Name: KEY
TA's name: (Roshni, Gil, Remi, Barbara, Libby, Kenny)
Student number:

**MID-TERM BIOL 300: October 2019**

***For all statistical tests, make sure that you clearly state your hypotheses. Unless otherwise stated, assume $\alpha = 0.05$. Show your work. Be as precise as possible about P-values.***

***By taking this test and putting your name above, you are declaring that your answers on this test are all your own work.***

***Make sure that your copy of the test includes 8 pages, including this one.***

| | Points |
|---|---|
| Q1 | 9 |
| Q2 | 6 |
| Q3 | 15 |
| Q4 | 10 |
| Q5 | 20 |
| Q6 | 10 |
| Q7 | 12 |
| Q8 | 18 |
| | 100 |

**1**. The Cook pine tree is native to New Caledonia but has been planted as an ornamental in many countries. The trees tend to lean to one side as they grow. Johns et al. (2017) measured the tilt in 256 individual trees around the world. They found that 233 trees leaned toward the equator (in the Northern hemisphere they lean south, and in the Southern hemisphere they lean north), whereas 23 leaned away from the equator.

**a.** (*3 points*)What is the best estimate of the proportion of Cook pine trees that lean toward the equator?

$\hat{p} = \frac{233}{256} =$

**Answer:** 0.910

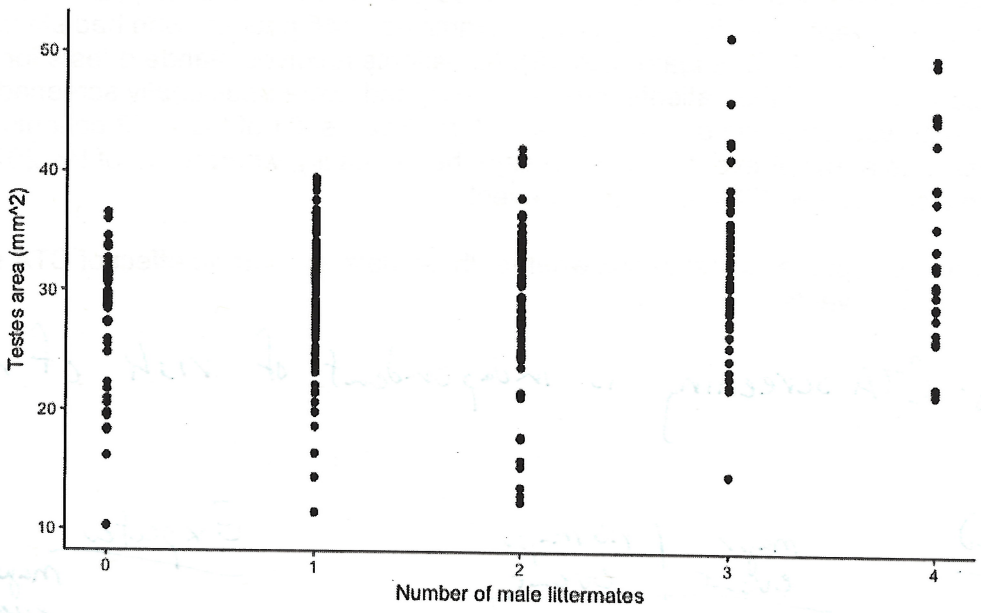
**b.** (*6 points*)What is the 95% confidence interval for the proportion of trees leaning toward the equator?

$p' = \frac{x+2}{N+4} = \frac{233+2}{256+4} = 0.904$

**Answer:** $0.868 < p < 0.940$

$p' \pm 1.96\sqrt{\frac{p'(1-p')}{n+4}} = 0.904 \pm 0.036$

**2.** (*3 points each*)




a. Identify the type of graph above.

scatterplot (strip chart also accepted)

b. What is the main function in R that can be used to make a graph like this?

ggplot

**3.** (*15 points*)New technical advances in diagnosis and screening are often expensive, and so their benefits must be evaluated. A study of a new screening procedure for heart disease, computerized tomographic angiography (CTA), examined 4146 patients who had stable chest pains (The SCOT-HEART Investigators, 2018). All patients received standard tests for coronary heart disease, but 2073 of the patients, randomly assigned, were additionally screened with CTA, which improves detection of heart disease. After 5 years, 81 of the 2073 patients receiving standard tests had a "major event" (death or major heart attack), whereas 48 of the 2073 patients who also received CTA had a major event.

Do an appropriate hypothesis test to ask whether these data support an effect of CTA screening on the risk of major events.

$H_0$: CTA screening is independent of risk of major events.

Observed

| | major event | no major event | |
|---|---|---|---|
| standard tests | 81 | 1992 | 2073 |
| CTA | 48 | 2025 | 2073 |
| | 129 | 4017 | |

Expected

| | major event | no major event |
|---|---|---|
| standard | 64.5 | 2008.5 |
| CTA | 64.5 | 2008.5 |

$$\chi^2 = \frac{(81-64.5)^2}{64.5} + \frac{(48-64.5)^2}{64.5} + \frac{(1992-2008.5)^2}{2008.5} + \frac{(2025-2008.5)^2}{2008.5} = 8.713$$

$df = (2-1)(2-1) = 1$

$\chi^2_{0.005,1} = 7.88$ ; $8.713 > 7.88$ ; therefore $P < 0.005$

We reject the null hypothesis; CTA screening is associated with a lower risk of major events.

**4.** (*10 points*)The time it takes an individual popcorn kernel to pop conforms to a normal distribution very closely. Assume that the mean time to popping is 165.0 seconds with a standard deviation of 92.1 seconds.

What is the probability that a kernel of popcorn would pop in less than 120 seconds?


**Answer:** 0.312

$$Z = \frac{(120 - 165)}{92.1} = -0.488$$

$$Pr[Z < -0.488] = Pr[Z > 0.488]$$

$$= 0.312$$

**5.** (*4 points each*) Just how slippery are banana peels on the floor? "Slipperiness" can be measured by the frictional coefficient. A small value for the coefficient indicates less friction (more slippery) than a larger value. Ice on steel has a frictional coefficient of about 0.02, whereas wood on concrete has a coefficient of 0.6. Below are 41 measurements of the frictional coefficients of banana skins on linoleum, with the slippery side down (Mabuchi et al. 2012).

0.029, 0.034, 0.036, 0.037, 0.039, 0.041, 0.041, 0.042, 0.043, 0.043, 0.045, 0.047, 0.047, 0.048, 0.048, 0.048, 0.053, 0.054, 0.054, 0.054, 0.055, 0.055, 0.056, 0.056, 0.057, 0.059, 0.060, 0.062, 0.062, 0.063, 0.064, 0.065, 0.069, 0.069, 0.070, 0.071, 0.075, 0.078, 0.080, 0.113, 0.125

$n = 41; \sum Y = 2.347; \sum Y^2 = 0.148515$

**a.** What is the mean frictional coefficient for banana skins on linoleum?

$\frac{2.347}{41} =$

**Answer:** 0.0572

**b.** What is the median frictional coefficient for banana skins on linoleum?

21st (middle) observation is 0.055

**Answer:** 0.055

**c.** What is the standard deviation of the frictional coefficient for banana skins on linoleum?

$s = \sqrt{\frac{0.148155 - 41\left(\frac{2.347}{41}\right)^2}{41-1}} =$

**Answer:** 0.0188

**d.** What is the 95% confidence interval for the mean banana peel frictional coefficient?

$\bar{Y} \pm t_{0.05(2),40}\sqrt{\frac{s^2}{n}}$

$= 0.0572 \pm 2.02\frac{0.0188}{41}$

**Answer:** $0.0513 < \mu < 0.0632$

**e.** Based on this confidence interval, are banana peels less slippery than ice?

yes (ice value not included in confidence interval given in (d))

**6.** (*10 points*) A standard deck of cards has 52 cards, half of which are red and half of which are black. Imagine that you deal out 5 separate cards from a well-shuffled deck. What is the probability that all five of these cards are red? What is the probability that all five cards are all the same color?

$\text{Prob}[1^{st} \text{ card is red}] = \frac{26}{52}$

$\text{Prob}[2^{nd} \text{ card is red} \mid 1^{st} \text{ is red}] = \frac{25}{51}$

$\vdots$

**Answer:** 0.0253 ; 0.0506

$\text{Prob}[1^{st} 5 \text{ cards are red}] = \frac{26}{52} \times \frac{25}{51} \times \frac{24}{50} \times \frac{23}{49} \times \frac{22}{48} = 0.0253$

$\text{Prob}[1^{st} 5 \text{ cards are all same color}] = \text{Pr}[1^{st} 5 \text{ cards are red}] + \text{Pr}[1^{st} 5 \text{ cards are black}]$

$= 0.0253 + 0.0253 = 0.0506$

**7.** (*3 points each*) Match the following R commands to their intended use:

F library()

A read.csv()

B x$y

C na.rm = TRUE

A. Load a data set from a file into a data frame

B. The column named y in data frame x

C. Ignore missing data

D. The column named x in data frame y

E. Save a dataset to a file

F. Load a package

**8.** (*1 point each*). For each of the following, **circle** increase, decrease, or stay the same to describe the change in the value at the left in response to the change indicated at the head of the column.

| | **Increase sample size** | **Increase α** | **Increase standard deviation** |
|---|---|---|---|
| **Type 1 error** | Increase<br>Decrease<br>Stay the same | Increase<br>Decrease<br>Stay the same | Increase<br>Decrease<br>Stay the same |
| **Type 2 error** | Increase<br>Decrease<br>Stay the same | Increase<br>Decrease<br>Stay the same | Increase<br>Decrease<br>Stay the same |
| **Power to detect 10 cm change in mean** | Increase<br>Decrease<br>Stay the same | Increase<br>Decrease<br>Stay the same | Increase<br>Decrease<br>Stay the same |
| **Sampling error (measured in same units as original values)** | Increase<br>Decrease<br>Stay the same | AMBIGUOUS – NOT MARKED<br>Increase<br>Decrease<br>Stay the same | Increase<br>Decrease<br>Stay the same |
| **Bias** | Increase<br>Decrease<br>Stay the same | Increase<br>Decrease<br>Stay the same | Increase<br>Decrease<br>Stay the same |
| **Width of confidence interval of mean** | Increase<br>Decrease<br>Stay the same | Increase<br>Decrease<br>Stay the same | Increase<br>Decrease<br>Stay the same |