Version: January 27, 2018 Spring 2018
EC200 Econometrics and Applications

Statistics Quiz

You have 75 minutes to complete this quiz. There are 30 total points. Please round to three decimal places when necessary.

Because the normal distribution table is slightly imprecise, you may have to round to the nearest value. You will not be penalized for whether you round up or down.

1. *(6 points)* Consider two random variables. D is the number of dogs owned by residents of La Brea county. C is the number of cats owned by the residents of La Brea county. Because of strict local ordinances, no La Brea resident is allowed to have more than 2 dogs and 2 cats.

Joint probability distribution of dog and cat ownership in La Brea county

	No cats	1 cat	2 cats
No dogs	0.34	0.15	0.08
1 dog	0.14	0.10	0.01
2 dogs	0.10	0.05	0.03

(a) Write or draw the marginal probability distribution of D. Write or draw the marginal probability distribution of C. *(2 points)*

C	0	1	2
P_C	0.58	0.30	0.12

D	0	1	2
P_D	0.57	0.25	0.18

(b) Compute $E(C)$ and $E(D)$ and interpret your results. *(2 points)*

$E[D] = \sum_d dP_D(d) = 0(0.57) + 1(0.25) + 2(0.18) = 0.61$

$E[C] = \sum_c cP_C(c) = 0(0.58) + 1(0.3) + 2(0.12) = 0.54$

The average number of dogs owned by La Brea residents is 0.61, and the average number of cats is 0.54.

(c) What is the expected number of dogs owned by La Brea residents with no cats? *(2 points)*

First calculate conditional probabilities

$P(D=0|C=0) = \frac{P(D=0,C=0)}{P(C=0)} = \frac{0.34}{0.58} = 0.586$

$P(D=1|C=0) = \frac{P(D=1,C=0)}{P(C=0)} = \frac{0.14}{0.58} = 0.241$

$P(D=2|C=0) = \frac{P(D=2,C=0)}{P(C=0)} = \frac{0.10}{0.58} = 0.172$

$$\begin{aligned} E[D|C=0] &= \sum_d dP_D(d|C=0) \\ &= 0(0.586) + 1(0.241) + 2(0.172) \\ &= 0.586 \end{aligned}$$

2. *(10 points)* The UVM development office has provided you with anonymous matches of starting salaries and GPAs for 121 graduating economics majors. The average starting salary for the 108 students was \$39,000 with a standard deviation of \$7,000.

(a) Construct a 95% confidence interval for the starting salary of all economics majors at your university/college. *Note that when $n > 100$, the t-distribution can be approximated with a z-distribution.* *(3 points)*

$$\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{121}}$$

$$39000 \pm 1.96\frac{7000}{\sqrt{121}}$$

$$39000 \pm 1247.273$$

$$[37752.727, 40247.273]$$

(b) Suppose that nationally, the average starting salary for economics graduates is \$40,500. Formally test the null hypothesis that the average starting salaries for UVM economics graduates are different from the national average at the 1-percent level. Make sure to report the null hypothesis, alternate hypothesis, decision rule, test statistic, and your conclusion. *(4 points)*

$H_0 : \mu = 40,500$
$H_1 : \mu \neq 40,500$

Reject null if $t > c_{\alpha/2} = 2.57$ or $t < -c_{\alpha/2} = -2.57$

$t = \frac{39000-40500}{7000/\sqrt{121}} = -2.357$

Do not reject the null hypothesis.

(c) Calculate the p-value of the test in part (b). Explain in one sentence what it means. *(3 points)*

The p-value represents the probability we get a value as extreme or more as the test statistic we calculated, if the null hypothesis is true.

$P(t < -2.357) + P(t > 2.357) = 2(1 - P(t < 2.357))$

$= 2(1 - 0.9909) = 0.0182$

This also means that we could reject any null hypothesis level at the 1.8% level or higher using a two-sided test

3. *(10 points)* The Sportsball Grand Championship is coming up! The Minnesota Aardvarks and playing the Scranton Beagles. In sportsball, whoever has the most points wins. In Sportsball, your team earns three points per “basket" (putting the ball in the other team's basket). Additionally, your team loses one point every time a player on your team is hit with the ball, called “getting tagged.”

   The number of baskets that the Aardvarks score per game (B) is distributed normally with a mean of 30 and a standard deviation of 20. The number of times they get tagged (T) is distributed normally with an mean of 60 and a standard deviation of 30. The covariance between B and T is -20.

   Note that even though baskets and tags are technically discrete, they happen frequently enough that they can be well-modeled by a continuous normal distribution. Recall that a linear combination of two normally distributed variables remains normal.

   (a) What is the expected value of the Aardvarks' total score per game (S)? *(2 points)*

   $E[S] = E[3B - T] = 3E[B] - E[T] = 3 * 30 - 60 = 30$

(b) What is the standard deviation of the Aardvark's total score per game? *(2 points)*

$$Var[S] = Var[3B - T] = 9Var[B] + Var[T] - 2(3)Cov(B,T)$$

$9(20^2) + 30^2 - 2(3)(-20) = 3600$0 $+ 900 -$ + 120

~~1140~~ — 4620

~~$\sigma_S = \sqrt{1140} = 33.764$~~ = 67.97

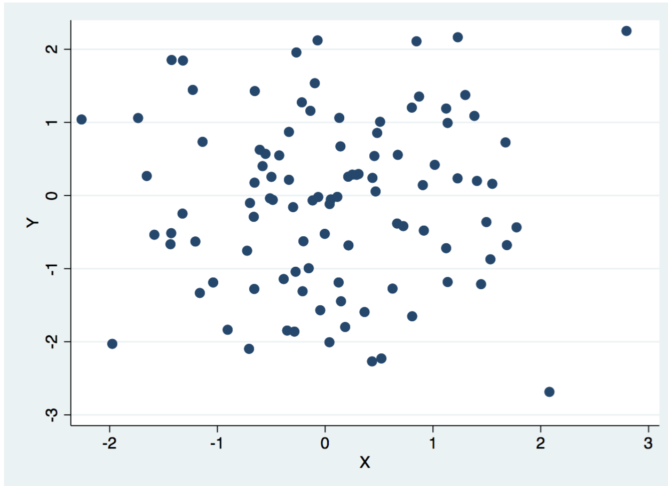
(c) What is the probability the Aardvarks score more than 40 **baskets** this game? *(3 points)*

$$P(B > 40) = P(Z > \frac{40 - 30}{20}) = P(Z > 0.5)$$
$$= 1 - P(Z < 0.5) = 1 - \Phi(0.5)$$
$$= 1 - 0.6915$$
$$= 0.3085$$

(d) What is the probability the Aardvarks end up with a negative total score his game? *(3 points)*

$$P(S > 5) = P(Z < \frac{0 - 30}{\sqrt{1184}}) = P(Z < -0.889)$$
$$= 1 - P(Z < 0.889) = 1 - \Phi(0.89)$$
$$= 1 - 0.8133$$
$$= 0.1867$$

4. *(4 points)* Consider the following scatter plot of two random variables, X and Y.



5. Based on the figure, would you say that $Cov(X, Y)$ is positive, negative, or zero? Justify your answer briefly for full credit. *(2 points)*

Eyeballing the graph, it looks like the covariance is around zero. This is because there appears to be no linear relationship between the two variable - if you were to draw a best fit line, it would look approximately flat.

6. Based on the figure and your previous answer, can you say that X and Y are statistically dependent, statistically independent, or is there not enough information? Justify your answer briefly for full credit. *(2 points)*

Even though the covariance is likely to be zero, we cannot say whether they are statistically independent or not. If they are statistically independent, then the covariance is zero, but the converse is not necessarily true. We can only rule out a linear relationship.

Total: 30
Subtotal: 30