Possible responses to the 2017 AP Stats Free Response questions, draft #1.

You can access the questions here.

Note: I construct these as a service for both students and teachers to start discussions. There is nothing "official" about these solutions. I certainly can't even guarantee that they are correct. They probably have typos and errors. If you catch some, let me know! But if they generate discussion and help others, then I've succeeded.

1. *Wolves*

a) What does positive, strong, linear mean in this context?

i. Positive: wolves with above-average lengths tended to have above-average weights, and wolves with below-average lengths tended to have below-average weights.

ii. Linear: The weights of wolves tended to increase by a roughly constant amount when wolf lengths increase by a constant amount. The overall pattern in the scatterplot would be modeled well by an equation whose graph is a straight line.

iii. Strong: The actual wolf lengths are very close to their predicted wolf lengths, based on a linear regression model to predict weight from length. In other words, predictions using a linear model would have relatively little error.

b. The slope is $35.02\frac{kilograms}{meter} = 3.5\frac{kilograms}{10cm}$ *. A wolf with a length 10 cm greater than another wolf is predicted to weight about 3.5 kilograms more than the other wolf.*

c. Here's the work to find actual weight of the wolf, in kg:

$$residual = actual - predicted.$$

$$-9.67 \approx actual - (-16.46 + 35.02(1.4))$$

$$-9.67 \approx actual - 32.568$$

$$22.898 \approx actual$$

2. ***Water Cups and Soft Drinks***

a)

Setup: We'll construct a 95% one-sample z interval for *p,* the true proportion of customers at this restaurant with a water cup who actually get a soft drink instead of water.

Conditions:

- We have a random sample of 80 customers, and we can trust that 80 is less than 10% of all customers with a water cup who might choose a soft drink over water when getting their beverage.
- $np = 23 \quad n(1-p) = 57$, which are both greater than 10, so the sampling distribution of p is approximately normal.

We have met the conditions for a one sample z test:

Mechanics:

Our 95% z interval for *p* interval is $\left(\frac{23}{80}\right) \pm (1.96)(\sqrt{\frac{\left(\frac{23}{80}\right)\left(\frac{57}{80}\right)}{80}} \approx (0.18832, 0.38668)$.

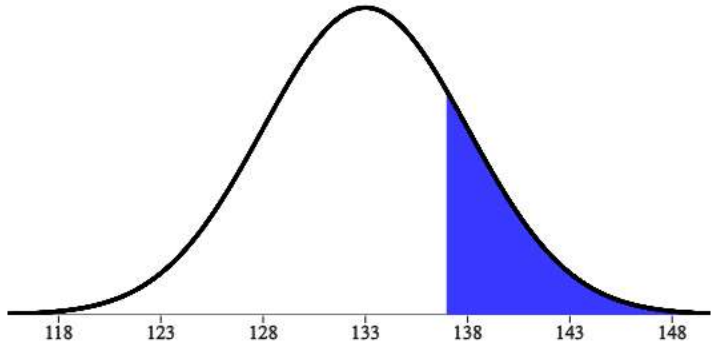
Conclusion:

We're 95% confident that between 18.8% and 38.7% of all customers who, having asked for a water cup when placing an order, actually get a soft drink instead of water.

b) Estimated loss: Multiply each endpoint of the interval by $(3000\ customers)(0.25\frac{dollars}{customer})$. This gives us an estimated loss between $141.24 and $290.01 for the restaurant in June.

3. **Melons.**

a)



$P(diameter > 137)$

$\approx P(Z > \frac{137-133}{5})$

$\approx P(Z > 0.8)$

≈ 0.2119

b)

$P(diameter > 137)$

$\approx (0.7)(0.2119) + (0.3)(.8413) \backslash$

$\approx .14833 + .25239$

≈ 0.40072

c)

$P(\, melon\ \ from\ J \mid diameter > 137)$

$\approx \dfrac{.14833}{0.40072} \approx 0.37016$

4. **Pottery.**

(a)

For chemical Z, the percents found across the three sites have similar centers (medians all near 7 percent).

Percents for Chemical Z appear roughly symmetric across all three sites.

Variability across the three sites differs. By using IQRs, sites I and III have the most variability (IQR = about 4 percent) and site II has the least (IQR = about 1 percent). When we also include rages as a measure of variability, then see that Site III (range = about 8 percent) has more variability than site III (range = about 6 percent).

(b-i) We can use box plots to estimate the lowest and highest possible sum (from the previous data collected) for the three chemicals at each site:

	Site I	Site II	Site III
Minimum	6+11+4 = 21	1.8+5+6 = 12.8	5+6+3 = 14
Maximum	8+15+10 = 33	4+7+8 = 19	7+8+11= 26
Does 20.5% fit?	Seems too low	Seems too high	Seems to fit.

Based on the data already collected, 20.5% only falls between the minimum possible and maximum possible sum at site III.

I pick site III.

(b – ii) Choose Chemical Y.

Of the three chemicals, it appears that the middle 50% for chemical X has overlap across the three sites (5-7.5 at I, 5-6.7 at II, and 5.8-6.2 at III). The same is true for chemical Z (5-9 at site I, 6.5-7.5 at site II, and 5-7 at site III).

But there is no overlap in the middle 50% for chemical Y (12-14 at Site I, 2.5-3.2 at Site II, and 5-9 at Site III).

5. ***Schizophrenia.***

To answer this question, I will perform a χ^2 *test for association, with a significance level of 0.05.*

Setup:

H_0: *There is no association between gender and age for the population of those being treated for schizophrenia.*

H_A: *There is an association between gender and age for the population of those being treated for schizophrenia.*

Mechanics: Expected counts are in parentheses

Obs (Exp)	20 to 29	30 to 39	40 to 49	50 to 59	Total
Women	**46(56.91)**	**40(36.22)**	**21(17.25)**	**12(8.62)**	**119**
Men	**53(42.09)**	**23(26.78)**	**9(12.75)**	**3(6.38)**	**88**
Total	**99**	**63**	**30**	**15**	**207**

The χ^2 statistic is $\sum_{all\ cells} \left(\frac{observed - expected}{expected}\right)^2 \approx 10.88$. *df = 3. P-value= 0.01*.

Conditions:

- We have a random sample from the population of men and women being treated for schizophrenia.
- Note that all expected counts are all at least 5 (the smallest is 6.38). The observed counts are not, but that is not required.
- Since 207 < 10% all patients treated for schizophrenia, we are ok on the independence check.

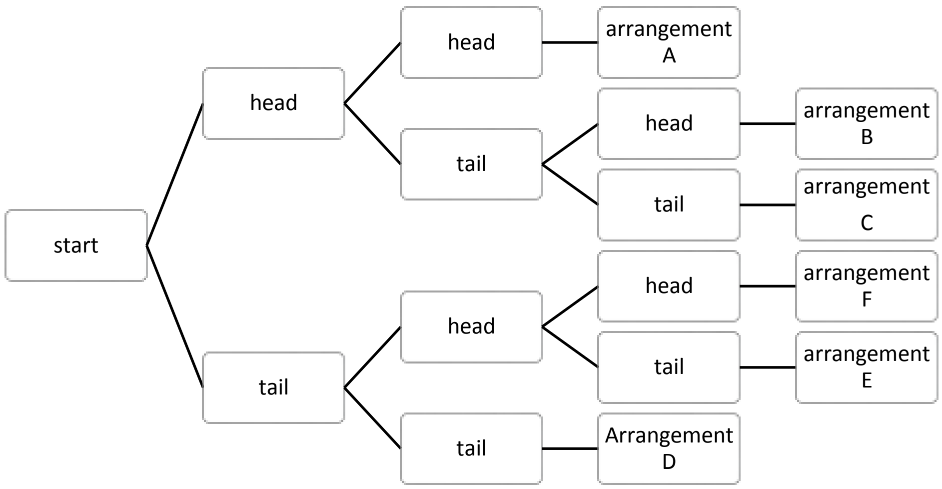
Conclusion:

Because out p-value is *0.01 < 0.05,* we have convincing evidence of an association between gender and age in the population of those being treated for schizophrenia.

6. ***Coins vs Chips***

a) The sequential coin flip method:

i) The assignment of treatments can be shown in a tree diagram. It could take two or three flips until a treatment group is filled.



Each head / tail has a probability of ½ occurring. Therefore:

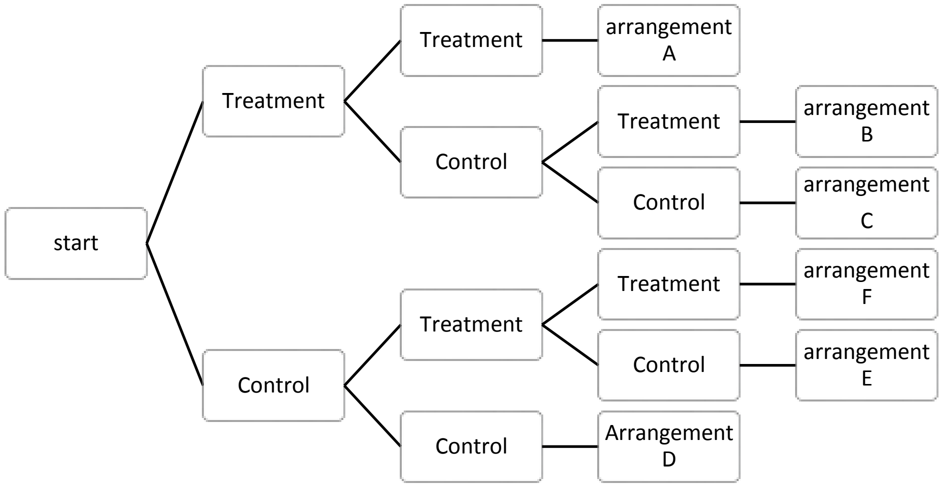
P(arrangement A) = $(1/2)(1/2) = 1/4$. *This is also the correct probability for arrangement D.*

$P(arrangmentB) = (1/2)(1/2)(1/2) = 1/8$. *This is also the correct probability for arrangement C, E, and F.*

Arrangement	A	B	C	D	E	F
Probability	0.25	0.125	0.125	0.25	0.125	0.125

ii) P (both in same group)
=P(arrangement A or D) = 0.25+0.25 = 0.50.

b) For the chip method, we get:



P(arrangement A) = $(1/2)(1/3) = 1/6$. *This is also the correct probability for arrangement D.*

$P(arrangmentB) = (1/2)(2/3)(1/2) = 2/12 = 1/6$ *. This is also the correct probability for arrangement C, E, and F.*

Arrangement	A	B	C	D	E	F
Prob	1/6	1/6	1/6	1/6	1/6	1/6

b-ii) *P(both men in same group) = P (arrangement A or D)* = 1/6 + 1/6 = 1/3.

c) Imagine the treatments assigned to each person in line, in sequence. We want every possible sequence of 8 grilled cheese and 8 salads to be equally likely. This means that (SSSSSSSSGGGGGGGG) (GGGGGGGGSSSSSSSS) is no more/ less likely than any other sequence. Otherwise, we might increase the likelihood of getting a lopsided assignment of treatments. We learned in parts (a) and (b) that the coin flip method makes arrangements where all males are placed in the same group more likely than the other arrangements. The chip method does prevents this from happening. Based on the results from (a) and (b), I would predict that the chip method is still better for this more complex situation.