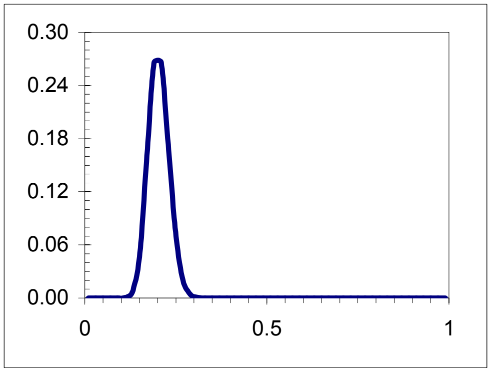
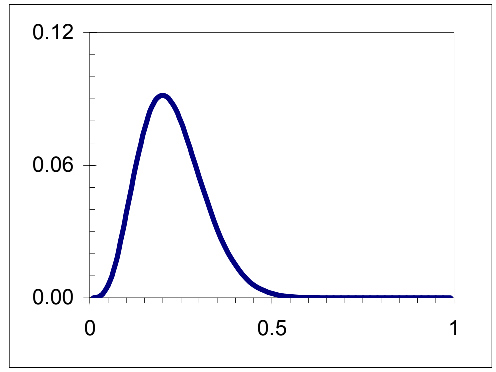
Biostatistics **22S:101** **Answers to Practice Exam 2**

1. A large population contains an unknown proportion (p) of black marbles. A sample of n=200 drawn scientifically from the population contained x=40 black marbles. Which picture shows the posterior distribution of the population proportion p? Justify your answer. **A. Because sep = sqrt(.2*.8/200) = .028**

A. Looks like this ($\mu = .2$ and $\sigma = .028$)



B. Looks like this ($\mu = .2$ and $\sigma = .089$)



2. An investigator wants to determine the proportion (p) of retirees who chose not to fill a prescription last year because it was too expensive. In a scientific sample of n=525, x = 98 respondents said that they had done this. State in words the posterior distribution of p. Obtain a 95% credible interval for p. What (approximately) is the probability that p exceeds .25 ? **95% CI: .153 to .220. P(p > .25 | data) = (Area above z = 3.72) = .0000.**

3. A prospective study recruited 10,000 smokers and 10,000 non-smokers aged 30 to 39 and followed them for 20 years. The cases of throat cancer are shown in this table.

	no Cancer	Cancer	Relative Frequency(%)
Smokers	9800	200	2.0%
Non-Smokers	9950	50	0.5%

What is the estimated relative risk of throat cancer for smokers vs. non-smokers? $\widehat{RR}$ **= 2.0/.5 = 4.0**

4. Two hundred forty dieters volunteered for a study of chromium picolinate, a fat-reducing dietary supplement. They were randomly assigned to receive placebo or chromium picolinate. One side effect is reduction in iron, a key component of hemoglobin. Here are the data on percents of subjects with lower iron after 8 weeks of treatment.

	Placebo		Chromium Picolinate
n	120		120
% with reduced iron	16%		29%
sep	0.033		0.041
$\hat{\Delta}$		**0.130**	
sed		**0.053**	
95% CI	**0.026**		**0.234**

- What are the mean difference and the standard error of the difference (SED) between the percents?
- Obtain a 95% credible interval for the difference.
- Is the difference statistically significant? **Yes.**

5. In a study to compare osteoporosis rates for men and women over the age of 70 it was observed that 6.9% of 25000 men and 67% of 27000 women had osteoporosis. The difference is 60.1 percentage points and the credible interval is 59.4 to 60.7. Is the difference significant? **Yes, zero is ruled out.**

6. In a randomized experiment, 400 kids brushed with baking powder and 400 brushed with toothpaste. 52 of the baking powder kids (13%) got cavities and 40 of the toothpaste kids (10%) got cavities. The difference is 3 percentage points. Obtain a 95% credible interval for the true difference. Is the difference significant?

$\hat{\Delta}$ = .030, SED = .023 95% CI: -.0142 to .0742 Not Significant.

7. An economic survey of a sample of 225 US wage earners showed an average of $\bar{x}$ =$23.50 was spent per week eating out. The standard deviation of the sample was reported to be s = $12.00. State the posterior distribution and obtain a 95% credible interval on the mean (μ) of all wage earners.

The posterior distribution is approximately normal with μ = $\bar{x}$ =$23.50, and σ = sem = 12/sqrt(225) = .80. The approximate 95% CI is 23.5 + 1.96x.80; i.e. from 21.9 to 25.1 .

8. A random roadside survey of 481 males and 138 females found that 77 males and 16 females had detectable amounts of alcohol by a breathalyzer test. Is the difference significant?

n	**481**	**138**	
x	**77**	**16**	
phat	**0.160**	**0.116**	
sep	**0.017**	**0.027**	
Dhat	**0.044**		
sed	**0.032**		
95% CI	**-0.019**	**0.107**	**Not Significant**

9. R. M. Lyle, reported a study in which healthy men aged 45 to 65 received either a calcium supplement or a placebo for 12 weeks. He reported, "The calcium group had significantly lower blood pressure compared with the placebo group." (Note: blood pressure is measured in millimeters of mercury, abbreviated mm Hg.)

Which of the following sets of data is consistent with Lyle's statement? Why?

A: Difference = 10 mm Hg with 95% credible interval 2.4 to 17.6. **(Consistent – rules out 0.)**

B: Difference = 20 mm Hg with 95% credible interval -5 to 45.

10. A sample drawn from a box of numbers with a fairly normal distribution has sample mean $\bar{x}$ = 16.5 and sample standard deviation s = 8.8. State the approximate posterior distribution of the box average (μ_{box})

a) if n=400 b) if n=36

a) Approximately normal with μ = 16.5 and σ = 0.44.
b) Approximately t(35) with μ = 16.5 and σ = 1.47.

11. An unknown quantity, which we will call η, has an approximately t(9) distribution with μ = 3.1 and σ = 0.6 Find the 95% credible interval for the unknown quantity.

95% Credible interval: μ + 2.26·σ ; i.e. from 1.74 to 4.46

12. One hundred male alcoholics suffering from secondary hypertension participated in a study to determine the efficacy of a new antihypertensive agent. The men were assigned at random to either the control group or the treatment group. Men in the control group received a placebo. Statistics for arterial pressure at 30 days post treatment for the 97 subjects who completed the study are shown in this Table.

Hypertension Study	Placebo	Treatment
n	22	23
mean ($\bar{x}$)	127.1	99.0
standard deviation (s)	24.08	8.81

State the approximate posterior distribution of the difference ($\Delta = \mu_{Pbo} - \mu_{Trt}$).

The posterior distribution is: **Approximately t(26.3) with μ = 28.1 and σ = 5.453**

95% Credible interval: **16.9** to **39.3**

	A	B	C
1	n	22	23
2	xbar	127.1	99
3	s	24.08	8.81
4	mu=xbar	127.1	99
5	sigma=sem	5.134	1.837
6	df=n-1	21	22
7	mu=deltaHat=b5-c5	28.1	
8	sigma=sqrt(b6^2+c6^2)	5.453	
9	df = B9^4/(B6^4/B7+C6^4/C7)	26.3	
10	t(26) - percentile	2.05	
11	95% CI	16.9	39.3

Formulas

$$sem = \frac{s}{\sqrt{n}}$$

$$sep = \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}$$

$$sed = \sqrt{(se_1)^2 + (se_2)^2}$$

$$Satterthwaite's\ df = \frac{(sed)^4}{\frac{(sem_1)^4}{df_1} + \frac{(sem_2)^4}{df_2}}$$