MATH 111: REVIEW FOR FINAL EXAM

SUMMARY STATISTICS

Spring 2005 exam: 1(A), 2(E), 3(C), 4(D)

Comments:

This is very simple, just enter the sample into a list in the calculator and go to STAT – CALC – 1-Var Stats. You will get

$\bar{x} = 4$
$Sx = 5.207$
$Med = 3$
$Q_3 = 7$

Note that Sx is the sample variance, *not* σx.

Practice problem:

1. Find mean, median, sample standard deviation, and first and third quartiles for this data set: $-2, 0, 0, 3, 5, 11, 12, 14$.

COMBINATORICS

> The number of ways to choose r out of n objects is
>
> $$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{n(n-1)\cdots(n-r+1)}{r!}$$
>
> where both numerator and denominator have r factors in the last expression.

On the calculator you compute $\binom{n}{r}$ by

n – MATH – Prb – nCr – r

so if you compute $\binom{9}{3}$, the display will read: 9 nCr 3   84

Spring 2005 exam: 5(C), 6(B), 8(C), 16(D)

Notes:

Problem 6. First choose the red blocks in $\binom{N}{R}$ ways, then the blue blocks in $\binom{N-R}{B}$ ways, then the green blocks in one way. Multiply and use the definition above to get answer (B).

Problem 16. It makes no difference if we instead of the last three consider the first three. The problem then becomes to compute the probability that 2 out of 3 are red and this is $\binom{5}{2}\binom{5}{1}/\binom{10}{3} = 0.417$, alternative (D). Also note that this problem is lumped together with problems on the binomial distribution but has nothing to do with the binomial distribution!

Practice problem:

2. An urn contains 10 black and 20 red marbles. You choose 8 without replacement. **(a)** In how many ways can this be done? **(b)** What is the probability that you get exactly 3 red marbles?

> If you do repeated choices, the total number of ways in which this can be done is the product of the number of choices in each step.

Spring 2005 exam: 7(E)

Notes:

Problem 7. You choose among 26 letters four times and 10 digits twice. The total number of ways in which this can be done is $26^4 \cdot 10^2 = 45,697,600$ (but there is probably a typo in (A) and this was intended to be the correct answer).

Practice problem:

3. Swedish license plates have three letters followed by three digits. The total number of letters used are 23. How many possible plates are there?

BASIC PROBABILITY CALCULATIONS

Remember that A *and* B is the same as $A \cap B$ and that A *or* B is the same as $A \cup B$.

> I. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
>
> II. A and B are independent if $P(A \cap B) = P(A)P(B)$
>
> III. A and B are mutually exclusive if $P(A \cap B) = 0$
>
> IV. The conditional probability of A given B is
> $P(A|B) = P(A \cap B)/P(B)$

Spring 2005 exam: 9(A), 10(B), 11(C), 12(D)

Notes:

Problem 9. Use I above.

Problem 10. Use IV above to get $P(A \cap B) = P(A|B)P(B) = 0.3$

Problem 11. Draw a Venn diagram and note that the area outside both A and B is 0.4.

Problem 12. Use III above.

Practice problem:

4(a) If $P(A) = 0.25$, $P(B) = 0.35$, and $P(A \cup B) = 0.5$, what is $P(A \cap B)$?
(b) If $P(A) = 0.5$, $P(B) = 0.8$, $P(A|B) = 0.6$, what is $P(A \cap B)$?

BINOMIAL DISTRIBUTION

> The binomial distribution has the parameters n (number of trials) and p (success probability). Its mean is np and its standard deviation is $\sqrt{np(1-p)}$

Spring 2005 exam: 13(B), 14(A)

Notes:

Problem 13. The values included in the range $6 \leq X < 10$ are $6, 7, 8$, and 9. Note that 10 is *not* included. The probability $P(6 \leq X < 10)$ is therefore the same as $P(5 < X \leq 9)$ and on the calculator this is computed as:

binomcdf(20, 0.4, 9) $-$ binomcdf(20, 0.4, 5) $=$ 0.630

Note the order: first n, then p, then the value.

Practice problem:

5. Let x be binomial with $n = 18$ and $p = 0.6$. Compute **(a)** $P(9 \leq x < 13)$ **(b)** the standard deviation of x.

THE NORMAL DISTRIBUTION

> The normal distribution has parameters μ (mean) and σ (standard deviation). It is a continuous distribution so there is no difference between $\leq$ and $<$ or $\geq$ and $>$ in probability calculations (unlike the binomial distribution).

Spring 2005 exam: 20(C), 21(A), 22(D)

Notes:

Problem 20. This is $P(x \leq 67)$ where x is normal with $\mu = 68$ and $\sigma = 2.5$. On the calculator this is computed as

normalcdf(0, 67, 68, 2.5) $= 0.345$

where the 0 represents $-\infty$. You just need to plug in some value that is so far from the mean 68 that it does not affect the probability. Note the order: first the values, then μ and σ.

Problem 21. First compute $P(67 \leq x \leq 69)$:

normalcdf(67, 69, 68, 2.5) $= 0.311$

which means that about 311 out of 1,000 are expected to be in this range.

Problem 22. You are here asked to find the value t which is such that $P(x \leq t) = 0.90$. You do this by

invNorm(0.90, 68, 2.5) $= 71.2$

Practice problem:

6. Let x have a normal distribution with mean 50 and standard deviation 3.1. Find **(a)** $P(x \leq 48)$ **(b)** $P(49 \leq x \leq 52)$ **(c)** the value t which is such that 5% of observations are larger than t.

THE SAMPLE MEAN (AVERAGE)

> If individual observations has mean μ and standard deviation σ, the sample mean of N observations has mean μ and standard deviation $\sigma/\sqrt{N}$.

Spring 2005 exam: 15(D), 23(B)

Comments:

Problem 15. The standard deviation here is $\sigma = \sqrt{np(1-p)} = \sqrt{20 \cdot 0.4 \cdot 0.6} = 2.191$ and the average of $N = 12$ observations has standard deviation $\sigma/\sqrt{N} = 2.191/\sqrt{12} = 0.632$.

Problem 23. The mean is $\mu = 68$ and the standard deviation $\sigma = 2.5$. The average $\bar{x}$ of $N = 36$ observations has mean 68 and standard deviation $\sigma/\sqrt{N} = 2.5/\sqrt{36} = 2.5/6$ and we get

$$P(67 \leq \bar{x} \leq 69) = \text{normalcdf}(67,\ 69,\ 68,\ 2.5/6) = 0.984$$

Practice problem:

7. **(a)** Find the standard deviation of the average of 14 observations in practice problem 5 above. **(b)** Find the probability that the average of 25 observations in practice problem 6 above is between 49 and 51.

CONFIDENCE INTERVALS

Confidence interval for unknown proportion p. If there are x out of n "successes" let $\hat{p} = x/n$.

$$p = \hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

Confidence interval for unknown mean in normal distribution (σ known). The sample mean of n observations is $\bar{x}$.

$$\mu = \bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

In both intervals, $z_{\alpha/2}$ is the value that has $\Phi(z_{\alpha/2}) = 1-\alpha/2$ and gives confidence level $1-\alpha$. If σ is unknown, it is replaced by s and $z_{\alpha/2}$ is replaced by $t_{\alpha/2}$ from the t distribution. The expression after the $\pm$ is called the *margin of error*.

Spring 2005 exam: 24(C), 26(A), 27(B)

Comments:

Problem 24. Do this on the calculator using "1-PropZInt" under STAT – TESTS:

x:35
n:100
C-Level: 0.96

which gives (0.252, 0.448).

Problem 26. The margin of error is $z_{\alpha/2}\sigma/\sqrt{n}$ which we want to be equal to 2. We are given $\sigma = 10$. To find $z_{\alpha/2}$, note that confidence level 99% gives $1-\alpha = 0.99$ and $1-\alpha/2 = 0.995$. To find $z_{\alpha/2}$ use "invNorm" under DISTR:

invNorm(0.995, 0, 1) = 2.576

and we get the equation

$$2.576 \cdot 10/\sqrt{n} = 2$$

which we solve to get $n = 165.89$ which we round up to 166 (you must *always* round *up* to make n large enough).

<u>Problem 27</u>. A *higher* confidence level means that you are *more likely* to catch the parameter in your interval and for this you need the interval to be *longer*, that is, the margin of error to be *larger*.

Practice problem:

8. (a) Out of 120 people tested for a new medicine, 46 developed headaches. Find a 92% confidence interval for the true proportion that will suffer headaches from this medicine. **(b)** Norwegian salmon have weights that are normally distributed with unknown mean μ and standard deviation $\sigma = 8$. A sample of eight salmon gave sample mean $\bar{x} = 25$. Find a 95% confidence interval for μ. **(c)** In (b), how large sample do we need to make the margin of error at most 2?

HYPOTHESIS TESTS

A null hypothesis is tested against an alternative hypothesis. The null hypothesis is always of the type "no difference" of "no effect" and the alternative hypothesis can be two-sided: "difference" or one-sided: "difference in a particular direction." The *significance level* is the probability to reject a true null hypothesis and the p-value measures the strength of evidence in support of the alternative hypothesis against the null hypothesis. The lower the p-value, the stronger the evidence.

On the calculator, to test $\mu = \mu_0$ in the normal distribution use the "Z-Test" if σ is known and the "T-Test" if σ is unknown and estimated by the sample standard deviation. To test $\mu_1 = \mu_2$ with equal but unknown variance in the two samples, use "2-SampTTest" and answer yes to "Pooled." To test $p = p_0$ use the "1-PropZTest".

Spring 2005 exam: 29(D), 30(B), 31(E), 32(B), 33(A)

Comments:

Problem 29. The alternative here is $\mu \neq \mu_0$ (no particular direction of difference). Since we are given a sample standard deviations, use the T-Test:

$\mu_0 : 750$
$\bar{x} : 790$
$Sx : 20$
$n : 4$

which gives p-value 0.028.

Problem 30. Since the group will take action if the breaking strength is lower than claimed, we should have the alternative hypothesis $\mu < \mu_0$, that is, the breaking strength is less than 1880 pounds.

Problem 31. A p-value of 0.35 indicates that there is not much evidence against the null hypothesis. None of the given alternatives A-D is correct.

Problem 32. The null hypothesis is that the proportion equals 75% and the alternative that it is not equal to 75%. Use "1-PropZTest":

$p_0 : 0.75$
$x : 7400$
$n : 10000$
prop$\neq p_0$

which gives p-value 0.021. Since this is smaller than 0.05, we can reject on the 5% level (but not on the 1% level).

Problem 33. The null hypothesis is that the mean weights are the same and the alternative that they are different. The standard deviations in the two populations are equal but unknown. You need to use the T-Test:

$\bar{x}1 = 548$
$Sx1 : 65$
$n1 : 9$
$\bar{x}2 = 497$
$Sx2 : 12$
$n2 : 4$
$\mu_1 \neq \mu_2$
Pooled: Yes

and you get p-value 0.156.

Practice problem:

9. (a) A type of tire is said to last more than 20,000 miles. A sample of 7 tires gave sample mean 23,000 and sample standard deviation 3,000. Data are normally distributed. Test the claim and find the p-value. Can you reject on level 1%? 5%? **(b)** Swedish bears used to have a mean weight of 1,200 pounds. To test if this has changed, nine bears were weighed which gave the sample mean 1,311. The standard deviation is known to be 120. Find the p-value. **(c)** A politician claims to have support from at least half the population. He bases his claim on an opinion poll where 450 out of 860 said that they supported him. Test his claim and give the p-value.

Answers to practice problems.

1. mean= 5.375, median = 4, standard deviation = 6.186, $Q_1 = 0$, $Q_3 =$ 11.5.

2. **(a)** $\binom{30}{8} = 5,852,925$ **(b)** $\binom{20}{3}\binom{10}{5}/\binom{30}{8} = 0.049$

3. $23^3 \cdot 10^3 = 12,167,000$

4. **(a)** $P(A \cap B) = 0.25 + 0.35 - 0.5 = 0.1$ **(b)** $P(A \cap B) = P(A|B)P(B) = 0.6 \times 0.8 = 0.48$

5. **(a)** $P(9 \leq x < 13) = P(8 < x \leq 12) = 0.657$ **(b)** 2.078

6. **(a)** 0.259 **(b)** 0.367 **(c)** 55.1

7. **(a)** $2.078/\sqrt{14} = 0.555$ **(b)** 0.893

8. **(a)** $(0.306, 0.461)$ **(b)** $(19.456, 30.544)$ **(c)** 62

9. **(a)** $p = 0.019$, no, yes **(b)** $p = 0.0055$ **(c)** 0.086 (not much evidence for his claim)