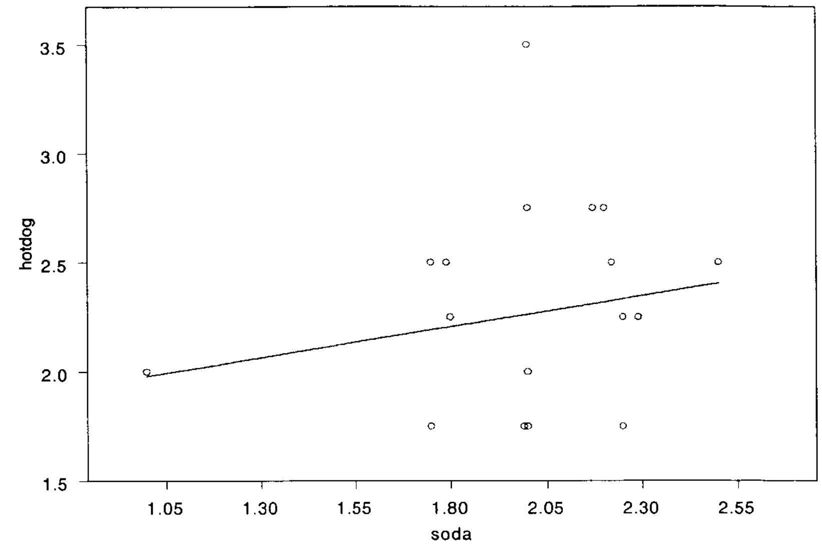
Review for exam 2

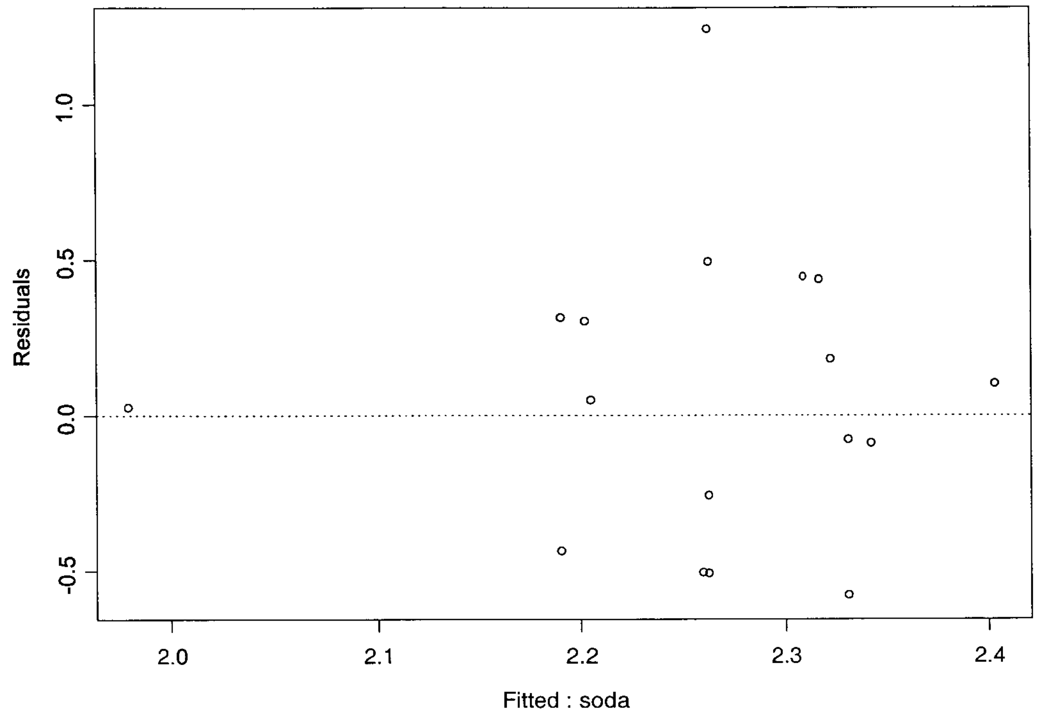


1. Can the price charged for a 16 ounce soda be used to predict the price charged for a hot dog in major league baseball stadiums?

$r = 0.1907093$

	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	1.6963	0.6750	2.5128	0.0212
soda	0.2831	0.3343	0.8468	0.4076

a) What is the increase in hotdog price for every $1 increase in soda price?
b) What percentage of the variation in hotdog price can be explained by soda price?
c) What is the predicted hotdog price for a soda that costs $2?
d) Why is the line drawn on the scatterplot the *least squares line*?



e) Does this residual plot reveal any problems? If so, what?

f) What are three other types of residual plots?

g) Residual =

h) What is wrong with predicting hot dog price for a $5 soda?

i) How do we detect influential points?

j) Why do we say soda price and hot dog are associated, but that an increase in soda prices does not cause an increase in hot dog prices?

2. Problem 6.23 on page 148:

Business

	Admit	Deny	Total
Male	480	120	600
Female	180	20	200
Total	660	240	800

Law

	Admit	Deny	Total
Male	10	90	100
Female	100	200	300
Total	110	290	400

Combined

	Admit	Deny	Total
Male	490	210	700
Female	280	220	500
Total	770	430	1200

a) For the combined table, what is the marginal distribution for admissions?

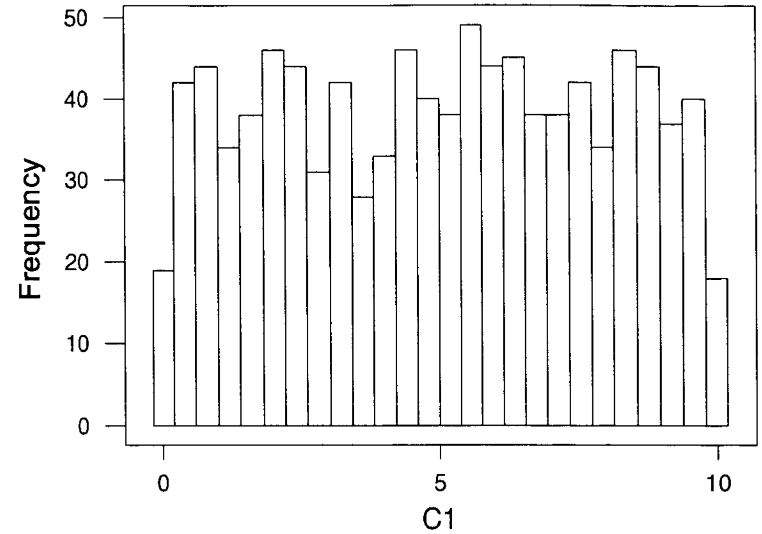
b) For the combined table what percentage are admitted?

c) . . . what is the conditional distribution for admissions given males?

d) . . . what percentage of the males are admitted?

e) Do these tables display Simpson's paradox? Why or why not?

3. Given this population with $\mu = 5.0$ and $\sigma = 3.16$, describe the sampling dist. of $\overline{x}$ for random samples of size n = 81:



a) Mean:

b) Standard deviation:

c) Shape:

Did we need to apply the Central Limit Theorem?

What does the Central Limit Theorem allow us to do?

d) What does this sampling distribution of $\overline{x}$ tell us?

e) Between what two values are the middle 68% of the $\overline{x}$'s?

f) The WAIS (Wechsler Adult Intelligence Scale) is a common "IQ test" for Adults. The scores for persons over 16 years of age are approximately normal with $\mu = 100$ and $\sigma = 15$.

a) What is the probability that a randomly chosen individual has a WAIS score of 105 or higher?

b) What are the mean and standard deviation of the sampling distribution of $\bar{x}$ for an SRS of 60 adults?

c) What is the probability that the average WAIS score of an SRS of 60 people is 105 or higher? (Do we need to apply CLT?)

5. Problem 13.16 page 335 "Give a 95% confidence interval for the mean..."

$\overline{x} \pm z^* \sigma / \sqrt{n}$ $\overline{x} = 30.40$ 95.0% CI: (26.06, 34.74)

Are the conditions met for computing this CI? Justify.

<u>True or false for problem 5:</u>

a) This interval tells us that reasonable values for μ are all numbers between 26.06 and 34.74.

b) The parameter of interest is the mean of the 10 data points in the sample.

c) The value of the sample mean is always contained in the confidence interval computed using its value.

d) 95% of the time, when sampling from this same population, we will get this confidence interval (26.06, 34.74).

e) The probability that μ is in this 95% confidence interval (26.06, 34.74) is 0.95.

f) *95% level of confidence* means that using this procedure, we will obtain confidence intervals that contain μ for 95% of all possible samples.

g) *95% confidence* means that using this procedure, we will obtain confidence intervals that contain μ 95% of the time.

h) 95% of the DMS odor thresholds are contained in the interval (26.06, 34.74).

i) We are 95% confident that $\overline{x}$ is contained in the confidence interval (26.06, 34.74).

j) We are 95% confident that the mean DMS odor threshold for all untrained students is somewhere between 26.06 and 34.74.

More True or false:

6. The shape of the histogram of sample data gets closer to the shape of the population as the sample size increases.
7. The shape of the sampling distribution of $\overline{x}$ is always normal.
8. The mean of the sampling distribution of $\overline{x}$ gets closer to μ as n increases.
9. The standard deviation of $\bar{x}$ (for $n > 1$) is always less than the standard deviation of the population.
10. Central Limit Theorem allows us to compute probabilities on data in a sample from a non-normal population whenever the sample is large and SRS.

11. Probabilities on individuals can only be computed using the standard normal table if the population is normally distributed.
12. Standard deviation of $\overline{x}$ is a shorter way of saying standard deviation of the sampling distribution of $\overline{x}$.
13. The mean of the data in a sample gets closer and closer to μ as n increases.
14. Standard deviation of $\overline{x}$ is computed using $\sigma / \sqrt{n}$.
15. 95% of all possible $\overline{x}$'s will be within 2 σ of μ.
16. 95% of all possible $\overline{x}$'s will be within 2 $\sigma / \sqrt{n}$ of μ.
17. The researcher usually wants to "prove" that the null hypothesis is true.
18. The null hypothesis is assumed to be true until the data via P-value provide evidence against it.
19. Large (> α) P-values "prove" that the null hypothesis is true.
20. We always reject H_0 if P-value < α.
21. In order to test whether the mean DMS odor threshold is higher for students than for trained wine tasters (about 25 micrograms per liter of wine), we test the hypotheses: H_0: $\overline{x} = 25$ vs. H_a: $\overline{x} > 25$.

22. If P-value > α, we believe H_a.
23. Results of a test of significance are statistically significant only if the P-value for the observed effect is < α.
24. If testing H_0: $\mu = 80$ vs. H_a: $\mu > 80$ with $z = 1.82$, the P-value is 0.0344.
25. If testing H_0: $\mu = 80$ vs. H_a: $\mu \neq 80$ with $z = 1.82$, the P-value is 0.0344.
26. P-value is the probability that the null hypothesis is true.
27. Two sided tests have "≠" in the alternative hypothesis.
28. "=" is never found in an alternative hypothesis.
29. Margin of error requires only confidence level, sample size and σ.