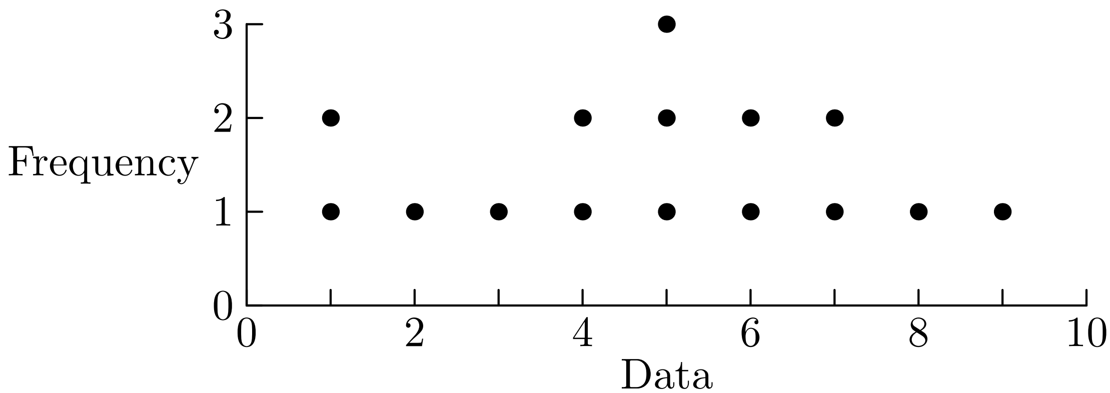
Final Exam Solutions

Math 112-I

Saturday, May 14, 2005

1. Three examples of a quantitative variable would be height, weight, and salary. Three examples of a categorical variable would be gender, blood type, and home state.

2.



3. Modified box-plots identify as outliers any points that are more that 1.5IQR away from the nearest quartile. These outliers are drawn as isolated points on the modified box-plot, and the whiskers are drawn out from the quartiles to the nearest farthest non-outliers.

4. We needed probability to describe the confidence level for a confidence interval, and to describe the p-value of a significance test.

5. We labeled each card with a baby's name, shuffled the cards, dealt them out, and recorded how many mothers got their correct baby back. We pooled our responses, and used the proportions to get estimates for the probabilities.

6. The margin of error for a confidence interval can be controlled through the choice of the confidence level and the sample size, both of which can be determined before the sample is taken.

7. There are three conditions:

   (a) That the data be from a simple random sample, which it is.

   (b) That $n\hat{p} \geq 10$. This is $17 \geq 10$, which is true.

   (c) That $n(1 - \hat{p}) \geq 10$. This is $145 \geq 10$, which is true.

8. The distance from the center (mean) of a normal distribution to either inflection point is the standard deviation.

9. The median is the center of the sorted data, and pays no attention to the extremes of the data. The mean, being found by summing the data and dividing by the sample size, changes with the addition of extremely low or extremely high values.

10. We check to see if the p-value is less than the significance level; if so, we reject the null hypothesis as being inconsistent with the sample data. Otherwise, if the p-value is over the significance level, the null hypothesis is consistent with the sample data.

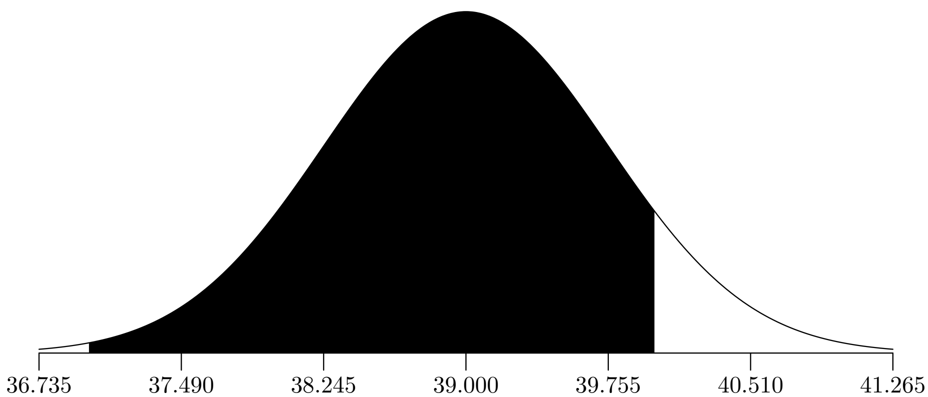
    All we can do is show that the null hypothesis is consistent with the sample data; if the null hypothesis were false, but close to being true, the sample data wouldn't necessarily show it.

11. We need to verify two technical conditions to apply the Central Limit Theorem.

    (a) That we are using a random sample, which is given in the statement of the question.

    (b) That $n \geq 30$ or the population is normal; $n = 86$ which is over 30.

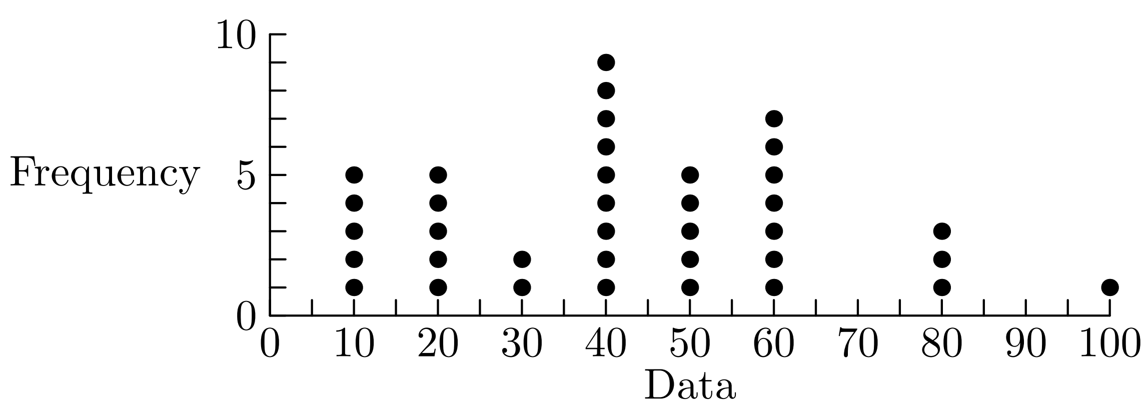
    According to the Central Limit Theorem, $\overline{x}$ is normal with mean $\mu_{\overline{x}} = \mu = 39$ and standard deviation $\sigma_{\overline{x}} = \sigma/\sqrt{n} = 7/\sqrt{86} = 0.755$.



    We would use the following *Fathom* command to find $\Pr(37 \leq \overline{x} \leq 40)$:

$$\texttt{normalCumulative}(40, 39, 0.755) - \texttt{normalCumulative}(37, 39, 0.755)$$

12. A granular variable should have evenly spaced gaps in its dot-plot. Here is an example:



13. In this test $\theta_0 = 0.3$. There are three technical conditions to check:

    (a) That we have a random sample, which is given.

    (b) That $n\theta_0 = (35)(0.3) = 10.5 \geq 10$, which is true.

    (c) That $n(1 - \theta_0) = 35(0.7) = 24.5 \geq 10$, which is true.

14. The two confidence intervals should have the same center, at $(25\% + 35\%)/2 = 30\%$. The 68% confidence interval goes out one standard error to either side, while the 99.7% confidence interval goes out three standard errors to either side. The margin for the second interval should be three times the margin for the first interval.

    The margin of error for the first confidence interval is $(35\% - 25\%)/2 = 5\%$, and the margin of error for the second confidence interval is $3(5\%) = 15\%$.

    The 99.7% confidence interval would be $(30\% - 15\%, 30\% + 15\%) = (15\%, 45\%)$.

15. We reject the null hypothesis when $p < \alpha$. Therefore every time we would reject the null hypothesis with $\alpha = 1\%$, we would also reject the null hypothesis with $\alpha = 5\%$. However, for any p-value between 1% and 5%, we would accept the null hypothesis if $\alpha = 1\%$ but not if $\alpha = 5\%$.

    Therefore we would reject the null hypothesis less with a significance level of 1% than with a significance level of 5%.

16. There were 339 under 30, 104 between 30 and 40, and 232 over 40 for a total of 675 in the sample.

    The marginal distribution for age groups is $339/675 = 50.2\%$ under 30, $104/675 = 15.4\%$ between 30 and 40, and $232/675 = 34.4\%$ over 40.

17. There are a total of 79 left-handed respondents.

    $33/79 = 41.8\%$ are under 30, $14/79 = 17.7\%$ are between 30 and 40, and $32/79 = 40.5\%$ are over 40.

18. We need to compute a and b to find the regression model $\hat{y} = a + bx$.

$$
\begin{aligned}
b &= r\frac{s_y}{s_x} \\
&= 0.65\left(\frac{5}{13}\right) \\
&= 0.250 \\
a &= \overline{y} - b\overline{x} \\
&= 11 - (0.250)45 \\
&= -0.250 \\
\hat{y} &= -0.250 + 0.250x
\end{aligned}
$$

19. When testing a medication or other form of treatment, it is important to measure how much of a difference the treatment makes. To do so, some of the respondents in the sample should receive no treatment while others should receive treatment. The group not receiving treatment is known as the control group.

20.

5	0
6	3569
7	5
8	134
9	489

21. There are twelve numbers. Once we have sorted the data, the median will lie between the sixth and seventh numbers. The lower quartile will lie between the third and fourth numbers, and the upper quartile will lie between the ninth and tenth numbers.

    The sorted data is

$$51, 59, 61, 63, 64, 68, 71, 75, 77, 78, 85, 96$$

    The five number summary is $(51, 62, 69.5, 77.5, 96)$.

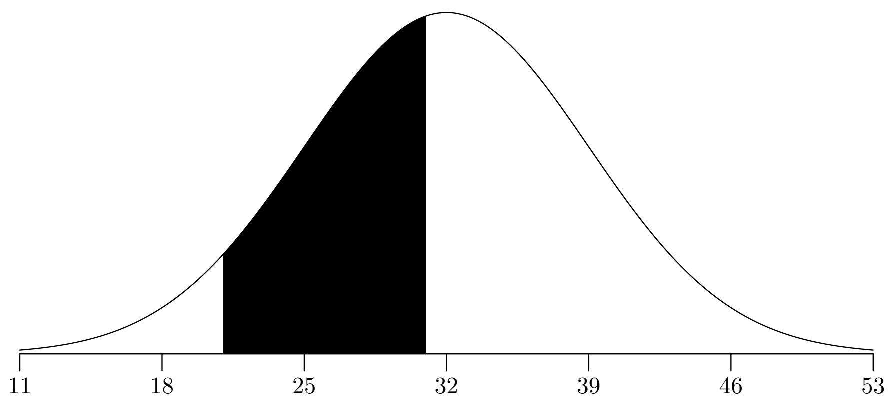
22.

$$\begin{aligned} z &= \frac{x - \overline{x}}{s} \\ &= \frac{38 - 34}{7} \\ &= 0.571 \end{aligned}$$

23. The IQR is resistant to outliers since it is the width of the middle two-fourths of the data, and ignores extreme values.

    The standard deviation is computed by averaging the squared deviations from the mean, and then taking a square root. Outliers will make the squared deviations grow larger.

24.



The *Fathom* commands to find $\Pr(21 \le X \le 31)$ are:

$$\texttt{normalCumulative}(31, 32, 7) - \texttt{normalCumulative}(21, 32, 7)$$

25. In an observational study, we can only establish whether or not there is a connection between two variables; we can't determine if one has an influence on the other, since we are not controlling for other (lurking) variables.

    In a controlled experiment, we can control for other variables, eliminating their influence, to determine if the explanatory variable really influences the response variable.