1) The RubbaDubDub Soap Company has two machines which wrap bars of RubbaDubDub hand soap in attractive, sanitary foil wrapping. The employees have nicknamed the machines "SpongeBob" and "Patrick." SpongeBob wraps 65% of the bars of soap, while Patrick wraps the rest. 2% of the bars wrapped by SpongeBob are wrapped improperly, while 4% of the bars wrapped by Patrick are wrapped improperly. If a bar of RubbaDubDub soap is selected at random, what is the probability that it is wrapped PROPERLY?

2) One person is randomly selected. Using the contingency table, find:

<table>
<thead>
<tr>
<th></th>
<th>men</th>
<th>women</th>
</tr>
</thead>
<tbody>
<tr>
<td>doctors</td>
<td>25</td>
<td>5</td>
</tr>
<tr>
<td>lawyers</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>teachers</td>
<td>25</td>
<td>65</td>
</tr>
</tbody>
</table>

a) P(doctor or man)  
d) P(man and lawyer)  
b) P(doctor or teacher)  
e) P(woman given doctor)  
c) P(man)

3) Find the expected gain or loss

a) for an insurance company, if a life insurance premium on a $100,000 policy is $400, and the chance of death during the term is .003.  
b) per ticket for a raffle with a grand prize of a $2000 television and two second prizes of a $250 mobile phone. A total of 1000 tickets were sold for $5 each.

4) Here is a set of sample data: 12, 14, 10, 12, 20, 15, 21, 16, 11, 12.  
Find the mean, median, range, variance, and standard deviation.

5) Assume that the data in Question 5 were drawn from a normally distributed population with unknown standard deviation. Construct a 95% confidence interval for the population mean.

6) rating | frequency  
---|---
0 | 3  
1 | 5  
2 | 12  
3 | 46  

a) Before doing any calculations, the mean for this data must be between what two values?  
b) Calculate the mean.

7) A binomial random variable has p = .56, n = 33, calculate the mean and standard deviation of this distribution.

8) A hat contains 2 red apples and 3 green apples. A bucket contains 7 red apples and 3 green apples.

a) A container is selected at random and an apple is drawn out. What is the probability that it will be a red apple?  
b) A green apple is selected from the bucket and eaten. A second apple is to be selected from the bucket. What is the probability that this second apple will be red?  
c) Two different apples are selected from the hat. What is the probability that both are green?

9) x | 0 | 1 | 2 | 3  
---|---|---|---|---
P(x) | .13 | .21 | .55 |  

a) P(at most 2)  
c) P(more than 0)  
e) E(x) or µ  
b) P(at least 1)  
d) P(less than 3)
10) A certain dataset of systolic blood pressure measurements has a mean of 80 and a standard deviation of 3. Assuming the distribution is bell-shaped:
   a) About how many of the 550 data values will fall between 71 and 89?
   b) Is a value of 85 considered to be ordinary (usual) or unusual?

11) Suppose that the weight of the adult male wombat is normally distributed with mean 8.6 pounds and standard deviation 1.1 pounds.
   a) What is the probability that a randomly selected adult male wombat will weigh at least 9.5 lbs?
   b) Rounded to the nearest .01 pound, what is the 85th percentile of adult male wombat weight?
   c) A sample of 50 wombats is chosen. What is the probability that its mean is less than 8.3 pounds?
   d) To conduct a new study to find the weight of wombats to within .2 pounds with 95% confidence, what is the minimum sample size necessary?

12) How many students do you need to estimate the mean GPA of Rancho Santiago Community College District students to within .1 with 98% confidence, if we assume that GPAs are between 1.0 and 4.0?

13) You are an environmentalist interested in the prevalence of acid rain (as measured by the pH of the rainwater) in the northeastern United States. Assume that the mean "normal" pH for pure rain falling through clean air is about 5.7. You randomly collect and test 39 samples of rainwater in the northeastern United States and calculate a sample mean pH of 5.5 with a known population standard deviation of .6.
   Explain why the initial requirements have been met and whether to use the t or z table.
   Using this data, test the claim that northeastern rainfall is more acidic than it should be, that is, that the mean pH is lower than the "normal" value of 5.7? Test using $\alpha = .05$.

14) Shortly after the Apollo 11 lunar landing on July 20, 1969 a survey revealed that 28% of Americans did not believe it had happened. You have a theory that people nowadays are just as cynical and distrustful of government as they were back then, so you randomly sample 75 people and find that 25 of them do not believe that the Apollo 11 landing actually happened.
   Explain why the initial requirements have been met.
   Test the claim that the percentage is the same as its 1969 value?

15) An eager but misguided statistics professor comes to class one day with 13 slices of bread, a knife, and a large jar of peanut butter. The professor distributes all of the bread to his students, generously smearing peanut butter on one side of each slice. He then instructs his students to toss their slices vigorously into the air. (One slice sticks to the ceiling, and may be ignored in our computations.)
   a) Of the 12 slices that do fall to the floor, however, calculate the probability that exactly 5 of them will fall peanut-butter-side-down. Assume that 60% of all slices will fall peanut-butter-side-down in the long run, and that they never land on their edges.
   b) What is the mean and standard deviation of this distribution?
   c) Find the probability that all 12 of them land peanut-butter-side-up.
   d) Find the probability that at least one of the 12 slices lands peanut-butter-side-up.
   e) Find the minimum sample size necessary to estimate the true percentage that will land peanut-butter side down to within 4% with 95% confidence if we assume the 60% figure is pretty reliable.
   f) We try this with 400 pieces of bread and 264 of them land peanut-butter side down. Estimate the population proportion with a 95% confidence interval.
   g) Assume the 60% figure: what is the probability that in a sample of 400, the proportion landing down > .66?
16) In how many ways can you select:
   a) three students from a class of 45 to represent the class?
   b) three different students from the 50 who apply to receive three different scholarships?
   c) a shirt and tie, if you have 20 shirts and 10 ties and they all look good together?

17) The higher the bowling score the better. The lower the golf score the better. Assume both are normally distributed.
   a) Suppose the mean bowling score is 155 with a standard deviation of 16 points. We will give a trophy for the best 5% of scores. What score must you get to receive a trophy?
   b) Suppose the mean golf score is 77 with a standard deviation of 3 strokes. We will give a trophy for the best 5% of scores. What score must you get to receive a trophy?
   c) Suppose the mean golf score is 77 with a standard deviation of 3 strokes. Would a golf score of 70 be ordinary, a mild outlier, or an extreme outlier?
   d) Suppose we have a sample of the Santa Ana Strikers' bowling scores. $Q_1 = 125$ and $Q_3 = 156$. Would it be usual or unusual to have a score of 200?
   e) Suppose the mean bowling score is 155 with a standard deviation of 16 points. What is the probability that in a sample of 40 bowling scores, the mean will be smaller than 150?

18) The histogram shows a distribution that is
   a) uniformly distributed       c) skewed left
   b) bell-shaped                  d) skewed right

19) This histogram shows a distribution that is
   a) uniformly distributed       c) skewed left
   b) bell-shaped                  d) skewed right

20) Which measure of central tendency is used for qualitative or categorical data?
   a) mean                        b) median                   c) mode
21) When the data has extreme highs or lows, which is the best measure of central tendency?
   a) mean  b) median  c) mode

22) Which is the only true statement?
   a) Mutually exclusive (disjoint) and independent mean the same thing.
   b) A conservative estimate of the standard deviation is the range divided by 6.
   c) A z-score is the number of standard deviations above the mean.
   d) The purpose of the initial condition requirements is to make sure the sampling distribution is bell-shaped.
   e) The median is always a data element.

23) Which is an example of cluster sampling? What are the others?
   a) Select every 25th person who enters Disneyland.
   b) Randomly choose 4 out of 16 CSU campuses and survey all statistics majors from those campuses.
   c) Randomly choose 100 statistics students from each of the 16 CSU campuses.
   d) Randomly choose 100 student ID numbers from students at SAC.

24) \( P(-.6 < z < .4) = \)
   a) .9297  b) .6189  c) .3811  d) .2743  e) .0703

25) \( P(z > \text{what number}) = .85, \text{what is the number?} \)
   a) -.56  b) .8023  c) .5596  d) 1.04  e) -1.04

26) Which is the only false statement?
   a) A least squares line can be constructed for any set of data points, but without correlation it is meaningless and cannot be used for predictions.
   b) The more confident you wish to be, the further apart the limits on a confidence interval.
   c) The z-table is used for hypothesis tests about proportions as long as np and n(1–p) are both \( \geq 5 \).
   d) If \( \alpha = .05 \) and the p-value is .06, you should reject equality.
   e) A hypothesis test to decide independence for a contingency table is more reasonable than the probability definition because we cannot tell if the difference is significant by just comparing them.

27) In a bell-shaped distribution with a mean of 25 and a standard deviation of 5, usual values would be between:
   a) 20 and 30  d) 15 and 35
   b) 10 and 40  e) 5 and 45
   c) 17.5 and 32.5

28) If Q1 = 40 and Q3 = 50, then usual values would be between:
   a) 25 and 65  c) 30 and 60
   b) 40 and 50  d) 20 and 70

29) If Q1 = 40 and Q3 = 50, then an outlier value would be:
   a) between 25 and 40  c) between 25 and 65
   b) between 50 and 65  d) lower than 25
30) Which of the two dotplots has a larger standard deviation?

![Dotplot](image)

a) Prozac  

b) PGeneric

31) Matching: Each is used once.

Decide which type of test would be most appropriate for each of these.

a) The percent that use the generic Prozac is 15% more than the percent who use the name brand Prozac  
   i) two means independent
   ii) matched pairs

b) Prozac costs more than $20 more than Celexa.  
   iii) ANOVA

   iv) two proportions

   v) contingency table

c) Generic medications cost less than their comparable name brands.

d) Income level (poor, middle, wealthy) and use of generics are related.

e) All generic medications for depression have the same cost.
Which hypotheses are most appropriate?  
Claim: One can predict the cost of the generic drug from the cost of its name brand version.

a) \( H_0 \): cost of generic and cost of name brand are independent  
   \( H_1 \): cost of generic and cost of name brand are dependent

b) \( H_0 \): cost of generic and cost of name brand are dependent  
   \( H_1 \): cost of generic and cost of name brand are independent

c) \( H_0 \): cost of generic and cost of name brand are linearly correlated  
   \( H_1 \): cost of generic and cost of name brand are not linearly correlated

d) \( H_0 \): cost of generic and cost of name brand are not linearly correlated  
   \( H_1 \): cost of generic and cost of name brand are linearly correlated

e) \( H_0 \): \( \mu_d = 0 \)  
   \( H_1 \): \( \mu_d \neq 0 \)

Which hypotheses are most appropriate?  
Claim: The cost of a generic medication is significantly lower than its name brand version.

a) \( H_0 \): cost of generic and cost of name brand are independent  
   \( H_1 \): cost of generic and cost of name brand are dependent

b) \( H_0 \): cost of generic and cost of name brand are dependent  
   \( H_1 \): cost of generic and cost of name brand are independent

c) \( H_0 \): cost of generic and cost of name brand are not linearly correlated  
   \( H_1 \): cost of generic and cost of name brand are linearly correlated

d) \( H_0 \): \( \mu_d = 0 \)  
   \( H_1 \): \( \mu_d > 0 \)

e) \( H_0 \): \( \mu_d = 0 \)  
   \( H_1 \): \( \mu_d > 0 \)

Which hypotheses are most appropriate?  
Claim: Psychology, Biology, and Criminal Justice majors have the same averages on the final.

a) \( H_0 \): \( \mu_1 - \mu_2 = 0 \)  
   \( H_1 \): \( \mu_1 - \mu_2 \neq 0 \)

b) \( H_0 \): \( \mu_d = 0 \)  
   \( H_1 \): \( \mu_d \neq 0 \)

c) \( H_0 \): \( \mu_p = \mu_b = \mu_c \)  
   \( H_1 \): \( \mu_p \neq \mu_b \neq \mu_c \)

d) \( H_0 \): \( \mu_p = \mu_b = \mu_c \)  
   \( H_1 \): at least one of the means is different from the others

e) \( H_0 \): major and scores are independent  
   \( H_1 \): major and scores are dependent
35) When the p-value is .0213 and the claim is the null hypothesis, which is the appropriate conclusion?
   a) We support our claim.
   b) We reject our claim.
   c) We do not reject our claim.
   d) We accept our claim.

36) \(H_0: p = .25\)
    \(H_1: p > .25\) (claim)
    p-value = .07

Which is the appropriate conclusion?
   a) We have enough evidence to prove that \(p = .25\)
   b) We would reject that the proportion is .25 and accept that it is greater than .25
   c) We do not have enough evidence to reject our claim that \(p = .25\)
   d) We do not have enough evidence to support our claim that \(p > .25\)

Is this a binomial experiments?  (a) yes        (b) no

37) Selecting balls with replacement from a box that contains 100 balls, 60 of which are blue and the rest are red.

38) Selecting balls without replacement from a box with 100 balls, 60 of which are blue and the rest are red.

39) Rolling a die many times.

40) Guessing on a true/false test.

The spreadsheet below is just the first 5 houses. For each scenario, decide which type of test is most appropriate.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27</td>
<td>48</td>
</tr>
<tr>
<td>2</td>
<td>49</td>
<td>63</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>51</td>
<td>72</td>
</tr>
<tr>
<td>5</td>
<td>42</td>
<td>55</td>
</tr>
</tbody>
</table>

41) In column A is the price (in hundreds of thousands of dollars) of a sample of houses in Anytown from ten years ago, and column B is the price (in hundreds of thousands of dollars) of a different sample of houses in Anytown from today.
   a) matched pairs
   b) two independent means
   c) correlation and regression

42) For a sample of homes in Anytown, in column A is the age of the home and in column B is the price of the home (in hundreds of thousands of dollars).
   a) matched pairs
   b) two independent means
   c) correlation and regression

43) In column A is the price (in hundreds of thousands of dollars) of a sample of houses in Anytown from ten years ago, and column B is the price (in hundreds of thousands of dollars) of a the same sample of houses in Anytown from today.
   a) matched pairs
   b) two independent means
   c) correlation and regression
Using the pie chart below for the work status for a sample of 391 students:

44) The percentage who are not employed is:
   a) slightly less than 25%
   b) slightly more than 50%
   c) about 33%
   d) slightly less than 50%

45) Roughly how many of the 391 students work full-time?
   a) 100
   b) 190
   c) 210
   d) 300