STT 351 11-5-07 Key

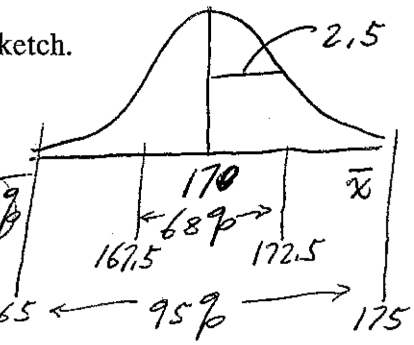
1. A sample of 100 is selected with replacement for score x = accumulated days of sick leave. Population mean mu = 170, sigma = 25.

a. Sketch approx dist of sample mean xBAR. Label mean and sd of your sketch.

$E\bar{x} = \mu$ (POP^N MEAN) $= 170$

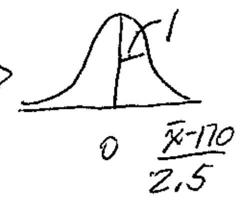
$\sigma_{\bar{x}} = SD\,\bar{x} = \sigma/\sqrt{n} = 25/\sqrt{100} = 2.5$

b. Give a 68% a 95% and a 2.5% region in your sketch (a).



c. Sketch approx dist of (xBAR – 170) / (25 / root(100)). Label mean and sd of your sketch.

STD VERSION OF $\bar{x}$ HAS MEAN 0 (DUE TO SUBTRATION OF 170) + HAS SD 1 (DUE TO DIVISION BY $\sigma_{\bar{x}} = \sigma/\sqrt{n}$) CLT ~ NORMAL DIST^N FOR $\bar{x}$ $\Rightarrow$



d. Give a 68% a 95% and a 2.5% region in your sketch (c).

$[2, \infty)$ (CAN USE 1.96 INSTEAD OF 2)

$[-2, 2]$

$[-1, 1]$

2. Refer to (1).

a. If we are to use the z-table to approximate P(xBAR < 187.6) we need a z-score. What is this z-score?

z-SCORE $\frac{187.6-170}{2.5} = \frac{17.6}{2.5} = 7.04$

b. Approximate (a).

ie $P(\bar{x} < 187.6)$

$= P\left(\frac{\bar{x}-170}{2.5} < \frac{187.6-170}{2.5}\right)$

CLT $\approx P\left(Z < \frac{187.6-170}{2.5}\right)$

$\doteq P(Z < 7.0\ldots) \approx 1$

3. A sample of 50 is selected with replacement for score x = accumulated days of sick leave. From this sample we find xBAR = 181.4 and sample sd s = 18.2.

a. Determine the sample-based estimated values of

population mean mu EST^d BY $\bar{x} = 181.4$

population sd sigma EST^d BY $s = 18.2$

expected value of xBAR $= \mu$ ESTIMATED BY $\bar{x} = 181.4$

sd of r.v. xBAR $= \sigma/\sqrt{n}$ ESTIMATED BY $s/\sqrt{n} = 18.2/\sqrt{50}$

b. Determine a 95% CI for population mean mu. $\bar{x} \pm 1.96\,\frac{18.2}{\sqrt{50}} = 181.4 \pm 1.96\,\frac{18.2}{\sqrt{50}}$

c. What claim is made for the CI methods employed in (b)? $P(\mu \text{ IN } 95\% \text{ CI FOR } \mu) \approx .95$

ie $P(\mu \text{ IN } \bar{x} \pm 1.96\, s/\sqrt{n}) \approx .95$

CANNOT SAY $P(\mu \text{ IN } 181.4 \pm 1.96\,\frac{18.2}{\sqrt{50}}) \approx .95$ IT IS 0 OR 1 !!

d. What is the ESTIMATED margin of error of xBAR?

EST^d MOE = HALF WIDTH OF 95% CI ABOVE

$= 1.96\,\frac{18.2}{\sqrt{50}}$ (TRUE MOE IS $1.96\,\sigma/\sqrt{n}$)

4. R.v. X, Y are independent with

E X = 10 Var X = 67

E Y = 30 Var Y = 44

a. Give E(2 X + 6 Y – 4X + 2 Y – 5) $= 2EX + 6EY - 4EX + 2EY - 5$ 10 30 67 44

b. Give Var(2 X + 6 Y – 4X + 2 Y – 5) $= Var(-2X + 8Y) \overset{INDEP}{=} 4 Var X + 64 Var Y$

c. Give sd(2 X + 6 Y – 4X + 2 Y – 5) $\sqrt{(b)}$

d. Give E (X Y²) $\overset{INDEP}{=} EX\ EY^2$ 44 30^2

$EY^2 = Var Y + (EY)^2$

SINCE $Var Y = EY^2 - (EY)^2$

5. Independent samples are selected from each of populations x, y

xBAR = 40 sx = 5 nx = 100

yBAR = 50 sy = 7 ny = 200

a. Give a 95% CI for mux – muy. $\bar{x} - \bar{y} \pm 1.96 \sqrt{\frac{s_x^2}{n_x} \oplus \frac{s_y^2}{n_y}}$

$(40-50) \pm \boxed{1.96 \sqrt{\frac{25}{100} \oplus \frac{49}{200}}}$

b. Determine the ESTD margin of error of xBAR – yBAR. ESTN MOE OF $\bar{x} - \bar{y}$

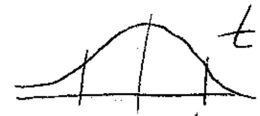
c. Determine the estimate of sd of r.v. xBAR – yBAR. $\sqrt{\frac{s_x^2}{n_x} \oplus \frac{s_y^2}{n_y}}$

d. What is the claim made for the CI (a)? $P\left(\mu_x - \mu_y \in \boxed{(\bar{x} - \bar{y}) \pm 1.96 \sqrt{\frac{s_x^2}{n_x} \oplus \frac{s_y^2}{n_y}}}\right) \approx .95$

6. A process is * in control. A sample of n = 5 yields

xBAR = 45.2 s = 12.4

a. Determine a 95% t-based CI for mux. $DF = n - 1 = 5 - 1 = 4$



$t_{CRITICAL} = 2.776$

$\bar{x} \pm 2.776\ s_x/\sqrt{n}$

$45.2 \pm 2.776\ 12.4/\sqrt{5}$

DF	$t_{CRITICAL}$ CENTRAL CUMULATIVE	95%
4		2.776
∞		1.96

b. What claim is made for (a)?

$P\left(\mu \text{ IN } \bar{x} \pm 2.776 \frac{s_{5\ SAMPLES}}{\sqrt{5}}\right) = .95$ ← WOULD BE EXACT W/ PERFECT TABLED t AND CALCULATIONS.

* VERY CRITICAL TO USE OF t.

c. Determine an n such that a 95% CI of the form $xBAR_n$ +/- 0.2 may be obtained.

FOR $n > 5$ DEPENDING ON $s_{n_0 = 5}$ WE HAVE AVAILABLE HYBRID CI $\bar{x}_n \pm 2.776\ s_{5\ samples}/\sqrt{n}$ ← 4 DF t

⇒ CHOSE n WITH $2.776 \frac{12.4}{\sqrt{n}} = 0.2$

$\Rightarrow n = \left(\frac{2.776\ 12.4}{0.2}\right)^2$

d. If the sample size is extended to n found in (a) what is the 95% CI for mux if the final sample mean $xBAR_n$ is equal to 46.1?

46.1 ± 0.2 AS DESIRED.

REALLY $46.1 \pm 2.776 \frac{12.4}{\sqrt{n}}$ BUT WE CHOSE n SO $\frac{2.776\ 12.4}{\sqrt{n}} \approx 0.2$

ROUNDING ERROR ONLY

7. A population has

mean age men = 55.4 Nmen = 500

mean age women = 62.4 Nwomen = 1000

FRAC. MEN IN $Pop^n = \frac{500}{500+1000} = \frac{1}{3}$

a. Determine the mean age of the population. $\mu = \frac{1}{3}\mu_{MEN} + \frac{2}{3}\mu_{WOMEN}$

$$= \frac{1}{3}(55.4) + \frac{2}{3}(62.4)$$

b. Give the proportionally stratified estimate of population mean age for sample data

xBARmen = 52.5

xBARwomen = 64.3

PROPORTIONALLY STRATIFIED SAMPLE HAS $\frac{1}{3}$ MEN AND $\frac{2}{3}$ WOMEN, SAME AS THE POPULATION DOES.

SO $\bar{x} = \frac{1}{3}(52.5) + \frac{2}{3}(64.3)$. WE DO NOT NEED TO KNOW n.

c. Give the sample mean age for the sample in (b).

SAME AS (b). THE PROP'L STRATIFIED ESTIMATE OF OVERALL POP^N MEAN μ IS JUST THE SAMPLE MEAN.

d. Give the number of men in the sample if the overall sample size in (b) is 300. $\frac{1}{3}(300) = 100$

USE $n = 300$

e. Determine a 95% CI for the overall population mean based on the stratified estimate (b) if the samples give

sample sd sMEN = 6.2

sample sd sWOMEN = 7.4

$\bar{x}$ FROM (b) ± 1.96 (EST'D SD OF $\bar{x}$)

ESTD SD OF $\frac{1}{3}\bar{x}_m + \frac{2}{3}\bar{x}_w$ IS $\sqrt{\left(\frac{1}{3}\right)^2\frac{s_m^2}{n_m} + \left(\frac{2}{3}\right)^2\frac{s_w^2}{n_w}}$

BECAUSE WE ARE SAMPLING EACH STRATUM!!

$$= \bar{x} \pm 1.96\sqrt{\frac{1}{9}\frac{6.2^2}{100} + \frac{7.4^2}{200}}$$

8. A sample of 45 pairs (x = time in kiln, y = strength of glaze) finds

xBAR = 2.5 sx = 0.9

yBAR = 6.1 sy = 1.3 r = 0.7

It is known that the population mean mux = 2.3.

a. The regression estimate of muy is

yBAR + (mux – xBAR) r sy / sx.

Give the estimated margin of error of this estimator.

$$1.96\sqrt{1-r^2}\,\frac{s_y}{\sqrt{n}}$$

$$= 1.96\sqrt{1-.7^2}\,\frac{1.3}{\sqrt{45}}$$

b. By what factor is the margin of error of the regression estimator narrower than the margin of error of yBAR? FACTOR IS $\sqrt{1-.7^2} \approx .7^+$ (ALSO)

c. If the estimator yBAR requires 100 samples to obtain a given precision in estimating muy around how many samples are required by the regression estimator?

$$\left(\frac{1}{\sqrt{1-.7^2}}\right)^2 \times 100 = \frac{1}{1-.49} \times 100 = 200$$

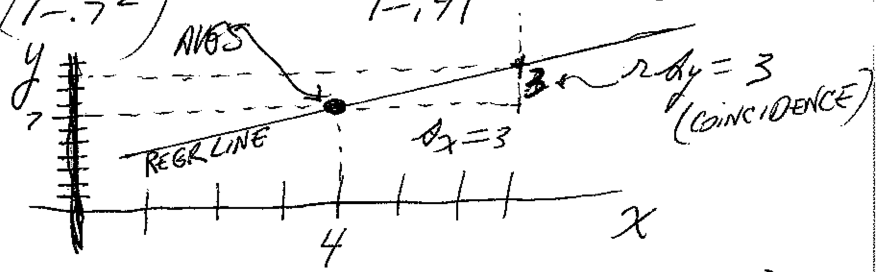
TO REDUCE BY FACTOR $.7^+$ REQUIRES n INCREASE BY FACTOR $\left(\frac{1}{.7^+}\right)^2$

9. Given data

xBAR = 4 sx = 3

yBAR = 7 sy = 5 r = 0.6

a. Sketch a regression line of y on x.



b. For each increase of 5 in x what is the average increase in y (according to the regression line)? SLOPE $= \frac{3}{3} = 1$

SO $x \to x+5$ PRODUCES $y \to y+5$ (ON REGR LINE) (y INCREASES 1:1 WITH x)

4 CATEGORIES

10. Manufactured parts are classified as Best, Good, Average, Worst. We wish to prepare a random sample of 40 parts in order to test the hypothesis that the probabilities of these categories have remained at their established levels 0.2, 0.3, 0.3, 0.2.

a. Determine the expected count for category Best. Is it at least 5? If so, none of the four categories will cause trouble with the 5-criterion, so we may proceed to sampling 40.

EXPECT $(0.2)40 = 8$ IN BEST (WITH RANDOM SAMPLE OF 40)

b. The sample data finds observed counts 4, 12, 14, 10. Determine the contribution of category Best to a chi-square test of the model.

$$\chi^2 = \sum_{4\text{ CELLS}} \frac{(O-E)^2}{E}$$

CATEGORY (CELL) "BEST" CONTRIBUTES $(4-8)^2/8 = \frac{16}{8} = 2$

c. Determine DF for test (b). $DF = k-1 = 4-1 = 3$ "RUBRIC"

GEN'L REASONING: FULL MODEL $p_1 p_2 p_3 p_4$ IS SPECIFIED BY p_1, p_2, p_3. NULL HYP MODEL HAS 0 FREE PARAMETERS SO $3-0=3$

FRACTIONS USUALLY RESULT FOR E_i

d. Use chi-square table to evaluate pSIG if the chi-square statistic works out to 6.1 (it does not).

$P_{SIG} = P(\chi^2_{3DF} > 6.1) > .100$ BY TABLE VII ("OFF" TABLE) — CORRECTION —

e. From (d) what action is taken if we test the null hypothesis H0 that the given model is correct? FOR $\alpha = .05$ (SAY)

REJ. $H_0: \{.2, .3, .3, .2\}$ IF $P_{SIG} < \alpha = .05$ (SAY). IN THIS CASE WE FAIL TO REJECT H_0 SINCE .10 IS NOT LESS THAN .05.

11. Same setup as (10) except we wish to test the null hypothesis H0: p1 = p4 and p2 = p3. That is, the probabilities are symmetric as were 0.2, 0.3, 0.3, 0.2. As in (10b) suppose observed counts 4, 12,14, 10. ← O_i

a. Determine the expected count for category Best. Hint: If the probabilities are symmetric we estimate

p1 = p4 = (4 + 10) / (2 40) = 7 / 40

p2 = p3 = (12 + 14) / (2 40) = 13 / 40 (all together sum to one).

FIRST ESTIMATE $\hat{p}_1 = \frac{4+10}{2(40)} = \frac{7}{40}$

THEN $E_{BEST} = \left(\frac{7}{40}\right)40 = 7$ (JUST AVG OF BEST + WORST).

b. Determine the contribution of category Best to the chi-square statistic.

$$\frac{(4-7)^2}{7} = 9/7.$$

c. Determine the DF of the chi-square test of H0: the probabilities are symmetric.

$DF = \#\text{PARAM FULL MODEL} - \#\text{PARAM } H_0 \text{ MODEL} = 3 - 1 = 2$

FULL MODEL $p_1 p_2 p_3 p_4$ IS SPECIFIED BY 3 PARAMETERS p_1, p_2, p_3.

H_0 MODEL $p_1 p_2 p_2 p_1$ IS SPECIFIED BY p_1 ONLY, SINCE $p_2 = (1-2p_1)/2$.

d. Are all expected counts at least 5?

BEST E_i IS 7 (SAME AS WORST)

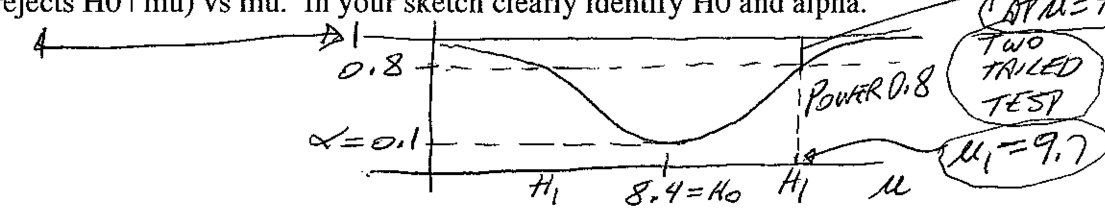
GOOD E_i IS 13 (AVG OF 12 AND 14)

YES!

12. A process is in statistical control. Based on a sample of n = 5 we will test the null hypothesis that mu = 8.4 versus the alternative that mu is no 8.4 (2-sided) at alpha = 0.1.

a. Sketch the curve P(test rejects H0 | mu) vs mu. In your sketch clearly identify H0 and alpha.

$\beta = 0.2$ AT $\mu = 9.7$



b. For your sketch, locate mu = 9.7 on the mu-axis if the power there is 0.8. Do this by eye. Clearly show the power in your sketch.

c. In (b) show the type two error beta at mu = 9.7.

13. Refer to 12.

a. Give the t-value (critical value) required from the table for this two-sided t-test with alpha = 0.1 and n = 5.

$n=5 \quad DF = n-1 = 4$ 2 TAILED TEST $P(|t_{CRITICAL}| > t_{CRITICAL}) = .1$

$t_{CRITICAL} = 2.132$

TABLE IV DF 4 .9 ← CENTRAL $\boxed{2.132}$

b. If the data is xBAR = 8.7, sx = 0.4 what is the t-statistic?

$t_{STATISTIC} = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{8.7 - 8.4}{0.4/\sqrt{5}}$ REJECT H_0 IF $|t_{STAT}| > 2.132$

c. If the t-statistic works out to 2.5 (it does not) what action will the test take and why?

2-TAILED

2 SIDED TEST → $|2.5| > 2.132$ SO REJECT $H_0: \mu = 8.4$ USING 2-SIDED t-TEST FOR $\alpha = 0.1$

d. If the t-statistic works out to –2.5 (it decidedly does not) what action will the test take and why?

$|-2.5| > 2.132$ SO REJECT H_0 AS WITH (c).

14. A sample of electronic modules i = 1 to 100 are each scored for

yi = reliability

x1i = 1 if supplier A, 0 if supplier B

x2i = insulation value of the circuit board

x3i = price

A least squares fit is made to the data finding estimated model

y = 0.4 – 0.1 x1 + 0.6 x2 + 0.2 x3 (i.e. betaHAT values 0.4, -0.1, 0.6, 0.2).

Taking the fitted model at face value,

"FITTED MODEL" = "ESTIMATED MODEL"

a. What is the indicated average change in reliability switching from supplier A to supplier B?

$-0.1\,X1$ GOES FROM (-0.1) (FOR A) TO $(-0.1)(0)$ (FOR B) SO SWITCHING TO SUPPLIER B INCREASES (MODEL) MEAN RELIABILITY 0.1.

b. What is the combined indicated average change in reliability switching from supplier A to supplier B while at the same time doubling insulation value of the circuit board? SINCE THE MODEL IS LINEAR IN X1, X2 THE (MODEL) EFFECTS ADD. ANS. $0.1 + 0.6\,(2X_2 - X_2) = 0.1 + 0.6\,X_2$ THIS DEPENDS ON X2.

c. What is the fitted value for a module supplied by B, with board insulation value 5.2 and price 1.34?

$\hat{y} = 0.4 - 0.1(0) + 0.6(5.2) + 0.2(1.34)$

d. Refer to (c). What is the residual for that module if its reliability (determined by testing it) is y = 3.51?

$y - \hat{y} = 3.51 - (c)$.

e. If this sample was with replacement and equal probability on modules on hand and if it were known that the population mean values for all such parts were

mu of x1 = fraction of parts supplied by A = 0.4

mu of x2 = average insulation value of all parts = 9.3

mu of x3 = mean price of all parts = 2.67

what would be the regression-based estimate of the population mean muy? i.e. $\hat{\mu}_{y\,REGRESSION}$

$\hat{\mu}_{y\,REGRESSION} = 0.4 - 0.1(0.4) + 0.6(9.3) + 0.2(2.67)$

SIMPLY INSERT INFORMATION ON X-MEANS INTO FITTED MODEL.

f. If the multiple correlation (i.e. the correlation between y and fitted y values) is 0.8 and the sample sd of ~~the residuals is~~ y 1.4 what is the estimated margin of error of (d)? CORRECTION

CORRECTION

$\sqrt{1-R^2}\,\dfrac{s_y}{\sqrt{n}} = \sqrt{1-.8^2}\,\dfrac{1.4}{\sqrt{100}}$ ← n=100 (ABOVE)