Review for Statistics Final Exam

1) Based on a random sample of 500 registered voters, a researcher estimates that 47% of all registered voters favor stronger restrictions on smoking in public places. State whether this is an example of descriptive or inferential statistics.

2) A meteorologist calculates the average daily precipitation in Phoenix, Arizona during the year 1998. State whether this is an example of descriptive or inferential statistics.

3) At one hospital in 1992, 674 women were diagnosed with breast cancer. Five years later, 88% of the Caucasian women and 83% of the African American women were still alive. Is this an example of an observational study or a designed experiment?

4) Is a person’s political affiliation an example of a qualitative or quantitative variable?

5) Is the speed of a car in miles per hour an example of a qualitative or quantitative variable?

6) The number of cars passing a busy intersection between 4:30 P.M. and 6:30 P.M. on a Monday is 2200. Is this data discrete or continuous?

7) A study discusses the height of mountains in a certain state. Is this an example of discrete data or continuous data?

8) The following data represent the number of cars passing through a toll booth during a certain time period over a number of days. Construct a dot plot for the following data:

   24  25  23  30  24  27  24  25  21  28  25  29  23  27

9) The following data show the number of laps run by each participant in a marathon. Construct an ordered stem-and-leaf diagram for the following given data:

   46  65  55  43  41  48  57  30  43  42  63  32  56  38  48

10) The highest temperatures ever recorded (in °F) in 32 different U.S. states are shown below. Find the five number summary and construct a boxplot for the given data set.

   100  100  105  105  106  106
   107  107  109  110  110  112
   112  112  114  114  114  115
   116  117  118  118  118  118
   118  119  120  121  122  125
   128  134

11) Given the following sample data, find the mean, median, mode, range, standard deviation and variance.

   29  23  47  43  27  43  28  33  40  29  54  36

12) The paired data below consist of the test scores of 6 randomly selected students and the number of hours they studied for the test. Obtain the linear correlation coefficient for the given data:

   Hours  5  10  4  6  10  9
   Score  64  86  69  86  59  87

13) The following are costs of advertising (in thousands of dollars) and the numbers of products sold (in thousands). Find the coefficient of determination.

   Cost  9  2  3  4  2  5  9  10
   Number 85  52  55  68  67  86  83  73
14) Use the following data to find the equation of the regression line. Round the final values to three decimal places.

<table>
<thead>
<tr>
<th>x</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>8</td>
<td>11</td>
<td>7</td>
<td>14</td>
<td>20</td>
</tr>
</tbody>
</table>

15) In one city, the probability that a person will pass his or her driving test on the first attempt is 0.62. There are 11 people selected at random from among those taking their driving test for the first time. What is the probability that among these 11 people,
(a) exactly 7 will pass their test?
(b) more than 7 will pass their test?
(c) at most 6 will pass their test?

16) Find the z-score for which
(a) the area under the standard normal curve to its left is 0.04.
(b) the area under the standard normal curve to its right is 0.65.

17) The average runner at a local college runs the mile in 4.5 minutes, with a standard deviation of 0.3 minutes. What is the probability that a person will run a mile in less than 4 minutes? Round to the nearest whole percent.

18) The mean monthly income of trainees at a local mill is $1100 with a standard deviation of $150. Find the probability that the average trainee from a group of 20 earns less than $1050 a month. Round to the nearest tenth of a percent.

19) A random sample of 105 light bulbs had a mean life of 591 hours. Assume that $\sigma = 25$ hours. Construct a 90% confidence interval for the mean life of all light bulbs of this type.

20) The amounts (in ounces) of juice in eight randomly selected juice bottles are:

15.9 15.6 15.4 15.1
15.9 15.5 15.8 15.2

Construct a 98% confidence interval for the mean amount of juice in all such bottles.

21) In the past, the mean running time for a certain type of flashlight battery has been 8.3 hours. The manufacturer has introduced a change in the production method and wants to perform a hypothesis test to determine whether the mean running time has increased as a result. The hypotheses are:

$H_0: \mu = 8.3$ hours
$H_a: \mu > 8.3$ hours

Select the meaning of a Type I error from the following statements:

A) Failing to reject the hypothesis that $\mu = 8.3$ hours when in fact $\mu > 8.3$ hours.
B) Concluding that $\mu > 8.3$ hours when in fact $\mu > 8.3$ hours.
C) Concluding that $\mu > 8.3$ hours when in fact $\mu = 8.3$ hours.
D) Failing to reject the hypothesis that $\mu = 8.3$ hours when in fact $\mu = 8.3$ hours.

Work the following hypothesis tests. Give both hypotheses, the test statistic and its name, the p-value, and the decision and conclusion. Also give the name of the hypothesis test as used by the calculator.

22) An article in a journal reports that 34% of American fathers take no responsibility for child care. A researcher claims that the figure is higher for fathers in the town of Littleton. A random sample of 239 fathers from Littleton yielded 96 who did not help with child care. Do the data provide sufficient evidence to conclude that in Littleton the proportion is higher than 0.347? Use a 0.05 significance level.
23) Applicants for a particular job, which involves extensive travel in Spanish speaking countries, must take a proficiency test in Spanish. The sample data below were obtained in a study of the relationship between the numbers of years applicants have studied Spanish and their score on the test.

<table>
<thead>
<tr>
<th>Number of years (x)</th>
<th>Score (y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>78</td>
</tr>
<tr>
<td>4</td>
<td>72</td>
</tr>
<tr>
<td>2</td>
<td>58</td>
</tr>
<tr>
<td>5</td>
<td>89</td>
</tr>
<tr>
<td>3</td>
<td>63</td>
</tr>
<tr>
<td>4</td>
<td>73</td>
</tr>
<tr>
<td>5</td>
<td>84</td>
</tr>
<tr>
<td>3</td>
<td>75</td>
</tr>
<tr>
<td>2</td>
<td>48</td>
</tr>
</tbody>
</table>

The equation of the regression line is \( \hat{y} = 31.55 + 10.90x \). The standard error of the estimate is approximately 5.651. At the 5% level of significance, do the data provide sufficient evidence to conclude that the number of years of study is useful as a predictor of score on the test?

24) The sample data below are the typing speeds (in words per minute) and reading speeds (in words per minute) of nine randomly selected secretaries.

<table>
<thead>
<tr>
<th>Typing Speed (x)</th>
<th>Reading Speed (y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>370</td>
</tr>
<tr>
<td>56</td>
<td>551</td>
</tr>
<tr>
<td>52</td>
<td>528</td>
</tr>
<tr>
<td>63</td>
<td>348</td>
</tr>
<tr>
<td>70</td>
<td>645</td>
</tr>
<tr>
<td>58</td>
<td>454</td>
</tr>
<tr>
<td>44</td>
<td>503</td>
</tr>
<tr>
<td>79</td>
<td>618</td>
</tr>
<tr>
<td>62</td>
<td>500</td>
</tr>
</tbody>
</table>

The sample linear coefficient is \( r = 0.352 \). At the 10% significance level, do the data provide sufficient evidence to conclude that typing speed and reading speed are linearly correlated so that the correlation differs from 0?

25) A researcher was interested in comparing the resting pulse rate of people who exercise regularly and people who do not exercise regularly. Independent random samples of 16 people aged 30 – 40 who do not exercise regularly (sample 1) and 12 people aged 30 – 40 who do exercise regularly (sample 2) were selected and the resting pulse rate of each person was measured. The summary statistics are as follows:

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>Sample 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 = 73.3 )</td>
<td>( x_2 = 68.4 )</td>
</tr>
<tr>
<td>( s_1 = 10.0 )</td>
<td>( s_2 = 8.1 )</td>
</tr>
<tr>
<td>( n_1 = 16 )</td>
<td>( n_2 = 12 )</td>
</tr>
</tbody>
</table>

Do the data provide sufficient evidence to conclude that the mean resting pulse rate of people who do not exercise regularly is greater than the mean resting pulse rate of people who exercise regularly? Perform a hypothesis test at the 2.5% level of significance.
26) A consumer magazine wants to compare the lifetimes of ballpoint pens of three different types. The magazine takes a random sample of pens of each type in the following table.

<table>
<thead>
<tr>
<th>Brand 1</th>
<th>Brand 2</th>
<th>Brand 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>260</td>
<td>181</td>
<td>238</td>
</tr>
<tr>
<td>218</td>
<td>240</td>
<td>257</td>
</tr>
<tr>
<td>184</td>
<td>162</td>
<td>241</td>
</tr>
<tr>
<td>219</td>
<td>218</td>
<td>213</td>
</tr>
</tbody>
</table>

Do the data indicate that there is a difference in mean lifetime for the three brands of ballpoint pens? Use $\alpha = 0.01$.

27) In 1990, workplace accidents were distributed on workdays as follows.

<table>
<thead>
<tr>
<th>Day</th>
<th>Mon</th>
<th>Tues</th>
<th>Wed</th>
<th>Thurs</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage</td>
<td>25%</td>
<td>15%</td>
<td>15%</td>
<td>15%</td>
<td>30%</td>
</tr>
</tbody>
</table>

In 1995, a random sample of 200 workplace accidents yielded the following data.

<table>
<thead>
<tr>
<th>Day</th>
<th>Mon</th>
<th>Tues</th>
<th>Wed</th>
<th>Thurs</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accidents</td>
<td>54</td>
<td>32</td>
<td>34</td>
<td>24</td>
<td>56</td>
</tr>
</tbody>
</table>

Do the data provide sufficient evidence to conclude that the distribution of workplace accidents in 1995 differs from the 1990 distribution? Perform the hypothesis test at the 0.01 level of significance.

28) A researcher finds that of 1,000 people who said that they attend a religious service at least once a week, 31 stopped to help a person with car trouble. Of 1,200 people interviewed who had not attended a religious service at least once a month, 22 stopped to help a person with car trouble. At the 0.05 significance level, do the data provide sufficient evidence to conclude that the two proportions are different?

29) A manufacturer makes steel rods that are supposed to have a mean length of 50 cm. A retailer suspects that the bars are running long. A sample of 47 bars is taken and their mean length is determined to be 48 cm. Using a 1% level of significance, perform a hypothesis test to determine whether the population mean is less than 50 cm. Assume that the population standard deviation is 3.6 cm.

30) A public bus company official claims that the mean waiting time for bus number 14 during peak hours is less than 10 minutes. Karen took bus number 14 during peak hours on 18 different occasions. Her mean waiting time was 7.9 minutes with a standard deviation of 1.5 minutes. At the 0.01 level of significance, test the claim that the mean is less than 10 minutes.

31) A car insurance company performed a study to determine whether an association exists between age and the frequency of car accidents. They obtained the following sample data.

<table>
<thead>
<tr>
<th>AGE GROUP</th>
<th>Number of accidents in past 3 yrs</th>
<th>Under 25</th>
<th>25 - 45</th>
<th>Over 45</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>74</td>
<td>90</td>
<td>84</td>
<td>248</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>19</td>
<td>8</td>
<td>12</td>
<td>39</td>
</tr>
<tr>
<td>more than 1</td>
<td></td>
<td>7</td>
<td>2</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>300</td>
</tr>
</tbody>
</table>

At the 1% significance level, do the data provide sufficient evidence to conclude that age and frequency of car accidents are independent?
32) The table below shows the weights of seven subjects before and after following a particular diet for two months.

<table>
<thead>
<tr>
<th>Subject</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>191</td>
<td>185</td>
<td>165</td>
<td>191</td>
<td>193</td>
<td>176</td>
<td>161</td>
</tr>
<tr>
<td>After</td>
<td>184</td>
<td>176</td>
<td>163</td>
<td>196</td>
<td>179</td>
<td>178</td>
<td>149</td>
</tr>
</tbody>
</table>

Do the data suggest that the diet is effective in reducing weight? Perform a hypothesis test at the 1% level of significance.

33) A real estate agent compares the selling prices of homes in two suburbs of Chicago to see whether there is a difference in price. The results of the study are shown here. Is there evidence to say there is a difference in the cost of a home in the two locations?

<table>
<thead>
<tr>
<th>Suburb 1</th>
<th>Suburb 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\bar{x} = 63,255)</td>
<td>(\bar{x} = 59,102)</td>
</tr>
<tr>
<td>(\sigma = 5602)</td>
<td>(\sigma = 4731)</td>
</tr>
<tr>
<td>(n = 35)</td>
<td>(n = 40)</td>
</tr>
</tbody>
</table>

34) A certain brand of apple juice needs to keep a mean content of 64.05 ounces in each bottle of juice. However, the filling machine may not be precise and we need to test and see if the assembly line is creating bottles with less than 64.05 ounces. Following are the amounts found in 32 bottles of apple juice. If you know that \(\sigma = \) 0.06 ounce, use a 0.01 level of significance to test the claim that the mean amount of apple juice in this brand is less than 64.05 ounces.

| 64.05 | 64.05 | 64.00 | 63.97 | 63.95 | 64.06 |
| 64.01 | 64.07 | 64.06 | 63.90 | 63.98 | 65.07 |
| 64.06 | 64.01 | 64.01 | 63.95 | 64.05 | 64.06 |
| 63.90 | 64.03 | 64.03 | 63.94 | 64.06 | 64.03 |
| 64.05 | 64.02 | 64.06 | 63.90 | 63.90 | 63.90 |
| 64.01 | 64.04 | | | | |

35) What sample size would be required to generate a 95% confidence interval for the community's mean annual real estate tax payment within $120? Assume the population standard deviation is $535.
Answers to Review for Final Exam

1) Inferential
2) Descriptive
3) Observational Study
4) Qualitative
5) Quantitative
6) Discrete
7) Continuous
8) [Bar chart of number of cars]

9) 3 | 0 2 8
    4 | 1 2 3 3 6 8 8
    5 | 5 6 7
    6 | 3 5

10) [Box plot with five number summary]

   Five Number Summary:
   Min Q1 Median Q3 Max
   100 108 114 118 134

11) mean = 36
    median = 34.5
    mode = 29, 43
    range = 31
    standard deviation = 9.44
    variance = 89.09

12) 0.224
13) 0.5009
14) \( y' = 5.074 + 0.753x \)
15) (a) 0.2423
    (b) 0.3455
    (c) 0.4122
16) (a) \( z = -1.75 \)
    (b) \( z = -0.385 \)
17) 0.05 or 5%
18) 0.068 or 6.8%
19) 587.0 to 595.0 hours
20) 15.225 to 15.875 ounces
21) C
22) \( H_0: p = 0.34 \)
    \( H_a: p > 0.34 \)
    Test statistic: \( z = 2.013 \)
    p-value = 0.0221
    Reject \( H_0 \)
    We do find sufficient evidence to support the researcher's claim that the proportion for fathers in Littleton is significantly higher than 34%.
    (1 prop z test)
23) $H_0$: $\beta = 0$
$H_a$: $\beta \neq 0$
Test statistic: $t = 6.252$
$p$-value = 0.000245
Reject $H_0$
We do find sufficient evidence to say the number of years of study is a useful predictor of test scores.
(in reg t test)

24) $H_0$: $\rho = 0$
$H_a$: $\rho \neq 0$
Test statistic: $t = 0.9947$
p = 0.353
Fail to reject $H_0$
We do not find sufficient evidence to say there is a significant correlation between typing speed and reading speed.
(in reg t test)

25) $H_0$: $\mu_1 = \mu_2$
$H_a$: $\mu_1 > \mu_2$
Test statistic: $t = 1.431$  \hspace{1cm} df = 25.8
$p$ = 0.0822
Fail to reject $H_0$
We do not find sufficient evidence to say people who don’t exercise have a significantly higher pulse rate than those who do exercise.
(2 sample t test)

26) $H_0$: $\mu_1 = \mu_2 = \mu_3$
$H_a$: At least two of the means are not equal
Test statistic: $F = 1.620$
p-value = 0.251
Fail to reject $H_0$
We do not find sufficient evidence to say there is a significant difference in the mean lifetime of at least two of the means of the brands.
(ANOVA)

<table>
<thead>
<tr>
<th>ANOVA Table</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>2744</td>
<td>2</td>
<td>1372</td>
<td>1.62</td>
</tr>
<tr>
<td>Error</td>
<td>7622.25</td>
<td>9</td>
<td>846.92</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10366.25</td>
<td>11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

27) $H_0$: The 1995 distribution of workplace accidents fits the 1990 distribution.
$H_a$: The 1995 distribution of workplace accidents does not fit the 1990 distribution.
Test Statistic: $\chi^2 = 1.227$
p-value = 0.8737
Fail to reject the null hypothesis
There is not sufficient evidence to conclude that the 1995 distribution of workplace accidents does not fit the 1990 distribution.

28) $H_0$: $p_1 = p_2$
$H_a$: $p_1 \neq p_2$
Test statistic: $z = 1.93$
p-value = 0.054
Fail to reject the null hypothesis
We do not find sufficient evidence to conclude that the two proportions are significantly different.
(2 prop z test)
29) \( H_0: \mu = 50 \text{ cm} \)
\( H_a: \mu < 50 \text{ cm} \)
Test statistic: \( z = -3.81 \)
\( p\)-value = 0.0000699
Reject the null hypothesis
There is sufficient evidence to say the mean length is significantly less than 50 cm.
(z test)

30) \( H_0: \mu = 10 \text{ minutes} \)
\( H_a: \mu < 10 \text{ minutes} \)
Test statistic: \( t = -5.94 \)
\( p\)-value = 0.00000808
Reject \( H_0 \)
We do find sufficient evidence to support the claim that the mean is significantly less than 10 minutes.
(t test)

31) \( H_0: \) Frequency of car accidents and age are independent.
\( H_a: \) Frequency of car accidents and age are dependent.
Test statistic: \( \chi^2 = 9.273 \)
\( p\)-value = 0.0546
Fail to reject \( H_0 \)
We do not find sufficient evidence to say frequency of car accidents and age are dependent.
(chi square test of independence)

32) \( H_0: \mu_d = 0 \)
\( H_a: \mu_d > 0 \)
Test statistic: \( t = 1.954 \)
\( p\)-value = 0.0493
Fail to reject the null hypothesis
We do not find sufficient evidence to say the diet significantly reduced weight for the subjects.
(t test) (independent samples)

33) \( H_0: \mu_1 = \mu_2 \)
\( H_a: \mu_1 \neq \mu_2 \)
Test statistic: \( z = 3.442 \)
\( p\)-value = 0.000058
Reject the null hypothesis
We do find sufficient evidence to say there is a significant difference in the cost of a home in the two locations.
(2 sample z test)

34) \( H_0: \mu = 64.05 \text{ ounces} \)
\( H_a: \mu < 64.05 \text{ ounces} \)
Test statistic: \( z = -1.237 \)
\( p\)-value = 0.1080
Fail to reject \( H_0 \)
We do not find sufficient evidence to say the mean content is significantly less than 64.05 ounces.
(z test)

35) \( n = 77 \)