(45)



**Biostatistics 200A Midterm**

(10 pts) 1. A researcher wishes to design a study to estimate the **sensitivity** of a new blood test for diagnosing a certain **disease.** The researcher's preliminary data suggests the sensitivity is approximately around 0.76. How many and what kind of subjects should be in the study so that with 95% confidence the estimate of sensitivity is within ± .04 of the true sensitivity?

$\varepsilon = 0.04$

True sensitivity

-0.02 +0.02

~~I believe this is the same as saying that we want our CI to be ± 0.04~~

(Circle one below and then fill in the blank)

8

438

(a) N = 21 subjects with disease who will be given the blood test

(b) N = ________ subjects without disease who will be given the blood test

(c) N = ________ subjects of whom half have disease and half do not, who will be given the blood test

positive test

based on preliminary data

$$\text{Sensitivity} = P(+ \mid D) \approx 0.76$$

person has disease is given

~~We want to control length of the confidence interval so that~~

Sensitivity should be binomially distributed w/ $\hat{p} = 0.76$

$$\cancel{CI = 0.76 \pm (\text{statistic})\left(\frac{s}{\sqrt{n}}\right)}$$

For an estimate of variance, let's use

$$Var(\hat{p}) = \frac{(1-\hat{p})\hat{p}}{N}$$

$$\cancel{(\text{statistic})\left(\frac{s}{\sqrt{n}}\right) = 0.04}$$

$$N = \left(\frac{Z_{0.975}}{\varepsilon}\right)\sqrt{\frac{(\hat{p})(1-\hat{p})}{N}}$$

$$= \left(\frac{1.96}{0.04}\right)\sqrt{\frac{(0.76)(0.24)}{N}} = 20.93$$

$N \geq 21$ And

~~N = 21 or more~~

$\alpha = 0.05$

$\varepsilon = 0.04$

$$N = \left(\frac{Z_{1-\frac{\alpha}{2}}\,\sigma}{\varepsilon}\right)^2$$

$$N = \left(\frac{Z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{p})}}{\varepsilon}\right)^2$$

$$N = \left(\frac{Z_{1-\frac{\alpha}{2}}^2 \frac{(1-\hat{p})\hat{p}}{N}}{\varepsilon^2}\right)$$

$$N = \sqrt{\frac{Z_{1-\frac{\alpha}{2}}^2 (1-\hat{p})\hat{p}}{\varepsilon^2}}$$

$n=85$

**(8 pts) 2.** Suppose we have a random sample of 85 blood pressure measurements from a population. Using this data it is calculated that the 90% confidence interval for the mean blood pressure is 117 to 133

$133-117=16$, so we know $\bar{X} \pm 8 = (117, 133)$

Suppose we want to perform the hypothesis test $H_0$: $\mu=130$ vs. $H_1$: $\mu \neq 130$. We want to determine the p-value.

(Circle one and fill in the blank)

8

(a) It is possible to calculate the p value from the available information and the p-value is approximately 0.30 (fill in)

(b) It is not possible to calculate the p-value from the information given. The additional information needed is

______________________________________________(fill in).

CI (90%):

$\bar{x} \pm 8 = (117, 133$

$\bar{X} = 125$

n is large enough for normal distribution

$(Z_{0.95})\left(\frac{s}{\sqrt{n}}\right) = 8 \rightarrow \frac{s}{\sqrt{n}} = \frac{8}{Z_{0.95}}$

$\cancel{s = \frac{8\sqrt{n}}{Z_{0.95}} = \frac{(8)\sqrt{85}}{Z_{0.95}}}$

p-value (two-sided) is probability of getting an $\bar{X}$ as or more extreme than the one we did, so we need a Z-score for our data.

$Z = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \rightarrow Z = \frac{125-130}{\left(\frac{8}{Z_{0.95}}\right)} = \frac{-5}{\left(\frac{8}{1.64}\right)}$

$= -1.025$

$p = 2P(Z < -1.025) \approx 0.30$

4. A computer simulation is performed in which 144 observations are generated from a right skewed probability distribution with population mean (or expectation) μ=49 and population variance $\sigma^2$=100. The sample mean ( $\bar{X}$ ) and sample variance ( $S^2$ ) from the 144 observations is computed.

$n=144$

The simulation is repeated 16 times. This will produce 16 values of both $\bar{X}$ and $S^2$

**(5 pts) A.** The sample variance of the 16 sample means ( the $\bar{X}$'s) will be computed. What is the expected value of that sample variance?

Answer: $\frac{\sigma^2}{n} = \frac{(100)}{144} = 69.44$

Expected value of the standard error of the means is $\frac{s}{\sqrt{n}}$, so expected variance of the means should be $\frac{s^2}{n}$. We have $\sigma$

4

**(5 pts) B.** The mean of the 16 values of $S^2$ will be computed. What is the expected value of this mean?

Answer: 100

Expected value of the sample SD is $\sigma^2$. We are simply taking means of these values, so we are not asking for standard error of means.

5

**(7 pts) D.** Suppose 95% confidence intervals for the population mean are calculated using the simulated data for each of the 16 simulations. We produce 16 confidence intervals. What is the probability that at least 1 of the 16 confidence intervals does **not** include the number 49?

These are 16 independent trials that we can think of as binomially distributed, with $p = 0.05$

↳ probability of "success" here is probability CI will not contain $\mu$

7

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X=0)$$

$$P(X=0) = \binom{16}{0}(0.05)^0(0.95)^{16}$$

$$= (0.95)^{16}$$

$$P(X \geq 1) = 1-(0.95)^{16} = 0.5599$$

$$\boxed{P(X \geq 1 = 0.5599}$$

**(8 pts) C.** Suppose in one simulation we plan to use the simulated data to test the null hypothesis $H_0$: μ =50 versus $H_1$: μ≠50 at significance level .05. What is the probability that the null hypothesis is rejected?

$\mu_0 = 50 \quad \mu_1 = 49$

8

I think, since we know $H_0$ is not true and should be rejected, this is a two-sided power calculation. We are being asked $P(\text{Rej. } H_0 | H_1) = P\left(\frac{\bar{X}-\mu_0}{\frac{s}{\sqrt{n}}} > Z_{0.975}\right)$ given $H_1$ $+ P\left(\frac{\bar{X}-\mu_0}{\frac{s}{\sqrt{n}}} < Z_{0.025}\right)$ given $H_1$

$$= P\left(\frac{\bar{X}-\mu_1}{s/\sqrt{n}} > Z_{0.975} + \frac{\mu_0-\mu_1}{s/\sqrt{n}}\right) + P\left(\frac{\bar{X}-\mu_0}{s/\sqrt{n}} < Z_{0.025} + \frac{\mu_0-\mu_1}{s/\sqrt{n}}\right) = P(Z > (1.96+1.2)) + P(Z < (-1.96+1.2))$$

$$P(Z>3.16) + P(Z<-0.76) = 0.0008 + 0.2236 = 0.2244. \quad \boxed{1-\beta = 0.2244}$$

**(7 pts) 5.** A computer simulation is performed in which 50 observations are generated from a NORMAL probability distribution with population mean (or expectation) $\mu$=36 and population standard deviation $\sigma$=4. From these 50 numbers, we determine the **proportion** of the 50 observations that exceed the number 40.

The simulation is repeated 400 times. In this way, we obtain 400 values for the **proportions**. The sample variance of the 400 proportions is calculated. What is the expected value of that sample variance?

A) That value is 6.72.

B) Cannot be determined from information given. The additional information needed is:

________________________________________________________.

5

We know that 68% of data points in a population will lie within $1\sigma$ of $\mu$. It follows that 32% will lie elsewhere. Given symmetry of $N(36,16)$, we can state that 16% of values will lie above 40. Thus, we can expect $8/50 = 0.16 = p$ to be the proportion. For the variance of the sample proportions, we must note that our expected value is NOT $\sigma^2$ of the original distribution. Importantly, we can observe that $npq > 5$, so we can approximate the results with a normal distribution $X \sim N(np, npq)$.

Thus, the expected value of variance for these samples is $\frac{npq}{N} = (50)(0.16)(0.84)$
$= 6.72$

$Var(\hat{p}) = \frac{pq}{N}$