2014 FR

1) (a) P(on campus) = 24/33 = 0.727

P(off campus) = 37/67 = 0.552

GRADING:
E = both calculations correct, with work shown (fraction)

P = one of the two calculations correct, with work shown

OR both correct answers, no work shown

OR both calculations of *only one* activity (not *at least* one) with work shown

(b) On campus residents are more likely to participate in extra curricular activities than off campus residents.

Though the % of students who participate in 2 or more activities are similar between the two groups, there is a large difference between the % of on and off campus students who participate in none (0.273 vs. 0.448). There is also a large difference between on and off campus residents with one activity (0.515 vs. 0.373).

This shows an association between living arrangements and activity level.

GRADING:
E = correctly comparing % between the two groups for at least 2 of the 3 categories.

P = correctly lists % for two of the 3 categories but does not compare them

OR correctly compares only 1 of the 3 categories

(c) We fail to reject Ho because the p-value of 0.23 > alpha = 0.05.

We have insufficient evidence that there is an association between the residential status and level of participation in extracurricular activities among students at the university.'

GRADING:
E = both sentences correct

P = one sentence correct

OVERALL GRADING:

4 = EEE

3 = EEP

2 = EEI EPI EPP PPP

1 = EII PPI

0 = anything else

2)
(a) P(all 3 are women) = (3/9)*(2/8)*(1/7)= 0.0119 = (1/84)

GRADING:
E = correct probability computed, showing work

P = correctly shows how the probability should be computed, but does not follow through

OR correct answer, no work

OR mistakenly assumes independence and gets (1/3) (1/3)(1/3) = 0.037

(b) Yes, there is reason to doubt the managers claim, since the chance of selecting all 3 women randomly is very small (0.012). I would doubt they were selected at random.

GRADING:

E = states the probablility is small, makes and appropriate decision, and does so in context

P = correct decision but no context

OR does not explicitly make a decision about whether there is reason to doubt the managers claim.

(c) No, this is not a correct simulation. The selection of the 3 people is done without replacement, but the simulation assumes replacement.

GRADING:
E = answers NO and states that the dice simulation assumes independence, and the actual selection is dependent.

P = only comments on one of the two things (dice or actual selection)

OR response is vague, incomplete, etc.

OVERALL GRADING:

4 = EEE

3 = EEP

2 = EEI EPI EPP PPP

1 = EII PPI

0 = anything else

3) $X \sim N(120, 10.5)$

(a) $P(X > 140) = 0.0284$

GRADING:
E: * identifies normal model and parameters
* correct notation
* correct answer

P : 2 of the 3 things above

(b) $\overline{x} \sim N(120, 6.062)$ n = 3

$P(\overline{x} > 140) = 4.84 \times 10^{-4} = 0.0005$

They are less likely to lose their funding if they use a sample of 3 days. The probability that the average of 3 days is more than 140 absences (0.0005) is less than the probability calculated in part (a).

GRADING:
E = Gives correct answer (less likely) AND:
* clearly gives the distribution of sample means
* indicates that the variability is smaller
* indicates that the distribution is centered at 120

P = missing 1 part

(c) P(not Tues, Wed, Thurs) = 2/5 = 0.4

Assuming each week is independent, the probability of not selecting Tues, Wed, or Thurs in 3 weeks is:
$(0.4)^3 = 0.064$

GRADING:
E = correct probability and shows work

P = correct prob, no work

4)
(a) the median is not affected by outliers or skewness, so it is often a better representation of the center of a distribution.

With income, a small number of very large incomes could dramatically increase the mean, but not the median. So the median is a better representation of the "typical" income level.

GRADING:
E = * describes how skewness/outliers affects mean but not median
* makes a relevant conclusion based on this

P = one of the two things above

(b) PART 1 (there are 2 letters given in B)
Method 2 is better than Method 1. Method 1 is voluntary, so the results obtained would not be considered representative of the population. However Method 2 is a random sample, so even though the sample size is smaller, the fact that it is random makes it a better representation of the population and the results more accurate.

GRADING:
E = * select Method 1
* give a reason that method 1 is bad
* give a reason that method 2 is good

P = missing one of the things above
I = select Method 1 as being better

PART 2

Method 1 would likely have more people with higher incomes respond, since it is voluntary. So most likely the average would be overestimated. Method 2 would likely have a mean income that is close to the correct mean income of the whole class since the sample is random.

GRADING:
E = * indicates that the incomes of responders would be different than that of non-responders
* indicates a direction for the way the statistic would change (over/under estimated).

5)

Step 1: States a correct pair of hypotheses

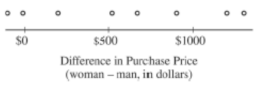
Ho: $\mu_d = 0$
Ha: $\mu_d > 0$

GRADING:
E = correct hypotheses (symbol, parameter, etc.)
P = one mistake

2 sample t test:

Ho: $\mu_M = \mu_W$
Ha: $\mu_M < \mu_W$

STEP 2: Identifies the correct procedure and checks appropriate conditions

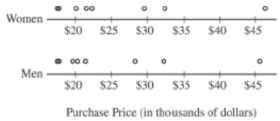


Conditions:

1- Paired Data	1- the prices that the menand women pay are measured on the same cars
2- Random	2- Stated random sample
3- 10% condition	3- There are more than 80 car models
4- Normality	4- The dotplot of the differences is symmetric, no outliers, therefore normal data

Conditions met, use t distribution and Paired t test

2 sample t test:



Conditions:

1) Random	1) stated random samples
2) Independent	2) Men and women car prices are independent (not true)
3) 10% condition	3) There are more than 80 men and 80 women buying cars
4) Normality	4) The dotplots are NOT symmetric, therefore data is not normal

Conditions not met but proceeding anyway with a t distribution for a 2 sample t test

GRADING:

E = all conditions checked and statement made
P = one mistake

STEP 3: Correct mechanics

Test statistic: $t = \dfrac{585 - 0}{\dfrac{530.71}{\sqrt{8}}} \approx 3.12.$ (3.1178)

P-Value:

P(t > 3.12) = 0.0084
df = 7

	Mean	Standard Deviation
Women	$25,926.25	$9,846.61
Men	$25,341.25	$9,728.60
Difference	$585.00	$530.71

GRADING:
E = test stat AND p-value correct
P = test stat OR p-value correct

2 sample t test:

$$t = \frac{25926.25 - 25341.25}{\sqrt{\frac{9846.61^2}{8} + \frac{9728.6^2}{8}}} = 0.1195$$

P(t > 0.1195) = 0.4533 df = 13.998 = 14

	Mean	Standard Deviation
Women	$25,926.25	$9,846.61
Men	$25,341.25	$9,728.60
Difference	$585.00	$530.71

STEP 4: States a correct conclusion

We reject Ho b/c p-value of 0.008 < alpha = 0.05.

We have sufficient evidence that the average difference between what men and women pay for cars is greater than 0. Therefore there is convincing evidence that, on average, women pay more for cars than men.

GRADING:
E = 1st AND 2nd sentence correct
P = 1st OR 2nd sentence correct

2 sample t test

We fail to reject Ho b/c p-value of 0.4533 > α = 0.05.

We have insufficient evidence that the average car price for women is greater than that of men.

TOTAL SCORE:

E = 1 pt
P = 1/2 pt
I = 0 pts

ROUND (no half point totals)

If you did a 2 sample t-test instead of paired, subtract 1 pt from their score

6)
(a) FCR = -1.595789 + 0.0372614(175) = 4.924956

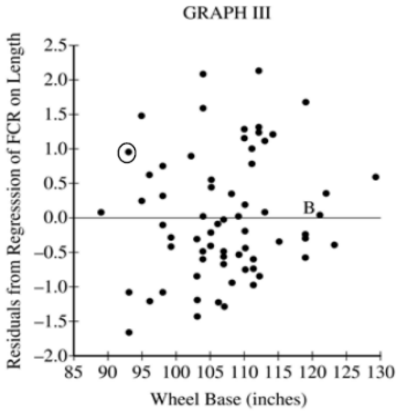
residual = 5.88 - 4.924956 = 0.955044 mpg

The residual means that the car's FCR is 0.96 gallons per 100 miles greater than would be predicted for a car its length.

GRADING:
E = * correct residual with supporting calculation
* correct interpretation in context

P = 1 of the two above things

(b) (i) see below



(ii) Point B shows a car whose actual FCR is very close to its predicted FCR. The predicted FCR is the one from the model using "length" as the explanatory variable.

GRADING:
E = * circles the correct point on the graph
* provides a reasonable interpretation of the car associated with point B having a residual near 0 that refers to predicting FCR from length.

P = one of the two components above

(c) Graph II has a moderate, positive, linear association between engine size and the residuals of FCR from length.

Graph III has a scattered (or weak, positive, linear) association between wheel base and residuals of FCR from length.

The association between engine size and the residuals is much stronger than the association between wheel base and residuals.

GRADING:
E = * description of form AND direction for both graphs
* description of strength of association of both graphs
* comparison between the two graphs

P = 2 of the 3 things above

(d) Engine size is the better choice.

There is a stronger association between engine size and the residuals (from predicting FCR from length. This indicates that engine size is more useful than wheel base for reducing the variability in FCR values that remain unexplained by the model.

GRADING:
E = correct choice and justification based on:
* the stronger association
* reducing the variability that remains unexplained in the model which predicts FCR from length

P = correct choice and justification based on one of the two things above.

#1-5 = ___ x 1.875 =
#6 = ___ x 3.125 = ___ +
50

60/100