
(d) (2 points) Calculate the following. Show work or write the calculator function used with inputs. Round test statistic to 2 places and the $p$ value to 4 places.
Degrees of freedom $=4 \quad$ Test statistic $\chi^{2}=108.87 \mathrm{p}$-value $=0$

$$
X^{2}-\text { Test }^{2}
$$

(e) (2 points) Make a conclusion in the context of the problem.

Reject $\mathrm{Ho}_{0}: p<\alpha=0.05$
There is convincing evidence that the proportions falling under the First Amendment rights categovies are not the same for teachers and students.
16. Do people feel hungrier after sampling a healthy food? A study randomly assigned volunteers into one of three groups. The people in the first group were asked to taste a snack that was billed as a new health bar containing high levels of protein, vitamins, and fiber. The people in the second group were asked to taste the same snack but were told it was a tasty chocolate bar with a raspberry center. After tasting the snack, participants were asked to rate their hunger level on a scale from 1 (not at all hungry) to 7 (very hungry). The people in the third group were asked to rate their hunger but were not given a snack. The data in the accompanying table are consistent with summary quantities given in the study (although the sample sizes in the actual study were larger). Do these data provide evidence that the mean hunger rating differs for at least two of the treatments (healthy snack, tasty snack, no snack)? Test the relevant hypotheses using a significance level of 0.01 .

| Treatment Group | Hunger Rating |  |  |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :---: |
| Healthy | 4 | 7 | 7 | 4 | 5 | 3 | 4 | 7 | 6 |  |
| Tasty | 4 | 1 | 3 | 2 | 6 | 2 | 5 | 3 | 4 |  |
| No Snack | 3 | 4 | 5 | 6 | 5 | 4 | 2 | 4 | 4 |  |

(a) (2 points) Determine the hypotheses
$H_{0}$ : mean hunger ratings are the same for the 3 treaturents
$H_{a}$ : at least two mean hunger ravings are different.
(b) (2 points) Answer the 4 key questions

Question type: Estimation o hypothesis test
Study type: Sample or experiment data

Type of Data: 1 or 2 variables? Categorical Numerical?
Number of samples:

$$
u_{1}=u_{2}=u_{3}=9
$$

(c) (2 points) Check the conditions that need to be met

- normal populations for each treatment
- random assignment to 3 treatment ts
- independent samples
- the 3 st. deviations do not differ significantly
(d) (2 points) Calculate the following. Write the calculator function used with inputs. Round test statistic to 2 places and the p value to 4 places.
Test statistic $F=3.86$ p-value $=0.0352$
$\operatorname{ANova}\left(L_{1}, L_{2}, L_{3}\right)$
(e) (2 points) Make a conclusion in the context of the problem.

Fail to reject $H_{0}: p>\alpha=0.01$
There is not convincing evidence thar the mean hunger ratings differ between the treatments.

