# STAT 145 Final Exam

Spring, 2021
Calvin University

Name:________________________

You must **justify your answers** with appropriate work to receive full credit. Even for calculations done on a calculator, you are responsible for writing in sufficient detail on the exam sheet what calculation is being performed.

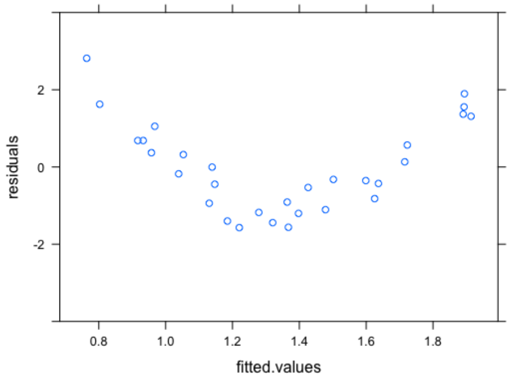
It has been my sincere pleasure to work with you this semester. Good luck on the exam, and may the summer break be productive, restful, joyful and enriching as befits your life in Christ.

Section 1. Multiple Choice *(2 pts each)*: Select the (one) best response.

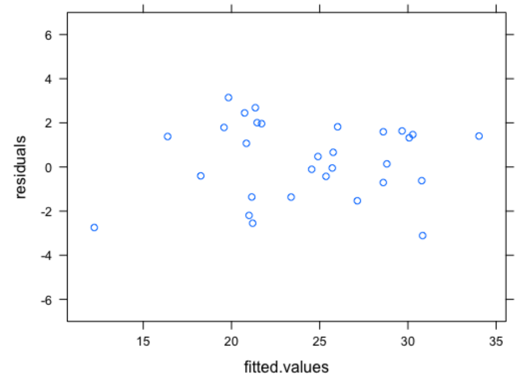
1. Which of the following is a categorical variable?
   - (a) Daily caloric intake
   - (b) Height
   - (c) Race
   - (d) Pulse rate
2. Which of the following divides a density curve (or histogram) of a skewed distribution into two pieces enclosing equal areas?
   - (a) mean
   - (b) median
   - (c) standard deviation
   - (d) interquartile range
3. Suppose we find that the $P$-value associated with a test statistic for a given sample is $P = 0.033$. This means that
   - (a) the probability that the null hypothesis is true is 3.3%.
   - (b) the probability that the alternative hypothesis is true is 3.3%.
   - (c) if the null hypothesis is true, the probability of getting a test statistic at least as extreme as ours is 3.3%.
   - (d) 3.3% of all possible samples will produce this value of the test statistic.
4. Which of the following is NOT a characteristic of all Normal distributions?
   - (a) Standard deviation is 1
   - (b) Bell-shaped distribution curve
   - (c) Total area under the distribution curve is 1
   - (d) About 95% of distribution lies with 2 standard deviations of the mean
5. A basic difference between an observational study and an experiment is that
   - (a) experiments can establish associations.
   - (b) experiments use randomization.
   - (c) statistical inference is used in experiments.
   - (d) values of explanatory variables are imposed, not simply measured, in experiments.
6. A Type II error occurs, in the context of hypothesis testing, when
   - (a) one rejects a null hypothesis that is true.
   - (b) one does not reject a null hypothesis that is false.
   - (c) one rejects a null hypothesis that is false.
   - (d) one does not reject a null hypothesis that is false.
7. Which of the following is NOT a general guideline for the creation of randomization samples for some statistical test?
   - (a) Use the data in the original sample.
   - (b) Reflect the way the original data were collected.
   - (c) Make it so the distribution of randomization statistics is centered at zero.
   - (d) Be consistent with the null hypothesis.

8. Which of the following plots of "residuals vs. fitted values" would tend to best *confirm* the assumptions of simple linear regression?

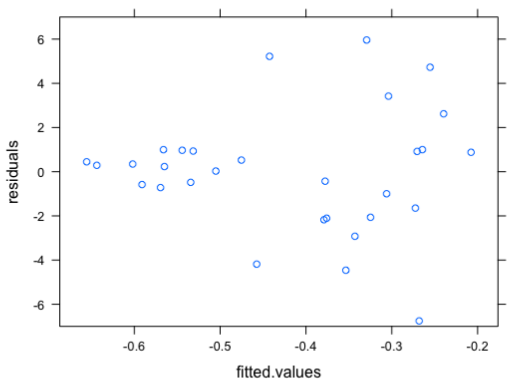
(a)



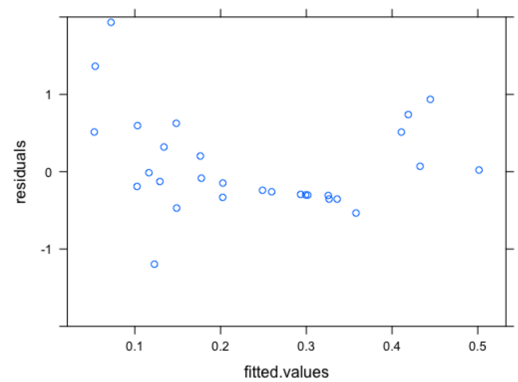

(c)




(b)




(d)




9. To cut in half the *standard deviation of the sampling distribution* (also known as the **standard error**) of the sample mean or proportion, it is generally necessary to _______________ the sample size.

(a) cut in half

(b) double

(c) reduce by 75%

(d) quadruple (make 4 times as large)

10. Which of the following quantities is *resistant* to the presence of outliers? [Imagine the effect on each of the named quantities if, to an existing list of numbers, you add one more that is a clear outlier.]

(a) standard deviation

(b) IQR

(c) Mean

(d) Range

11. When the technician taking the response measurements in an experiment is unaware of the treatments assigned to individual subjects/units, this is done with the intention to

(a) reduce bias.

(b) reduce the amount of variability in the response variable.

(c) increase randomness.

(d) none of the above.

12. Which of the following is a standard way to display the relationship between *two* quantitative variables?

    (a) Bar graph

    (b) Histogram

    (c) Scatterplot

    (d) Contingency table

13. Suppose we find that a 95% confidence interval for the mean SAT-Verbal score for high school seniors 495 ± 10. This means that

    (a) the method used to construct the confidence interval, when used repeatedly, will produce an interval that contains the population mean in 95% of cases over the long haul.

    (b) 95% of high school seniors have SAT-Verbal scores that lie in this range.

    (c) 95% of random samples drawn from high school seniors would produce confidence intervals containing 495.

    (d) the margin of error of our estimate is 0.95.

14. A poll of 1800 American residents age 18 or over asked if they approved of the Predident's job performance. Of these 1800, 41% said "Yes", 52% said "No", and 7% were undecided. In this situation, the 52% is a

    (a) population

    (b) parameter

    (c) statistic

    (d) sample

15. Experiments should use randomization to assign subjects to treatments in order to

    (a) avoid confounding the effect of a treatment with lurking variables.

    (b) make the experiment more sensitive to differences among treatments.

    (c) make subjects unaware of which treatment they are receiving.

    (d) allow the use of random number tables.

16. Suppose an article in a biology journal reports a result that is significant at the 1% level. Are such results always, never, or sometimes significant at the 5% level?

    (a) always

    (b) sometimes

    (c) never

    (d) not enough information to decide

17. Which of the following is NOT generally true of the regression line $y = b_0 + b_1 x$?

    (a) The *observed* $y$-values all fall on this line.

    (b) The line is chosen so that the sum of squares of residuals is as small as possible.

    (c) The line should be used with caution in extrapolating beyond the region of $x$-data.

    (d) The slope of the line and the correlation coefficient have the same sign (i.e., both positive or both negative).

18. It has been reported that 30% of all new businesses fail. Assuming this value perfectly represents the situation in the country at large, the 30% is a

    (a) population

    (b) parameter

    (c) statistic

    (d) sample

19. If the 95% confidence interval for a particular population mean $\mu$ is $3.1 \pm 3.85$, what can you accurately say about the $P$-value for an hypothesis test with $\mathbf{H}_0:\ \mu = 0,\ \mathbf{H}_a:\ \mu \neq 0$?

    (a) The $P$-value is larger than 0.1.

    (b) The $P$-value is larger than 0.05.

    (c) The $P$-value is less than 0.05.

    (d) It is impossible to tell which of these three is correct from this information.

20. As sample size increases, which of the following is NOT true?

    (a) The standard error of the sample mean becomes smaller.

    (b) The sampling distribution of the sample mean becomes increasingly *normal*.

    (c) The width of corresponding 95% confidence intervals decreases.

    (d) The chance of rejecting a true null hypothesis at the 5% level decreases.

21. There is the potential for students to be confused over the purpose of one-way ANOVA tests. If, say, we have 5 different groups, each with its own mean, we might systematically choose two groups at a time (Groups A and B together, then Groups A and C, . . . , Groups D and E, resulting in ten such pairings in all), and use 2-sample $t$ tests to tell whether the evidence against equality of sample means is significant at the 5% level or not. The reason for **not** doing this without first having obtained a significant result from 1-way ANOVA is

    (a) because it would be tedious to carry out ten different 2-sample $t$ tests.

    (b) because, even if all population means are equal, the chance of committing a Type I error at least once in these ten 2-sample $t$ tests is far greater than 5%.

    (c) because, for 2-sample $t$, the categorical variable "Which group?" cannot serve as the explanatory variable.

    (d) because one should always use a *matched-pairs* $t$ test, not a 2-sample $t$ test, for such pairwise comparisons.

22. Which of the following is NOT generally a characteristic of the correlation coefficient $r$?

    (a) $r$ is between $(-1)$ and 1

    (b) $r$ equals 1 or $(-1)$ only for perfect straight-line relationships.

    (c) $r$ has units (such as *meters*, *minutes*, etc.) that match those of the response variable.

    (d) $r$ requires two variables which are quantitative.

23. At right I have depicted a contingency table involving the variables `Rh-factor` and `blood type`. This problem has nothing to do with blood, nor with the data in this table. Its presence is simply to remind you of what a contingency table does/is.

| | Rh-factor | |
|---|---|---|
| Type | Neg. | Pos. |
| A | 32 | 161 |
| B | 9 | 44 |
| AB | 4 | 13 |
| O | 29 | 181 |

Suppose you have a contingency table with exactly 4 cells, two in the first row and two in the second. One statistical procedure you might use in such a setting is a chi-square test for association between categorical variables; but another equally-useful procedure that applies in such settings is

(a) a model utility test.

(b) a chi-square goodness-of-fit test.

(c) a 2-sample $t$ test.

(d) a 2-proportion test.

## Section 2. Short Answer

24. ANOVA TABLE.

(a) *(3 pts)* Fill in the missing entries in the ANOVA table below.

| Source of Variation | df | SS | MS | F | $P$-value |
|---|---|---|---|---|---|
| Group | 4 | 98.2 | | | 0.0345 |
| Error | 18 | 134.5 | | | |

(b) *(3 pts)* How many groups are represented in this data set? How many subjects/cases?

(c) *(4 pts)* What conclusion is reached at the 5% significance level? [You should state, generically, appropriate null and alternative hypotheses as part of your answer.]

25. *(4 pts each)* WHAT DO I DO? In each of the following situations, pretend you want to know some information and you are designing a statistical study to find out about it. Give the following **three** pieces of information for each: (i) **what variables** you would need to have in your dataset, (ii) whether these variables are **categorical** or **quantitative**, and (iii) **what statistical procedure** you would use to analyze the results.

   Select your procedures from the following list: 1-proportion, 2-proportion, 1-sample $t$, 2-sample $t$, paired-$t$, chi square goodness-of-fit, chi square test for association of categorical variables, 1-way ANOVA. or model utility test.

   Any procedure may appear as your answer more than once, or be omitted entirely. If two procedures are possible, you need only give one.

   (a) You want to compare how long two different types of lightbulbs last before burning out.

   (b) You wonder whether there is any difference in success rates (defined by whether participants are still smoking one year later) between three different smoking cessation programs.

   (c) You want to see whether three different smoking cessation programs lead to different amounts of weight gain among women who successfully quit smoking using these programs.

   (d) You want to know whether Americans prefer peas over carrots.

   (e) You want to know whether the amount of time spent listening to iPods or similar devices with earphones is associated with standard hearing test scores.

| | **Variable(s)** | **Procedure** |
|---|---|---|
| (a) | | |
| (b) | | |
| (c) | | |
| (d) | | |
| (e) | | |

26. *(2 pts each)* R COMMANDS. Suppose there is a data frame called **Cars**. The *first few lines* of this data frame appear below.

```
head(Cars)
```

| Make | Model | Type | Drive | CityMPG | HwyMPG | Weight | Size |
|---|---|---|---|---|---|---|---|
| Subaru | Forester | SUV | AWD | 24 | 32 | 3370 | Small |
| Cadillac | CTS | Sedan | AWD | 18 | 26 | 3915 | Midsized |
| Toyoto | Sequioa | 7Pass | RWD | 12 | 18 | 6025 | Large |
| Scion | xB | Wagon | FWD | 22 | 28 | 3120 | Small |
| Audi | A3 | Sedan | FWD | 23 | 33 | 3135 | Small |

Write an R command (complete and ready to be executed) that would produce the desired information for each of the following scenarios. Do not use `favstats()` in any of your answers.

(a) Give the number of cases in the data frame.

(b) Give the mean `Weight` among vehicles in the data frame.

(c) Display the frequencies of cars of different `Type`.

(d) Produce side-by-side box plots of `CityMPG` for each value of `Drive`.

(e) Make a two-way table on variables `Drive` and `Size`.

(f) Give regression line coefficients using `Weight` as explanatory, `HwyMPG` as response.

(g) Select out the part of the data frame consisting of those cars with `CityMPG` under 20.

27. After marginal totals are added, the two-way table sought in part (e) of the last problem appears at right.

| | Drive | | | |
|---|---|---|---|---|
| Size | AWD | FWD | RWD | Total |
| Large | 7 | 13 | 9 | 29 |
| Midsized | 9 | 20 | 5 | 34 |
| Small | 9 | 30 | 8 | 47 |
| Total | 25 | 63 | 22 | 110 |

(a) *(4 pts)* What contribution to the $\chi^2$-statistic comes from the (`RWD`, `Midsized`) cell?

(b) *(3 pts)* The output below comes from applying the `chisq.test()` command to this table. What is the null hypothesis of this test, and what conclusion should we draw?

```
X-squared = 4.1017, df=4, p-value = 0.3924
```

28. *(6 pts)* Difference of proportions. Suppose 18 of 37 males sampled indicate they have tried a banned substance; the number is 15 out of 41 females sampled. If this data is to be used to test hypotheses

$$\mathbf{H}_0:\ p_M - p_F = 0 \qquad \text{vs.} \qquad \mathbf{H}_a:\ p_M - p_F \neq 0,$$

calculate and label the *standardized test statistic*.

29. *(6 pts)* DIFFERENCE OF MEANS. In a study of the effect of college student employment on academic performance, a sample of 184 students who were simultaneously employed had an average gpa of 3.12 with standard deviation 0.485. Among 114 students who were otherwise unemployed, the average gpa was 3.23 with standard deviation 0.524. Taking the subscripts $U$ and $E$ for Unemployed and Employed students respectively, find a 95% confidence interval for $\mu_U - \mu_E$, the difference in mean gpa. The critical value for the 95% level is $t^* = 1.981$.

30. Student volunteers at Ohio State University were randomly assigned a number of cans of beer to drink. Thirty minutes later a police officer measured their blood alcohol content (BAC) in grams of alcohol per deciliter of blood (g/dl). Here are data for the first 6 participants, arranged in rows (so as to fit the page better) instead of the usual columns. (So, Student 1 drank 5 beers and had a measured BAC of 0.10.)

| | 1 | 2 | 3 | 4 | 5 | 6 | ... |
|---|---|---|---|---|---|---|---|
| Beers | 5 | 2 | 9 | 8 | 3 | 7 | ... |
| BAC | 0.10 | 0.03 | 0.19 | 0.12 | 0.04 | 0.10 | ... |

Some regression-related output (the `summary()` command was used) appears below.

```
              Estimate   Std. Error   t value   PR(>|t|)
(Intercept)    -0.0127       0.0126     -1.00       0.33
Beers           0.0180       0.0024      7.48   0.000003

Residual standard error:  1.02 on 14 degrees of freedom
Multiple R-squared:  0.8
F-statistic:  55.9 on 1 and 14 DF, p-value:  0.000003
```

(a) *(2 pts)* Inspect the output carefully. How many participants/cases are in the full dataset?

(b) *(2 pts)* What portion of the variability in `BAC` values among participants in the study is *explained* by a linear model involving number of `Beers` consumed?

(c) *(3 pts)* What is the slope of the regression line? Explain, in context, what this slope value tells us.

(d) *(2 pts)* Find the *correlation* between `BAC` and `Beers`.

(e) *(2 pts)* Suggest at least one other variable not considered here that could help explain some of the variability in `BAC` *not* accounted for by our linear model with `Beers`.

(f) *(3 pts)* Why is it important that the students were randomly assigned a number of beers to drink?

(g) *(2 pts)* What is the residual for Student/Case #5?

(h) *(3 pts)* State the null and alternative hypotheses for the **model utility test**. What conclusion do we draw regarding this test?

31. In a study to test the user-interface of her latest computer program, Cindy selects 20 students at random and has them perform a series of tasks as quickly as they can.

| Command | Result |
|---|---|
| `pt(3.64, 19)` | `0.999129` |
| `pt(3.64, 18)` | `0.999064` |
| `qt(.90, 18)` | `1.330391` |
| `qt(.95, 18)` | `1.734064` |
| `qt(.90, 19)` | `1.327728` |
| `qt(.95, 19)` | `1.729133` |

(a) *(5 pts)* Suppose, among the 20 students, the mean time to complete the tasks is 43.2 seconds, with a standard deviation of 5.4 seconds. Find a 90% confidence interval for the mean time it takes to complete the series of tasks. Several `RStudio` commands (not all of which are useful to this situation) have been provided along with their results.

(b) *(4 pts)* Cindy may, as an alternate method to the one used above for finding a confidence interval, employ bootstrapping. Describe how she might physically produce a *single bootstrap sample*.

(c) *(3 Extra Credit pts)* Suppose Cindy simulates a bootstrap distribution using the command in `RStudio`

```
manyMeans = do(10000) * mean(resample(timeToCompleteTasks))
```

Give a follow-up command (or commands) which would produce left- and right-endpoints of a 90% **bootstrap percentile confidence interval**. [Note: `manyMeans` will have just one column, called `mean`.]

32. For a given 4-week period in New York City, there were 9,144 births recorded. Of these, 3,067 occurred on the weekends (Sat. or Sun.), and the rest on weekdays. If one were curious whether there is any difference on the rate of births (as measured in births per day) on weekends from that during weekdays, perhaps this data can help settle the question, at least in regards to the city of New York.

    (a) *(3 pts)* Indicate by name a type of test that is tailored to this situation. Does it seem such a test is appropriate here? Explain briefly.

    (b) *(3 pts)* State null and alternative hypotheses suited to the test you named in the previous part..

    (c) *(5 pts)* Calculate a test statistic appropriate for the test you named in the part (a).

    (d) *(2 pts)* Write an `RStudio` command that uses the test statistic of the previous part and will produce the corresponding $P$-value.

# STAT 145 Final Exam

Spring, 2021
Calvin University

Name:______________________

You must **justify your answers** with appropriate work to receive full credit. Even for calculations done on a calculator, you are responsible for writing in sufficient detail on the exam sheet what calculation is being performed.

It has been my sincere pleasure to work with you this semester. Good luck on the exam, and may the summer break be productive, restful, joyful and enriching as befits your life in Christ.

# Answer Key for Exam $\boxed{\text{A}}$

Section 1. Multiple Choice *(2 pts each)*: Select the (one) best response.

1. Which of the following is a categorical variable?
   - (a) Daily caloric intake
   - (b) Height
   - $\boxed{\text{(c)}}$ Race
   - (d) Pulse rate
2. Which of the following divides a density curve (or histogram) of a skewed distribution into two pieces enclosing equal areas?
   - (a) mean
   - $\boxed{\text{(b)}}$ median
   - (c) standard deviation
   - (d) interquartile range
3. Suppose we find that the $P$-value associated with a test statistic for a given sample is $P = 0.033$. This means that
   - (a) the probability that the null hypothesis is true is 3.3%.
   - (b) the probability that the alternative hypothesis is true is 3.3%.
   - $\boxed{\text{(c)}}$ if the null hypothesis is true, the probability of getting a test statistic at least as extreme as ours is 3.3%.
   - (d) 3.3% of all possible samples will produce this value of the test statistic.
4. Which of the following is NOT a characteristic of all Normal distributions?
   - $\boxed{\text{(a)}}$ Standard deviation is 1
   - (b) Bell-shaped distribution curve
   - (c) Total area under the distribution curve is 1
   - (d) About 95% of distribution lies with 2 standard deviations of the mean
5. A basic difference between an observational study and an experiment is that
   - (a) experiments can establish associations.
   - (b) experiments use randomization.
   - (c) statistical inference is used in experiments.
   - $\boxed{\text{(d)}}$ values of explanatory variables are imposed, not simply measured, in experiments.
6. A Type II error occurs, in the context of hypothesis testing, when
   - (a) one rejects a null hypothesis that is true.
   - $\boxed{\text{(b)}}$ one does not reject a null hypothesis that is false.
   - (c) one rejects a null hypothesis that is false.
   - (d) one does not reject a null hypothesis that is false.

7. Which of the following is NOT a general guideline for the creation of randomization samples for some statistical test?

   (a) Use the data in the original sample.

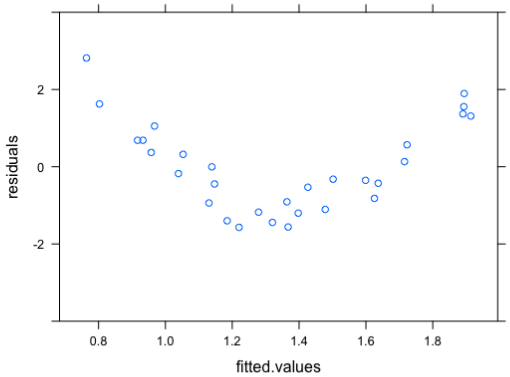
   (b) Reflect the way the original data were collected.

   $\boxed{\text{(c)}}$ Make it so the distribution of randomization statistics is centered at zero.

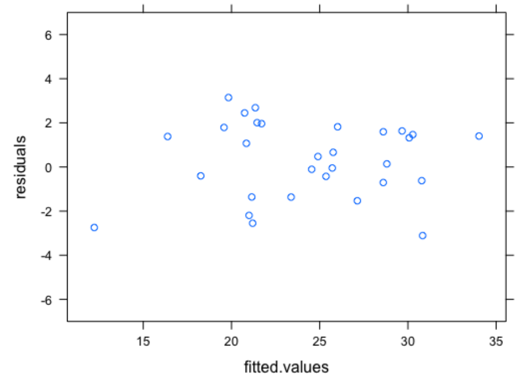
   (d) Be consistent with the null hypothesis.

8. Which of the following plots of "residuals vs. fitted values" would tend to best *confirm* the assumptions of simple linear regression?

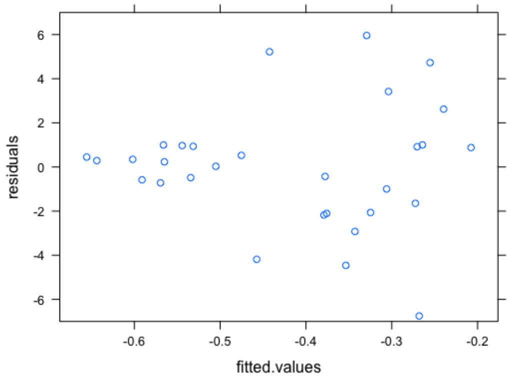
(a)



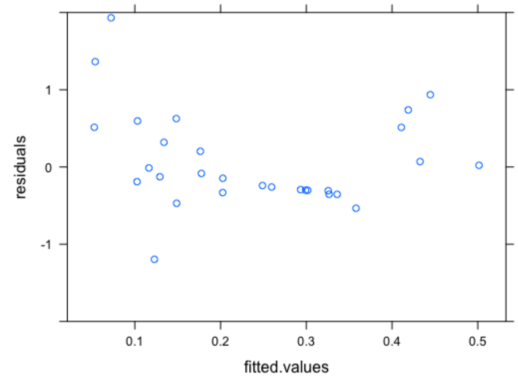

(c)




(b)




(d)




9. To cut in half the *standard deviation of the sampling distribution* (also known as the **standard error**) of the sample mean or proportion, it is generally necessary to ______________ the sample size.

   (a) cut in half

   (b) double

   (c) reduce by 75%

   $\boxed{\text{(d)}}$ quadruple (make 4 times as large)

10. Which of the following quantities is *resistant* to the presence of outliers? [Imagine the effect on each of the named quantities if, to an existing list of numbers, you add one more that is a clear outlier.]

   (a) standard deviation

   $\boxed{\text{(b)}}$ IQR

   (c) Mean

   (d) Range

11. When the technician taking the response measurements in an experiment is unaware of the treatments assigned to individual subjects/units, this is done with the intention to

    $\boxed{\text{(a)}}$ reduce bias.

    (b) reduce the amount of variability in the response variable.

    (c) increase randomness.

    (d) none of the above.

12. Which of the following is a standard way to display the relationship between *two* quantitative variables?

    (a) Bar graph

    (b) Histogram

    $\boxed{\text{(c)}}$ Scatterplot

    (d) Contingency table

13. Suppose we find that a 95% confidence interval for the mean SAT-Verbal score for high school seniors 495 ± 10. This means that

    $\boxed{\text{(a)}}$ the method used to construct the confidence interval, when used repeatedly, will produce an interval that contains the population mean in 95% of cases over the long haul.

    (b) 95% of high school seniors have SAT-Verbal scores that lie in this range.

    (c) 95% of random samples drawn from high school seniors would produce confidence intervals containing 495.

    (d) the margin of error of our estimate is 0.95.

14. A poll of 1800 American residents age 18 or over asked if they approved of the Predident's job performance. Of these 1800, 41% said "Yes", 52% said "No", and 7% were undecided. In this situation, the 52% is a

    (a) population

    (b) parameter

    $\boxed{\text{(c)}}$ statistic

    (d) sample

15. Experiments should use randomization to assign subjects to treatments in order to

    $\boxed{\text{(a)}}$ avoid confounding the effect of a treatment with lurking variables.

    (b) make the experiment more sensitive to differences among treatments.

    (c) make subjects unaware of which treatment they are receiving.

    (d) allow the use of random number tables.

16. Suppose an article in a biology journal reports a result that is significant at the 1% level. Are such results always, never, or sometimes significant at the 5% level?

    $\boxed{\text{(a)}}$ always

    (b) sometimes

    (c) never

    (d) not enough information to decide

17. Which of the following is NOT generally true of the regression line $y = b_0 + b_1 x$?

    $\boxed{\text{(a)}}$ The *observed* $y$-values all fall on this line.

    (b) The line is chosen so that the sum of squares of residuals is as small as possible.

    (c) The line should be used with caution in extrapolating beyond the region of $x$-data.

    (d) The slope of the line and the correlation coefficient have the same sign (i.e., both positive or both negative).

18. It has been reported that 30% of all new businesses fail. Assuming this value perfectly represents the situation in the country at large, the 30% is a

(a) population
$\boxed{\text{(b)}}$ parameter
(c) statistic
(d) sample

19. If the 95% confidence interval for a particular population mean $\mu$ is $3.1 \pm 3.85$, what can you accurately say about the $P$-value for an hypothesis test with $\mathbf{H}_0: \mu = 0$, $\mathbf{H}_a: \mu \neq 0$?

(a) The $P$-value is larger than 0.1.
$\boxed{\text{(b)}}$ The $P$-value is larger than 0.05.
(c) The $P$-value is less than 0.05.
(d) It is impossible to tell which of these three is correct from this information.

20. As sample size increases, which of the following is NOT true?

(a) The standard error of the sample mean becomes smaller.
(b) The sampling distribution of the sample mean becomes increasingly *normal*.
(c) The width of corresponding 95% confidence intervals decreases.
$\boxed{\text{(d)}}$ The chance of rejecting a true null hypothesis at the 5% level decreases.

21. There is the potential for students to be confused over the purpose of one-way ANOVA tests. If, say, we have 5 different groups, each with its own mean, we might systematically choose two groups at a time (Groups A and B together, then Groups A and C, . . . , Groups D and E, resulting in ten such pairings in all), and use 2-sample $t$ tests to tell whether the evidence against equality of sample means is significant at the 5% level or not. The reason for **not** doing this without first having obtained a significant result from 1-way ANOVA is

(a) because it would be tedious to carry out ten different 2-sample $t$ tests.
$\boxed{\text{(b)}}$ because, even if all population means are equal, the chance of committing a Type I error at least once in these ten 2-sample $t$ tests is far greater than 5%.
(c) because, for 2-sample $t$, the categorical variable "Which group?" cannot serve as the explanatory variable.
(d) because one should always use a *matched-pairs* $t$ test, not a 2-sample $t$ test, for such pairwise comparisons.

22. Which of the following is NOT generally a characteristic of the correlation coefficient $r$?

(a) $r$ is between $(-1)$ and 1
(b) $r$ equals 1 or $(-1)$ only for perfect straight-line relationships.
$\boxed{\text{(c)}}$ $r$ has units (such as *meters*, *minutes*, etc.) that match those of the response variable.
(d) $r$ requires two variables which are quantitative.

23. At right I have depicted a contingency table involving the variables `Rh-factor` and `blood type`. This problem has nothing to do with blood, nor with the data in this table. Its presence is simply to remind you of what a contingency table does/is.

| | Rh-factor | |
|---|---|---|
| Type | Neg. | Pos. |
| A | 32 | 161 |
| B | 9 | 44 |
| AB | 4 | 13 |
| O | 29 | 181 |

Suppose you have a contingency table with exactly 4 cells, two in the first row and two in the second. One statistical procedure you might use in such a setting is a chi-square test for association between categorical variables; but another equally-useful procedure that applies in such settings is

(a) a model utility test.

(b) a chi-square goodness-of-fit test.

(c) a 2-sample $t$ test.

$\boxed{\text{(d)}}$ a 2-proportion test.

## Section 2. Short Answer

24. ANOVA Table.

(a) *(3 pts)* Fill in the missing entries in the ANOVA table below.

| Source of Variation | df | SS | MS | F | $P$-value |
|---|---|---|---|---|---|
| Group | 4 | 98.2 | | | 0.0345 |
| Error | 18 | 134.5 | | | |

(b) *(3 pts)* How many groups are represented in this data set? How many subjects/cases?

(c) *(4 pts)* What conclusion is reached at the 5% significance level? [You should state, generically, appropriate null and alternative hypotheses as part of your answer.]

**Answer:** (a)

| Source of Variation | df | SS | MS | F | $P$-value |
|---|---|---|---|---|---|
| Group | 4 | 98.2 | 24.55 | 3.286 | 0.0345 |
| Error | 18 | 134.5 | 7.47 | | |
| Total | 22 | 232.7 | | | |

(b) There are $4 + 1 = 5$ groups and $22 + 1 = 23$ cases.

(c) For ANOVA, we have hypotheses

$$\mathbf{H}_0: \ \mu_1 = \mu_2 = \mu_3 = \mu_4,$$

and alternative hypotheses that at least one of these population means is different from others. With $P = 0.018$, we reject this null hypothesis at the 5% significance level, concluding that not all of the group means are identical.

25. *(4 pts each)* What do I do? In each of the following situations, pretend you want to know some information and you are designing a statistical study to find out about it. Give the following **three** pieces of information for each: (i) **what variables** you would need to have in your dataset, (ii) whether these variables are **categorical** or **quantitative**, and (iii) **what statistical procedure** you would use to analyze the results.

Select your procedures from the following list: 1-proportion, 2-proportion, 1-sample $t$, 2-sample $t$, paired-$t$, chi square goodness-of-fit, chi square test for association of categorical variables, 1-way ANOVA. or model utility test.

Any procedure may appear as your answer more than once, or be omitted entirely. If two procedures are possible, you need only give one.

(a) You want to compare how long two different types of lightbulbs last before burning out.

(b) You wonder whether there is any difference in success rates (defined by whether participants are still smoking one year later) between three different smoking cessation programs.

(c) You want to see whether three different smoking cessation programs lead to different amounts of weight gain among women who successfully quit smoking using these programs.

(d) You want to know whether Americans prefer peas over carrots.

(e) You want to know whether the amount of time spent listening to iPods or similar devices with earphones is associated with standard hearing test scores.

| | **Variable(s)** | **Procedure** |
|---|---|---|
| (a) | | |
| (b) | | |
| (c) | | |
| (d) | | |
| (e) | | |

**Answer:**

| | **Variable(s)** | **Procedure** |
|---|---|---|
| (a) | light bulb type (categorical)<br>time until burns out (quant) | 2-sample *t*<br>or 1-way ANOVA |
| (b) | cessation program (categorical)<br>quit smoking? (categorica) | chi-square test for association<br>or Fisher test |
| (c) | cessation program (categorical)<br>weight gain (quantitative) | 1-way ANOVA<br>or Kruskal-Wallis test |
| (d) | preferred veggie: carrots or peas (categorical) | 1-proportion |
| | | or binomial test |
| (e) | hearing test score (quantitative)<br>time listening to i-Pod (quantitative) | simple linear regression |

26. *(2 pts each)* R Commands. Suppose there is a data frame called **Cars**. The *first few lines* of this data frame appear below.

```
head(Cars)
```

| Make | Model | Type | Drive | CityMPG | HwyMPG | Weight | Size |
|---|---|---|---|---|---|---|---|
| Subaru | Forester | SUV | AWD | 24 | 32 | 3370 | Small |
| Cadillac | CTS | Sedan | AWD | 18 | 26 | 3915 | Midsized |
| Toyoto | Sequioa | 7Pass | RWD | 12 | 18 | 6025 | Large |
| Scion | xB | Wagon | FWD | 22 | 28 | 3120 | Small |
| Audi | A3 | Sedan | FWD | 23 | 33 | 3135 | Small |

Write an R command (complete and ready to be executed) that would produce the desired information for each of the following scenarios. Do not use `favstats()` in any of your answers.

(a) Give the number of cases in the data frame.

(b) Give the mean `Weight` among vehicles in the data frame.

(c) Display the frequencies of cars of different `Type`.

(d) Produce side-by-side box plots of `CityMPG` for each value of `Drive`.

(e) Make a two-way table on variables `Drive` and `Size`.

(f) Give regression line coefficients using `Weight` as explanatory, `HwyMPG` as response.

(g) Select out the part of the data frame consisting of those cars with `CityMPG` under 20.

27. After marginal totals are added, the two-way table sought in part (e) of the last problem appears at right.

| | Drive | | | |
|---|---|---|---|---|
| Size | AWD | FWD | RWD | Total |
| Large | 7 | 13 | 9 | 29 |
| Midsized | 9 | 20 | 5 | 34 |
| Small | 9 | 30 | 8 | 47 |
| Total | 25 | 63 | 22 | 110 |

(a) *(4 pts)* What contribution to the $\chi^2$-statistic comes from the (`RWD`, `Midsized`) cell?

(b) *(3 pts)* The output below comes from applying the `chisq.test()` command to this table. What is the null hypothesis of this test, and what conclusion should we draw?

```
X-squared = 4.1017, df=4, p-value = 0.3924
```

28. *(6 pts)* Difference of proportions. Suppose 18 of 37 males sampled indicate they have tried a banned substance; the number is 15 out of 41 females sampled. If this data is to be used to test hypotheses

$$\mathbf{H}_0: \ p_M - p_F = 0 \qquad \text{vs.} \qquad \mathbf{H}_a: \ p_M - p_F \neq 0,$$

calculate and label the *standardized test statistic*.

29. *(6 pts)* DIFFERENCE OF MEANS. In a study of the effect of college student employment on academic performance, a sample of 184 students who were simultaneously employed had an average gpa of 3.12 with standard deviation 0.485. Among 114 students who were otherwise unemployed, the average gpa was 3.23 with standard deviation 0.524. Taking the subscripts $U$ and $E$ for Unemployed and Employed students respectively, find a 95% confidence interval for $\mu_U - \mu_E$, the difference in mean gpa. The critical value for the 95% level is $t^* = 1.981$.

30. Student volunteers at Ohio State University were randomly assigned a number of cans of beer to drink. Thirty minutes later a police officer measured their blood alcohol content (BAC) in grams of alcohol per deciliter of blood (g/dl). Here are data for the first 6 participants, arranged in rows (so as to fit the page better) instead of the usual columns. (So, Student 1 drank 5 beers and had a measured BAC of 0.10.)

| | 1 | 2 | 3 | 4 | 5 | 6 | ... |
|---|---|---|---|---|---|---|---|
| Beers | 5 | 2 | 9 | 8 | 3 | 7 | ... |
| BAC | 0.10 | 0.03 | 0.19 | 0.12 | 0.04 | 0.10 | ... |

Some regression-related output (the `summary()` command was used) appears below.

```
              Estimate   Std. Error   t value   PR(>|t|)
(Intercept)    -0.0127       0.0126     -1.00       0.33
Beers           0.0180       0.0024      7.48   0.000003

Residual standard error:  1.02 on 14 degrees of freedom
Multiple R-squared:  0.8
F-statistic:  55.9 on 1 and 14 DF, p-value:  0.000003
```

(a) *(2 pts)* Inspect the output carefully. How many participants/cases are in the full dataset?

(b) *(2 pts)* What portion of the variability in `BAC` values among participants in the study is *explained* by a linear model involving number of `Beers` consumed?

(c) *(3 pts)* What is the slope of the regression line? Explain, in context, what this slope value tells us.

(d) *(2 pts)* Find the *correlation* between `BAC` and `Beers`.

(e) *(2 pts)* Suggest at least one other variable not considered here that could help explain some of the variability in `BAC` *not* accounted for by our linear model with `Beers`.

(f) *(3 pts)* Why is it important that the students were randomly assigned a number of beers to drink?

(g) *(2 pts)* What is the residual for Student/Case #5?

(h) *(3 pts)* State the null and alternative hypotheses for the **model utility test**. What conclusion do we draw regarding this test?

**Answer:** (a) To read in the data, we would execute the command

```
beerStudy = read.csv("http://www.calvin.edu/ scofield/beersAndBAC.csv")
```

(b)
```
summary(lm(BAC ~ Beers, data=beerStudy))
xyplot(lm(BAC ~ Beers, data=beerStudy, type=c("p","r")))
```

(c) There are $14 + 2 = 16$ participants/cases.

(d) The regression line has slope $= 0.018$. This means that for each extra drink consumed, the blood alcohol level 30 minutes hence rises approximately 0.018.

(e) It is important in order to reduce bias: we provide no opportunity for participants or experimenters to *choose* which subjects drink more and which drink less.

(f) Approximately 80% of the variability in `BAC` levels is explained by this model with Beers.

(g) Perhaps a participant's `weight` could be such a variable.

(h) The correlation is $r = \sqrt{.8} \doteq 0.894$.

(i) We have

$$\text{residual} \;=\; \text{observed} - \text{predicted} \;=\; 0.04 - [-0.0127 + (0.018)(3)] \;=\; -0.0013.$$

(j) The hypotheses are

$$\mathbf{H}_0: \;\beta = 0 \quad \text{(or the number of beers has no linear association with BAC)}$$
$$\mathbf{H}_a: \;\beta \neq 0 \quad \text{(or there is a linear association).}$$

With a $P$-value of $3 * 10^{-6}$, we reject $\mathbf{H}_0$.

31. In a study to test the user-interface of her latest computer program, Cindy selects 20 students at random and has them perform a series of tasks as quickly as they can.

| Command | Result |
|---|---|
| `pt(3.64, 19)` | `0.999129` |
| `pt(3.64, 18)` | `0.999064` |
| `qt(.90, 18)` | `1.330391` |
| `qt(.95, 18)` | `1.734064` |
| `qt(.90, 19)` | `1.327728` |
| `qt(.95, 19)` | `1.729133` |

(a) *(5 pts)* Suppose, among the 20 students, the mean time to complete the tasks is 43.2 seconds, with a standard deviation of 5.4 seconds. Find a 90% confidence interval for the mean time it takes to complete the series of tasks. Several `RStudio` commands (not all of which are useful to this situation) have been provided along with their results.

(b) *(4 pts)* Cindy may, as an alternate method to the one used above for finding a confidence interval, employ bootstrapping. Describe how she might physically produce a *single bootstrap sample*.

(c) *(3 Extra Credit pts)* Suppose Cindy simulates a bootstrap distribution using the command in `RStudio`

```
manyMeans = do(10000) * mean(resample(timeToCompleteTasks))
```

Give a follow-up command (or commands) which would produce left- and right-endpoints of a 90% **bootstrap percentile confidence interval**. [Note: `manyMeans` will have just one column, called `mean`.]

**Answer:** (a) Here, the 90% CI is

$$43.2 \pm (1.729)\frac{5.4}{\sqrt{20}} \doteq 43.2 \pm (1.729)\frac{5.4}{4.4721} \doteq 43.2 \pm 2.0877,$$

or $[41.112, 45.288]$.

(b) She could write each participant's time on a single card. Shuffling them up, she could draw one card and record the time on that card. She would then replace the card in the deck, shuffling them again and drawing another. She would repeat this process until she had drown (with replacment) 20 cards. The twenty numbers now recorded would then be used to calculate one single mean.

(c) `quantile(manyMeans$result, c(0.05, 0.95))`

32. For a given 4-week period in New York City, there were 9,144 births recorded. Of these, 3,067 occurred on the weekends (Sat. or Sun.), and the rest on weekdays. If one were curious whether there is any difference on the rate of births (as measured in births per day) on weekends from that during weekdays, perhaps this data can help settle the question, at least in regards to the city of New York.

   (a) *(3 pts)* Indicate by name a type of test that is tailored to this situation. Does it seem such a test is appropriate here? Explain briefly.

   (b) *(3 pts)* State null and alternative hypotheses suited to the test you named in the previous part..

   (c) *(5 pts)* Calculate a test statistic appropriate for the test you named in the part (a).

   (d) *(2 pts)* Write an `RStudio` command that uses the test statistic of the previous part and will produce the corresponding $P$-value.

**Answer:** (a) Either a chi-square goodness of fit can be used (appropriate since the expected counts—namely 2612.57 and 6531.43—are so high), or a 1-proportion test, which also is appropriate since $np$, $n(1-p)$ are both much higher than 10.

(b) Either a chi-square goodness of fit can be used (in which case the hypotheses should be

$$\mathbf{H}_0: \; p_{\text{weekday}} = \frac{5}{7} \text{ and } p_{\text{weekend}} = \frac{2}{7} \qquad \text{vs.} \qquad \mathbf{H}_a: \text{ neither } p_{\text{weekend}} \text{ nor } p_{\text{weekday}} \text{ are as proposed in } \mathbf{H_0}.$$

One can also give, as answer, that the procedure is 1-proportion. with $p = 2/7$.

(c) For the goodness-of-fit test, the test statistic is

$$\sum_j \frac{(O_j - E_j)^2}{E_j} \doteq 110.66.$$

For the 1-proportion test it is

$$Z = \frac{(3067/9144) - (2/7)}{\sqrt{\frac{(2/7)(1-2/7)}{9144}}} \doteq \frac{0.3354 - 0.2857}{0.004724} \doteq 10.06.$$

(d) For the goodness-of-fit test, the `RStudio` command is `1 - pchisq(110.66, 1)`
For the 1-proportion test, the `RStudio` command is `2 * (1 - pnorm(10.06))`.

# STAT 145 Final Exam

Spring, 2021
Calvin University

Name:______________________

You must **justify your answers** with appropriate work to receive full credit. Even for calculations done on a calculator, you are responsible for writing in sufficient detail on the exam sheet what calculation is being performed.

It has been my sincere pleasure to work with you this semester. Good luck on the exam, and may the summer break be productive, restful, joyful and enriching as befits your life in Christ.

Section 1. Multiple Choice *(2 pts each)*: Select the (one) best response.

1. Which of the following is NOT a general guideline for the creation of randomization samples for some statistical test?

   (a) Use the data in the original sample.

   (b) Reflect the way the original data were collected.

   (c) Make it so the distribution of randomization statistics is centered at zero.

   (d) Be consistent with the null hypothesis.

2. When the technician taking the response measurements in an experiment is unaware of the treatments assigned to individual subjects/units, this is done with the intention to

   (a) reduce bias.

   (b) reduce the amount of variability in the response variable.

   (c) increase randomness.

   (d) none of the above.

3. If the 95% confidence interval for a particular population mean $\mu$ is $3.1 \pm 3.85$, what can you accurately say about the $P$-value for an hypothesis test with $\mathbf{H}_0: \ \mu = 0, \ \mathbf{H}_a: \ \mu \neq 0$?

   (a) The $P$-value is larger than 0.1.

   (b) The $P$-value is larger than 0.05.

   (c) The $P$-value is less than 0.05.

   (d) It is impossible to tell which of these three is correct from this information.

4. There is the potential for students to be confused over the purpose of one-way ANOVA tests. If, say, we have 5 different groups, each with its own mean, we might systematically choose two groups at a time (Groups A and B together, then Groups A and C, . . . , Groups D and E, resulting in ten such pairings in all), and use 2-sample $t$ tests to tell whether the evidence against equality of sample means is significant at the 5% level or not. The reason for **not** doing this without first having obtained a significant result from 1-way ANOVA is

   (a) because it would be tedious to carry out ten different 2-sample $t$ tests.

   (b) because, even if all population means are equal, the chance of committing a Type I error at least once in these ten 2-sample $t$ tests is far greater than 5%.

   (c) because, for 2-sample $t$, the categorical variable "Which group?" cannot serve as the explanatory variable.

   (d) because one should always use a *matched-pairs* $t$ test, not a 2-sample $t$ test, for such pairwise comparisons.

5. A poll of 1800 American residents age 18 or over asked if they approved of the Predident's job performance. Of these 1800, 41% said "Yes", 52% said "No", and 7% were undecided. In this situation, the 52% is a

   (a) population

   (b) parameter

   (c) statistic

   (d) sample

6. Which of the following is NOT generally a characteristic of the correlation coefficient $r$?
   (a) $r$ is between $(-1)$ and 1
   (b) $r$ equals 1 or $(-1)$ only for perfect straight-line relationships.
   (c) $r$ has units (such as *meters, minutes*, etc.) that match those of the response variable.
   (d) $r$ requires two variables which are quantitative.
7. As sample size increases, which of the following is NOT true?
   (a) The standard error of the sample mean becomes smaller.
   (b) The sampling distribution of the sample mean becomes increasingly *normal*.
   (c) The width of corresponding 95% confidence intervals decreases.
   (d) The chance of rejecting a true null hypothesis at the 5% level decreases.
8. Which of the following divides a density curve (or histogram) of a skewed distribution into two pieces enclosing equal areas?
   (a) mean
   (b) median
   (c) standard deviation
   (d) interquartile range
9. At right I have depicted a contingency table involving the variables `Rh-factor` and `blood type`. This problem has nothing to do with blood, nor with the data in this table. Its presence is simply to remind you of what a contingency table does/is.

   Rh-factor

   | Type | Neg. | Pos. |
   |---|---|---|
   | A | 32 | 161 |
   | B | 9 | 44 |
   | AB | 4 | 13 |
   | O | 29 | 181 |

   Suppose you have a contingency table with exactly 4 cells, two in the first row and two in the second. One statistical procedure you might use in such a setting is a chi-square test for association between categorical variables; but another equally-useful procedure that applies in such settings is
   (a) a model utility test.
   (b) a chi-square goodness-of-fit test.
   (c) a 2-sample $t$ test.
   (d) a 2-proportion test.
10. To cut in half the *standard deviation of the sampling distribution* (also known as the **standard error**) of the sample mean or proportion, it is generally necessary to ____________ the sample size.
   (a) cut in half
   (b) double
   (c) reduce by 75%
   (d) quadruple (make 4 times as large)
11. It has been reported that 30% of all new businesses fail. Assuming this value perfectly represents the situation in the country at large, the 30% is a
   (a) population
   (b) parameter
   (c) statistic
   (d) sample

12. Suppose an article in a biology journal reports a result that is significant at the 1% level. Are such results always, never, or sometimes significant at the 5% level?

    (a) always

    (b) sometimes

    (c) never

    (d) not enough information to decide

13. Suppose we find that a 95% confidence interval for the mean SAT-Verbal score for high school seniors $495 \pm 10$. This means that

    (a) the method used to construct the confidence interval, when used repeatedly, will produce an interval that contains the population mean in 95% of cases over the long haul.

    (b) 95% of high school seniors have SAT-Verbal scores that lie in this range.

    (c) 95% of random samples drawn from high school seniors would produce confidence intervals containing 495.

    (d) the margin of error of our estimate is 0.95.

14. Which of the following is a standard way to display the relationship between *two* quantitative variables?

    (a) Bar graph

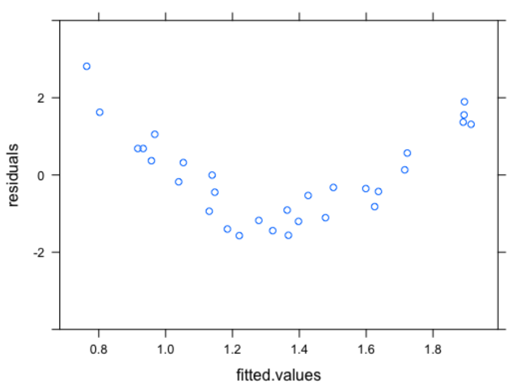
    (b) Histogram

    (c) Scatterplot

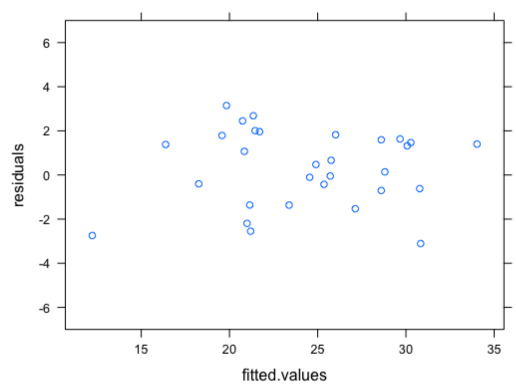
    (d) Contingency table

15. Which of the following plots of "residuals vs. fitted values" would tend to best *confirm* the assumptions of simple linear regression?

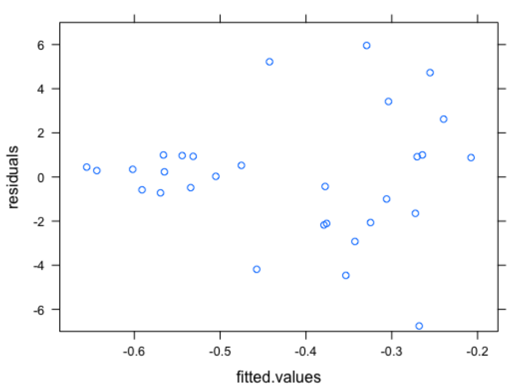
(a)



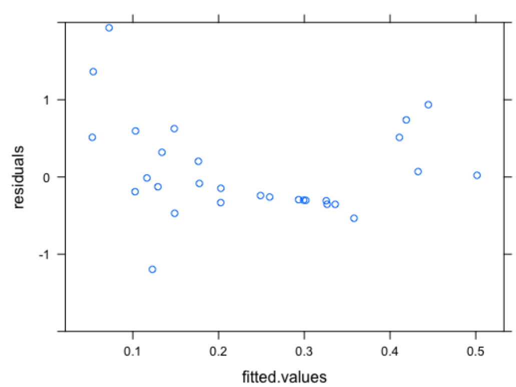

(c)




(b)




(d)



16. Experiments should use randomization to assign subjects to treatments in order to

    (a) avoid confounding the effect of a treatment with lurking variables.

    (b) make the experiment more sensitive to differences among treatments.

    (c) make subjects unaware of which treatment they are receiving.

    (d) allow the use of random number tables.

17. Suppose we find that the $P$-value associated with a test statistic for a given sample is $P = 0.033$. This means that

    (a) the probability that the null hypothesis is true is 3.3%.

    (b) the probability that the alternative hypothesis is true is 3.3%.

    (c) if the null hypothesis is true, the probability of getting a test statistic at least as extreme as ours is 3.3%.

    (d) 3.3% of all possible samples will produce this value of the test statistic.

18. Which of the following quantities is *resistant* to the presence of outliers? [Imagine the effect on each of the named quantities if, to an existing list of numbers, you add one more that is a clear outlier.]

    (a) standard deviation

    (b) IQR

    (c) Mean

    (d) Range

19. Which of the following is NOT a characteristic of all Normal distributions?

    (a) Standard deviation is 1

    (b) Bell-shaped distribution curve

    (c) Total area under the distribution curve is 1

    (d) About 95% of distribution lies with 2 standard deviations of the mean

20. Which of the following is a categorical variable?

    (a) Daily caloric intake

    (b) Height

    (c) Race

    (d) Pulse rate

21. A Type II error occurs, in the context of hypothesis testing, when

    (a) one rejects a null hypothesis that is true.

    (b) one does not reject a null hypothesis that is false.

    (c) one rejects a null hypothesis that is false.

    (d) one does not reject a null hypothesis that is false.

22. Which of the following is NOT generally true of the regression line $y = b_0 + b_1 x$?

    (a) The *observed* $y$-values all fall on this line.

    (b) The line is chosen so that the sum of squares of residuals is as small as possible.

    (c) The line should be used with caution in extrapolating beyond the region of $x$-data.

    (d) The slope of the line and the correlation coefficient have the same sign (i.e., both positive or both negative).

23. A basic difference between an observational study and an experiment is that

(a) experiments can establish associations.

(b) experiments use randomization.

(c) statistical inference is used in experiments.

(d) values of explanatory variables are imposed, not simply measured, in experiments.

## Section 2. Short Answer

24. ANOVA TABLE.

(a) *(3 pts)* Fill in the missing entries in the ANOVA table below.

| Source of Variation | df | SS | MS | F | $P$-value |
|---|---|---|---|---|---|
| Group | 4 | 98.2 | | | 0.0345 |
| Error | 18 | 134.5 | | | |

(b) *(3 pts)* How many groups are represented in this data set? How many subjects/cases?

(c) *(4 pts)* What conclusion is reached at the 5% significance level? [You should state, generically, appropriate null and alternative hypotheses as part of your answer.]

25. *(4 pts each)* WHAT DO I DO? In each of the following situations, pretend you want to know some information and you are designing a statistical study to find out about it. Give the following **three** pieces of information for each: (i) **what variables** you would need to have in your dataset, (ii) whether these variables are **categorical** or **quantitative**, and (iii) **what statistical procedure** you would use to analyze the results.

   Select your procedures from the following list: 1-proportion, 2-proportion, 1-sample $t$, 2-sample $t$, paired-$t$, chi square goodness-of-fit, chi square test for association of categorical variables, 1-way ANOVA. or model utility test.

   Any procedure may appear as your answer more than once, or be omitted entirely. If two procedures are possible, you need only give one.

   (a) You want to compare how long two different types of lightbulbs last before burning out.

   (b) You wonder whether there is any difference in success rates (defined by whether participants are still smoking one year later) between three different smoking cessation programs.

   (c) You want to see whether three different smoking cessation programs lead to different amounts of weight gain among women who successfully quit smoking using these programs.

   (d) You want to know whether Americans prefer peas over carrots.

   (e) You want to know whether the amount of time spent listening to iPods or similar devices with earphones is associated with standard hearing test scores.

| | Variable(s) | Procedure |
|---|---|---|
| (a) | | |
| (b) | | |
| (c) | | |
| (d) | | |
| (e) | | |

26. *(2 pts each)* R Commands. Suppose there is a data frame called **Cars**. The *first few lines* of this data frame appear below.

```
head(Cars)
```

```
    Make     Model  Type Drive CityMPG HwyMPG Weight     Size
  Subaru  Forester   SUV   AWD      24     32   3370    Small
Cadillac       CTS Sedan   AWD      18     26   3915 Midsized
  Toyoto   Sequioa 7Pass   RWD      12     18   6025    Large
   Scion        xB Wagon   FWD      22     28   3120    Small
    Audi        A3 Sedan   FWD      23     33   3135    Small
```

Write an R command (complete and ready to be executed) that would produce the desired information for each of the following scenarios. Do not use `favstats()` in any of your answers.

(a) Give the number of cases in the data frame.

(b) Give the mean `Weight` among vehicles in the data frame.

(c) Display the frequencies of cars of different `Type`.

(d) Produce side-by-side box plots of `CityMPG` for each value of `Drive`.

(e) Make a two-way table on variables `Drive` and `Size`.

(f) Give regression line coefficients using `Weight` as explanatory, `HwyMPG` as response.

(g) Select out the part of the data frame consisting of those cars with `CityMPG` under 20.

27. After marginal totals are added, the two-way table sought in part (e) of the last problem appears at right.

| Size | Drive AWD | FWD | RWD | Total |
|---|---|---|---|---|
| Large | 7 | 13 | 9 | 29 |
| Midsized | 9 | 20 | 5 | 34 |
| Small | 9 | 30 | 8 | 47 |
| Total | 25 | 63 | 22 | 110 |

(a) *(4 pts)* What contribution to the $\chi^2$-statistic comes from the (`RWD`, `Midsized`) cell?

(b) *(3 pts)* The output below comes from applying the `chisq.test()` command to this table. What is the null hypothesis of this test, and what conclusion should we draw?

```
X-squared = 4.1017, df=4, p-value = 0.3924
```

28. *(6 pts)* Difference of proportions. Suppose 18 of 37 males sampled indicate they have tried a banned substance; the number is 15 out of 41 females sampled. If this data is to be used to test hypotheses

$$\mathbf{H}_0:\ p_M - p_F = 0 \qquad \text{vs.} \qquad \mathbf{H}_a:\ p_M - p_F \neq 0,$$

calculate and label the *standardized test statistic*.

29. *(6 pts)* Difference of means. In a study of the effect of college student employment on academic performance, a sample of 184 students who were simultaneously employed had an average gpa of 3.12 with standard deviation 0.485. Among 114 students who were otherwise unemployed, the average gpa was 3.23 with standard deviation 0.524. Taking the subscripts $U$ and $E$ for Unemployed and Employed students respectively, find a 95% confidence interval for $\mu_U - \mu_E$, the difference in mean gpa. The critical value for the 95% level is $t^* = 1.981$.

30. Student volunteers at Ohio State University were randomly assigned a number of cans of beer to drink. Thirty minutes later a police officer measured their blood alcohol content (BAC) in grams of alcohol per deciliter of blood (g/dl). Here are data for the first 6 participants, arranged in rows (so as to fit the page better) instead of the usual columns. (So, Student 1 drank 5 beers and had a measured BAC of 0.10.)

| | 1 | 2 | 3 | 4 | 5 | 6 | ... |
|---|---|---|---|---|---|---|---|
| Beers | 5 | 2 | 9 | 8 | 3 | 7 | ... |
| BAC | 0.10 | 0.03 | 0.19 | 0.12 | 0.04 | 0.10 | ... |

Some regression-related output (the `summary()` command was used) appears below.

```
               Estimate   Std. Error   t value   PR(>|t|)
(Intercept)    -0.0127       0.0126     -1.00       0.33
Beers           0.0180       0.0024      7.48   0.000003

Residual standard error:  1.02 on 14 degrees of freedom
Multiple R-squared:  0.8
F-statistic:  55.9 on 1 and 14 DF, p-value:  0.000003
```

(a) *(2 pts)* Inspect the output carefully. How many participants/cases are in the full dataset?

(b) *(2 pts)* What portion of the variability in `BAC` values among participants in the study is *explained* by a linear model involving number of `Beers` consumed?

(c) *(3 pts)* What is the slope of the regression line? Explain, in context, what this slope value tells us.

(d) *(2 pts)* Find the *correlation* between `BAC` and `Beers`.

(e) *(2 pts)* Suggest at least one other variable not considered here that could help explain some of the variability in `BAC` *not* accounted for by our linear model with `Beers`.

(f) *(3 pts)* Why is it important that the students were randomly assigned a number of beers to drink?

(g) *(2 pts)* What is the residual for Student/Case #5?

(h) *(3 pts)* State the null and alternative hypotheses for the **model utility test**. What conclusion do we draw regarding this test?

31. In a study to test the user-interface of her latest computer program, Cindy selects 20 students at random and has them perform a series of tasks as quickly as they can.

| Command | Result |
|---|---|
| `pt(3.64, 19)` | `0.999129` |
| `pt(3.64, 18)` | `0.999064` |
| `qt(.90, 18)` | `1.330391` |
| `qt(.95, 18)` | `1.734064` |
| `qt(.90, 19)` | `1.327728` |
| `qt(.95, 19)` | `1.729133` |

(a) *(5 pts)* Suppose, among the 20 students, the mean time to complete the tasks is 43.2 seconds, with a standard deviation of 5.4 seconds. Find a 90% confidence interval for the mean time it takes to complete the series of tasks. Several `RStudio` commands (not all of which are useful to this situation) have been provided along with their results.

(b) *(4 pts)* Cindy may, as an alternate method to the one used above for finding a confidence interval, employ bootstrapping. Describe how she might physically produce a *single bootstrap sample*.

(c) *(3 Extra Credit pts)* Suppose Cindy simulates a bootstrap distribution using the command in `RStudio`

```
manyMeans = do(10000) * mean(resample(timeToCompleteTasks))
```

Give a follow-up command (or commands) which would produce left- and right-endpoints of a 90% **bootstrap percentile confidence interval**. [Note: `manyMeans` will have just one column, called `mean`.]

32. For a given 4-week period in New York City, there were 9,144 births recorded. Of these, 3,067 occurred on the weekends (Sat. or Sun.), and the rest on weekdays. If one were curious whether there is any difference on the rate of births (as measured in births per day) on weekends from that during weekdays, perhaps this data can help settle the question, at least in regards to the city of New York.

    (a) *(3 pts)* Indicate by name a type of test that is tailored to this situation. Does it seem such a test is appropriate here? Explain briefly.

    (b) *(3 pts)* State null and alternative hypotheses suited to the test you named in the previous part..

    (c) *(5 pts)* Calculate a test statistic appropriate for the test you named in the part (a).

    (d) *(2 pts)* Write an `RStudio` command that uses the test statistic of the previous part and will produce the corresponding $P$-value.

# STAT 145 Final Exam

Spring, 2021
Calvin University

Name:______________________

You must **justify your answers** with appropriate work to receive full credit. Even for calculations done on a calculator, you are responsible for writing in sufficient detail on the exam sheet what calculation is being performed.

It has been my sincere pleasure to work with you this semester. Good luck on the exam, and may the summer break be productive, restful, joyful and enriching as befits your life in Christ.

# Answer Key for Exam $\boxed{\text{B}}$

Section 1. Multiple Choice *(2 pts each)*: Select the (one) best response.

1. Which of the following is NOT a general guideline for the creation of randomization samples for some statistical test?

   (a) Use the data in the original sample.

   (b) Reflect the way the original data were collected.

   $\boxed{\text{(c)}}$ Make it so the distribution of randomization statistics is centered at zero.

   (d) Be consistent with the null hypothesis.

2. When the technician taking the response measurements in an experiment is unaware of the treatments assigned to individual subjects/units, this is done with the intention to

   $\boxed{\text{(a)}}$ reduce bias.

   (b) reduce the amount of variability in the response variable.

   (c) increase randomness.

   (d) none of the above.

3. If the 95% confidence interval for a particular population mean $\mu$ is $3.1 \pm 3.85$, what can you accurately say about the $P$-value for an hypothesis test with $\mathbf{H}_0\colon\ \mu = 0,\ \mathbf{H}_a\colon\ \mu \neq 0$?

   (a) The $P$-value is larger than 0.1.

   $\boxed{\text{(b)}}$ The $P$-value is larger than 0.05.

   (c) The $P$-value is less than 0.05.

   (d) It is impossible to tell which of these three is correct from this information.

4. There is the potential for students to be confused over the purpose of one-way ANOVA tests. If, say, we have 5 different groups, each with its own mean, we might systematically choose two groups at a time (Groups A and B together, then Groups A and C, . . . , Groups D and E, resulting in ten such pairings in all), and use 2-sample $t$ tests to tell whether the evidence against equality of sample means is significant at the 5% level or not. The reason for **not** doing this without first having obtained a significant result from 1-way ANOVA is

   (a) because it would be tedious to carry out ten different 2-sample $t$ tests.

   $\boxed{\text{(b)}}$ because, even if all population means are equal, the chance of committing a Type I error at least once in these ten 2-sample $t$ tests is far greater than 5%.

   (c) because, for 2-sample $t$, the categorical variable "Which group?" cannot serve as the explanatory variable.

   (d) because one should always use a *matched-pairs* $t$ test, not a 2-sample $t$ test, for such pairwise comparisons.

5. A poll of 1800 American residents age 18 or over asked if they approved of the Predident's job performance. Of these 1800, 41% said "Yes", 52% said "No", and 7% were undecided. In this situation, the 52% is a

   (a) population

   (b) parameter

   $\boxed{\text{(c)}}$ statistic

   (d) sample

6. Which of the following is NOT generally a characteristic of the correlation coefficient $r$?

   (a) $r$ is between $(-1)$ and 1

   (b) $r$ equals 1 or $(-1)$ only for perfect straight-line relationships.

   $\boxed{\text{(c)}}$ $r$ has units (such as *meters*, *minutes*, etc.) that match those of the response variable.

   (d) $r$ requires two variables which are quantitative.

7. As sample size increases, which of the following is NOT true?

   (a) The standard error of the sample mean becomes smaller.

   (b) The sampling distribution of the sample mean becomes increasingly *normal*.

   (c) The width of corresponding 95% confidence intervals decreases.

   $\boxed{\text{(d)}}$ The chance of rejecting a true null hypothesis at the 5% level decreases.

8. Which of the following divides a density curve (or histogram) of a skewed distribution into two pieces enclosing equal areas?

   (a) mean

   $\boxed{\text{(b)}}$ median

   (c) standard deviation

   (d) interquartile range

9. At right I have depicted a contingency table involving the variables `Rh-factor` and `blood type`. This problem has nothing to do with blood, nor with the data in this table. Its presence is simply to remind you of what a contingency table does/is.

   Rh-factor

   | Type | Neg. | Pos. |
   |---|---|---|
   | A | 32 | 161 |
   | B | 9 | 44 |
   | AB | 4 | 13 |
   | O | 29 | 181 |

   Suppose you have a contingency table with exactly 4 cells, two in the first row and two in the second. One statistical procedure you might use in such a setting is a chi-square test for association between categorical variables; but another equally-useful procedure that applies in such settings is

   (a) a model utility test.

   (b) a chi-square goodness-of-fit test.

   (c) a 2-sample $t$ test.

   $\boxed{\text{(d)}}$ a 2-proportion test.

10. To cut in half the *standard deviation of the sampling distribution* (also known as the **standard error**) of the sample mean or proportion, it is generally necessary to ______________ the sample size.

   (a) cut in half

   (b) double

   (c) reduce by 75%

   $\boxed{\text{(d)}}$ quadruple (make 4 times as large)

11. It has been reported that 30% of all new businesses fail. Assuming this value perfectly represents the situation in the country at large, the 30% is a

   (a) population

   $\boxed{\text{(b)}}$ parameter

   (c) statistic

   (d) sample

12. Suppose an article in a biology journal reports a result that is significant at the 1% level. Are such results always, never, or sometimes significant at the 5% level?

    $\boxed{\text{(a)}}$ always
    (b) sometimes
    (c) never
    (d) not enough information to decide

13. Suppose we find that a 95% confidence interval for the mean SAT-Verbal score for high school seniors $495 \pm 10$. This means that

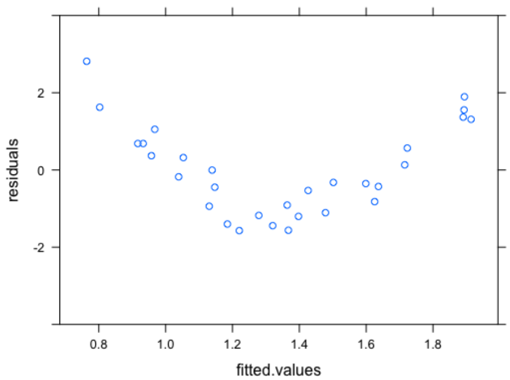
    $\boxed{\text{(a)}}$ the method used to construct the confidence interval, when used repeatedly, will produce an interval that contains the population mean in 95% of cases over the long haul.
    (b) 95% of high school seniors have SAT-Verbal scores that lie in this range.
    (c) 95% of random samples drawn from high school seniors would produce confidence intervals containing 495.
    (d) the margin of error of our estimate is 0.95.

14. Which of the following is a standard way to display the relationship between *two* quantitative variables?

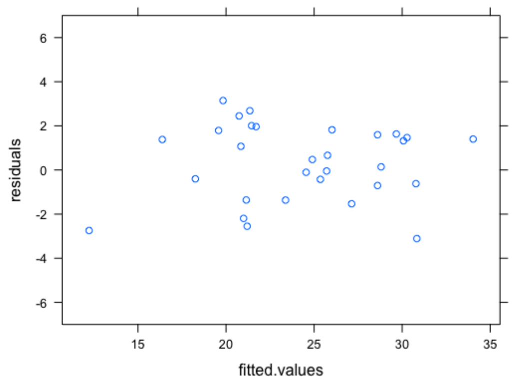
    (a) Bar graph
    (b) Histogram
    $\boxed{\text{(c)}}$ Scatterplot
    (d) Contingency table

15. Which of the following plots of "residuals vs. fitted values" would tend to best *confirm* the assumptions of simple linear regression?

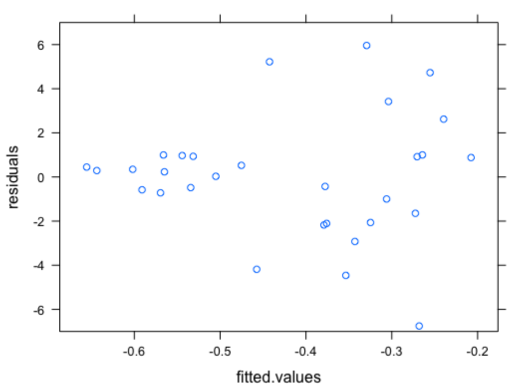
(a)



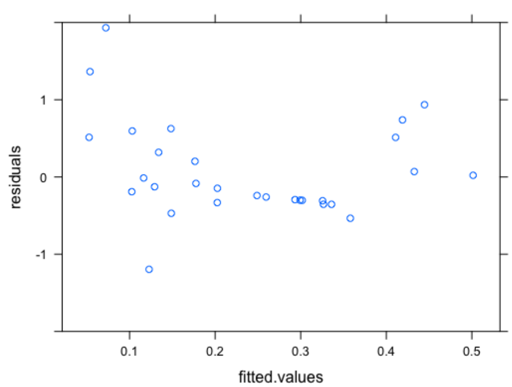

(c)




(b)




(d)



16. Experiments should use randomization to assign subjects to treatments in order to

    $\boxed{\text{(a)}}$ avoid confounding the effect of a treatment with lurking variables.

    (b) make the experiment more sensitive to differences among treatments.

    (c) make subjects unaware of which treatment they are receiving.

    (d) allow the use of random number tables.

17. Suppose we find that the $P$-value associated with a test statistic for a given sample is $P = 0.033$. This means that

    (a) the probability that the null hypothesis is true is 3.3%.

    (b) the probability that the alternative hypothesis is true is 3.3%.

    $\boxed{\text{(c)}}$ if the null hypothesis is true, the probability of getting a test statistic at least as extreme as ours is 3.3%.

    (d) 3.3% of all possible samples will produce this value of the test statistic.

18. Which of the following quantities is *resistant* to the presence of outliers? [Imagine the effect on each of the named quantities if, to an existing list of numbers, you add one more that is a clear outlier.]

    (a) standard deviation

    $\boxed{\text{(b)}}$ IQR

    (c) Mean

    (d) Range

19. Which of the following is NOT a characteristic of all Normal distributions?

    $\boxed{\text{(a)}}$ Standard deviation is 1

    (b) Bell-shaped distribution curve

    (c) Total area under the distribution curve is 1

    (d) About 95% of distribution lies with 2 standard deviations of the mean

20. Which of the following is a categorical variable?

    (a) Daily caloric intake

    (b) Height

    $\boxed{\text{(c)}}$ Race

    (d) Pulse rate

21. A Type II error occurs, in the context of hypothesis testing, when

    (a) one rejects a null hypothesis that is true.

    $\boxed{\text{(b)}}$ one does not reject a null hypothesis that is false.

    (c) one rejects a null hypothesis that is false.

    (d) one does not reject a null hypothesis that is false.

22. Which of the following is NOT generally true of the regression line $y = b_0 + b_1x$?

    $\boxed{\text{(a)}}$ The *observed* $y$-values all fall on this line.

    (b) The line is chosen so that the sum of squares of residuals is as small as possible.

    (c) The line should be used with caution in extrapolating beyond the region of $x$-data.

    (d) The slope of the line and the correlation coefficient have the same sign (i.e., both positive or both negative).

23. A basic difference between an observational study and an experiment is that

(a) experiments can establish associations.

(b) experiments use randomization.

(c) statistical inference is used in experiments.

$\boxed{\text{(d)}}$ values of explanatory variables are imposed, not simply measured, in experiments.

## Section 2. Short Answer

24. ANOVA TABLE.

(a) *(3 pts)* Fill in the missing entries in the ANOVA table below.

| Source of Variation | df | SS | MS | F | $P$-value |
|---|---|---|---|---|---|
| Group | 4 | 98.2 | | | 0.0345 |
| Error | 18 | 134.5 | | | |

(b) *(3 pts)* How many groups are represented in this data set? How many subjects/cases?

(c) *(4 pts)* What conclusion is reached at the 5% significance level? [You should state, generically, appropriate null and alternative hypotheses as part of your answer.]

**Answer:** (a)

| Source of Variation | df | SS | MS | F | $P$-value |
|---|---|---|---|---|---|
| Group | 4 | 98.2 | 24.55 | 3.286 | 0.0345 |
| Error | 18 | 134.5 | 7.47 | | |
| Total | 22 | 232.7 | | | |

(b) There are $4 + 1 = 5$ groups and $22 + 1 = 23$ cases.

(c) For ANOVA, we have hypotheses

$$\mathbf{H}_0: \ \mu_1 = \mu_2 = \mu_3 = \mu_4,$$

and alternative hypotheses that at least one of these population means is different from others. With $P = 0.018$, we reject this null hypothesis at the 5% significance level, concluding that not all of the group means are identical.

25. *(4 pts each)* WHAT DO I DO? In each of the following situations, pretend you want to know some information and you are designing a statistical study to find out about it. Give the following **three** pieces of information for each: (i) **what variables** you would need to have in your dataset, (ii) whether these variables are **categorical** or **quantitative**, and (iii) **what statistical procedure** you would use to analyze the results.

Select your procedures from the following list: 1-proportion, 2-proportion, 1-sample $t$, 2-sample $t$, paired-$t$, chi square goodness-of-fit, chi square test for association of categorical variables, 1-way ANOVA. or model utility test.

Any procedure may appear as your answer more than once, or be omitted entirely. If two procedures are possible, you need only give one.

(a) You want to compare how long two different types of lightbulbs last before burning out.

(b) You wonder whether there is any difference in success rates (defined by whether participants are still smoking one year later) between three different smoking cessation programs.

(c) You want to see whether three different smoking cessation programs lead to different amounts of weight gain among women who successfully quit smoking using these programs.

(d) You want to know whether Americans prefer peas over carrots.

(e) You want to know whether the amount of time spent listening to iPods or similar devices with earphones is associated with standard hearing test scores.

| | **Variable(s)** | **Procedure** |
|---|---|---|
| (a) | | |
| (b) | | |
| (c) | | |
| (d) | | |
| (e) | | |

**Answer:**

| | Variable(s) | Procedure |
|---|---|---|
| (a) | light bulb type (categorical)<br>time until burns out (quant) | 2-sample *t*<br>or 1-way ANOVA |
| (b) | cessation program (categorical)<br>quit smoking? (categorica) | chi-square test for association<br>or Fisher test |
| (c) | cessation program (categorical)<br>weight gain (quantitative) | 1-way ANOVA<br>or Kruskal-Wallis test |
| (d) | preferred veggie: carrots or peas (categorical) | 1-proportion |
| | | or binomial test |
| (e) | hearing test score (quantitative)<br>time listening to i-Pod (quantitative) | simple linear regression |

26. *(2 pts each)* R COMMANDS. Suppose there is a data frame called **Cars**. The *first few lines* of this data frame appear below.

```
head(Cars)
```

| Make | Model | Type | Drive | CityMPG | HwyMPG | Weight | Size |
|---|---|---|---|---|---|---|---|
| Subaru | Forester | SUV | AWD | 24 | 32 | 3370 | Small |
| Cadillac | CTS | Sedan | AWD | 18 | 26 | 3915 | Midsized |
| Toyoto | Sequioa | 7Pass | RWD | 12 | 18 | 6025 | Large |
| Scion | xB | Wagon | FWD | 22 | 28 | 3120 | Small |
| Audi | A3 | Sedan | FWD | 23 | 33 | 3135 | Small |

Write an R command (complete and ready to be executed) that would produce the desired information for each of the following scenarios. Do not use `favstats()` in any of your answers.

(a) Give the number of cases in the data frame.

(b) Give the mean `Weight` among vehicles in the data frame.

(c) Display the frequencies of cars of different `Type`.

(d) Produce side-by-side box plots of `CityMPG` for each value of `Drive`.

(e) Make a two-way table on variables `Drive` and `Size`.

(f) Give regression line coefficients using `Weight` as explanatory, `HwyMPG` as response.

(g) Select out the part of the data frame consisting of those cars with `CityMPG` under 20.

27. After marginal totals are added, the two-way table sought in part (e) of the last problem appears at right.

| | Drive | | | |
|---|---|---|---|---|
| Size | AWD | FWD | RWD | Total |
| Large | 7 | 13 | 9 | 29 |
| Midsized | 9 | 20 | 5 | 34 |
| Small | 9 | 30 | 8 | 47 |
| Total | 25 | 63 | 22 | 110 |

(a) *(4 pts)* What contribution to the $\chi^2$-statistic comes from the (`RWD`, `Midsized`) cell?

(b) *(3 pts)* The output below comes from applying the `chisq.test()` command to this table. What is the null hypothesis of this test, and what conclusion should we draw?

```
X-squared = 4.1017, df=4, p-value = 0.3924
```

28. *(6 pts)* Difference of proportions. Suppose 18 of 37 males sampled indicate they have tried a banned substance; the number is 15 out of 41 females sampled. If this data is to be used to test hypotheses

$$\mathbf{H}_0:\ p_M - p_F = 0 \qquad \text{vs.} \qquad \mathbf{H}_a:\ p_M - p_F \neq 0,$$

calculate and label the *standardized test statistic*.

29. *(6 pts)* Difference of means. In a study of the effect of college student employment on academic performance, a sample of 184 students who were simultaneously employed had an average gpa of 3.12 with standard deviation 0.485. Among 114 students who were otherwise unemployed, the average gpa was 3.23 with standard deviation 0.524. Taking the subscripts $U$ and $E$ for Unemployed and Employed students respectively, find a 95% confidence interval for $\mu_U - \mu_E$, the difference in mean gpa. The critical value for the 95% level is $t^* = 1.981$.

30. Student volunteers at Ohio State University were randomly assigned a number of cans of beer to drink. Thirty minutes later a police officer measured their blood alcohol content (BAC) in grams of alcohol per deciliter of blood (g/dl). Here are data for the first 6 participants, arranged in rows (so as to fit the page better) instead of the usual columns. (So, Student 1 drank 5 beers and had a measured BAC of 0.10.)

| | 1 | 2 | 3 | 4 | 5 | 6 | ... |
|---|---|---|---|---|---|---|---|
| Beers | 5 | 2 | 9 | 8 | 3 | 7 | ... |
| BAC | 0.10 | 0.03 | 0.19 | 0.12 | 0.04 | 0.10 | ... |

Some regression-related output (the `summary()` command was used) appears below.

```
              Estimate   Std. Error   t value   PR(>|t|)
(Intercept)    -0.0127       0.0126     -1.00       0.33
Beers           0.0180       0.0024      7.48   0.000003

Residual standard error:  1.02 on 14 degrees of freedom
Multiple R-squared:  0.8
F-statistic:  55.9 on 1 and 14 DF, p-value:  0.000003
```

(a) *(2 pts)* Inspect the output carefully. How many participants/cases are in the full dataset?

(b) *(2 pts)* What portion of the variability in `BAC` values among participants in the study is *explained* by a linear model involving number of `Beers` consumed?

(c) *(3 pts)* What is the slope of the regression line? Explain, in context, what this slope value tells us.

(d) *(2 pts)* Find the *correlation* between `BAC` and `Beers`.

(e) *(2 pts)* Suggest at least one other variable not considered here that could help explain some of the variability in `BAC` *not* accounted for by our linear model with `Beers`.

(f) *(3 pts)* Why is it important that the students were randomly assigned a number of beers to drink?

(g) *(2 pts)* What is the residual for Student/Case #5?

(h) *(3 pts)* State the null and alternative hypotheses for the **model utility test**. What conclusion do we draw regarding this test?

**Answer:** (a) To read in the data, we would execute the command

`beerStudy = read.csv("http://www.calvin.edu/ scofield/beersAndBAC.csv")`

(b) `summary(lm(BAC ~ Beers, data=beerStudy))`
`xyplot(lm(BAC ~ Beers, data=beerStudy, type=c("p","r")))`

(c) There are $14 + 2 = 16$ participants/cases.

(d) The regression line has slope $= 0.018$. This means that for each extra drink consumed, the blood alcohol level 30 minutes hence rises approximately 0.018.

(e) It is important in order to reduce bias: we provide no opportunity for participants or experimenters to *choose* which subjects drink more and which drink less.

(f) Approximately 80% of the variability in `BAC` levels is explained by this model with Beers.

(g) Perhaps a participant's `weight` could be such a variable.

(h) The correlation is $r = \sqrt{.8} \doteq 0.894$.

(i) We have

$$\text{residual} \;=\; \text{observed} - \text{predicted} \;=\; 0.04 - [-0.0127 + (0.018)(3)] \;=\; -0.0013.$$

(j) The hypotheses are

$$\begin{aligned} \mathbf{H}_0: \;& \beta = 0 \quad \text{(or the number of beers has no linear association with BAC)} \\ \mathbf{H}_a: \;& \beta \neq 0 \quad \text{(or there is a linear association).} \end{aligned}$$

With a $P$-value of $3 * 10^{-6}$, we reject $\mathbf{H}_0$.

31. In a study to test the user-interface of her latest computer program, Cindy selects 20 students at random and has them perform a series of tasks as quickly as they can.

| Command | Result |
|---|---|
| `pt(3.64, 19)` | `0.999129` |
| `pt(3.64, 18)` | `0.999064` |
| `qt(.90, 18)` | `1.330391` |
| `qt(.95, 18)` | `1.734064` |
| `qt(.90, 19)` | `1.327728` |
| `qt(.95, 19)` | `1.729133` |

(a) *(5 pts)* Suppose, among the 20 students, the mean time to complete the tasks is 43.2 seconds, with a standard deviation of 5.4 seconds. Find a 90% confidence interval for the mean time it takes to complete the series of tasks. Several `RStudio` commands (not all of which are useful to this situation) have been provided along with their results.

(b) *(4 pts)* Cindy may, as an alternate method to the one used above for finding a confidence interval, employ bootstrapping. Describe how she might physically produce a *single bootstrap sample*.

(c) *(3 Extra Credit pts)* Suppose Cindy simulates a bootstrap distribution using the command in `RStudio`

```
manyMeans = do(10000) * mean(resample(timeToCompleteTasks))
```

Give a follow-up command (or commands) which would produce left- and right-endpoints of a 90% **bootstrap percentile confidence interval**. [Note: `manyMeans` will have just one column, called `mean`.]

**Answer:** (a) Here, the 90% CI is

$$43.2 \pm (1.729)\frac{5.4}{\sqrt{20}} \doteq 43.2 \pm (1.729)\frac{5.4}{4.4721} \doteq 43.2 \pm 2.0877,$$

or $[41.112, 45.288]$.

(b) She could write each participant's time on a single card. Shuffling them up, she could draw one card and record the time on that card. She would then replace the card in the deck, shuffling them again and drawing another. She would repeat this process until she had drown (with replacment) 20 cards. The twenty numbers now recorded would then be used to calculate one single mean.

(c) `quantile(manyMeans$result, c(0.05, 0.95))`

32. For a given 4-week period in New York City, there were 9,144 births recorded. Of these, 3,067 occurred on the weekends (Sat. or Sun.), and the rest on weekdays. If one were curious whether there is any difference on the rate of births (as measured in births per day) on weekends from that during weekdays, perhaps this data can help settle the question, at least in regards to the city of New York.

(a) *(3 pts)* Indicate by name a type of test that is tailored to this situation. Does it seem such a test is appropriate here? Explain briefly.

(b) *(3 pts)* State null and alternative hypotheses suited to the test you named in the previous part..

(c) *(5 pts)* Calculate a test statistic appropriate for the test you named in the part (a).

(d) *(2 pts)* Write an `RStudio` command that uses the test statistic of the previous part and will produce the corresponding $P$-value.

**Answer:** (a) Either a chi-square goodness of fit can be used (appropriate since the expected counts—namely 2612.57 and 6531.43—are so high), or a 1-proportion test, which also is appropriate since $np$, $n(1-p)$ are both much higher than 10.

(b) Either a chi-square goodness of fit can be used (in which case the hypotheses should be

$$\mathbf{H}_0:\ p_{\text{weekday}} = \frac{5}{7} \text{ and } p_{\text{weekend}} = \frac{2}{7} \qquad \text{vs.} \qquad \mathbf{H}_a:\ \text{neither } p_{\text{weekend}} \text{ nor } p_{\text{weekday}} \text{ are as proposed in } \mathbf{H_0}.$$

One can also give, as answer, that the procedure is 1-proportion. with $p = 2/7$.

(c) For the goodness-of-fit test, the test statistic is

$$\sum_j \frac{(O_j - E_j)^2}{E_j} \doteq 110.66.$$

For the 1-proportion test it is

$$Z = \frac{(3067/9144) - (2/7)}{\sqrt{\frac{(2/7)(1-2/7)}{9144}}} \doteq \frac{0.3354 - 0.2857}{0.004724} \doteq 10.06.$$

(d) For the goodness-of-fit test, the `RStudio` command is `1 - pchisq(110.66, 1)`
For the 1-proportion test, the `RStudio` command is `2 * (1 - pnorm(10.06))`.