CONF-770874--1

ERROR CONTOURS FOR STUDENT'S t UNDER NON-NORMALITY

K. O. Bowman, Union Carbide Corporation
J. J. Beauchamp, Union Carbide Corporation
L. R. Shenton, University of Georgia

1. INTRODUCTION

Geary (*Biometrika* 34, 1947) introduced a differential series for the density function of Student's t under non-normality where the sample population is assumed to have finite moments up to a certain order. The series is obtained by the use of a differential operator on the density of t under normality, where the operator has the effect of replacing one set of cumulants by another.

If $\{K_r\}$ and $\{K'_r\}$ are the cumulants of t under the two regimes (normality and non-normality), then the discrepancies $\{K_r - K'_r\}$ can be ordered with respect to sample size, resulting in an ordering in powers of $1/\sqrt{n}$ (n = sample size) of the modified density and cumulative distribution function of t. Geary carried this out to order n^{-2}. Geary's n^{-1} formulas for the necessary probability integrals (t > 0) are

$$\int_{-\infty}^{-t} f(x)\,dx \sim \int_{-\infty}^{-t} T(x)\,dx + A_1/\sqrt{n} + A_2/n \qquad (1)$$

for the lower probability, and

$$\int_{t}^{\infty} f(x)\,dx \sim \int_{t}^{\infty} T(x)\,dx - A_1/\sqrt{n} + A_2/n, \qquad (2)$$

for the upper probability, where

i. f(t) is the density of t under non-normality;
ii. T(t) is the density of t under normality;
iii. $A_1 = (T_0/2 + T_2/3)\sqrt{\beta_1}$,
$A_2 = (T_1 + 2T_3/3 + T_5/18)\beta_1 + (3 - \beta_2)\,T_3/12$,
with $\sqrt{\beta_1}$ = skewness, β_2 = kurtosis,
$T_s = d^sT(t)/dt^s$.

Relative error contours comparing Geary's n^{-1} and n^{-2} approximations of the upper and lower modified probability levels for $\alpha = 0.05$ (i.e. the value of approximations (1) and (2) when t satisfies

$$\int_{-\infty}^{-t} T(x)\,dx = \int_{t}^{\infty} T(x)\,dx = \alpha \;)$$

and n = 15, 20, and 25 are given for sets of mixtures of normal distributions, and Pearson distributions, in the $(\sqrt{\beta_1}, \beta_2)$ plane. Additional comparisons are given comparing these approximations from the results of a Monte Carlo simulation for a wide range of mixtures of normal distributions.

Details concerning the validity of the Geary formulas for samples of $n \geq 25$ and for the probability levels 0.05 and 0.95 are given in a paper by the present authors (*International Statistical Review*, 1978).

2. FIGURES AND TABULATION

Under normality, the density of t is

$$T(t,n) = \Gamma(\tfrac{1}{2}n)\,(1 + t^2/(n-1))^{-\frac{1}{2}n} \,/\{\Gamma(\tfrac{1}{2}n-\tfrac{1}{2})\sqrt{(\pi(n-1))}\}, \qquad (3)$$

and the rth derivative (Geary, 1947) of

$$T_1(t,n) = (1+t^2/(n-1))^{-\frac{1}{2}n},$$

is

$$T_1^{(r)}(t,n) = \frac{(-1)^r\Gamma(n+r)}{\Gamma(n)\,(n-1)^r}\,\Phi(t,n)\,(1+t^2/(n-1))^{-\frac{1}{2}(n+2r)} \qquad (4)$$

where

$$\Phi(t,n) = \sum_{s=0}^{[\frac{1}{2}r]} \frac{(-1)^s r^{(2s)} t^{r-2s}}{2.4\ldots(2s)}\, n_s\,,$$

and $n_s = (n-1)^s/\{(n+1)(n+3)\ldots(n+2s-1)\}$,

$r^{(s)} = r(r-1)\ldots(r-s+1)$.

Geary's formula for the denisty of t under non-normality takes the form

$$f(t) = T(t,n) + n^{-\frac{1}{2}} \sum_{s=0}^{1} a_s T_{2s+1} + n^{-1} \sum_{s=1}^{3} b_s T_{2s} + n^{-3/2} \sum_{s=1}^{4} c_s T_{2s+1} + n^{-2} \sum_{s=1}^{6} d_s T_{2s} \qquad (5)$$

where a_s, b_s, c_s, d_s are functions of the cumulant discrepancies (to certain orders of magnitude in the sample size n), of t under non-normality and normality. For example (Geary, 1947)

$a_0 = -\frac{1}{2}\sqrt{\beta_1}$

$a_1 = (\sqrt{\beta_1})/3$,

$b_1 = 1 + \beta_1$

etc. For a specified level α and sample size n, the normal value of t substituted in (5) gives the approximate modified value of α.

DISCLAIMER

DISCLAIMER

From the studies upon which the I.S.R. paper is based, we have constructed error charts (Figures 1 and 2) to enable assessments of the probability levels of t to be made for samples of n = 25, 20 and 15 of the upper and lower 0.05 normal values for samples from mixture of normal distributions and Pearson distributions.

3. USE OF FIGURES

The Geary formula (5) turns out to be a very good approximation when the sample size $n \geq 25$ for the assessments of the upper and lower 0.05 levels for populations defined by skewness and kurtosis (limited to the approximate ranges $0 < \sqrt{\beta_1} < 1.4$, $2 < \beta_2 < 4.6$). Even samples as small as 15 may be used in (5) with acceptable results (Table 1).

We define $G_1(\alpha)$ as Geary's approximation of the modified probability level α to order n^{-1}. Similarly $G_2(\alpha)$ is Geary's approximation of the α-level using terms to order n^{-2}. The relative error is defined by

$$Re(\alpha) = 100\,\frac{(G_2(\alpha) - G_1(\alpha))}{G_1(\alpha)}$$

Being given $Re(\alpha)$ and $G_1(\alpha)$ (for the latter see (1) and (2)) we can determine $G_2(\alpha)$ approximately. It must be emphasized that we have in mind $\alpha = 0.05$ for samples of at least 15, and populations confirmed to the ranges of $\sqrt{\beta_1}$, β_2 displayed in Figure 1.

Example (i)

Population: Normal Mixture $\sqrt{\beta_1} = 0.4$ $\beta_2 = 2.6$

	Lower ($\alpha = 0.05$)	Upper ($\alpha = 0.05$)
n = 15	Re = 3.3	Re = -7.0

Calculate T, T_1, T_2, T_3, T_4, and T_5 for n = 15 and $\alpha = 0.05$; thus t = -1.345 for lower and 1.345 for upper. Then (5)

$G_1 = 0.0514$ (lower) $\quad G_1 = 0.0414$ (upper)

These are computed from (1) and (2). We deduce the assessments

$G_2^* = 0.0634$ (lower) $\quad G_2^* = 0.0385$ (upper)

which to 4 decimal places agree with G_2.

Example (ii)

Population: Normal Mixture $\sqrt{\beta_1} = 1.0$ $\beta_2 = 2.2$

	Lower ($\alpha = 0.05$)	Upper ($\alpha = 0.05$)
Re	3.4	-25.5
G_1	0.0830	0.0329
G_2^*	0.0858	0.0245
G_2(from (5))	0.0858	0.0245

Further examples are given in Table 2.

4. CONCLUDING REMARKS

We have found satisfactory assessments of the modified lower and upper 0.05 levels for t for $n \geq 15$ in sampling from Pearson and Normal Mixture distribution. These are based on the error charts (Figure 1 and 2) along with the first order Geary formulas in (1) and (2), which can easily be calculated on a pocket computer. The question of the robustness of these results under different classes of populations is under investigation.

If readers find the diagrams difficult to use for interpolation purposes, they may request enlarged versions from the authors (limited number available).

REFERENCES

Bowman, K. O., Beauchamp, J. J., and Shenton, L. R., *The Distribution of the t-Statistics Under Non-normality*, Int. Stat. Rev., to appear.

Geary, R. L., (1947) *Testing for Normality*, Biometrika 34, 209-242.

The first two author's research was sponsored by the U.S. Energy Research and Development Administration under contract with the Union Carbide Corporation, Nuclear Division.

Table 2

Error Approximants for Normal Mixtures (n=15)

$\sqrt{\beta_1}$	β_2	Lower (α=0.05) Re	Upper (α=0.05) Re	Lower G_2^*	Lower G_2	Upper G_2^*	Upper G_2
1.4	4.2	7.5	-38.4	0.1072	0.1072	0.0182	0.0182
1.4	3.8	6.3	-38.9	0.1061	0.1061	0.0182	0.0181
1.4	3.4	4.8	-39.8	0.1048	0.1048	0.0179	0.0180
0.4	2.2	3.1	- 7.5	0.0634	0.0634	0.0384	0.0384
1.0	4.2	6.2	-22.5	0.0874	0.0875	0.0250	0.0250
1.0	3.0	5.7	-25.0	0.0874	0.0876	0.0245	0.0245

TABLE 1

COMPARISON FOR N = 15
FROM MIXTURE OF NORMAL DISTRIBUTIONS OF MONTE CARLO (M.C.)
AND GEARY'S $N^{-1}(G_1)$ and $N^{-2}(G_2)$ APPROXIMATIONS TO THE
LOWER AND UPPER 5% PROBABILITY OF THE t-DISTRIBUTION

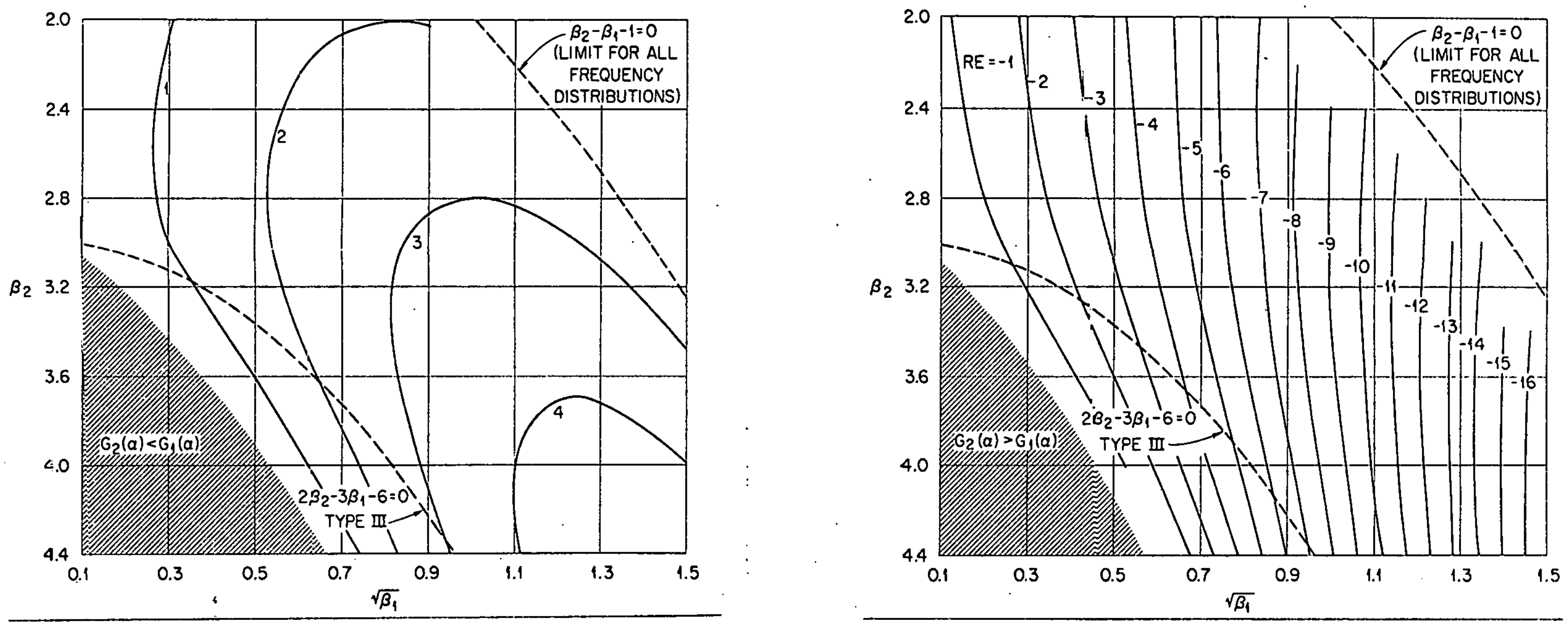
LOWER					UPPER				
β_2 ↓	$\sqrt{\beta_1}$ →	0.2	0.6	1.2	β_2 ↓	$\sqrt{\beta_1}$ →	0.2	0.6	1.2
2.2	M.C.	0.057	0.071		2.2	M.C.	0.955	0.965	
	G_1	0.057	0.071			G_1	0.956	0.967	
	G_2	0.056	0.068			G_2	0.954	0.962	
2.6	M.C.	0.056	0.073		2.6	M.C.	0.959	0.965	
	G_1	0.056	0.071			G_1	0.956	0.967	
	G_2	0.055	0.068			G_2	0.955	0.962	
3.0	M.C.	0.057	0.071	0.101	3.0	M.C.	0.956	0.965	0.977
	G_1	0.056	0.071	0.096		G_1	0.956	0.967	0.979
	G_2	0.055	0.068	0.091		G_2	0.955	0.962	0.969
3.4	M.C.	0.055	0.070	0.098	3.4	M.C.	0.955	0.965	0.977
	G_1	0.054	0.070	0.096		G_1	0.955	0.966	0.979
	G_2	0.055	0.068	0.091		G_2	0.955	0.962	0.969
3.8	M.C.	0.052	0.069	0.100	3.8	M.C.	0.953	0.965	0.978
	G_1	0.050	0.069	0.097		G_1	0.952	0.966	0.979
	G_2	0.055	0.068	0.091		G_2	0.955	0.963	0.969
4.2	M.C.	0.053	0.066	0.104	4.2	M.C.	0.952	0.963	0.977
	G_1	0.041	0.067	0.098		G_1	0.948	0.965	0.979
	G_2	0.055	0.067	0.091		G_2	0.955	0.963	0.969

(The entries refer to the modified upper and lower 0.05 levels of Student's t replacing the upper and lower 0.05 levels under normality; the normal mixture population has 4 parameters (two means, a common variance, and a mixing constant) uniquely determined by $\sqrt{\beta_1}$, β_2)

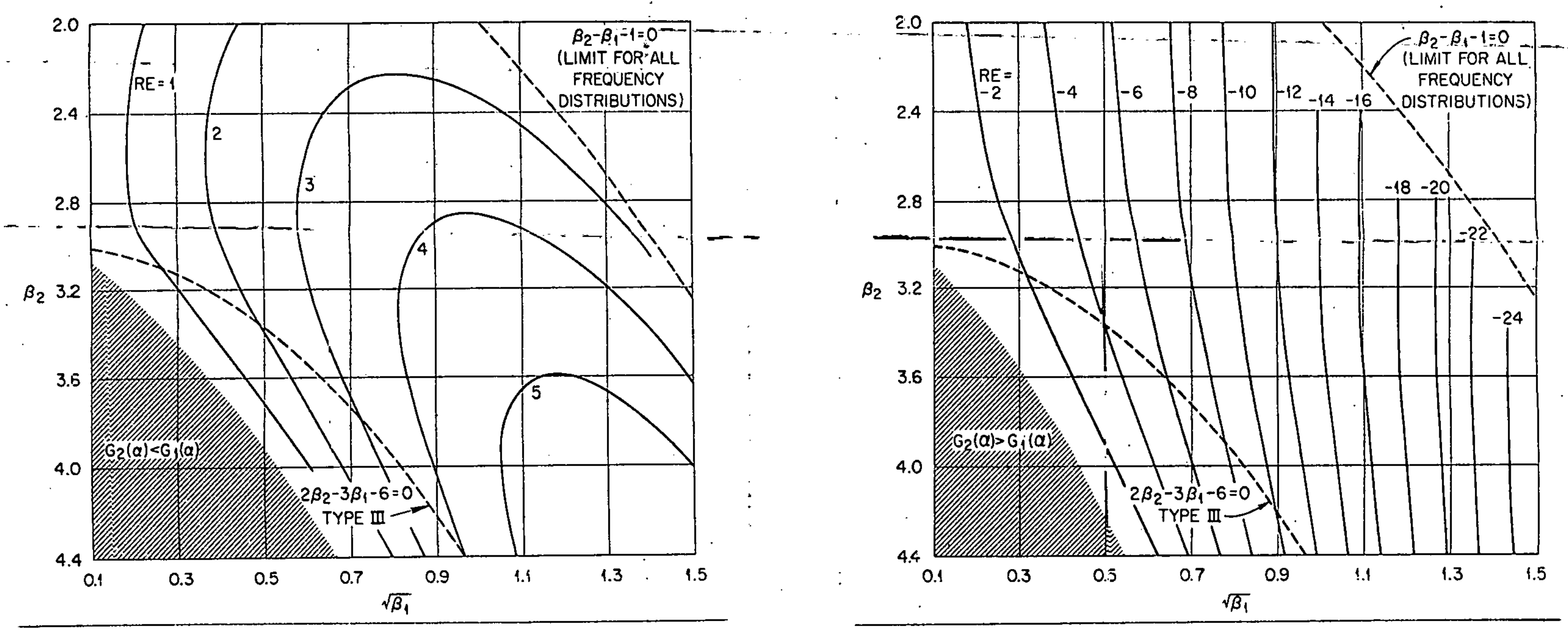
Figure 1 Relative Error Contours Comparing Geary's n^{-1} and n^{-2} Approximations for Probability Levels of Student's t on Sampling from Normal Mixtures (2 components, equal variance, 2 means, and mixing constant)

α = 0.05 (LOWER) α = 0.05 (UPPER)

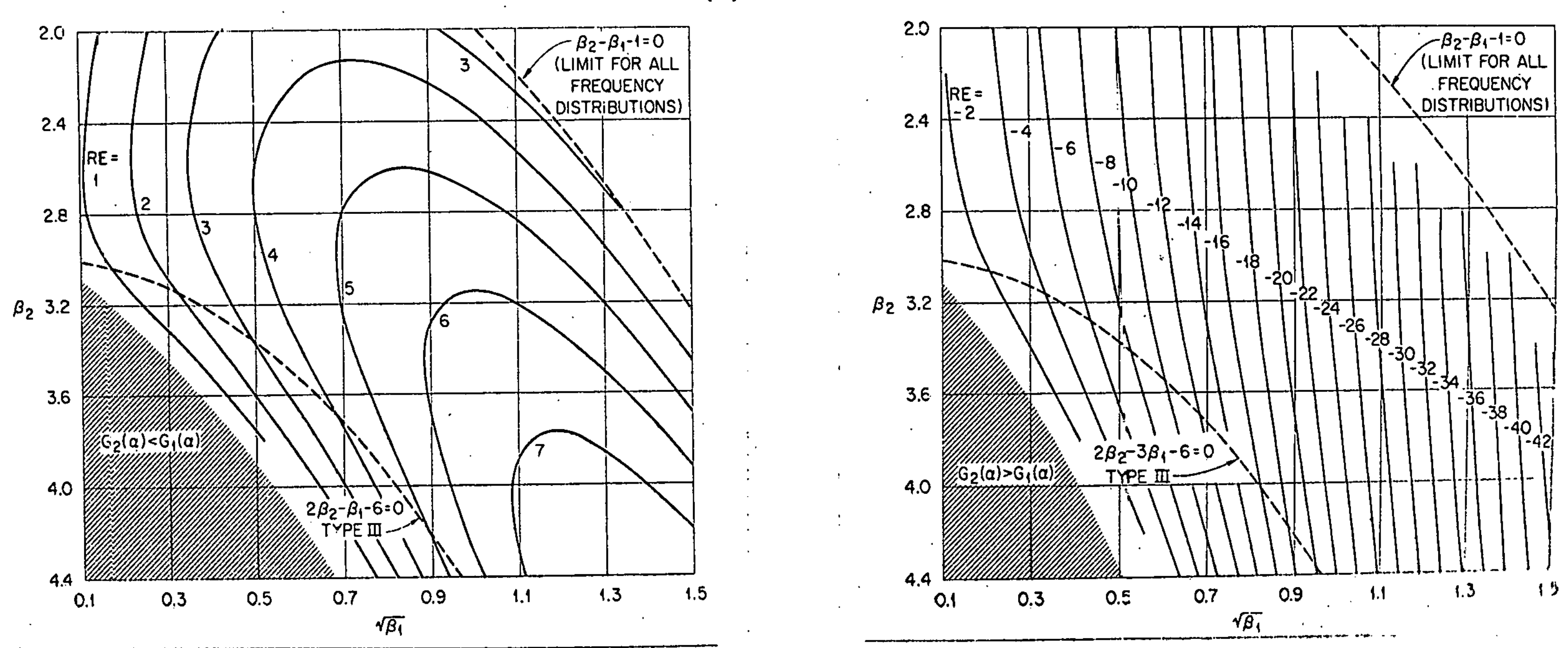
(a) n = 25



(b) n = 20



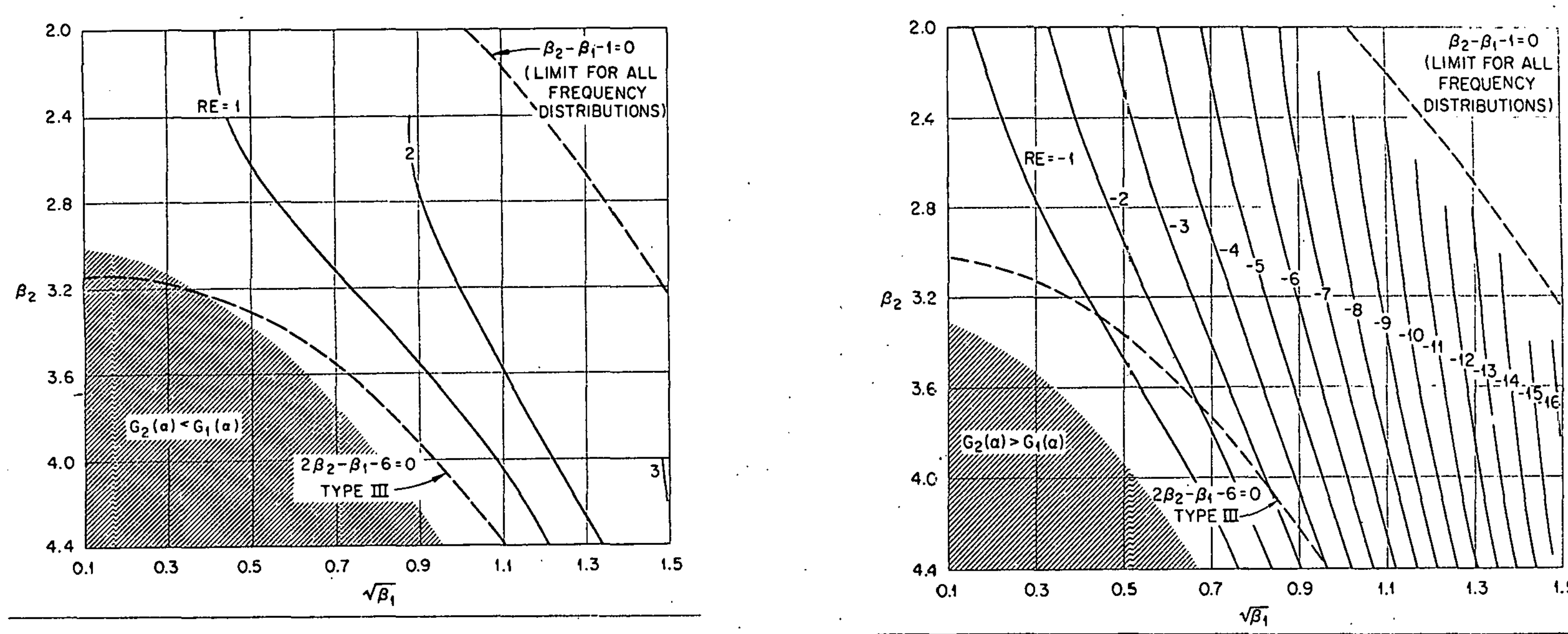
(c) n = 15



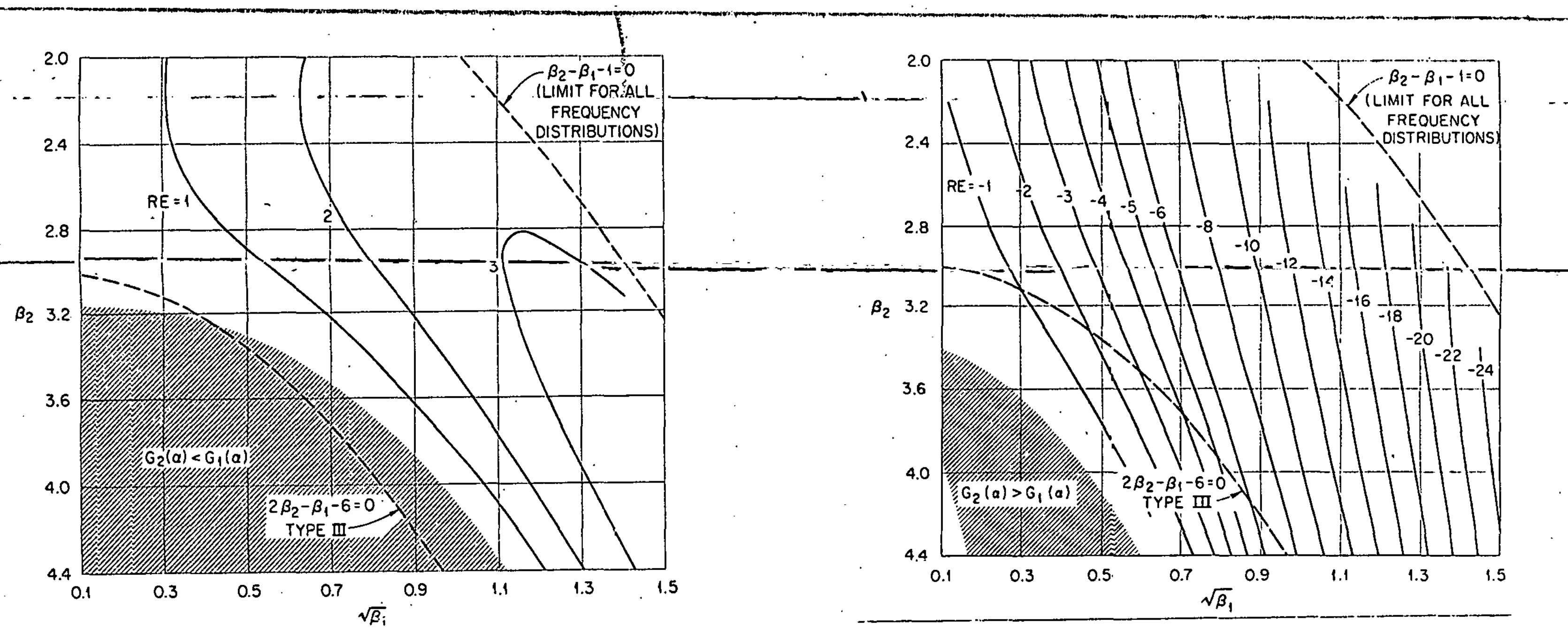
<u>Figure 2</u> Relative Error Contours Comparing Geary's n^{-1} and n^{-2} Approximations for Probability Levels of Student's t on Sampling from Pearson Populations

α = 0.05 (LOWER) α = 0.05 (UPPER)

(a) n = 25



(b) n = 20



(c) n = 15