Q1. Suppose that at the north pole, every single day it either snows or is windy (or both). A meteorological station finds that it snows for 75% of the days and is windy for 50% of the days. If we know that it is snowing today at the north pole, what is the probability that it is also windy?

A) 25%
B) 33.3% ***
C) 50%
D) 60%
E) 75%
F) 95%
G) 99.9%

Q2. Meshuggah Cafe hires three dishwashers. Abe washes 40% of the dishes and breaks only 1% of them. Bush and Clinton each wash 30% of the dishes, and Bush breaks only 1% of his, but Clinton breaks 3% of the dishes he washes. You go to Meshuggah and hear a dish break at the sink. What is the probability Clinton is on the job?

A) 3%
B) 30%
C) 33.3%
D) 50%
E) 56.3% ***
F) 65.5%
G) 77.8%
H) 93.1%
Q3. An archer is able to hit the bull’s eye 60% of the time. If she shoots 6 arrows, what is the probability that her first bull’s eye comes on the fourth or fifth arrow? <Ch 17 Q14c>

A) 5.38% ***
B) 7.23%
C) 13.57%
D) 30%
E) 33.33%
F) 45.43%
G) 50%
H) 60%

Q4. Assume that the duration of human pregnancies can be described by a Normal model with mean 266 days and standard deviation 16 days. Suppose that a certain obstetrician is currently providing prenatal care to 60 pregnant women. What is the probability that the mean duration of these patients’ pregnancies will be less than 260 days? <Ch 18 Q25d>

A) 0.18% ***
B) 1.21%
C) 1.75%
D) 2.01%
E) 2.50%
F) 2.75%
G) 2.90%
H) 3.60%

Q5. You roll a die, winning nothing if the number of spots is odd, $1 for a 2 or a 4, and $10 for a 6. You play 50 times. What is the probability that you win at least $100? <Ch 18 Q29c>

A) 13.3%
B) 25%
C) 28.3%
D) 33.3%
E) 45.5%
F) 50% ***
G) 57.4%
H) 65.3%
Q6. In a random sample of 226 WUSTL students, 20 reported being “only” children (with no siblings). With 95% confidence, estimate the proportion of WUSTL students that are only children. <Ch19 Q23>

A) 8.85%
B) (7.43%, 10.27%)
C) (6.12%, 11.58%)
D) (5.15%, 12.55%)
E) (4.81%, 12.89%)
F) (4.14%, 13.56%)
G) (3.15%, 14.55%)
H) (0%, 95%)

Q7. The seller of a loaded die claims that it will favor the outcome 6. We don’t believe that claim, and (having nothing better to do) we roll the die 200 times to test an appropriate hypothesis. Our P-value turns out to be 0.02. Which conclusion is appropriate? <Ch20 Q4>

A) There’s a 2% chance that the die is fair.
B) There’s a 98% chance that the die is fair.
C) There’s 2% chance that a loaded die could randomly produce the results we observed, so its reasonable to conclude that the die is loaded.
D) There’s 2% chance that a loaded die could randomly produce the results we observed, so its reasonable to conclude that the die is fair.
E) There’s 2% chance that a fair die could randomly produce the results we observed, so its reasonable to conclude that the die is loaded.
F) There’s 2% chance that a fair die could randomly produce the results we observed, so its reasonable to conclude that the die is fair.

Q8. You have an idea for a startup company making cheap alarm clocks for students, that needs to be thrown against the wall to switch off the alarm. You randomly sample 80 students and 20 of them say they would buy your product. You have decided to invest money in the startup only if more than 15% of students agree to buy your product. How should you proceed?

A) Invest money in startup.
B) Do not invest money in startup.
Q 9. A company has surveyed a stratified sample of its employees to find out how many might take advantage of a program to help people quit smoking. You may assume that the sampling strategy was properly random and the data gathering methodology avoided biases. The table below shows the results. For which of the groups could we use our methods of inference to determine a 95% confidence interval for the proportion of employees who would participate in the stop smoking program? <Ch21Q25>

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of Employees</th>
<th>Number Surveyed</th>
<th>Percent to Participate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laborers</td>
<td>6235</td>
<td>300</td>
<td>9%</td>
</tr>
<tr>
<td>Clerical</td>
<td>1520</td>
<td>200</td>
<td>7%</td>
</tr>
<tr>
<td>Management</td>
<td>342</td>
<td>25</td>
<td>8%</td>
</tr>
</tbody>
</table>

A) Only for Laborers ***  
B) Only for Clerical  
C) Only for Management  
D) Clerical and Management  
E) Laborers and Clerical  
F) Laborers and Management  
G) All of them

Q10. The toothpaste company Whitewash would like to claim in its advertisement that its toothpaste kills more germs than its competitor Germicide. If the lawyers of Whitewash are worried about being sued for incorrect claims in its advertisement, should it increase or decrease the value of \( \alpha \) in its sampling procedures? <Ch23 23a>

A) Increase  
B) Decrease***

Q11. Students investigating the packaging of potato chips purchased 10 bags of Lay’s Ruffles marked with a net weight of 28.3 gms. They carefully weighed the contents of each bag, recording the following weights (in grams):

29.3, 28.2, 29.1, 28.7, 28.9, 28.5, 28.1, 28.2, 29.0, 29.1

Calculate a confidence interval for mean weight of such bags of chips. <Ch23 Q27>

A) (28.71, 29.37)  
B) (28.4, 29.02)***  
C) (28.12, 29.3)  
D) (27.51, 29.91)  
E) (27.01, 30.41)  
F) (26.58, 31.14)  
G) (26.04, 31.79)
Q12. To determine if men and women perform differently on questions requiring abstract visualization a surveyor takes a random sample of 50 men and 50 women and gives each 25 questions to obtain the following statistics:

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean score</td>
<td>19.39</td>
<td>17.91</td>
</tr>
<tr>
<td>SD of score</td>
<td>2.52</td>
<td>3.39</td>
</tr>
</tbody>
</table>

Assume all conditions necessary for inference are met. Using a 95%-confidence interval, what can you conclude? <Ch24 Q14>

A) The mean score for men is more than the mean score for women by between (0.12, 2.14) points
B) The mean score for men is more than the mean score for women by between (0.19, 2.31) points
C) The mean score for men is more than the mean score for women by between (0.22, 2.50) points
D) The mean score for men is more than the mean score for women by between (0.29, 2.67) points ***
E) The mean score for men is more than the mean score for women by between (0.32, 2.58) points
F) The mean score for women is more than the mean score for men by between (0.32, 3.2) points.

Q13. Newspaper headlines recently announced a decline in math scores among high school seniors. In 2000, 15109 seniors tested scored a mean of 147 points. Four years earlier, 7537 seniors had averaged 150 points. The standard error of the difference in the mean scores for the two groups was 1.22. Have the math scores declined significantly? (Use t* = 1.96) <Ch24 Q23>

A) Yes***
B) No
Q14. To see if the number of meals families eat together has changed over the years, ten households selected randomly were surveyed in 1980. Ten different households selected randomly were surveyed thirty years later to get the following data of number of meals shared per week. Is there a significant decrease in the mean number of meals families eat together? Assume conditions for inference are met and estimate this difference with a confidence interval.

<table>
<thead>
<tr>
<th></th>
<th>1980</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>5</td>
</tr>
</tbody>
</table>

A) (-0.77, 4.37) so no significant decrease ***
B) (-0.77, 4.37) so there is a significant decrease
C) (1.32, 2.28) so no significant decrease
D) (1.32, 2.28) so there is a significant decrease
E) (-2.28, -1.32) so no significant decrease
F) (-2.28, -1.32) so there is a significant decrease
G) (1.92, 2.01) so no significant decrease
H) (1.92, 2.01) so there is a significant decrease

Q15. To verify the claim that “Fish is brain food”, 10 students were randomly selected and given an IQ test. They were then prescribed a daily dosage of fish oil and tested again after a month’s time to see if there was any improvement. Assume conditions for inference are met and find a confidence-interval for the mean increase in IQ. <Ch25>

<table>
<thead>
<tr>
<th></th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>102</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>110</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>98</td>
<td>117</td>
</tr>
<tr>
<td></td>
<td>115</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>107</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>87</td>
<td>92</td>
</tr>
</tbody>
</table>

A) (0.323, 2.477)***
B) (0.513, 3.141)
C) (0.914, 2.941)
D) (1.314, 2.513)
E) (2.415, 5.433)
F) (-2.921, 4.134)
G) (-4.132, 5.332)
H) (-11.01, 8.214)
Q16. Census data in Springfield indicates that of the youth 29% are White, 28% Black, 31% Latino, 9% Asian and the remaining of Other ethnicities. Assume that Springfield Elementary has 600 students with the following demographics: 31% White, 27% Black, 29% Latino, 11% Asian and the remaining clubbed under Others. If we assume that the Springfield Elementary student body is representative of the ethnic composition of Springfield, what is the likelihood (P-value) of the observed demographics? <Ch26 Q7>

A) 0.013
B) 0.166***
C) 0.231
D) 0.262
E) 0.52
F) 0.714
G) 0.821
H) 0.913

Q17. A survey was taken about a bill that would make health checkups mandatory once a year for each citizen. People were classified into homeless, poor (but not homeless), rich and super-rich to get the following data:

<table>
<thead>
<tr>
<th></th>
<th>Homeless</th>
<th>Poor</th>
<th>Rich</th>
<th>Super-rich</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of people surveyed</td>
<td>300</td>
<td>350</td>
<td>500</td>
<td>200</td>
</tr>
<tr>
<td>% supporting Health Bill</td>
<td>69%</td>
<td>70%</td>
<td>65%</td>
<td>63%</td>
</tr>
</tbody>
</table>

We test to check if there is any association between support for Health Bill and financial status. If we assume that there is no association, what is the likelihood (P-value) of the observed proportions? <Review Part VI Q43>

A) 0.131
B) 0.230***
C) 0.291
D) 0.341
E) 0.413
F) 0.459
G) 0.513
H) 0.621
Q18. Two different professors teach an introductory Statistics course. The table shows the distribution of final grades they reported. We wonder whether one of these professors is an “easier” grader.

<table>
<thead>
<tr>
<th></th>
<th>Prof Alpha</th>
<th>Prof Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>B</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>C</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>D</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>F</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

A) The distribution of grades for Prof Alpha is significantly different from the distribution for Prof Beta.
B) The distribution of grades for Prof Alpha is not significantly different from the distribution for Prof Beta.***

Q19. In an investigation of environmental causes of disease, data were collected on the annual Mortality rate (deaths per 100,000) for males in 62 large towns in England and Wales. In addition, the water hardness was recorded as the Calcium concentration (in ppm) in the drinking water. Assuming the conditions for regression analysis are true, we get the following data:

*Dependent variable: Mortality*

*R-squared = 43%*

*s = 143 with 62 - 2 = 60 degrees of freedom*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE(Coeff)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1676</td>
<td>29.30</td>
</tr>
<tr>
<td>Calcium</td>
<td>-3.23</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Find a confidence interval for the slope of the true line relating Calcium concentration and Mortality. (Use \( t^* = 2.00 \))

A) (-3.11, -3.35)
B) (-4.19, -2.27) ***
C) (-4.92, -1.58)
D) (-5.13, -1.33)
E) (-5.71, -1.42)
F) (-6.24, -0.13)
A scientist claims that the number of UFO sightings all over the world have been increasing linearly over the years. To justify his claim he presents the following data on number of sightings in St Louis over the years:

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of sightings</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>10</td>
</tr>
<tr>
<td>2001</td>
<td>12</td>
</tr>
<tr>
<td>2002</td>
<td>13</td>
</tr>
<tr>
<td>2003</td>
<td>17</td>
</tr>
<tr>
<td>2004</td>
<td>17</td>
</tr>
<tr>
<td>2005</td>
<td>20</td>
</tr>
<tr>
<td>2006</td>
<td>21</td>
</tr>
<tr>
<td>2007</td>
<td>24</td>
</tr>
<tr>
<td>2008</td>
<td>25</td>
</tr>
<tr>
<td>2009</td>
<td>29</td>
</tr>
</tbody>
</table>

Assuming that sightings in St Louis are a correct representation of sightings all over the world (and that other conditions of regression analysis hold) find a confidence interval for the slope of the true line relating the number of sightings with the year (Independent variable: year). (Use $s_e = 0.8788$ and $t^* = 2.306$)

A) (0.13, 3.89)
B) (0.51, 3.51)
C) (0.78, 3.24)
D) (1, 3)
E) (1.23, 2.79)
F) (1.79, 2.24)***
G) (1.90, 2.12)
Q21. A healthy cereal should be low in both calories and sodium. Data for 77 cereals were examined and judged acceptable for inference. The Regression analysis is shown.

Dependent variable is: Sodium  
R-squared = 9.0%  
s = 80.49 with 77 - 2 = 75 degrees of freedom

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE[Coeff]</th>
<th>t-ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>21.4143</td>
<td>51.47</td>
<td>0.416</td>
<td>0.6786</td>
</tr>
<tr>
<td>Calories</td>
<td>1.2936</td>
<td>0.4738</td>
<td>2.73</td>
<td>0.0079</td>
</tr>
</tbody>
</table>

Is there any association between the number of calories and the sodium content of cereals? Do you think this association is strong enough to be useful?

A) There is no association.  
B) There is an association, but not strong enough to be useful.***  
C) There is an association and it is strong enough to be useful.

Q22. An experimenter wants to test if washing hands with different soaps makes any difference on the average number of germs killed. He washes his hands with 4 brands of soaps and with water. He washes 8 times randomly with each brand and water. Assuming conditions for ANOVA test are met, he finds the mean squared treatment and the mean squared error. If we assume that in fact washing with any soap is just as effective as washing with water, what is the P-value when the mean squared treatment is twice the mean squared error?

A) 0.012  
B) 0.083  
C) 0.116***  
D) 0.298  
E) 0.351  
F) 0.393  
G) 0.5  
H) 0.724
Q23. Three new brands of tea were introduced at a student cafe. Each brand was sampled randomly by three students and judged on a scale from 1 to 10. From the following partially complete ANOVA table, determine the F-ratio.<Ch28 Q13>

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>17.300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>0.460</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>17.769</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A) 20.131  
B) 37.132  
C) 51.391  
D) 71.253  
E) 84.234  
F) 95.253  
G) 100  
H) 112.778***

Q24. A statistics course was taught in three large sections - morning, afternoon and evening. Ten students were randomly selected from each of the sections and their score recorded below. Determine whether the three sections have mean scores that are significantly different.

<table>
<thead>
<tr>
<th>Morning</th>
<th>Afternoon</th>
<th>Evening</th>
</tr>
</thead>
<tbody>
<tr>
<td>76</td>
<td>74</td>
<td>58</td>
</tr>
<tr>
<td>94</td>
<td>68</td>
<td>82</td>
</tr>
<tr>
<td>78</td>
<td>46</td>
<td>52</td>
</tr>
<tr>
<td>50</td>
<td>76</td>
<td>66</td>
</tr>
<tr>
<td>84</td>
<td>92</td>
<td>40</td>
</tr>
</tbody>
</table>

A) The three sections have mean scores that are significantly different  
B) The three sections have mean scores that are not significantly different***
Q25. A cyclist who plans on buying a new cycle runs an experiment. He tries out 4 different brands of cycles and he cycles each of them 5 times in hilly terrain, 5 times on gravel and 5 times on tarred roads. He records the time taken to cover 3 miles in each situation and fits an ANOVA Model with interaction. How many degrees of freedom does the error sum of squares have?

A) 6  
B) 18  
C) 36  
D) 48***  
E) 54  
F) 59  
G) 60

TRUE OR FALSE:

Q26. For a given sample size, higher confidence means smaller margin of error.

A) True  
B) False***

Q27. A certain drug company must be extremely sure that its drugs have no side-effects. It tests its drugs with the null hypothesis being that the drugs have no side-effects and the alternate being side-effects. The experiment can be set up so as to minimize either Type I Error or Type II Error but not both. The experimenter chooses to minimize Type I Error. This is the correct strategy.

A) True  
B) False***

Q28. Environmentalists concerned about the impact of high-frequency radio transmissions on birds found that there was no evidence of a higher mortality rate among hatchlings in nests near cell towers. They based this conclusion on a test using $\alpha = 0.05$ (and null hypothesis of no evidence). They would have made the same decision at $\alpha = 0.01$. <Ch 21 Q4>

A) True***  
B) False

Q29. Assuming that all necessary conditions for ANOVA are true, if the null hypothesis of equal means is true, the mean squared treatment is an approximation of the population variance.

A) True***  
B) False
Q30. An experiment is conducted with 2 factors A and B. A has 5 levels and B has 2. The following interaction plot is obtained. From this, you can conclude that most likely, the two factors interact.

A) True***
B) False

Formula: \[ SE(b_1) = \frac{s_e}{\sqrt{n-1}s_x} \]

All the best!