Name: Key

Mathematical Theory of Statistics, 01:640:481:01, Spring 2006, Prof. Sills

Exam 2

DIRECTIONS:

1. You may use one standard letter size sheet of notes that you have prepared in advance and the statistical tables provided.
2. You may NOT use other notes or books.
3. You may NOT use a calculator. However, the questions have been carefully written to make sure all necessary arithmetic computations are easy.
4. To ensure maximum credit, please show all work. You may use the backs of pages for scatch work, if necessary. In problems which involve a hypothesis test, this means you should sketch the appropriate probability denisity curve, clearly indicating the rejection region and where the test statistic lies in relation to it, and state the final decision, "reject H_0" or "fail to reject H_0."

1. (15 pts) Consider the data:

$$\underbrace{15, 23, 23, 30, 32,}\ \underbrace{34, 34, 38, 41, 68}$$

(a) Draw a stem-and-leaf plot.

6	8
5	
4	1
3	0 2 4 4 8
2	3 3
1	5

(b) Give the "five number summary" (min, Q1, Q2, Q3, max).

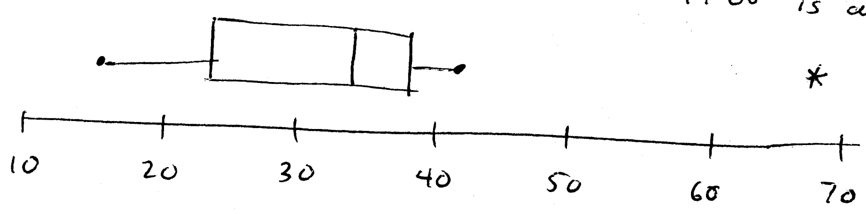
15, 23, 33, 38, 68

(c) Sketch a box and whisker plot, indicating any outliers.

$IQR = 38 - 23 = 15$

$38 + \frac{3}{2} \cdot 15 = 38 + 22\frac{1}{2} = 60\frac{1}{2} < 68$

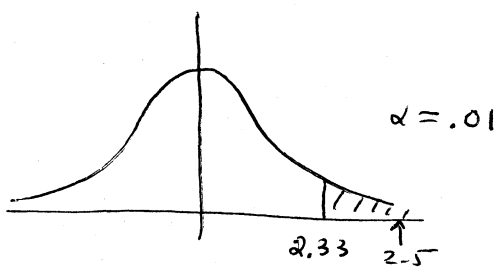
$\therefore$ 68 is an outlier



2. (12 pts) The breaking strengths of cables produced by a manufacturer is known to be normally distributed with a mean of 1800 pounds and a standard deviation of $\sigma = 100$ pounds. By a new technique in the manufacturing process, it is claimed that the breaking strength can be increased. To test this claim, a sample of 25 cables is tested and found to have a mean of 1850 pounds. Assuming that the new technique has no effect on the variability, formulate and perform an appropriate hypothesis test at the $\alpha = 0.01$ level of significance. Also find the P-value for this test.

$H_0 : \mu \overset{(\le)}{=} 1800$

$H_1 : \mu > 1800$



$$T.S. = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{1850 - 1800}{100/5} = \frac{50}{20} = 2.5 > 2.33$$

$\therefore$ Reject H_0.

$$\begin{array}{r} 5000 \\ 4938 \\ \hline 0062 \end{array}$$

$$P \text{ Value} = P(z > 2.5) = .5 - .4938 = .0062$$

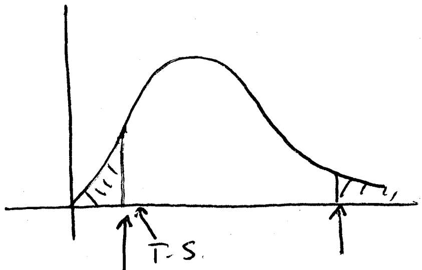
3. (10 pts) Two normal populations are to be studied. We wish to determine whether there is sufficient evidence to conclude that the populations have unequal variances. A sample of size 26 from the first population has a variance of 100, while a sample of size 21 from the second population has a variance of 200. Formulate and perform an appropriate hypothesis test that the $\alpha = 0.1$ level of significance.

$H_0: \sigma_1^2 = \sigma_2^2$ or $H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$

$H_1: \sigma_1^2 \neq \sigma_2^2$ $H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$

$n_1 = 26$ $n_2 = 21$

$s_1^2 = 100$ $s_2^2 = 200$



$T.S. = \frac{s_1^2}{s_2^2} = \frac{100}{200} = 0.5$

$f_{.95, 25, 20}$ $f_{.05, 25, 20}$

$= \frac{1}{f_{.05, 20, 25}} = \frac{1}{2.01} < .05$

$\boxed{\text{Fail to reject } H_0.}$

4. (13 pts) A sample of 50 SAT verbal scores of students at Adams High School has a mean of 540 and a standard deviation of 100. A sample of 200 SAT verbal scores of students from Beck High School has a mean of 500 and a standard deviation of 200. Assume that SAT verbal scores are normally distributed.

Perform the hypothesis test

$$H_0 : \mu_A = \mu_B$$
$$H_1 : \mu_A \neq \mu_B$$

at the $\alpha = 0.01$ level of significance. Also find the P-value for this test.

$\bar{x}_A = 540$ $\qquad$ $\bar{x}_B = 500$

$S_A = 100$ $\qquad$ $S_B = 200$

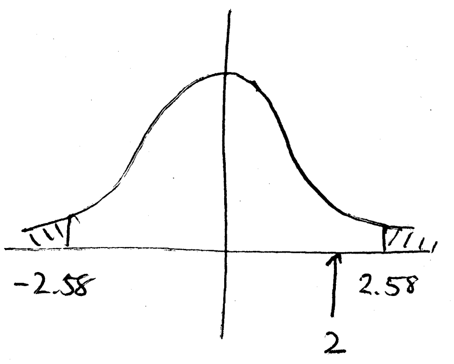
$n_A = 50$ $\qquad$ $n_B = 200$

n large;

use $S_A \approx \sigma_A$

$S_B \approx \sigma_B$

$$T.S. = \frac{\bar{x}_A - \bar{x}_B - 0}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} = \frac{540 - 500}{\sqrt{\frac{100 \cdot 100}{50} + \frac{200 \cdot \cancel{200}}{\cancel{200}}}} = \frac{40}{\sqrt{400}} = \frac{40}{20} = 2$$



$\boxed{\text{Fail to reject } H_0}$

.4991
.5000
.4772
.0228

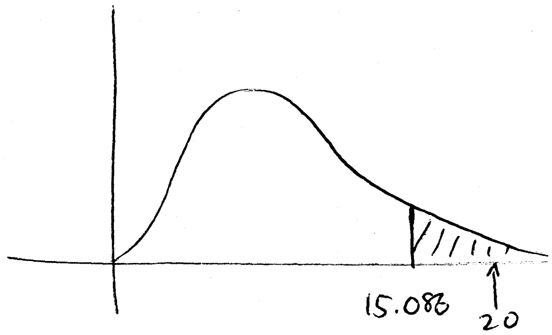
P Value $= 2P(Z > 2.00) = 2(.5 - .4772) = .0456$

5. (10 pts) In the general U.S. population, SAT verbal scores are normally distributed with a mean of 500 and a standard deviation of 100. A guidance counselor at Central High School suspects that the standard deviation of SAT verbal scores at his school, denoted σ_C, is in fact larger than 100. A sample of $n = 6$ SAT math scores from students at Central High School is taken and found to have a mean of 530 and a standard deviation of 200. Formulate and perform an appropriate hypothesis test at the significance level $\alpha = 0.01$.

$H_0: \sigma_c \overset{(\leq)}{=} 100$

$H_1: \sigma_c > 100$ $\quad\quad n = 6$

$$T.S. = \frac{(n-1)s^2}{100^2} = \frac{5 \cdot 200 \cdot 200}{100 \cdot 100} = 20.$$



$\boxed{\text{Reject } H_0}$

6. (20 pts) A discrete random variable X is said to have a *negative binomial distribution* with parameters p and r if its probability density function is given by

$$f(x) = \begin{cases} \frac{(x-1)!}{(r-1)!(x-r)!} p^r (1-p)^{x-r} & \text{if } x = r,\ r+1,\ r+2, \ldots \\ 0 & \text{elsewhere,} \end{cases}$$

where r is a positive integer and p is a real number with $0 \leqq p \leqq 1$. It can be shown by standard techniques that the mean and variance of such a random variable X are given by

$$\mu_X = \frac{r}{p} \qquad \sigma_X^2 = \frac{r(1-p)}{p^2},$$

although you are not asked to do so here.

Use the method of moments to derive estimators for the parameters p and r.

$$\mu_1' = \mu = \frac{r}{p} \qquad \mu_2' = \sigma^2 + \mu^2 = \frac{r^2}{p^2} + \frac{r(1-p)}{p^2}$$

$$= \frac{r^2 + r - rp}{p^2} = \frac{r(r+1-p)}{p^2}$$

$$\left.\begin{array}{l} \mu_1' \approx m_1' = \bar{x} \\ \mu_2' \approx m_2' \end{array}\right\} \Rightarrow \quad p\bar{x} \approx r$$

$$\therefore m_2' \approx \frac{p\bar{x}(p\bar{x}+1-p)}{p^2} = \frac{\bar{x}(p\bar{x}+1-p)}{p}$$

$$\therefore m_2' \approx \bar{x}^2 + \frac{\bar{x}}{p} - \bar{x}$$

$$\Rightarrow \frac{\bar{x}}{p} \approx m_2' - \bar{x}^2 + \bar{x} \Rightarrow p \approx \frac{\bar{x}}{m_2' - \bar{x}^2 - \bar{x}}$$

$$\Rightarrow \boxed{\hat{p} = \frac{\bar{x}}{m_2' - \bar{x}^2 + \bar{x}}}$$

$$\boxed{\text{and } \hat{r} = \hat{p}\bar{x} = \frac{\bar{x}^2}{m_2' - \bar{x}^2 + \bar{x}}}$$

7. (20 pts) An important special case of the negative binomial distribution (encountered in the preceding problem) occurs when the parameter r is equal to 1. This special case is called the *geometric distribution* with parameter p. Accordingly, a random variable X which has a geometric distribution with parameter p, where p is real and $0 \leqq p \leqq 1$, has probability density function is given by

$$f(x) = \begin{cases} p(1-p)^{x-1} & \text{if } x = 1,\ 2,\ 3, \ldots \\ 0 & \text{elsewhere.} \end{cases}$$

In this case we may write $X \sim \text{Geom}(p)$.

(a) Find the maximum likelihood estimator for p. Let $X_1, \ldots, X_n \sim Geom(p)$

$$L(p) = \left(p(1-p)^{x_1-1}\right)\left(p(1-p)^{x_2-1}\right) \cdots \left(p(1-p)^{x_n-1}\right)$$

$$= p^n (1-p)^{\Sigma x_i - n}$$

$$\ln L(p) = n \ln p + (\Sigma x_i - n) \ln(1-p)$$

$$\frac{d}{dp} \ln L(p) = \frac{n}{p} - \frac{\Sigma x_i - n}{1-p}$$

$$\frac{d}{dp} \ln L(p) = 0 \iff \frac{n}{p} = \frac{\Sigma x_i - n}{1-p} \iff \frac{1-p}{p} = \frac{\Sigma x_i - n}{n}$$

$$\iff \frac{1-p}{p} = \bar{x} - 1 \iff 1-p = p\bar{x} - p \iff 1 = p\bar{x} \iff p = \frac{1}{\bar{x}}$$

Thus $p = 1/\bar{x}$ is a critical pt of $\ln L(p)$ and $\therefore$ of $L(p)$

since $L(p)$ is a conts fn of p on $[0,1]$

the abs max occurs at endpt or crit. pt.

$L(0) = 0$

$L(1) = 0$

$L(\frac{1}{\bar{x}}) > 0 \quad \therefore \quad \hat{p} = \frac{1}{\bar{x}}$

(b) Show that if $X \sim \text{Geom}(p)$, the corresponding moment generating function is

$$M_X(t) = \frac{pe^t}{1-(1-p)e^t}.$$

$$M_X(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} p (1-p)^{x-1}$$

$$= e^t p + e^{2t} p(1-p) + e^{3t} p (1-p)^2 + \ldots$$

$$= \frac{pe^t}{1-(1-p)e^t}$$ (by summation of a geometric series)

(provided $|(1-p)e^t| = (1-p)e^t < 1$.

i.e. $e^t < \frac{1}{1-p}$

i.e. $t < -\ln(1-p)$.)

WARNING: The extra credit is *difficult.* Do not attempt unless you are satisfied with your performance on the main part of the exam and have plenty of time left over!

8. **(Extra credit)**

The *Pell sequence* $0, 1, 2, 5, 12, 29, \ldots$ is defined by

$$\begin{aligned} p_0 &= 0, \\ p_1 &= 1, \\ p_{n+1} &= 2p_n + p_{n-1} \quad \text{if } n \geqq 1. \end{aligned}$$

Let $P(x)$ denote the ordinary power series generating function for p_n, i.e. define

$$P(x) := \sum_{n=0}^{\infty} p_n x^n.$$

Express $P(x)$ as a rational function of x.

$$P(x) = P_0 + P_1 x + P_2 x^2 + P_3 x^3 + \cdots$$
$$= 0 + x + 2x^2 + 5x^3 + \cdots$$

$$P_{n+1} = 2P_n + P_{n-1} \quad \text{if } n \geq 1$$

$$\Rightarrow P_{n+1} x^n = 2P_n x^n + P_{n-1} x^n$$

$$\Rightarrow \sum_{n=1}^{\infty} P_{n+1} x^n = 2\sum_{n=1}^{\infty} P_n x^n + \sum_{n=1}^{\infty} P_{n-1} x^n$$

$$\Rightarrow (P_2 x + P_3 x^2 + P_4 x^3 + \ldots) = 2P(x) + (P_0 x + P_1 x^2 + P_2 x^3 + \ldots)$$

$$\Rightarrow \frac{1}{x}(P_2 x^2 + P_3 x^3 + P_4 x^4 + \ldots) = 2P(x) + x(P_0 + P_1 x + P_2 x^2 + \ldots)$$

$$\Rightarrow \frac{1}{x}(P(x) - x) = 2P(x) + xP(x)$$

$$\Rightarrow \frac{P(x) - x}{x} = (x+2)P(x)$$

$$\Rightarrow P(x) - x = x(x+2)P(x)$$

$$\Rightarrow -x = (x^2 + 2x)P(x) - P(x)$$

$$\Rightarrow -x = P(x)[x^2 + 2x - 1]$$

$$\Rightarrow \boxed{P(x) = \frac{-x}{x^2+2x-1} = \frac{x}{1-2x-x^2}}$$

EVEN STRONGER WARNING: The warning before problem 8 applies doubly to this problem.

9. (**EXTRA extra credit**) Use your solution the problem 8 to prove that a closed form formula for p_n is given by

$$p_n = \frac{(1+\sqrt{2})^n - (1-\sqrt{2})^n}{2\sqrt{2}}.$$

$$P(x) = \frac{x}{1-2x-x^2} = \frac{x}{[1-(1+\sqrt{2})x][1-(1-\sqrt{2})x]} = \frac{A}{1-(1+\sqrt{2})x} + \frac{B}{1-(1-\sqrt{2})x}$$

for some A, B. Use partial fractions:

$$x = A(1-(1-\sqrt{2})x) + B(1-(1+\sqrt{2})x)$$

$$\Rightarrow x = (A+B) - x(A(1-\sqrt{2}) + B(1+\sqrt{2}))$$

So $A+B=0 \Rightarrow B=-A$.

$$A(1-\sqrt{2}) + B(1+\sqrt{2}) = -1$$

$$\Rightarrow A(1-\sqrt{2}) - A(1+\sqrt{2}) = -1$$

$$\Rightarrow A - A\sqrt{2} - A - A\sqrt{2} = -1$$

$$\Rightarrow A = \frac{1}{2\sqrt{2}} \text{ \& } B = \frac{-1}{2\sqrt{2}}$$

Thus $P(x) = \sum_{n=0}^{\infty} p_n x^n = \frac{x}{1-2x-x^2} = \frac{1}{2\sqrt{2}}\left(\frac{1}{1-(1+\sqrt{2})x} - \frac{1}{1-(1-\sqrt{2})x}\right)$

$$= \frac{1}{2\sqrt{2}}\left(\sum_{n=0}^{\infty}(1+\sqrt{2})^n x^n - \sum_{n=0}^{\infty}(1-\sqrt{2})^n x^n\right)$$

$$= \sum_{n=0}^{\infty} \frac{(1+\sqrt{2})^n - (1-\sqrt{2})^n}{2\sqrt{2}} x^n$$

Extracting coefficient of x^n we find

$$p_n = \frac{(1+\sqrt{2})^n - (1-\sqrt{2})^n}{2\sqrt{2}}$$ as desired.