Stat 328 Exam 1

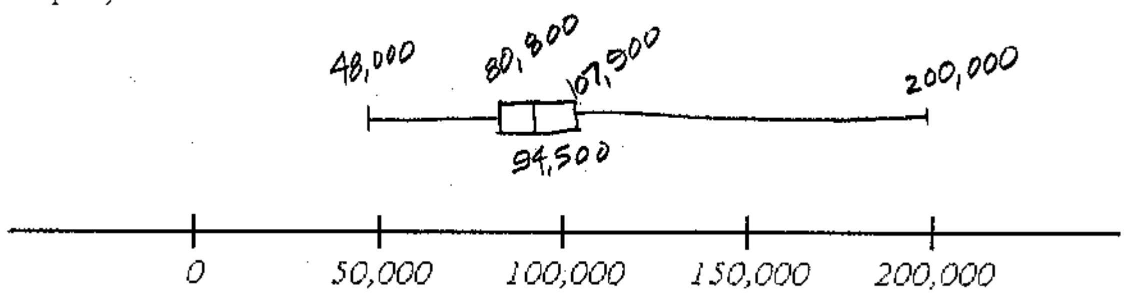
Summer 2003 **Prof. Vardeman**

1. Attached to this exam are two JMP reports summarizing sale prices (in the period 1/1/02 through 3/14/03) of 1-story houses in Ames, Iowa built 1945 and earlier.

2pts **a)** How would you describe the shape of the first histogram? What feature of the plot most immediately captures your attention?

It is mostly bell-shaped with a couple of extreme values (outliers) on the right side. It is the extreme values that first catch our attention (and rightly so).

3pts **b)** Use the first JMP report and make a box plot for summarizing these $n = 47$ sale prices. Draw it above the scale below. (Write the values of the "5 number summary" above appropriate features of the plot.)



Closer examination of the data set revealed that the two largest sale prices were for the same (and actually, very odd) house, sold twice within 3 months. In the interest of representing "Ames houses" rather than "Ames real estate transactions" Vardeman deleted the first price of this house from the data set and produced the 2nd JMP report. **Henceforth use this 2nd report.**

5pts **c)** Strictly speaking the houses represented here are not a random sample of 1-story Ames houses built 1945 or before. But if one treats them as such, what are 90% confidence limits for the mean fair market value of such homes in Ames? (Plug in, but you don't need to simplify.)

Use $\bar{x} \pm t \frac{s}{\sqrt{n}}$ Here d.f. $= 46 - 1 = 45$ and (from a computer package) $t = 1.6794$ So 90% limits are

$$94{,}719.57 \pm \underbrace{1.6794 \frac{23{,}279.09}{\sqrt{46}}}_{5{,}764.23}$$

2pts **d)** It would probably not be smart to use the prediction limits formula from class to predict another home value from this population. Why?

Because of the one very large value in the data set, it does not appear safe to model these values as normal ... and the prediction limits really do require normality to ensure their reliability (unlike the confidence limits for a mean, they are not robust)[1]

A citizen group contends that the Ames Assessor has consistently over-valued properties in the city. The group checks the assessed values at the times of sale for the 46 homes represented in the two JMP reports. They find that the differences

$$d = (assessed\ value) - (sale\ price)$$

had sample mean $\bar{d} = 4,250$ and sample standard deviation $s_d = 3,050$ and that the histogram of differences was very bell-shaped.

6pts e) What are 90% prediction limits for the difference d for the next home of this type sold in Ames? Is it clear that this house will be one over-valued by the assessor? Explain.

Use $\bar{d} \pm t s_d \sqrt{1+\frac{1}{n}}$ This is

$$4,250 \pm \underbrace{1.6794\,(3,050)\sqrt{1+\tfrac{1}{46}}}_{5,178}$$

This interval contains both positive and negative values. Negative values are ones where the assessment is too low. So it is not clear that this house will be over-valued.

6pts f) Is there clear evidence that on average, assessed values exceed sale prices for homes of this type in Ames? Show the details of an appropriate significance test and find and interpret a p-value.

$H_0: \mu_d = 0$

$H_a: \mu_d > 0$

Use the test statistic

$$T = \frac{\bar{d}-0}{\frac{s_d}{\sqrt{n}}} = \frac{4,250-0}{\frac{3050}{\sqrt{46}}} = 9.45$$

p-value is $P[\text{a } t_{45} \text{ r.v.} > 9.45] =$ very small

less than .0005

This is clear evidence that on average assessed values exceed sale prices for homes of this type in Ames.

2. A bank issues credit cards to its customers. Suppose that the random variable

X = the balanced carried over from June to July

by a randomly selected customer of this bank can be roughly described with the probability function specified in the table below.

Balance Carried X	0	1000	2000	3000
Probability	.5	.3	.1	.1

5pts **a)** Finish filling in the table and then compute the mean μ_X and standard deviation σ_X for the balance carried forward by a single randomly selected bank customer.

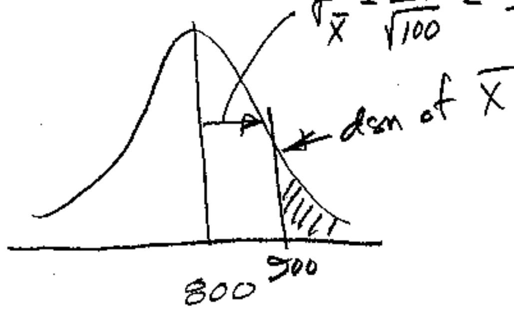
$$\mu_X = 0(.5) + 1000(.3) + 2000(.1) + 3000(.1) = 800$$

$$\sigma_X^2 = (0-800)^2(.5) + (1000-800)^2(.3) + (2000-800)^2(.1) + (3000-800)^2(.1)$$

and $\sigma_X = \sqrt{\text{above}} = 979.80$

6pts **b)** Use your mean and standard deviation from a) and approximate the probability that the sample mean balance carried over on $n = 100$ randomly selected accounts exceeds \$900. (If you were unable to do part a), you may use the *incorrect* values $\mu_X = 1100$ and $\sigma_X = 1700$ here.)

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{100}} = 97.98$$



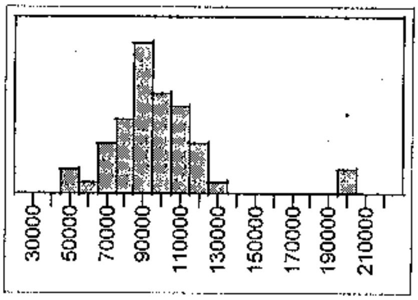
$$z = \frac{900 - 800}{97.98} = 1.02$$

So $P[\bar{X} > 900] = P[\text{a std normal r.v.} > 1.02]$

$= 1 - .8461$

$= .1539$

Distributions
Sales Price



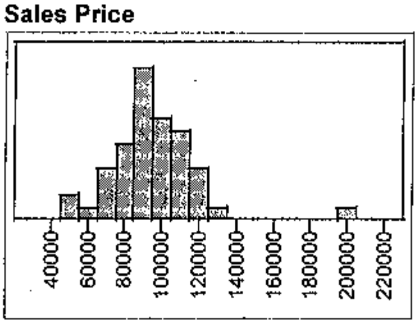
Quantiles

100.0%	maximum	200000
99.5%		200000
97.5%		199000
90.0%		117900
75.0%	quartile	107900
50.0%	median	94500
25.0%	quartile	80800
10.0%		66600
2.5%		49000
0.5%		48000
0.0%	minimum	48000

Moments

Mean	96959.574
Std Dev	27676.058
Std Err Mean	4036.968
upper 95% Mean	105085.57
lower 95% Mean	88833.579
N	47

Distributions
Sales Price



Quantiles

100.0%	maximum	195000
99.5%		195000
97.5%		183275
90.0%		116720
75.0%	quartile	105725
50.0%	median	94250
25.0%	quartile	80725
10.0%		66400
2.5%		48875
0.5%		48000
0.0%	minimum	48000

Moments

Mean	94719.565
Std Dev	23279.086
Std Err Mean	3432.3139
upper 95% Mean	101632.6
lower 95% Mean	87806.53
N	46