Note Title 3/29/2010

# STT 200 3pm 3-29-10

## REVIEW FOR EXAM 3.

1. YOU WILL NEED TO KNOW ALL RELEVANT FORMULAS.
2. 21 MULT. CHOICE QUESTIONS.
3. $t$, $\chi^2$ TABLES PROVIDED.

$t$-BASED CI FOR $\mu$ $\quad DF = n-1$

$P(\mu \text{ IN } \bar{x} \pm t^{DF}_{\alpha} s/\sqrt{n}) = .95$

$t^{DF}_{\alpha}$ : $11, .95$

IF CALC ARE ALL TO INFINITE PRECISION

FOR ALL $n > 1$

If $n = 12$

| DF | $t_{11}$ |
|---|---|
| 11 | |
| Conf | 95% |

PROVIDED POPULATION DIST IS NORMAL.

* DO KNOW ALSO $\sigma = \sqrt{\frac{(x_1-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n-1}}$ (ROOT)

(SAMPLE S.D.)

CENTRAL LIMIT THEOREM

SAYS SUMS (OR AVGS) OF INDEPENDENT R.V. $\approx$ NORMAL R.V. (UNDER CONDITIONS)

$\Rightarrow \approx$ $n \to \infty$

WITH-REPL (INDEP)

$P(\mu \text{ IN } \bar{x} \pm z \frac{\sigma}{\sqrt{n}}) \to$ CONF LEVEL

e.g. $1.96 \leftrightarrow .95$

WITHOUT REPL. (USE FPC)

$P(\mu \text{ IN } \bar{x} \pm z \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}})$

EXAMPLE 1. z-BASED CI. SUPPOSE WE SAMPLE 49 STUDENTS FROM CLASS ($N = 120$, $n = 49$) (49 "LARGE").

SAY WE FIND GRADE AVG (IN SAMPLE) $\bar{x} = 2.83$

AND SCORES $x_1 = 2.11$, $x_2 = 1.83$, $\cdots$, $x_{49} = 3.89$

CALC SAMPLE SD $s = \sqrt{\frac{(2.11-2.83)^2 + \cdots + (3.89-2.83)^2}{49-1}} = 1.61$

**DO USE A CALCULATOR**

THE FOR THIS SAMPLE (IF DRAWN W/ WITHOUT REPL.)

↪ 95% z-BASED CI FOR $\mu$

$$\bar{x} \pm z \, s/\sqrt{n} \sqrt{\frac{N-n}{N-1}}$$

"MARGIN OF ERROR"

$$2.83 \pm 1.96 \, \frac{1.61}{\sqrt{49}} \sqrt{\frac{120-49}{120-1}}$$

$\{ \quad , \quad \}$

$-$ $+$

THIS INTERVAL IN NUMERICAL –

$\sigma_{\bar{x}}$ WITHOUT REPL $= \frac{\sigma_{pop}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

USE $s$ INSTEAD OF $\sigma_{pop}$

$P(\mu \text{ IN } (2.83 \pm \cdots)) = 0 \text{ or } 1$

---

TWO POPULATION COUNTERPARTS.

SAMPLE OF 600 FROM HOMEOWNERS.
OF WHICH 120 ARE "IN MORTGAGE TROUBLE"
THINK $\hat{p}_1 = 120/600$

3 BASED CI FOR $p$ $\quad \frac{120}{600} \pm 2\sqrt{\frac{\frac{120}{600}\frac{480}{600}}{600}}$

$\hat{p} \pm 1.96$ "

[INDEPENDENTLY OF THIS]

SAMPLE 400 RENTERS OF WHICH 82 ARE

IN RISK OF "LOSING THEIR RENTAL."

$$\hat{p}_2 = 82/400$$

NOW THE CI FOR $p_1 - p_2$

$$(\hat{p}_1 \ominus \hat{p}_2) \pm z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} \oplus \hat{p}_2\hat{q}_2/n_2}$$

$$\left(\frac{120}{600} - \frac{82}{400}\right) \pm 1.96\sqrt{\frac{120}{600}\frac{480}{600}/600 \oplus \frac{82}{400}\frac{318}{400}/400}$$

$P(p_1 - p_2$ IN ABOVE TYPE $) \sim .95$

$n_1, n_2$ LARGE

---

FOR $\mu_1 - \mu_2$ $\quad (\bar{x}_1 \ominus \bar{x}_2) \pm z\sqrt{\sigma_1^2/n_1 \oplus \sigma_2^2/n_2}$

VARIANTS

$$(\bar{x}_1 \ominus \bar{x}_2) \pm z \sqrt{\sigma_1^2/n_1 \frac{N-n}{N-1} \oplus \sigma_2^2/n_2 \frac{N-n}{N-1}}$$

AS REQ'D.

THERE'S EVEN A $t$-VARIANT FOR SMALL $n_1, n_2$.

---

$\chi^2$ (CHI-SQUARE) TESTS.

GEN'L $\chi^2$ STATISTIC $\sum_{\text{CELLS (CATEGORIES)}}^{1} (O-E)^2/E$

GOODNESS OF FIT of MODEL: $E$ $M$ $\neq$ SORTING BASED ON

SAMPLE $n = 400$ ~~PEOPLE~~ MEN

DO YOU USE PRODUCT X ?

MODEL: $M$ $M^c$ $\Rightarrow$ EXPECTED COUNTS IN 400

$.25$ $.75$

| $m_{DO}$ | $m_{NOT}$ |
|---|---|
| 100 | 300 |

SUPPOSE WE SAMPLE FINDING OBSERVED COUNTS:

| $m_{DO}$ | $m_{NOT}$ |
|---|---|
| 110 | 210 |

| | $m_{DO}$ | $m_{NOT}$ |
|---|---|---|
| O | 110 | 210 |
| E | 100 | 300 |

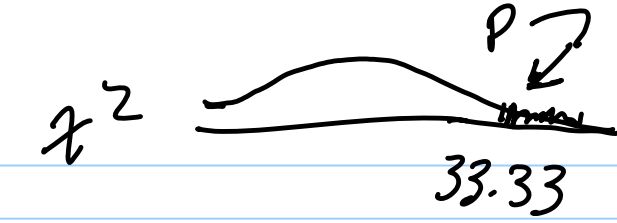
$$\sum \frac{(O-E)^2}{E} = \frac{(110-100)^2}{100} + \frac{(210-300)^2}{300}$$

TAKE CARE

$$= 25 + \frac{25}{3} = 33.33$$

IF LARGE, ITS EVIDENCE AGAINST MODEL

$DF = \#CELLS - 1 = 2 - 1 = 1$




So P-VALUE $<< 0.00001$

.00001
19.5114

MEANS THAT IT IS EXTREMELY RARE TO SEE $\chi^2_{DF=1} > 33.33$

YOU CAN CONTINUE TO BELIEVE IN MODEL .25 .75

BUT TO DO SO YOU MUST BELIEVE ALSO THAT AN EXTREME RARITY HAS JUST HAPPENED.

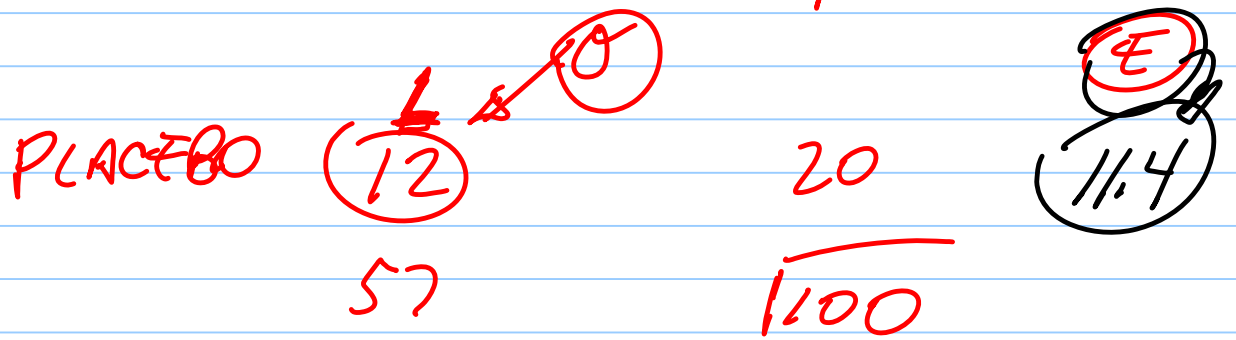
* DO REMEMBER $\chi^2$ REQUIRES THAT ALL EXPECTED COUNTS $E \geq 5$.

SECOND $X^2$ EXAMPLE. TEST OF ~~INDEPENDENCE~~: EQUALITY OF DISTRIBUTIONS.

| | | L | M | H | RESPONSE | OBSERVED COUNTS |
|---|---|---|---|---|---|---|
| FIX 20 | PLACEBO | 12 | 4 | 4 | 20 FIX | |
| FIX 50 | LOW DOSE MED | ~~30~~ 25 | 20 | 5 | 50 FIX | |
| FIX 30 | HIGH DOSE MED. | ~~20~~ 20 | 6 | 8 | 30 FIX | |
| (SUBJECTS) | | 57 | ~~26~~ | 17 | | 0 |

FOR THE TEST OF HOMOGEITY (ROW DISTNS SAME) OR THE TEST OF INDEPENDENCE

THE "EXPECTED COUNTS" DO DEPEND ON THE DATA.



E FOR THIS CELL $\frac{\text{ROW TOT} \cdot \text{COL TOT}}{\text{TOT}} = \frac{20 \cdot 57}{100} = 11.4$

THE CONTRIBUTION OF CELL "PLACEBO L" TO $\chi^2$:

$$\frac{(O-E)^2}{E} = \frac{(12-11.4)^2}{11.4} = \frac{.36}{11.6} \sim .3$$

DO LIKEWISE FOR ALL 9 CELLS, & TOTAL:

APPLICABLE DF $= (R-1)(C-1) = (3-1)(3-1) = 4$ DF

---

SO-CALLED STANDARDIZED RESIDUALS (FOR EACH CELL)

$\frac{O-E}{\sqrt{E}}$ ~ LIKE z-SCORE FOR EACH CELL.

IF YOU SEE THIS LIKE $\sqrt{.2}$ OR $-4.7$ $\Rightarrow$ SOME CELLS ARE POSSIBLY DOMINATING THE $\chi^2$.