

مقایسه چند تقریب برای میانگین و واریانس توزیع بتا

غلامحسین شاهکار^۱ حامد رضا طارقیان^۱

چکیده

روش ارزیابی و بازنگری پروژه‌ها که در زمان‌بندی و کنترل پروژه‌های عمده و پیچیده مورد استفاده قرار می‌گیرد؛ مبتنی بر توزیع بتاست. به منظور برآورد میانگین و واریانس زمان اجرای پروژه به صورتی ساده و عملی روشهای مختلفی پیشنهاد شده‌اند، که به طور تقریبی میانگین و واریانس توزیع بتا را تخمین می‌زنند. در این مقاله متداولترین این روشها را با یکدیگر مقایسه نموده، نشان می‌دهیم که در شرایط مختلف تقریب پیرسن - توکی نسبت به سایر تقریبه‌ها از دقت بیشتری برخوردار است.

واژه‌های کلیدی: زمان‌بندی و کنترل پروژه، توزیع بتا.

۱. مقدمه

روش ارزیابی و بازنگری پروژه‌ها یا پرت^۲ در اواخر دهه ۱۹۵۰ میلادی در جهت تسریع اتمام پروژه‌ها مربوط به ساخت موشکهای پولاریس ابداع شد [۹]. امروزه استفاده از این روش در زمان‌بندی و کنترل پروژه‌های عمده و مهم در زمینه‌های گوناگون از جمله پروژه‌های تحقیق و توسعه، تولید صنعتی، و ساخت و سازهای عمرانی و توسعه‌های متداول شده است. چنین پروژه‌هایی اکثراً از تعداد بسیاری فعالیت تشکیل شده که از طریق روابط منطقی متفاوتی که حاکم بر آنهاست به یکدیگر ارتباط پیدا می‌کنند. به این معنا که هر چند تعدادی از این فعالیتها می‌توانند همزمان و به صورت موازی انجام شوند، اما شروع تعدادی دیگر در گرو انجام یک یا چند فعالیت پیشیناز آنهاست. زمان تکمیل پروژه عبارت از کوتاهترین زمان ممکن برای تکمیل کلیه فعالیتها در چارچوب قوانین منطقی حاکم بر آنهاست. به دلیل وجود برخی عوامل احتمالی در اجرای فعالیتهای یک پروژه، زمان انجام هر فعالیت منفک از پروژه باید به

صورت احتمالی مدل‌بندی شود. از اینرو زمان تکمیل پروژه یک متغیر تصادفی است که توزیع احتمالی و گشتاورهای آن موردنظر برنامه‌ریزان پروژه است.

شبکه زمان‌بندی پرت یک گراف جهت‌دار غیرچرخشی است که بیانگر فعالیتها، روابط منطقی حاکم بر آنها و اطلاعاتی در مورد زمان شروع و پایان پروژه است. پالهای شبکه نشان‌دهنده فعالیتها و گره‌ها بیانگر واقعه و مبین زمان تکمیل یا شروع مراحل مختلف پروژه است. در یک شبکه پرت با n گره، گره‌های 1 و n به ترتیب مبین واقعه شروع و خاتمه پروژه‌اند. همه پالهای منتهی به یک گره، نشان‌دهنده فعالیتهای پیشینازی برای انجام فعالیتهایی است که از آن گره منشعب می‌شوند. مثلاً در شبکه شکل ۱، فعالیت متناظر با پال ۵-۴ زمانی می‌تواند شروع شود که هر دو فعالیت متناظر با پالهای ۴-۲ و ۴-۳ به اتمام رسیده باشند (شکل ۱).

اگر زمان اجرای هر فعالیت مشخص باشد، در این صورت هر مسیر منشعب از گره آغازین (شروع پروژه) و منتهی به گره پایانی (خاتمه پروژه) دارای طول زمانی است برابر با مجموع زمان انجام فعالیتهای واقع بر آن مسیر. طول هر یک از این مسیرها کران پایینی

^۱ گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

^۲ program Evaluation and Review Technique- PERT

۳. رویکردی که در آن بجای a و b (نقاط فرین α و β) توزیع (بتا) از سایر کسرکهای توزیع بتا مثل $X(0/1)$ ، $X(0/50)$ ، $X(0/90)$ ، $X(0/99)$ استفاده می کنند. مراجع [۱۱]، [۱۲] و [۱۳] از جمله روشهای ارائه شده در این رویکرد هستند.

۴. رویکردی که در آن سعی می شود تا تقریبهایی برای برآورد میانگین و واریانس زمان اجرای فعالیتها ارائه شوند که مبتنی بر هیچ توزیع آماری نباشند. [۱۴].

۵. رویکردی که در آن به منظور نمونه گیری از توزیع زمان اجرای فعالیتها از شیبه سازی مونت کارلو استفاده می شود. کاربرد شیبه سازی در این زمینه، در مراجع مختلفی از جمله [۱] و [۱۷] گزارش شده است.

به طور کلی، در رویکردهای فوق دست کم سه عامل می تواند در بروز خطای برآورد توزیع T و پارامترهای آن دخیل باشد. عامل اول به چگونگی تقریب زدن آن توزیع خاص مثلا بتا برای زمان اجرای هر یک از فعالیتها مربوط می شود. عدم آگاهی و دقت شخص خیره در ارائه سه برآورد زمانی عامل دیگری در بروز خطاست. عامل سوم به چگونگی استفاده از روابط تقریبی پیشنهادی بجای روابط دقیق برای محاسبه پارامترهای توزیع مربوط می شود. و بالاخره، در نظر نگرفتن امکان وجود مسیرهای متفاوت بحرانی در شبکه زمان بندی عامل دیگری برای بروز خطا محسوب می گردد.

فرض کنید متغیر تصادفی X_i مبین زمان انجام i امین فعالیت واقع بر تنها مسیر بحرانی یک شبکه زمان بندی باشد. مسیر بحرانی شامل k فعالیت است، لذا $E(T) = E(\sum_{i=1}^k X_i)$. به فرض آن که X_i ها مستقل و μ_i و σ_i^2 به ترتیب میانگین و واریانس X_i باشد، داریم

$$\mu_T = \sum_{i=1}^k \mu_i \quad , \quad \sigma_T^2 = \sum_{i=1}^k \sigma_i^2 \quad (1)$$

توزیع تجمعی T که آن را با $F_T(t)$ نشان می دهیم تحت شرایطی تقریبا نرمال می باشد. از اینرو تعیین $F_T(t)$ منوط به برآورد پارامترهای آن، یعنی μ_i و σ_i^2 است.

ما در این مقاله به معرفی و مقایسه اهم روشهای ارائه شده در رویکردهای معمولتر، یعنی رویکردهای ۱ و ۳ می پردازیم. در انجام این کار فرض می کنیم که توزیع زمان انجام فعالیتهای مفروض از

برای زمان تکمیل پروژه است. از اینرو در هر پروژه، زمان تکمیل پروژه را برابر با طولانی ترین مسیر واقع بین گره های ۱ و n در نظر می گیرند. این مسیر را مسیر بحرانی^۲ می نامند. اگر در شبکه شکل ۱ زمان اجرای هر فعالیت را میانگین دامنه داده شده بگیریم، آنگاه طول چهار مسیر منشعب از گره ۱ و منتهی به گره ۵ که عبارتند از: ۱-۲-۵، ۱-۲-۴-۵، ۱-۳-۴-۵ و ۱-۳-۵ به ترتیب برابر است با ۱۰، ۱۴، ۱۳ و ۱۳. بنابراین مسیر بحرانی مسیر ۱-۲-۴-۵ است.

از آنجا که طول هر یال (زمان انجام هر فعالیت) متغیری است تصادفی، لذا طول مسیر بحرانی، T نیز خود یک متغیر تصادفی است و اطلاع از امید ریاضی، $E(T)$ و واریانس آن، $Var(T)$ برای انجام محاسبات مربوط به پارامترهای توزیع زمان تحویل پروژه و همچنین سایر محاسبات در شبکه احتمالی ضروری است.

اگر در شبکه احتمالی شکل ۱ فرض کنیم که زمان اجرای هر فعالیت با احتمال مساوی، برابر یکی از دو کمیت داده شده است، در این صورت وقتی زمان اجرای همه فعالیتها مقدار پیشینه خود را اختیار می کنند، طول مسیرهای چهارگانه به ترتیب ۱۸، ۱۹، ۲۰، ۲۰ می شود. در این حالت دو مسیر ۱-۳-۵ و ۱-۳-۴-۵ بحرانی اند و با محاسبه میانگین طول مسیرهای بحرانی در $2^7 = 128$ حالت، می توان $E(T)$ را پیدا کرد.

معمولا یک شبکه زمان بندی پرت شامل هزاران فعالیت است، از اینرو تعیین توزیع دقیق طول مسیر بحرانی به روش بالا عملا غیر ممکن است. به همین دلیل رویکردهای متفاوتی برای بدست آوردن مقادیر تقریبی برای $E(T)$ و $Var(T)$ پیشنهاد شده اند. این رویکردها را می توان به طور کلی به پنج دسته به شرح زیر تقسیم کرد.

۱. رویکرد مبتنی بر اندیشه اولیه پرت که در آن تابع توزیع زمان اجرای فعالیتها، بتا فرض می شود. روشهای ارائه شده در این رویکرد را می توان در مراجع [۳]، [۴]، [۹] و [۱۵] ملاحظه نمود.

۲. رویکردی که در آن توزیع زمان اجرای فعالیتها را به صورتهای دیگر از جمله توزیع منطقی [۱۰] و توزیع نرمال قطع شده [۸] فرض می کنند.

شایان ذکر است که نمی‌توان تقریبهای (۲) و (۳) را مستقیماً از (۴) نتیجه گرفت و برای این منظور لازم است تا محدودیت‌هایی در مورد انتخاب پارامترهای p و q قائل شد. روابط (۲) و (۳) تنها در شرایطی که $p = 3 \pm \sqrt{2}$ و $q = 3 - \sqrt{2}$ باشد، برقرارند [۱۳]. علاوه بر آن، برای دقت روابط (۲) و (۳)، لازم است تا چولگی توزیع بتا به $\pm \sqrt{\sqrt{2}}$ یا صفر تنزل پیدا کند [۱۴].

ما در این مقاله هشت تقریب مختلف پیشنهاد شده برای پارامترهای توزیع بتا را از نظر دقت آنها با هم مقایسه می‌کنیم. این هشت تقریب را که همگی آنها نیاز به سه برآورد زمان اجرا از طرف شخص خیره دارند؛ در جدول ۱ آورده‌ایم. تقریب اول همان تقریب اصلی پرت یعنی در نظر گرفتن روابط (۲) و (۳) برای میانگین و واریانس توزیع بتاست. تقریب دوم از طرف محققین مختلفی از جمله [۷] پیشنهاد شده است. در این تقریب، بجای a و b یا $x(0)$ و $x(1)$ در روابط (۲) و (۳) از $x(0/99)$ و $x(0/1)$ استفاده می‌شود. تقریبهای سوم و چهارم که از طرف مؤلفان مرجع [۶] پیشنهاد شده‌اند مبتنی بر توزیع آزاد برای محاسبه میانگین و واریانس‌اند [۱۳]، در حالی که تقریب چهارم شیوه ساده محاسبه میانگین است و در پیوست B مرجع [۱۱] به آن اشاره شده است. در تقریب پنجم بجای میانگین در رابطه (۲) از میانه استفاده شده است [۱۶]. در تقریب ششم، مؤلف چگونگی استفاده از (۲) و (۳) را مشروط به مقدار نما قرار داده است. به این معنا که اگر نما بزرگتر یا مساوی $0/13$ باشد، آن‌گاه از تقریب اول استفاده می‌شود، در غیر این صورت برای محاسبه میانگین و واریانس لازم است تا از رابطه جدیدی استفاده گردد [۳]. در تقریب هفتم از نما و نقاط فرین $x(0)$ و $x(1)$ استفاده می‌شود [۴]. بالاخره در تقریب هشتم برآورد میانگین بستگی به میزان چولگی توزیع بتا دارد. به این معنا که اگر $0/87 \leq m_x \leq 0/13$ ، میانگین از یک دستور و در صورتی که $m_x \geq 0/87$ یا $m_x \leq 0/13$ باشد، میانگین از دستورات دیگری محاسبه می‌شوند [۱۵].

۳. مبنای مقایسه و بررسی نتایج حاصل

نتایجی را که ذیلاً ارائه می‌دهیم مبتنی بر توزیع بتای استاندارد با پارامترهای p و q هستند. تابع چگالی احتمال، میانگین، واریانس و نمای بتای استاندارد به قرار زیرند

شبکه زمان‌بندی به صورت بتاست و شخص خیره با آگاهی لازم و دقت کافی برآوردهای سه‌گانه را ارائه می‌دهد. همچنین فرض می‌کنیم که خطای مربوط به احتمال وجود بیش از یک مسیر بحرانی در شبکه قابل اغماض است.

۲. برآورد زمان اجرای هر فعالیت

به منظور برآورد زمان اجرای هر فعالیت (T) در شبکه زمان‌بندی پرت، از شخصی خیره خواسته می‌شود تا برای هر فعالیت سه برآورد زمانی ارائه دهد. شخص خیره با در نظر گرفتن شرایط محیط اجرایی پروژه، کوتاهترین زمان اجرا، محتمل‌ترین زمان اجرا و طولانی‌ترین زمان اجرای هر فعالیت را به ترتیب تحت عناوین برآوردهای خوشبینانه (a)، محتمل (m) و بدبینانه (b) ارائه می‌دهد. با در دست داشتن این سه برآورد، میانگین و واریانس زمان اجرای هر فعالیت به صورت زیر محاسبه می‌شوند.

$$E(T) = \frac{a + 4m + b}{6} \quad (2)$$

$$\sigma^2(t) = \frac{(b-a)^2}{36} \quad (3)$$

تقریبهای فوق که به تقریبهای اصلی پرت معروف‌اند مبتنی بر این فرض‌اند که t در قلمرو $[a, b]$ توزیع بتا و نمای m دارد. چگالی احتمال بتا در دامنه $[a, b]$ برای زمان اجرای فعالیت (T) عبارت است از

$$f(t) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{(t-a)^{p-1}(b-t)^{q-1}}{(b-a)^{p+q-1}}, \quad (4)$$

$$a < t < b, p > 0, q > 0$$

که در آن a و b قلمرو T و p و q پارامترهای توزیع بتا می‌باشند. در این حالت میانگین و واریانس زمان اجرای فعالیت عبارتند از

$$\mu_T = a + \frac{(b-a)p}{p+q} \quad (5)$$

$$\sigma_T^2 = \frac{(pq)(b-a)^2}{(p+q)^2(p+q+1)} \quad (6)$$

میانگینها و واریانسهای واقعی را با استفاده از روابط 5 و 6 محاسبه نموده‌ایم. کسرکها را با توجه به اینکه $F^{-1}(a) = x(a)$ است و با استفاده از توزیع p امین آماره ترتیبی توزیع یکنواخت یعنی

$$F[x(a)] = \sum_{k=p}^{p+q-1} \binom{p+q-1}{k} [x(a)]^k [1-x(a)]^{p+q-1-k} \quad (11)$$

محاسبه کرده‌ایم. آنگاه متوسط میانگینهای واقعی و تقریبی و همچنین متوسط واریانسهای واقعی و تقریبی این توزیعها را تعیین و نمودارهای آنها را در زیر آورده‌ایم. برای ملاحظه چگونگی اختلاف مقادیر واقعی و تقریبی، در مواردی که نمودار، اختلاف بین میانگین مقادیر واقعی و تقریبی را به درستی نشان نداده است، متوسط خطای مطلق برای میانگین (W) و متوسط خطای مطلق برای واریانس (V) را به صورت عددی زیر هر نمودار مشخص کرده‌ایم. همان‌طور که از نمودارها ملاحظه می‌شود تقریب A3 در مقایسه با سایر تقریبها از خطای مطلق کمتری برخوردار است.

همان‌طور که انتظار آن را داشتیم بدون توجه به مقادیر نسبی p و q به دلیل تقارن شرایط $p \leq q$ و $p \geq q$ ، همواره روش A3 یعنی روش پیرسن-توکی نسبت به سایر روشهای تقریبی از دقت بیشتری برخوردار است.

$$f_X(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1} (1-x)^{q-1}, 0 \leq x \leq 1, \quad (7)$$

$$p > 0, q > 0$$

$$\mu = \frac{p}{p+q} \quad (8)$$

$$\sigma^2 = \frac{pq}{(p+q)^2(p+q+1)} \quad (9)$$

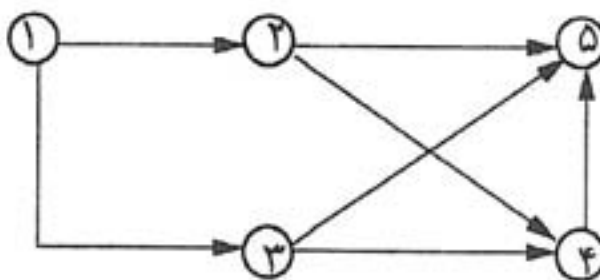
$$x_m = \frac{p-1}{p+q-2} \quad (10)$$

برای مقادیر مختلف p و q تابع چگالی بتا یا نسبت به میانگین تقارن و یا به چپ و راست چوله می‌باشد. اگر p کوچکتر از q باشد، در این صورت منحنی تابع چگالی آن به راست چوله است. این دسته از توزیعها وضعیتی از چگونگی زمان اجرای فعالیتها را به نمایش می‌گذارند که در آن زمان واقعی انجام فعالیت به برآورد بدیشانه آن یعنی b نزدیکتر است. در صورتی که q کوچکتر از p باشد تابع چگالی به چپ چوله است و مبین وضعیتی است که در آن زمان واقعی انجام فعالیت به برآورد خوشینانه یعنی a نزدیکتر است.

بنا به پیشنهاد مرجع [6]، به ازای ترکیبهای مختلف $p, q = 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 30, 60$ تابع توزیع بتا را که در آنها $p \leq q$ بوده است، در نظر گرفته و برای آنها مقادیر میانگین و واریانس توزیعهای بتا را بر طبق روابط تقریبی ارائه شده در جدول 1 محاسبه نموده‌ایم. سپس به ازای هر p و q ی مفروض

شکل 1- نمونه‌ای از یک شبکه احتمالی زمان‌بندی

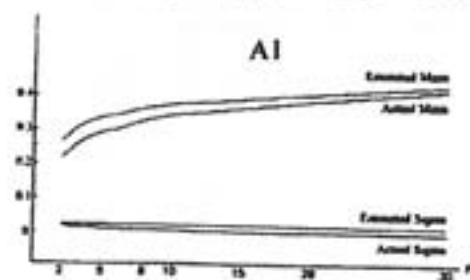
دامنه زمان اجرا	فعالیت
(1, 5)	1-2
(2, 8)	1-3
(6, 10)	2-4
(1, 13)	2-5
(2, 8)	3-4
(4, 12)	3-5
(2, 4)	4-5



جدول ۱- هشت تقریب ارائه شده برای میانگین و واریانس توزیع بتا

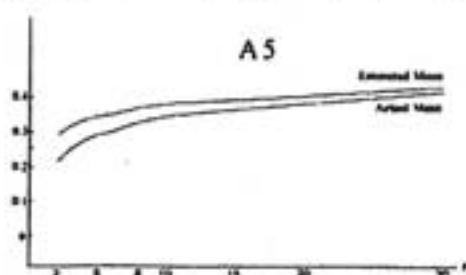
تقریب واریانس	تقریب میانگین	روش
$\hat{\sigma}^2 = \left[\frac{x(1/0) - x(0/0)}{6} \right]^2$	$\hat{\mu} = \frac{x(0/0) + 2x_m + x(1/0)}{6}$	A۱
$\hat{\sigma}^2 = \left[\frac{x(0/99) - x(0/01)}{6} \right]^2$	$\hat{\mu} = \frac{x(0/01) + 2x_m + x(0/99)}{6}$	A۲
$\hat{\sigma}^2 = .063 [x(0/50) - \hat{\mu}]^2 + .185 [x(0/05) - \hat{\mu}]^2 + [x(0/95) - \hat{\mu}]^2$	$\hat{\mu} = .063x(0/50) + .185 [x(0/05) + x(0/95)]$	A۳
$\hat{\sigma}^2 = .4 [x(0/50) - \hat{\mu}]^2 + .3 [x(0/1) - \hat{\mu}]^2 + [x(0/90) - \hat{\mu}]^2$	$\hat{\mu} = .4x(0/50) + .3 [x(0/1) + x(0/9)]$	A۴
ارائه نشده است.	$\hat{\mu} = \frac{x(0/0) + 2x(0/50) + x(1/0)}{6}$	A۵
$\hat{\sigma}^2 = \frac{x_m^2(1-x_m)}{1+x_m}, x_m < .13$	$\hat{\mu} = \frac{2}{2 + \frac{1}{x_m}}, x_m < .13$ $\hat{\mu} = \frac{x(0/0) + 2x_m + x(1/0)}{6}, x_m \geq .13$	A۶
$\hat{\sigma}^2 = \frac{[x(1/0) - x(0/0)]^2}{1268} (+ 22 + 81 \frac{x_m - x(0/0)}{x(1/0) - x(0/0)} - 81 \left[\frac{x_m - x(0/0)}{x(1/0) - x(0/0)} \right]^2)$	$\hat{\mu} = \frac{2x(0/0) + 9x_m + 2x(1/0)}{13}$	A۷
$\hat{\sigma}^2 = \frac{1}{36}$	$\hat{\mu} = \frac{36x_m(1-x_m) + 1}{36x_m(1-x_m) + 2}, .13 \leq x_m \leq .87$	A۸
$\hat{\sigma}^2 = x_m^2(1-x_m), x_m \leq .13$	$\hat{\mu} = \frac{2x_m}{1 + 2x_m}, x_m \leq .13$	
$\hat{\sigma}^2 = x_m(1-x_m), x_m \geq .87$	$\hat{\mu} = \frac{1}{3 - 2x_m}, x_m \geq .87$	

شکل ۲- نمودار مقادیر متوسط میانگین و واریانس واقعی و تخمینی در روشهای مختلف



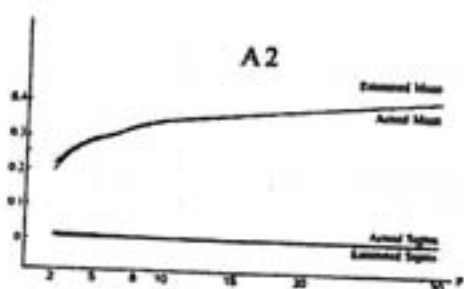
$$W = 0/0399$$

$$V = 0/0125$$



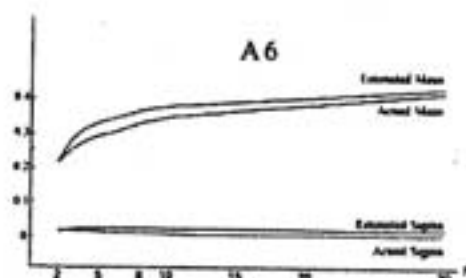
$$W = 0/0505$$

$$V = -$$



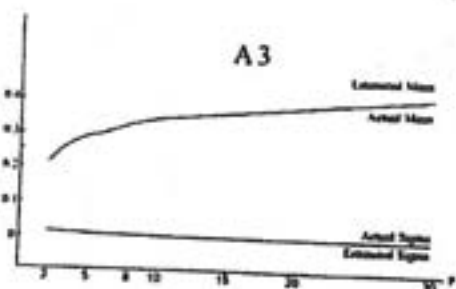
$$W = 0/0063$$

$$V = 0/0053$$



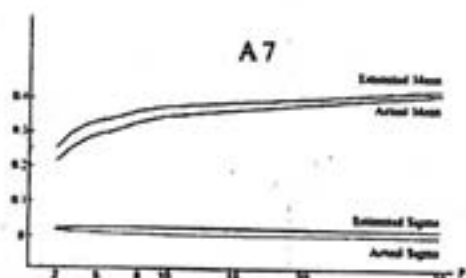
$$W = 0/0279$$

$$V = 0/0110$$



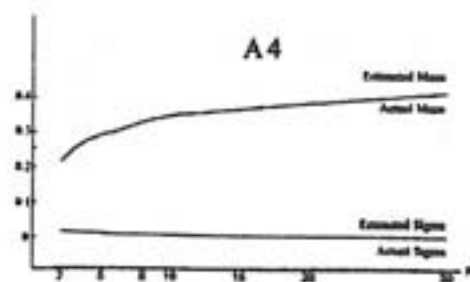
$$W = 0/0000$$

$$V = 0/0001$$



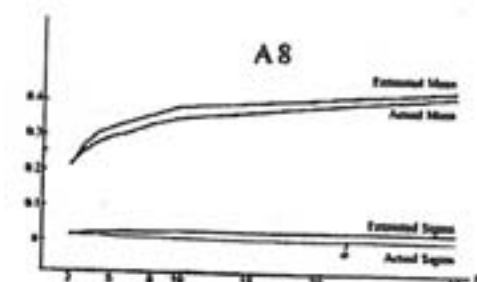
$$W = 0/0256$$

$$V = 0/0166$$



$$W = 0/0001$$

$$V = 0/0004$$



$$W = 0/0234$$

$$V = 0/0132$$

مراجع

- [1] Badiru, A.B., 1991. *A Simulation Approach to PERT Network Analysis*, Simulation, Vol.57, pp.245-255.
- [2] Donaldson, W.A., 1965. *The Estimation of the Mean and Variance of the PERT Activity Time*, Operations Research, Vol.13, pp.382-385.
- [3] Farnum, N.R. and Stanton, L.W., 1987. *Some Results Concerning the Estimation of Beta Distribution Parameters in PERT*, Journal of the Operations Research Society, Vol.38, pp.287-290.
- [4] Golenco, D.I., 1988. *On the Distribution of Activity Time in PERT*, Journal of the Operations Research Society, Vol.39, pp.767-771.
- [5] Grubbs, F.E., 1962. *Attempts to Validate Certain PERT Statistics or Picking on PERT*, Operations Research, Vol.10, pp.912-915.
- [6] Keefer, D.L. and Bodily, S.E., 1983. *Three-Point Approximate for Continuous Random Variables*, Management Science, Vol.29, pp.595-609.
- [7] Kerzner, H., 1992. *Project Management n: A System Approach to Planning Scheduling and Controlling*, Van Nostrand Reinhold, New York.
- [8] Kotiah, T. And Wallace, N., 1973. *Truncated Normal Distribution in PERT*, Management Science, Vol.20, pp.44-49.
- [9] Malcolm, D.G., Roseboom, J.H., Clark, C.E. and Fazer, W., 1959. *Application of a Technique for Research and Development Program Evaluation*, Operation Research, Vol.7, pp.646-669.
- [10] McCrimmon, K. and Ryavec, C., 1964. *An Analytical Study of the PERT Assumptions*, Operations Research, Vol.12, pp.16-37.
- [11] Megill, R.E., 1977. *An Introduction to Risk Analysis*, Tulsa, Petroleum Publishing Company.
- [12] Meredith, J.R. and Mantel, S.J., 1989. *Project Management: A managerial Approach*, 2nd Ed., New York, Wiley.
- [13] Pearson, E.S. Tukey, J.W., 1965. *Approximate Means and Standard Deviations Based on Distances Between Percentage Points of Frequency Curves*, Biometrics, Vol.52, pp.533-546.
- [14] Perry, C. And Greig, I.D., 1975. *Estimating the Mean and Variance of the Subjective Distributions in PERT and Decision Analysis*, Management Science, Vol.21, pp.1477-1480.
- [15] Premachandra, I.M., 2001. *An Approximation of the Activity Duration Distribution in PERT*, Computers & Operations Research, Vol.28, No. 5, pp.443-452.
- [16] Troutt, M.D., 1989. *On the Generality of the PERT Average Time Formula*, Decision Science, Vol.20, pp.410-412.
- [17] Van Slyke, and Richard, M., 1963. *Monte Carlo Methods and the PERT Problems*, Operations Research, Vol.11, pp.839-860.