True or False

1. T
2. F: The null may or may not be true.
3. F: If short sales is allowed, then the statement is false. Simple example:
   Suppose $R_1 \sim N(0.10, 0.25), R_2 \sim N(0.02, 0.01), \text{cov}(R_1, R_2) = 0$
   Suppose a portfolio consists of two $R_1$ and minus one $R_2$.
   That is, the portfolio has a short position of $R_2$.
   \[
   R = 2R_1 - R_2
   \]
   Then the variance of the portfolio is
   \[
   \text{var}(R) = \text{var}(2R_1 - R_2) = 4 \text{var}(R_1) + \text{var}(R_2) = 4 \times 0.25 + 0.01 = 1.01
   \]
4. F
5. T: Sum of any iid random variable is approximately distributed as a normal distribution by the central limit theorem. Since $Y$ is sum of $n$ iid Bernoulli variables, $Y \sim B(n,p)$. Mean of $Y$ is $np$ and variance of $Y$ is $np(1-p)$.
6. T
7. F: Independent implies the correlation is zero but not vice versa.
8. T: Unbiased estimator has the property that the expected value of the estimator is equal to the true parameter.
9. T
10. F: t distribution has a fatter tail than a normal distribution. In other words, the probability that the variable is close to 0 is less than that for a standard normal distribution.

II. Multiple choice

1. a, c
2. d: $s$ estimates the standard deviation of the error term. R square means the percent of the variation in $Y$ explained by the $X$’s.
3. b:
4. a
Let's define $R$ as the equally weighted portfolio of $R_1$ and $R_2$.

$$R = 0.5R_1 + 0.5R_2$$

Then

$$E(R) = 0.5E(R_1) + 0.5E(R_2) = 0.06$$

$$\text{var}(R) = 0.25 \text{var}(R_1) + 0.25 \text{var}(R_2) + 2 \times 0.5 \times 0.5 \text{cov}(R_1, R_2) = 0.00878$$

5. c: Probability of making an error is .0074. We need 135 students to expect to find an error because 135 times 0.0074 is approx. 1.

6. a.

$$\beta = \frac{\text{cov}(Y, X)}{\text{var}(X)}$$, where $Y$ is stock returns and $X$ is S&P500

$$\beta = \frac{.00028094}{.00079117} = .355$$

7. a

$$.46 \pm 2 \sqrt{\frac{.46(1-.46)}{1200}} = .46 \pm .029$$

8. d

$$\text{test statistic} = \left| \frac{.46 - .50}{\sqrt{.5(1-.5)}} \right| = 2.771$$

9. c

$$\bar{x} \pm 2 \frac{s}{\sqrt{n}} = 40.5 \pm 2 \frac{4}{\sqrt{50}} = 40.5 \pm 1.13$$

10. d

$$\text{test statistic} = \left| \frac{42-40.5}{\sqrt{50}} \right| = 2.65$$

11. c

Probability that you reject the claim given that it is true
$$= \text{Probability that both dice turn out to be 6 given that his claim is true}$$
$$= \text{Probability that both dice turn out to be 6 given that the both dice are fair}$$
Probability that both dice turn out to be 6 given that each die has 1/6 probability of getting 6
= 1/6 * 1/6 = 1/36

12. b

Probability that you reject his claim given that his claim is false
= Probability that both dice turn out to be 6 given that his claim is false
= Probability that both dice turn out to be 6 given that the probability of a 6 is ½ for each die
= ½ * ½ = 1/4

13. a, e
Mean and variance changes if the variable is multiplied by a number other than –1. If multiplied by –1, the mean changes but variance stays the same.

14. d.

\[ Y = 1 + 2X + \epsilon \]
If \( X = 2 \), then \( Y = 1 + 4 + \epsilon = 5 + \epsilon \)
Thus \( Y \sim N(5, 4) \)

15. e
Regression finds
a) the linear combination of X’s that is maximally correlated with Y or
b) the linear combination of X’s that has the highest R-square or
c) the linear combination of X’s that minimizes the sum of squared errors

III. Long answer questions

1.

a. estimate for \( \sigma = s = 8334 \)
b. R-square = 74.6%
c. Coefficient of ABS = 3217. Thus the sales price increases by 3217.
d. \(-12,348 + 223 \times 140 + 0 = 18,872\)
e. \( Y = \) predicted sales price + residual
\( 15,500 = 18,872 + \text{residual} \)
residual = -3,372
f. \( 18,872 \pm 2 \times s \)
\( = 18,872 \pm 2 \times 8334 \)
g. test stat = 2.04, Since test stat is greater than 2, we can reject that it’s unrelated.
h. \( b \pm 2 \times SE(b) = 223.14 \pm 2 \times 13.28 \)
2.  
a. \(0.006149 \times -3.12 = -0.0192\)  
ii. \(-0.05546 / -1.87 = 0.0297\)  
i. more complex, less seniority, less satisfied job  
j. \(\frac{-0.4155 - (-0.75)}{0.1581} = 2.115\), reject the hypothesis  
k. predicted = 4.44 -0.0192 * 40 - 0.0555 * 5 - .415 * 2 = 2.559  
\(95\%\) interval = (-0.084, 5.202)  
l. p-value = 0.000  
f. yes because p-value is less than 0.05  

3.  
a. The probability of testing positive given the person doesn’t have a disease is 2%  
b. \(1 - .9 = .1\)  
c.  
<table>
<thead>
<tr>
<th>(X = 0)</th>
<th>(Y = 0)</th>
<th>(Y = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X = 0)</td>
<td>.970</td>
<td>.001</td>
</tr>
<tr>
<td>(X = 1)</td>
<td>.020</td>
<td>.009</td>
</tr>
</tbody>
</table>

d. \(\text{Pr}(Y = 1|X = 1) = .009 / .029 = .31\)  
e. Probability of having disease given the test is positive is 31%. This is too low for a reliable test.  

4.  
\[t = \frac{0.04739 - 0.05}{0.00104} = 2.51\]  
The test says the average thickness is different from 0.05 inches. Should not buy it from this supplier.  
P-value = 0.0068 * 2 = .0136  

5.  
a. 95\% CI = 0.536 \pm 2 \times \sqrt{\frac{0.536(1-0.536)}{250}} = 0.536 \pm 2 \times 0.032 = (0.472,.600)  
The true proportion lies between .472 and .600 with 95\% chance.  

b. No. The CI includes the range less than 50\%. So CI doesn’t support manager’s claim.  

6. a. p-value is 0.129 from the report. Should not reject the normal hypothesis.
b. $s = 27.990$ and the CI is $(23.341, 34.967)$ from the report.

7. One can build a regression model to test the lawyer’s claim. Specifically, one can regress percentage of non-votes on voting system and other demographic factors. But one should be aware of the fact that regression doesn’t tell the causality but only implies the relationship between the variables.