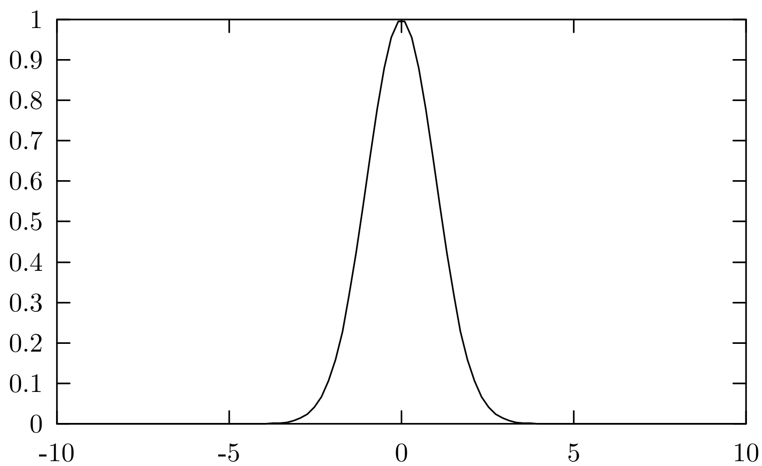
Final Exam Solutions

Math 114-D

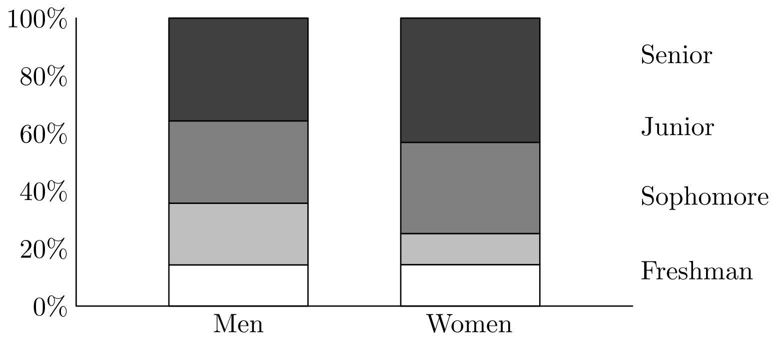
Thursday, May 13, 1999

1. (a) The graph of a normal distribution is bell-shaped.

   (b) Here is the graph of one possible normal distribution:



2. The percentages for Men are 14%, 21%, 29%, and 36%. The percentages for Women are 14%, 11%, 32%, and 43%.



3. Using `TInterval`, the 96% confidence interval for μ is $(5.2501, 12.75)$.

4.

	Over 30	Under 30
Men	67%	33%

	Over 30	Under 30
Women	57%	43%

	Men	Women
Over 30	55%	45%

	Men	Women
Under 30	45%	55%

5.

877655	1	001289
993	2	
52	3	67
2	4	027
	5	0

6. The Central Limit Theorem states that the sum or average of repeated, independent measurements of the same variable will have approximately a normal distribution. For our degree of accuracy, the sum must be of at least 30 items.

7. A measure of spread is resistant to outliers if its value does not change when if the least value is decreased or the greatest value is increased. Of the measures that we have discussed in class, only the inter-quartile range is resistant to outliers.

8. We are testing $H_0 : \theta = 0.3$ versus $H_a : \theta < 0.3$. Using `1-PropZTest` with $p_0 = 0.3$, $x = 25\%(1200) = 300$, and $n = 1200$, we have our test statistic $z = -3.78$ and our p-value $= 7.9 \times 10^{-5}$. Since the p-value is less than our 1% significance level, we reject H_0 and conclude that the company is hiring less than 30% of its older applicants long-term, i.e., that there is age discrimination.

9. (a) Using `1-PropZInt` with $x = 43\%(2000) = 860$ and $n = 2000$, we have $(0.40726, 0.45274) = (40.7\%, 45.3\%)$ as our 96% confidence interval for θ.

   (b) The margin-of-error is $\frac{1}{2}(0.45274 - 0.40726) = 0.02274 = 2.3\%$.

10. The regression line is $y = 0.65 + 0.55x$, and the correlation coefficient for the data is $r = 0.9723$.

11. The standard deviation of sample data is the square root of the average of the square of the differences between each value and the sample mean. It is a measure of how spread out the sample data is.

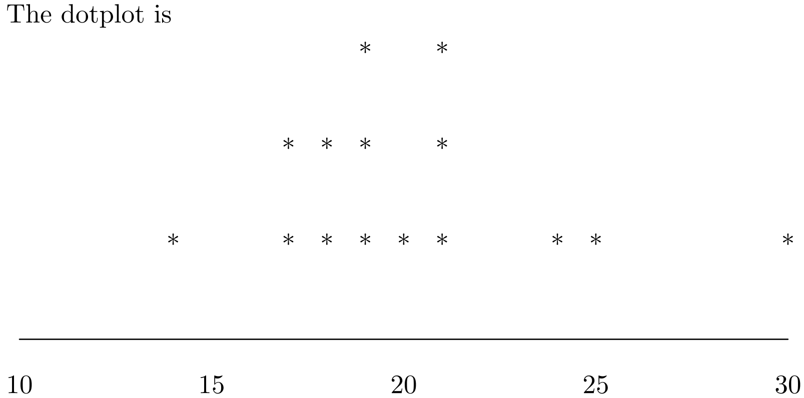
12. We look for the probability that $\hat{p} > 0.23$:

$$
\begin{aligned}
\theta &= 0.2 \\
\sigma_{\hat{p}} &= \sqrt{\frac{\theta(1-\theta)}{n}} \\
&= \sqrt{\frac{0.2(1-0.2)}{500}} \\
&= 0.01789 \\
\Pr(\hat{p} > 0.23) &= \Pr\left(z > \frac{0.23-0.2}{0.01789}\right) \\
&= \Pr(z > 1.68) \\
&= 1 - \Pr(z \le 1.68) \\
&= 1 - 0.9535 \\
&= 0.0465 \\
&= 4.65\%
\end{aligned}
$$

13. The respondents to such surveys are listeners who care strongly enough about the issue to call. They are not representative of the population at large, which includes non-listeners and people who don't care strongly about the issue.

14. We are testing $H_0 : \theta_1 = \theta_2$ versus $H_a : \theta_1 < \theta_2$, where θ_1 is the proportion of teenagers that listen to AM radio, and θ_2 is the proportion of adults over 40 who listen to AM radio. Using `2-PropZTest` with $x_1 = 13\%(500) = 65$, $n_1 = 500$, $x_2 = 21\%(300) = 63$, and $n_2 = 300$, we have our test statistic $z = -2.99$ and our p-value $= 0.0014$ which is less than our significance level of 5%. We reject H_0, and conclude that a smaller proportion of all teenagers listen to AM radio then that of all adults over 40.

15. The dotplot is



The distribution is skewed right, with two peaks near the center.

16. We are testing $H_0 : \mu = 0$ versus $H_a : \mu < 0$. Using `T-Test`, we have our test statistic $t = -1.31$, and our p-value $= 0.1037$. Since the p-value is larger than our significance level 1%, we accept H_0, and conclude that there is no significant evidence here that the average temperature in January is below 0.

17. We have 99.5% in the center of the normal distribution between $-z_*$ and z_*, leaving 0.5% for the left and right tails. The left tail has 0.25% in it, and the area to the left of z_* is $0.25\% + 99.5\% = 99.75\% = 0.9975$. Using Table I, $z_* = 2.81$.

18. The regression model is $y = 6.892 + 0.2371x$. The largest positive residual is 2.7362, and it occurs at $x = 10$, $y = 12$.

19. We solve for n:

$$\begin{aligned} n &\geq \frac{0.25z_*^2}{\text{margin}^2} \\ &= \frac{0.25(1.96)^2}{.004^2} \\ &= 60{,}025 \end{aligned}$$

20. Outliers are points that are far away from the rest of the sample data. For single variables, we name as outliers points that are more than 1.5IQR away from the nearest quartile. For paired data, we name as outliers points that have large residuals.

    Outliers need to be examined to see if they result from measurement error, or if they don't belong in the sample.