CEE110

Practice Exam

Please work together and enjoy pulling it all together!!

Probability

Probability Example 1)

Consider a deck of 52 different cards.
A) How many outcomes are there if Player 1 through 3 are dealt a card (it matters who gets which card)?

$52 \times 51 \times 50$ $\qquad P_{2,52} = \frac{52!}{49!} = 52 \times 51 \times 50$

$\boxed{P_{2,52} = 132{,}600}$

B) What if all three cards are given to the same player. How many outcomes are there now?

$\binom{n}{k} = \frac{n!}{(3!)(n-k)!} = \frac{132{,}600}{6}$

$\boxed{\binom{52}{3} = 22{,}100}$

Probability Example 2)

Five out of twelve items in a shipment are defective. $\quad \varnothing\varnothing\varnothing\varnothing\varnothing$ OOOOO OO
Four items are chosen and tested. *X*, the number of defective items, is recorded.
4 chosen

What is the sample space, and please give the PMF of *X*.

$P(X=0) = \frac{\binom{7}{4}}{\binom{12}{4}}$ $\qquad P(X=1) = \frac{\binom{5}{1}\binom{7}{3}}{\binom{12}{4}}$

$P(X=2) = \frac{\binom{5}{2}\binom{7}{2}}{\binom{12}{4}}$ $\qquad P(X=3) = \frac{\binom{5}{3}\binom{7}{1}}{\binom{12}{4}}$

$P(X=4) = \frac{\binom{5}{4}}{\binom{12}{4}}$

Probability Example 3]

In our chapter notes, we have this blurb:
If you sit down right now and try to see if you can flip 26 heads in a row, there is a one in 73 million chance you will have that outcome. However, if 300 million people were to have a contest to see who could roll 26 heads, the chances are about 98% that someone would roll them!!

Are these numbers correct? Please show the calculations.

$(.5)(.5) \times \ldots\ldots$

$(.5)^{26} = 1.49 \times 10^{-8}$

$1 - 1.49 \times 10^{-8} = 0.999999985$

$1 - (.999999985)^{300,000,000}$

$1 - .011$

$\boxed{98.8\%}$

Z transformation to Normal Distribution

CLT $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

In Ch 5 (p. 223 and others) in the text, we learned that the mean of a random sample is a random variable that has a normal distribution with mean μ and variance σ^2/n.

To transform to standard normal distribution, we do a Z transformation, which means we substract the mean and divide by the standard deviation.

Products are being tested for quality. Each unit undergoes a stress test and the time to failure is recorded. The time to failure is a random variable with the true population μ = 8.3 seconds and STDEV of 1.6 seconds. For a random sample of 50 units, what is the probability that the sample average is between 8.0 and 9.0 seconds?

$$P(8.0 < \bar{X} < 9.0)$$

$$\boxed{\bar{X} \sim N\left(8.3, \frac{1.6^2}{50}\right)}$$

$$\sim N(8.3, 0.051)$$

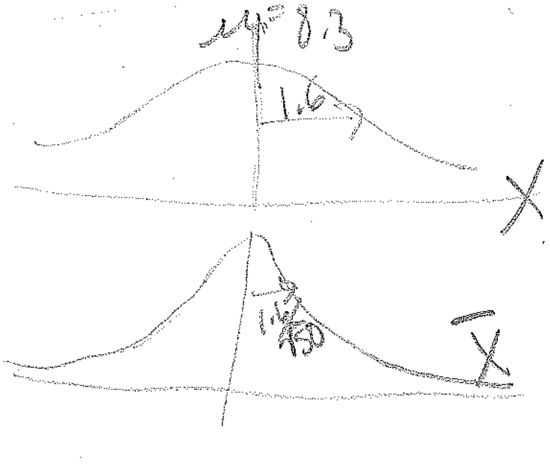
$$P\left(\frac{8-8.3}{\sqrt{.051}} < Z < \frac{9-8.3}{\sqrt{.051}}\right)$$

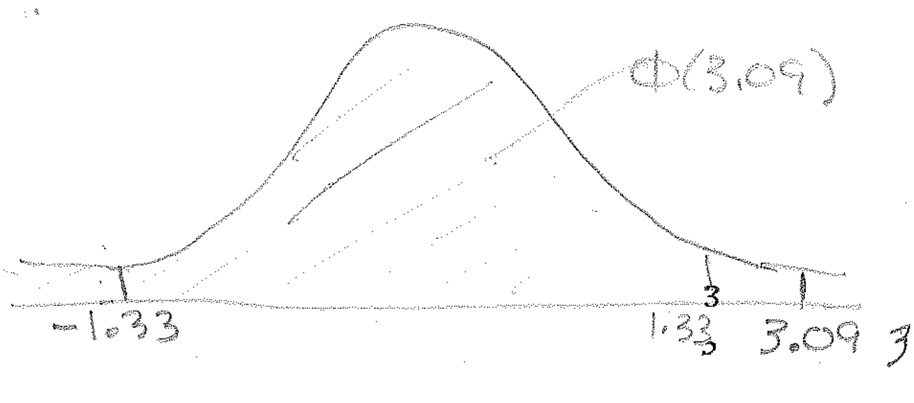
$$P\left(\frac{-0.3}{0.226} < Z < \frac{0.7}{0.226}\right)$$

$$\Phi(-1.33) = 1 - \Phi(1.33)$$
$$= 1 - .9082$$
$$= 0.0918$$

$$P(-1.33 < Z < 3.09)$$

$$\Phi(3.09) - (1 - \Phi(1.33))$$
$$= 0.999 - .0918$$

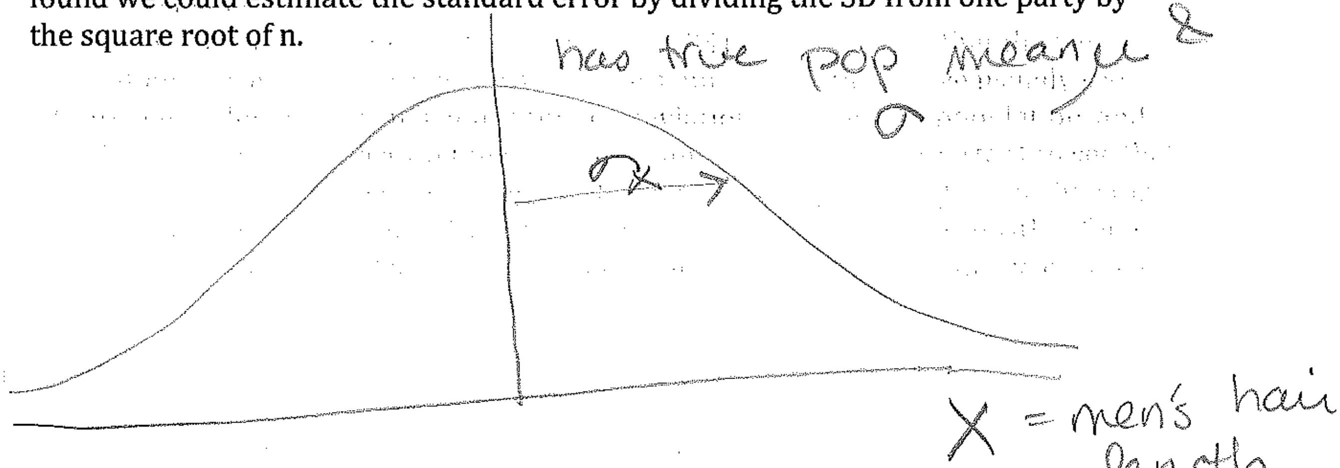


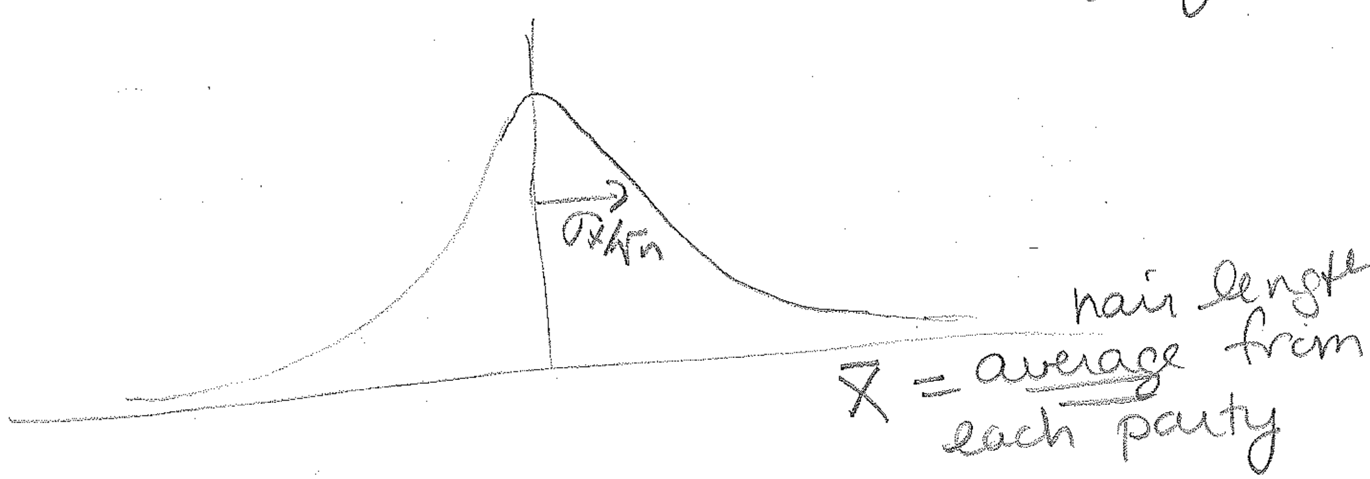


3

Further explanation:
This problem and problem 5.4-2 in the text (p. 224) are unusual in that the true population parameters are given, rather than sample statistics, as we usually see. We are given the true mean μ and standard deviation of the entire population, and asked to calculate statistics on the mean of a sample X_bar. It's important to see that X_bar is itself a random variable, and it's distribution is going to be tighter than that for the whole population of X. The standard dev of the new distribution, that for X-bar, is going to be the SD for the original population divided by the square root of n.

This is exactly analogous to the hair length example that we spent a lot of time on in class. Remember that men's hair length as the X had a pretty broad distribution (large standard deviation), but we looked at X_bar from a party as a random variable. This random variable had a tighter distribution, and we calculated it outright by taking the SD from a number of parties. Indeed the SD on the estimate of the means (also called standard error) was smaller than the SD for each party. We found we could estimate the standard error by dividing the SD from one party by the square root of n.





Ex 1 CI for Proportions

Example 2) Confidence Interval on a proportion

Imagine in your lab you are curious about how many of the bacteria you cultured from some retail meat samples are resistant to a certain antibiotic. You test 200 different bacteria cultured from a certain type of mean, and the fraction resistant is 0.7.

What is the 90% confidence interval on your estimate?

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

$\hat{p} = 0.7$

$\alpha = 0.10 \quad \alpha/2 = 0.05 \quad Z_{\alpha/2} = 1.65$

$0.7 \pm 1.65\sqrt{\frac{(.7)(.3)}{200}} = 0.7 \pm 1.65\sqrt{0.00105}$

$0.7 \pm 1.65(.032) = .053$

What is the 95% confidence interval?

$\alpha = .05 \quad \alpha/2 = .025 \quad Z_{\alpha/2} = 1.96$

$0.7 \pm 1.96\sqrt{\frac{(.7)(.3)}{200}} = 0.7 \pm 0.064$

$\boxed{= 0.64 \text{ to } 0.76}$

What would be the 95% confidence interval for the same observed fraction, but for 600 isolates?

$0.7 \pm 1.96\sqrt{\frac{(.7)(.3)}{600}}$

$\boxed{.7 \pm .037}$

Example 3) Confidence Interval on a mean, small sample size

The price of a certain surgical procedure varies due to different complications between patients.

A simple random sample of thirteen patients yielded the following total costs:

$48,000
$32,800
$34,200
$23,700
$31,300
$28,700
$39,700
$34,000
$33,800
$43,500
$50,400
$24,200
$43,800

Average = $34,931 $\qquad \bar{X} = 34{,}931$

Stdev = $7,350 $\qquad$ \$7,350 $\qquad S = 7350$

What is the 99% confidence interval on this? $\qquad \nu = n - 1 = 12$

Desired % CI

t depends on ν & % CI

$\alpha = .01$

$\alpha/2 = .005$

$t_{12,.005}$ Goto $t_{12,\alpha=.005}$ (99.5%)

$t = 3.055$

$$CI = \$34{,}931 \pm 3.055 \left(\frac{7350}{\sqrt{12}}\right)$$

$$\$34931 \pm 6{,}482$$

Testing of Hypotheses

Testing hypotheses Example 1) Large sample mean

Recall our example with products are being tested for quality. Each unit undergoes a stress test and the time to failure is recorded. For our first sample of 50 units, the average time is 8.3 seconds, with a standard deviation of 0.6.

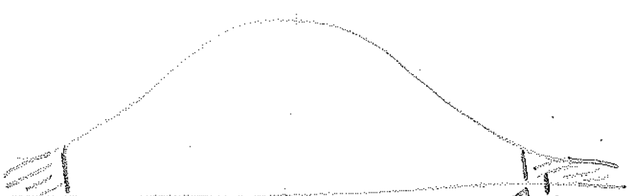
We can now test the hypothesis that our true population mean μ equals some specified value. The null hypothesis is that $\mu = 8$ seconds, so our $\mu_0 = 8$ seconds.

Test statistic:

$$Z_{obs} = \frac{\bar{X} - \mu_0}{SE(\bar{X})} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

$$Z_{obs} = \frac{8.6 - 8}{2/\sqrt{50}} = \frac{0.6}{0.28} = 2.12$$

If null hypoth were true, what is probabil. that we'd see an observation?



2.12

95% Conf 2 tailed

We are past the cutoff, so we reject the null.

$\alpha/2 = .05$ 90% conf. 2-tailed 1.65 Reject null

$\alpha/2 = .005$ 99% conf 2 tailed 2.57 Can't reject null.

Testing Hypotheses Example Z: Proportions

For proportions, the hull hypothesis is that the observed proportion in the sample, p, is equal to some specified value p_0. We saw in class that these specified values are arbitrary -- they are chosen by the curiosity of the investigator. You can test the likelihood that your observed value is statistically indistinguishable from whatever number you'd like to test.

The null hypothesis: H_0: $p = p_0$

Z test statistic is:

Recall our example from the confidence interval section. You tested 200 different bacteria cultured from a certain type of mean, and the fraction resistant is 0.7.

This time let's test with 95% confidence whether our observed proportion, 0.7, is indistinguishable from 0.75 (an arbitrary point).

The alternate hypothesis in this case could be that our proportion differs from 0.75, which would be a two-tailed test.

$$Z_{H_0} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$\hat{p} = 0.70$

H_0: $p_0 = .75$

H_a: $p \neq .75$

$$Z_{obs} = \frac{.7 - .75}{\sqrt{\frac{(.75)(.25)}{200}}} = \frac{.05}{.031} = 1.63$$

-1.63

$\Phi(1.63)$

For 95% conf., 2 tailed cut-off is 1.96

Keep null.

Testing hypotheses Example 3)

Recall from the confidence interval section that the price of a certain surgical procedure varies due to different complications between patients.

A simple random sample of thirteen patients yielded the following total costs:

[illegible]
$32,800
$34,200
$23,700
[illegible]
[illegible]
$39,700
$34,000
$33,800
$43,500
[illegible]
$24,200
$43,800

$$T_{H_0} = \frac{\hat{\theta} - \theta_0}{S_{\hat{\theta}}}$$

Average = $34,931
[illegible] 7350

Test the hypothesis that H_0: μ = $34,000 versus H_a: $\mu \neq$ $34,000 at α = 0.05.

$$T_{H_0} = \frac{34{,}931 - 34000}{7350/\sqrt{13}} = \frac{931}{2039} = 0.457$$

$t_{12,0.025} = 2.179$
($t_{12,97.5\%}$)

8

[illegible]

We can calculate confidence intervals on the difference between two populations, and we can test the hypothesis that the difference between the populations is zero, or is greater than, less than, or equal to a specified value.

Comparing Two Populations Example 1: Proportions

$$\frac{140}{200} = .7$$

Our meat gave a proportion resistant of 0.7 with n = 200.
We'd like to compare that to an organic meat sample, which gave a fraction resistant [illegible]

$p_2 = 0.4$ $\quad n = 205$ $\quad \frac{x}{205} = .4$ $\quad \frac{82}{205} = .4$

A) We first learned to calculate the confidence $\boxed{\text{interval on a difference}}$ between two populations.

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{S.E.(\hat{p}_1 - \hat{p}_2)} \qquad SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$SE = \sqrt{\frac{.7(.3)}{200} + \frac{(.4)(.6)}{205}}$$

$$p_1 - p_2 = (\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} SE(\hat{p}_1 - \hat{p}_2)$$

$$p_1 - p_2 = .3 \pm 1.96\,(0.0471)$$
$$= .3 \pm .092$$

✓ 95% 2 tailed

$$\hat{p} = \frac{140 + 82}{200 + 205}$$

B) We can also test the null hypothesis that they are equal. Because we are setting up the hypothesis test as if the null hypothesis were true (and then we see if we [illegible] need to pool our data for the standard error.

$$\hat{p} = \frac{222}{405} =$$

Our null hypothesis is that H_0: $p_1 - p_2 = \Delta_0$

[illegible] case, the test statistic is simplified because the numerator is just the difference between your two observed proportions. However, you can see how you could test whether the difference were equal to a non-zero Δ_0; the Δ_0 would be subtracted in the numerator.

Need a pooled SE_0

$$SE_0(\hat{p}_1 - \hat{p}_2) = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$\hat{p} = \frac{222}{405} = 0.55$

$Z = \frac{\hat{p}_1 - \hat{p}_2}{SE_0(\hat{p}_1 - \hat{p}_2)}$

$Z = \frac{0.3}{0.0471} = 6.36$ Reject null!

Comparing Two Populations Example 2) Sample means

We previously saw for a particular hospital:
$n_1 = 13$
$Average_1 = \$34,931$
$SD_1 = \$7,350$

For a second hospital:
$n_2 = 14$
$Average_2 = \$29,800$
$SD_2 = \$1,200$

For a third hospital:
$n_3 = 15$
$Average_3 = \$37,200$
$SD_3 = \$6,900$

A) Please compare hospital 1's average cost with that of hospital 2.

$\frac{(7350^2)}{(1200)^2} = 37$ So not equal can't pool

use SS to get v

$$v = \left[\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \right] = \frac{\left(\frac{7350^2}{13} + \frac{1200^2}{14} \right)^2}{\frac{(7350^2/13)^2}{12} + \frac{(1200^2/14)^2}{13}} = \boxed{13}$$

$$T_v = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{n_2}}} = \frac{34931-29800}{\sqrt{\frac{7350^2}{13} + \frac{1200^2}{14}}} = \boxed{2.49}$$

2.49

$T_{13, 97.5} = 2.160$

we fall here

$T_{13, 99\%}$ 2.650

→ For 95% certainty ($\alpha = .05$, $\alpha/2 = .025$), we are in reject null hypoth ⇔ can say they're different

→ For 98% certain (α .02, $\alpha/2 = .01$) cannot reject null.

B) Please compare hospital 1's average cost with that of hospital 3.

$$\frac{7350^2}{6900^2} = 1.13 \quad \text{So equal variance}$$

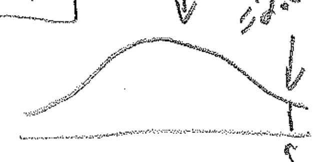
(2) $$\bar{X}_1 - \bar{X}_2 \pm t_{n_1+n_2-2, \alpha/2} \sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

(1) $$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

$$= \sqrt{\frac{(12)(7350^2) + (14)(6900^2)}{13+15-2}} = \boxed{50877} \; S_p$$

our result $= .12$

cutoff $= 2.056$



$$\frac{\bar{X}_1 - \bar{X}_2 \; \cancel{(-\mu_1 - \mu_2)}}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{34931 - 37200}{\sqrt{50877^2\left(\frac{1}{13} + \frac{1}{14}\right)}} = \boxed{0.12} \sim T_{n_1+n_2-2} \sim T_{26} \approx 2.056$$

can't reject null.