University of Chicago

Business 41000: Business Statistics

Collection of True/False questions

1 Probability

1. If $P(A|B) = 0.5$ and $P(B) = 0.5$, then the events A and B are necessarily independent.

   False. This is not necessarily true. We need more information about each event to definitively say so.

2. The Illinois state lottery introduces a new game. The numbers 1 to 20 are drawn at random without replacement (no two numbers can be the same). You win if you correctly identify all four numbers in exact order. The probability that you win is 1 in $124,750$.

   False, the probability that you win is 1 in 116,280 since $20 * 19 * 18 * 17 = 116,280$.

3. If two events are independent, then $p(A|B) = p(B|A)$.

   False, $p(A) = p(A|B) \neq p(B|A) = p(B)$

4. A kitchen has two cookie jars. The first jar contains 10 ginger snaps and 10 chocolate chip cookies. The second contains an unknown proportion of chocolate chip cookies. If you wish to eat a chocolate chip cookie you should be indifferent to selecting a cookie at random from either jar.

   True, we do not know the proportion of chocolate chip cookies in the second jar.

5. A box has three drawers; one contains two gold coins, one contains two silver coins, and one contains one gold and one silver coin. Assume that one drawer is selected randomly and that a randomly selected coin from that drawer turns out to be gold. Then the probability that the chosen drawer contains two gold coins is 50%.

   False.

   Knowing that we have a gold coin, there is 2/3 chance of being in the 2 gold coin drawer and a 1/3 chance of being in the 1 gold coin drawer. Therefore, the probability that the chosen drawer contains two gold coins is 2/3.

6. Suppose that $P(A) = 0.4, P(B) = 0.5$ and $P(A \cup B) = 0.7$ then $P(A \cap B) = 0.3$.

   False. $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.2$

7. If $P(A \cup B) = 0.5$ and $P(A \cap B) = 0.5$, then $P(A) = P(B)$.

   True. We know that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ *and so* $P(A) + P(B) = 1$. *Also* $P(A) \geq P(A \cap B) = 0.5$ *and so* $P(A) = P(B) = 0.5$

8. The following data on age and martial status of 140 customers of a Bondi beach night club were taken

Age	Single	Married
Under 30	77	14
Over 30	28	21

   Given this data, age and martial status are independent.

   False. We can compute conditional probabilities as $P(S|U_{30}) = 0.8462, P(M|U_{30}) = 0.1638$ *and* $P(S|O_{30}) = 0.5714, P(M|O_{30}) = 0.4286$. *The events are not independent as the conditional probabilities are not the same.*

9. If $P(A \cap B) = 0.5$ and $P(A) = 0.1$, then $P(B|A) = 0.1$.

False. $P(B|A) = P(A \cap B)/P(A)$ *would be greater than one.*

10. In a group of students, 45% play golf, 55% play tennis and 70% play at least one of these sports. Then the probability that a student plays golf but not tennis is 15%.

True. $P(G) = 0.45, P(T) = 0.55, P(G \cup T) = 0.70$ *implies* $P(G \cap \bar{T}) = 0.15$

11. The following probability table related age with martial status

Age	Single	Not Single
Under 30	0.55	0.10
Over 30	0.20	0.15

Given these probabilities, age and martial status are independent.

False. $P(X > 30 \cap Y = \text{married}) = 0.15 \neq -.35 \times 0.25 = P(X > 30)P(Y = \text{married})$

12. Thirty six different kinds of ice cream can be found at Ben and Jerry's. There are 58,905 different combinations of four choices of ice cream.

True. Combinations are given by $36!/(36-6)!6! = 58,905$.

13. If two events A and B are independent then both $P(A|B) = P(A)$ and $P(B|A) = P(A)$.

False. $P(B|A) = P(A)$. *Note that the statement is true if (and only if)* $P(A) = P(B)$.

14. Suppose that for a certain Caribbean island the probability of a hurricane is 0.25, the probability of a tornado is 0.44 and the probability of both occurring is 0.22. Then the probability of a hurricane or a tornado occurring is 0.05.

False. $P(H \cup T) = P(H) + P(T) - P(H \cap T) = 0.25 + 0.44 - 0.22 = 0.47$

15. If $P(A|B) = 0.6$ and $P(B) = 0.2$ then $P(A \cap B) = 0.12$.

True. $P(A \cap B) = P(A|B) \cdot P(B)$

16. If $P(A \cap B) \geq 0.10$ then $P(A) \geq 0.10$.

True. $A \cap B$ *is a subset of* A *and so* $P(A) \geq P(A \cap B) = 0.10$.

17. If A and B are mutually exclusive events, then $P(A|B) = 0$.

True. By definition, if A *and* B *are mutually exclusive events then* $P(A \cap B) = 0$ *and so* $P(A|B) = P(A \cap B)/P(B) = 0$.

2 Bayes Rule

1. Suppose that there's a 5% chance that it snows tomorrow and a 80% chance that the Chicago bears play their football game tomorrow given that it snows. The probability that they play tomorrow is then 80%.

   False, 99%

$$P(Play) = P(Play|Snow)P(Snow) + P(Play|NoSnow)P(NoSnow)$$

$$= 0.8 * 0.05 + 1 * 0.95 = 0.99$$

2. Bayes' rule states that $p(A|B) = p(B|A)$.

   False. Recall Bayes' rule:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

3. If $P(A \cap B) = 0.4$ and $P(B) = 0.8$, then $P(A|B) = 0.5$.

   True. $P(A|B) = P(A \cap B)/P(B)$ *by Bayes rule.*

3 Variance and Covariance

1. If the sample covariance between two variables is one, then there must be a strong linear relationship between the variables

   False. $Cov(X, Y) = Corr(X, Y)\sqrt{Var(X) * Var(Y)}$, *so knowing* $Cov(X, Y) = 1$ *doesn't inform you about the correlation*

2. If X and Y are independent random variables, then $Var(2X - Y) = 2Var(X) - Var(Y)$.

   False. $Var(2X - Y) = 4Var(X) + Var(Y)$

3. The sample variance is unaffected by outlying observations.

   False. We have the following formula for the sample variance:

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

   If x_i is large relative to $\bar{x}$ then it has an undue influence.

4. Suppose that a random variable X can take the values $\{0, 1, 2\}$ all with equal probability. Then the expected and variance of X are both 1.

   False

$$E[X] = \frac{1}{3} * 0 + \frac{1}{3} * 1 + \frac{1}{3} * 2 = 1$$

$$\begin{aligned} Var[X] &= \frac{1}{3}(0-1)^2 + \frac{1}{3}(1-1)^2 + \frac{1}{3}(2-1)^2 \\ &= \frac{1}{3} + 0 + \frac{1}{3} = \frac{2}{3} \end{aligned}$$

5. The maximum correlation is 1 and the minimum is 0.

   False, the maximum is 1 and the minimum in -1.

6. For independent random variables X and Y, we have $var(X-Y) = var(X) - var(Y)$.

   False. $Var(X-Y) = Var(X) + Var(Y)$

7. If the correlation between X and Y is zero then the standard deviation of $X+Y$ is the square root of the sum of the standard deviations of X and Y.

   False. *Using the plug in rule, the standard deviation of $X+Y$ is the square root of the sum of the variances of X and Y.*

8. It is always true that the standard deviation is less than the variance

   False. $s_X \geq Var(X)$ *if the* $Var(X) \leq 1$.

9. If the correlation between X and Y is $r = -0.81$ and if the standard deviations are $s_X = 20$ and $s_Y = 25$, respectively, then the covariance is $Cov(X,Y) = -401$.

   False. Using the formula relating covariance and correlation, namely $Cov(X,Y) = rs_Xs_Y$ *gives* $Cov(X,Y) = -0.81 \times 20 \times 25 = -405$.

10. If we drop the largest observation from a sample, then the sample mean and variance will both be reduced.

False. For example, $100, 101, 102$ *has mean* 101 *and variance* 0.333*, but* $100, 101$ *has mean* 100.5 *and variance* 0.5*.*

11. Suppose X and Y are independent random variables and $Var(X) = 6$ and $Var(Y) = 6$. Then $Var(X + Y) = Var(2X)$.

False. First, $Var(X + Y) = Var(X) + Var(Y)$ *(as* $Cov(X, Y) = 0$*) and so* $Var(X + Y) = 2Var(X)$*. Secondly* $Var(2X) = 4Var(X)$

12. Let investment X have mean return 5% and a standard deviation of 5% and investment Y have a mean return of 10% with a standard deviation of 6%. Suppose that the correlation between returns is zero. Then I can find a portfolio with higher mean and lower variance then X.

True. For example, consider the portfolio $P = 0.5X + 0.5Y$*. This has expected return* $\mu_P = 0.5 \times 5 + 0.5 \times 0.10 = 7.5\%$*. As the correlation is zero, its variance is given by* $\sigma_P^2 = 0.5^2 \times 0.0025 + 0.5^2 * 0.0036 = 0.001525 < Var(X)$*.*

4 Expectation

1. LeBron James makes 85% of his free throw attempts and 50% of his regular shots from the field (field goals). Suppose that each shot is independent of the others. He takes 20 field goals and 10 free throws in a typical game. He gets one point for each free throw and two points for each field goal assuming no 3-point shots. The number of points he expects to score in a game is 28.5.

   True.

   $E(points) = 2 * E(\#FGmade) + 1 * E(\#FTmade) = 2 \times 20 \times 0.5 + 1 \times 10 \times 0.85 = 28.5$

   This holds even if FG and FT are dependent.

2. Suppose that you have a one in a hundred chance of hitting the jackpot on a slot machine. If you play the machine 100 times then you are certain to win.

   False. The expected waiting time until you hit the jackpot is 100 times, but since each outcome of the game is independent of all previous outcomes, there is no guarantee about the number of plays until a jackpot it hit.

3. The expected value of the sample mean is the population mean, that is $E\left(\bar{X}\right) = \mu$.

   True. *The expected value of the sample does equal the true value.*

4. The expectation of X minus $2Y$ is just the expectation of X minus twice the expectation of Y, that is $E(X - 2Y) = E(X) - 2E(Y)$.

True. The plug-in rule states that $E(aX + bY) = aE(X) + bE(Y)$. *Plug in* $A = 1$ *and* $B = -2$. *Hence* $E(X - 2Y) = E(X) + E(-2Y) = E(X) - 2E(Y)$

5. A firm believes it has a 50-50 chance of winning a $80,000 contract if it spends $5,000 on a proposal. If the firm spends twice this amount, it feels its chances of winning improve to 60%. If the firm wants to maiximsie its expected value then it should spend $10,000 to try and gain the contract.

True. E(spend $5k) = 0.5·80000 - 5000 = 35000

E(spend $10k) = 0.6·80000 - 10000 = 38000

6. $E(X + Y) = E(X) + E(Y)$ only if the random variables X and Y are independent.

False. By the plug-in rule, this relation holds irrespective of whether X *and* Y *are independent.*

5 Binomial Distribution

1. Suppose that you toss a fair coin with probability 0.5 a head. The probability of getting five heads is a row is less than three percent.

   False. Given a fair coin with 50% probability for a head, $0.5^5 = 0.0312. > 0.03$

2. Suppose that you toss a biased coin with probability 0.25 of getting a head. The probability of getting five heads out of ten tosses is less than thirty percent.

   True This follows from a binomial distribution with $n = 10$.
   $Prob(5H) = \left(\begin{array}{c} 10 \\ 5 \end{array} \right) (0.25)^5 (0.75)^5 = 0.0583$

3. Suppose that you toss a coin 5 times. Then there are 10 ways of getting 3 heads.

   True. We see this by computing:

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = 10$$

4. The probability of observing three heads out of five tosses of a fair coin is 0.6.

   False. $P(3\ heads) = 0.5^5 = 0.03125$. 10 *such combinations are possible so the probability is* 0.3125.

5. A mortgage bank knows from experience that 2% of residential loans will go into default. Suppose it makes 10 such loans, then the probability that at least one goes into default is 95%.

   False. Using $P(\text{at least one}) = 1 - P(\text{none})$ *gives* $1 - (0.98)^{10} = 0.1829$.

6. Jessica Simpson is not a professional bowler and 40% of her bowling swings are gutter balls. She is planning to take 90 blowing swings. The mean and standard deviation of the number of gutter balls is $\mu = 36$ and $\sigma = 3.65$.

   False. From a Binomial distribution $E(X) = np = 90 \times 0.4 = 36$ *and the standrad deviation* $s_X = \sqrt{np(1-p)} = \sqrt{21.6} = 4.65$.

7. The probability of at least one head when tossing a fair coin 4 times is 0.9375.

   True. $P(\text{at least one head}) = 1 - (0.5)^4 = 0.9375$.

8. The Red Sox are to play the Yankees in a seven game series. Assume that the Red Sox have a 50% chance of winning each game, with the results being independent of each other. Then the probability of the series ending 4-3 in favor of the Red Sox is $0.5^7 = 0.0078$.

   False. The total wins of the Red Sox is a binomial random variable with $n = 7$ *and* $p = 0.5$. *The probability of* 4 *wins is then* $\binom{7}{4}0.5^4 0.5^{7-4} = 35 \cdot (0.5)^7 = 0.27$. *Alternatively, if the series is stopped as soon as one time wins 4 matches, the probability of a 4-3 outcome in favor of the Red Sox is* $\binom{6}{3}0.5^3 0.5^{6-3} \cdot 0.5 = 0.16$ *since the teams should first reach a 3-3 tie.*

9. Suppose that X is Binomially distributed with $E(X) = 5$ and $Var(X) = 2$, then $n = 10$ and $p = 0.5$.

   False. The variance of a $Bin(10, 0.5)$ random variable is $np(1-p) = 2.5$ and not 2.

10. If X is a Bernoulli random variable with probability of success, p, then its variance is $V(X) = p(1-p)$.

    True. The Bernoulli random variable is the building block of the binomial ($n = 1$) and hence has variance $p(1-p)$

11. Historically 15% of chips manufactured by a computer company are defective. The probability of a random sample of 10 chips containing exactly one defect is 0.15.

    False. The probability of one defectuce is given by the binomial probability $10 \times 0.15 \times (0.85)^9 = 0.3474$.

6 Poisson Distribution

1. Arsenal are playing Burnley at home in an English Premier League (EPL) game this weekend. They are favourites to win. They have a Poisson distribution for the number of goals they will score with a mean rate of 2.5 per game. Given this, the odds of Arsenal scoring at least two goals is greater than 50%.

   True. Here $S \sim Pois(2.5)$*. We need* $P(S \geq 2) = 1 - P(S \leq 1) = 1 - 0.2873 = 0.7127$*. In* `R` *we have* $ppois(1, 2.5) = 0.28729$*.*

2. Arsenal are playing Liverpool at home in an EPL game this weekend. You think that the number of goals to be scored by both teams follow a Poisson distribution with rates 2.2 and 1.6 respectively. Given this, the odds of a scoreless $0-0$ draw are $45-1$.

   False Assuming each scoring's ability is independent, we have that:

$$
\begin{aligned}
Pr(Arsenal = 0) &= \frac{(2.2)^0 e^{-2.2}}{0!} = e^{-2.2} \\
Pr(Liverpool = 0) &= \frac{(1.6)^0 e^{-1.6}}{0!} = e^{-1.6} \\
\longrightarrow \quad Pr(Arsenal = 0 \; and \; Liverpool = 0) &= e^{-2.2} \times e^{-1.6} = e^{-3.8}
\end{aligned}
$$

   The odds are: $O = (1 - e^{-3.8})/e^{-3.8}$ or 43.7 to 1.

3. Suppose your website gets on average 2 hits per hour. Then the probability of at least one hit in the next hour is 0.135.

   False, $dpois(0, 2) = 0.135 = e^{-2}$ *so that* $P(hit > 1) = 1 - P(hit = 0) = 0.865$

4. The soccer team Manchester United scores on average two goals per game. Given that the distribution of goals is Poisson, the chance that they score two or less goals is 87%

False. We have that :

$$E[g] = 2 \quad \Rightarrow \quad g \sim Pois(\lambda = 2)$$

So, we find the chance that they score two or fewer goals is:

$$P(g \leq 2) = 0.6767$$

In `R`: `ppois(2,2)=0.67667`

7 Normal Distribution

1. For any normal random variable, X, we have $\mathbb{P}(\mu - \sigma < X < \mu + \sigma) = 0.64$.
   [Hint: You may use $pnorm(1) = 0.841$]

   False. $pnorm(1) = 0.841$ *and* $2 \times pnorm(1) - 1 = 0.6828 > 0.64$.

2. Suppose that the annual returns for Facebook stock are normally distributed with a mean of 15% and a standard deviation of 20%. The probability that Facebook has returns greater than 10% for next year is 60%

   True. Let R denote returns. Convert to standard normal and compute

$$\mathbb{P}\left(\frac{R-\mu}{\sigma} > \frac{0.1-\mu}{\sigma}\right) = \mathbb{P}(Z > -0.25) = 0.5987$$

   where $\mu = 0.15$ *and* $\sigma = 0.2$.

3. Consider the standard normal random variable $Z \sim N(0,1)$. Then the random variable $-Z$ is also standard normal.

   True. If $X \sim N(0,1)$, then $-X \sim N\Big(-1(0), (-1)^2 \times 1\Big) = N(0,1)$. Moreover, we have that the pdf of Z and $-Z$ is the same:

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2} z^2 = \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}(-z)^2 = f(-z)$$

   Lastly, we can observe that all the even moments are the same due to symmetry and all the odd moments are zero. Thus, they are the same distribution.

4. A local bank experiences a 2% default rate on residential loans made in a certain city. Suppose that the bank makes 2000 loans. Then the probability of more than 50 defaults is 25 percent.

False The probability of having more than 50 defaults can be calculated using a Normal approximation to the $Bin(2000, 0.02) \approx N(2000 \times 0.02, 2000 \times 0.02 \times 0.98)$. So, the mean is 40 and the standard deviation is 6.261. Therefore,

$$Pr(default > 50) = Pr\Big(\frac{default - 40}{6.261} - \frac{50 - 40}{6.261}\Big) = Pr(z > 1.5972) \approx .055$$

5. The Binomial distribution can be approximated by a normal distribution when the number of trials is large.

True, if n is large enought, a reasonable approximation to $B(n, p)$ is $N(np, np(1-p))$.

6. Let $X \sim N(5, 10)$. Then $P(X > 5) = \frac{1}{2}$.

True, since the mean is 5, then half the density will be to the right of 5.

7. Shaquille O'Neal has a 55% chance of making a free throw in Basketball. Suppose he has 900 free throws this year. Then the chance he makes more than 500 free throws is 45%

False. $\mu = np = 495$, $\sigma = \sqrt{np(1-p)} = 14.92$. *Then* $P(X > 500) = P(Z > 0.335) = 1 - 0.631 = 36.9\%$

8. Suppose that the random variable $X \sim N(-2, 4)$ then $-2X \sim N(4, 16)$.

True. We have $-2X \sim N(4, 16)$.

9. Mortimer's steak house advertises that it is the home of the 16 ounce steak. They claim that the weight of their steaks is normally distributed with mean 16 and standard deviation 2. If this is so, then the probability that a steak weights less that 14 ounces is 16%.

   True. We have $P(X < 14) = P\left(\frac{X-16}{2} < \frac{14-16}{2}\right) = P(Z < -1) = 0.1587$

10. Advertising costs for a 30-second commercial are assumed to be normally distributed with a mean of $10,000$ and a standard deviation of 1000. Then the probability that a given commercial costs between 9000 and $10,000$ is 50%.

    False. Required probability is $0.5 - P(Z < -1) = 0.3413$

11. In a sample of 120 Zagat's ratings of Chicago restaurants, the average restaurant had a rating of 19.6 with a standard deviation of 2.5. If you randomly pick a restaurant, the chance that you pick one with with a rating over 25 is less than 1%.

    False. $P(X > 25) = 1 - P(Z < 2.16) = 0.0154$

12. A hospital finds that 20% of its bills are at least one month in arrears. A random sample of 50 bills were taken. Then the probability that less than 10 bills in the sample were at least one month in arrears is 50%

    True. Using normal approximation, $n = 50, p = 0.2$, *have* $X \sim N(10, 8)$.

13. A Chicago radio station believes 30% of its listeners are younger than 30. Out of a sample of 500 they find that 250 are younger than 30. This data supports their claim at the 1% level.

False. $n = 500$, $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$ *and* $H_0 : p = 0.3$ *versus* $H_1 : p \neq 0.3$. *Yields a 99% confidence interval of* $(0.1835, 0.4165)$. *Sample proportion doesn't lie in confidence interval so the data does not support their claim. Or could use a Z-test.*

14. The probability that a standard normal distribution is more than 1.96 standard deviations from the mean is 0.05.

True. In a normal distribution, 2.5% of the population is more than 1.96 standard deviations below the mean, and 2.5% is more than 1.96 standard deviations above the mean. Hence 5% of the distribution is more than 1.96 standard deviations from the mean.

15. A Chicago radio station believes 30% of its listeners are younger than 30. Out of a sample of 500, the probability that more than 200 are under 30 is 0.25.

False. The number of listeners under 30 is a binomial variable with $n = 500$ *and* $p = 0.3$. *This can be approximated by a* $N(np, np(1-p))$. *The Z-value associated with 200 is* $Z = \frac{200-150}{\sqrt{105}} = 4.88$ *and* $1 - F(z) = 0$. *So there is a zero probability of finding more than* 200 *listeners under* 30.

16. Suppose that the amount of money spent at Disney World is normally distributed with a mean of \$60 and a standard deviation of \$15. Then approximately 45% of people spend more than \$70 per visit.

False. $z_{70} = \frac{70-60}{15} = 0.67$ and 1-F$(z_{70}) = 0.25$. Hence, about 25% of people spend more than \$70 per visit.

17. A Normal distribution with mean 4 and standard deviation 3.6 will provide a good approximation to a Binomial random variable with parameters $n = 40$ and $p = 0.10$.

False. For $n \geq 40$, we can approximate a Binomial distribution $Bin(n, p)$ with a normal $N(np, np(1-p))$ In this case, we get $np = 4$ and variance $40 \times 0.1 \times 0.9 = 3.6$ or a standard deviation of 1.90.

18. If X is normally distributed with mean 3 and variance 9 then the probability that X is greater than 1 is 0.254.

False. The stanardized Z-score is given by $Z = (1-3)/3 = -0.667$. From tables $P(X > 1) = P(Z > -0.667) = P(Z < 0.667) = 0.7486$

8 Hypothesis Testing

1. Given a random sample of 2000 voters, 800 say they will vote for Hillary Clinton in the 2016 US Presidential Election. At the 95% level, I can reject the null hypothesis that Hillary has an evens chance of winning the election.

   True. We have $H_0 : p = 0.5$ *and* $\hat{p} = 0.4$. *The t-statistic is* $T = (\hat{p} - 0.5)/\sqrt{\frac{p\times(1-p)}{n}} = -9$. *Which is less than our threshold* $qnorm(0.025) = -1.96$. *We have enough evidence to reject the null hypothesis.*

2. The average movie is Netflix's database has an average customer rating of 3.1 with a standard deviation of 1. The last episode of *Breaking Bad* had a rating of 4.7 with a standard deviation of 0.5. The $p - value$ for testing whether Breaking Bad's rating is statistical different from the average is a lot less than 1%.

   False The formula for a T-ratio is

   $$T = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s^2_{x_1}}{n_1} + \frac{s^2_{x_2}}{n_2}}}$$

   However, since we do not know n_1 and n_2 we cannot get the T-ratio and thus p-value.

3. A chip manufacturer needs to add the right amount of chemicals to make the chips resistant to heat. On average the population of chips needs to be able to withstand heat of 300 degrees. Suppose you have a random sample of 30 chips with a mean of 305 and a standard deviation of 8. Then you can reject the null hypothesis $H_0 : \mu = 300$ versus $H_a : \mu > 300$ at the 5% level.

   True. Since the sample size is $n \geq 30$ *the t test is effectively the same as the z test. Therefore, we find the following z value:*

   $$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{305 - 300}{8/\sqrt{30}} = 3.4233$$

Here we use a one-sided test:

$$z_\alpha = 1.65$$

Since $z > z_\alpha$, we can reject H_0 in favor of H_a at the 5% level.

4. Zagats rates restaurants on food quality. In a random sample of 100 restaurants you observe a mean of 20 with a standard deviation of 2.5. Your favorite restaurant has a score of 25. This is statistically different from the population mean at the 5% level.

   True. *We can calculate the t statistic as $(25 - 20)/(2.5/10) > 1.96$. Hence we reject the null hypothesis.*

5. The t-score is used to test whether a null hypothesis can be rejected.

   True. *The t score can be used to test whether a null hypothesis can be rejected or not. Alternately, a z score can be used.*

6. An oil company introduces a new fuel that they claim has on average no more than 100 milligrams of toxic matter per liter. They take a sample of 100 liters and find that $\bar{X} = 105$ with a given $\sigma_X = 20$. Then there is evidence at the 5% level that their claim is wrong.

   True. This is a one-side test $H_0 : \mu_x \leq 100$ versus $H_1 : \mu_X > 100$. The Z-score is $Z = \frac{105-100}{20/\sqrt{100}} = 2.5$. The critical value for a 1-sided test at the 5% level is 1.64. Therefore you can reject H_0 are the 5% level.

7. A wine producer claims that the proportion of customers who cannot distinguish his product from grape juice is at most 5%. For a sample of 100 people he finds that 10 fail the taste test. He should reject his null hypothesis $H_0 : p < 0.05$ at the 5% level.

   True. $H_0 : p \geq 0.05$ and $H_1 : p < 0.05$

$z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.1-0.05}{\sqrt{\frac{0.05\times0.95}{100}}} = 2.3 \ and \ x \ z_\alpha = 1.64$

We reject the null-hypothesis since $|z| > z_\alpha$

9 Confidence Intervals

1. There is much discussion of the effects of second-hand smoke. In a survey of 500 children who live in families where someone smokes, it was found that 10 children were in poor health. A 95% confidence interval for the probability of a child living in a smoking family being in poor health is then 2% to 4%.

   False The 95% confidence interval should be:

   $$\left[.02 - 1.96\sqrt{\frac{.02(1-.02)}{500}}, .02 + 1.96\sqrt{\frac{.02(1-.02)}{500}}\right]$$

   So, the lower bound on this interval will definitely be less than 2 percent.

2. You are finding a confidence interval for a population mean. Holding everything else constant, an interval based on an unknown standard deviation will be wider than one based on a known standard deviation no matter what the sample size is.

   True Because you need to use the same data to estimate the standard deviation, and thus contains more noise or error. In other words, we are comparing between z distribution and t distribution.

3. There is a 95% probability that a normal random variable lies between $\mu \pm \sigma$.

   False. There is a 95% probability that it lies between $\mu \pm 1.96\sigma$.

4. A recent CNN poll found that 49% of 10000 voters said they would vote for Obama versus Romney if that was the election in November 2012. A 95% confidence interval for the true proportion of voters that would vote for Obama is then 0.49 ± 0.03.

False. The confidence interval for the proportion is given by:

$$\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n}$$

This gives:

$$0.49 \pm 1.96\sqrt{(0.49)(0.51)/10000} = 0.49 \pm 0.009798$$

5. A mileage test for a new electric car model called the "Pizzazz" is conducted. With a sample size of $n = 30$ the mean mileage for the sample is 36.8 miles with a sample standard deviation of 4.5. A 95% confidence interval for the population mean is $(32.3, 41.3)$ miles.

False. $CI = 36.8 \pm 1.96\frac{4.5}{\sqrt{30}} = (35.19, 38.41)$

6. In a random sample of 100 NCAA basketball games, the team leading after one quarter won the game 72 times.

   Then a 95% confidence interval for the proportion of teams leading after the first quarter that go on to win is approximately $(0.6, 0.84)$.

False. The 95% CI is approximately equal to

$$.72 \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.72 \pm 1.96 \times 0.045 = (.63, .81)$$

7. For the same sample, a 95% prediction interval for a particular team winning is also $(0.6, 0.84)$.

False. The prediction interval is equal to

$$0.72 \pm 2\sqrt{\hat{p}(1-\hat{p})}\sqrt{1+1/n}$$

In this case, that contains the entire interval (0,1).

8. In playing poker in Vegas, from 100 hours of play, you make an average of \$50 per hour with a standard deviation of \$10. A 95% confidence interval for your mean gain per hour is approximately \$ $(48, 52)$.

True. The CI is given by $50 \pm 2 \times \frac{10}{\sqrt{100}} = (48, 52)$

9. If 27 out of 100 respondents to a survey state that they drink Pepsi then a 95% confidence interval for the proportion p of the population that drinks Pepsi is $(0.26, 0.28)$.

False. The 95% confidence interval for the proportion is given by $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n}$*. Which here gives* $0.27 \pm 1.96\sqrt{0.27 \times 0.73/100}$ *or the interval* $(0.18, 0.36)$.

10. The p-value is the probability that the Null hypothesis is true.

False. The p-value is the probability of the observing something more extreme than the observed sample assuming the null hypothesis is true.

10 Sampling Distributions

1. The Central Limit Theorem states that the distribution of a sample mean is approximately Normal.

   True. The standard deviation of the sample mean is $\frac{\sigma}{\sqrt{n}}$. *The CLT has (approximately)* $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$, *where* μ *and* σ^2 *are the population true value.*

2. The Central Limit Theorem states that the distribution of the sample mean $\bar{X}$ is Normally distributed for large samples.

   True. If X_i is a random sample (or *iid*) with mean μ, then $\sqrt{n}(\bar{X} - \mu) \longrightarrow^d N(0, \sigma^2)$

3. The Central Limit Theorem guarantees that the distribution of $\bar{X}$ is constant.

   False. For some set of independent random variables, the mean of a sufficiently large number of them (each with finite mean and variance) will have a distribution which is Normal.

4. The Central Limit Theorem says that the distribution of the sample mean is approximately normal.

   True. *CLT says that for large sample sizes, the mean of the sample is distributed normally and approaches the true population mean*

5. The sample mean, $\bar{x}$, approximates the population mean for large random samples.

True. The CLT states that the $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$*, or the best guess for the population mean is* $\bar{x}$*.*

6. The Central Limit Theorem allows us to approximate the distribution of a sample average by a Normal distribution.

True. The Central Limit Theorem says that the distribution of the sample average is approximately normal, specifically $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$*.*

7. In a Markov process, tomorrow's value is independent of today's.

False. Tomorrow's value is only dependent on today's by the Markov property.

11 Descriptive Statistics

1. The trimmed mean of a dataset is more sensitive to outliers than the mean.

   False. The trimmed mean is the average of the observations deleting the outer 5% in the tails. Hence it is less sensitive to outliers than the sample mean

2. The sample mean of a dataset must be larger than its standard deviation

   False. Many datasets have a negative mean and the standard deviation must be positive.

3. Selection bias is not a problem when you are estimating a population mean.

   False. When sampling to estimate a population mean, you need to be certain to select a random sample, or selection bias may effect your results. The Dewey/Truman election example is the classic example of selection bias

4. The kurtosis of a distribution is not sensitive to outliers.

   False. *Kurtosis depends on each observation in the distribution and is extremely sensitive to the outliers.*