روش های باز نمونه گیری[1]

باز نمونه گیری بوت استرپ

در بسیاری از مسائل آمار علاقه مند به دانستن توزیع مقادیر حاصل از نمونه تصادفی بر گرفته از جامعه هستیم. اگر توزیعی که نمونه ها از آن گرفته می شوند معلوم باشد نظریه را می توانیم به توزیع نمونه گیری تعمیم دهیم. در مواردی که اطلاعی در مورد توزیع داده های تحت مطالعه در اختیارمان نمی باشد چکار می توانیم انجام دهیم؟ اِفرِن (۱۹۷۹) استفاده از بوت استرپ را برای این منظور پیشنهاد کرد. بر اساس این ایده نمونه بهترین راهنما در مورد توزیع نمونه گیری است. بر مبنای بازنمونه گیری با جایگذاری بر اساس نمونه اصلی می توانیم نمونه بوت استرپ را تولید کرده و با بکار گیری توزیع تجربی برآوردگرمان در حجم زیادی از نمونه های بوت استرپ می توانیم فواصل اطمینان بوت استرپی و آزمون های معنی داری انجام دهیم.

نکته: تفاوت مهم نمونه های تصادفی شده[2] و نمونه های بوت استرپ در این است که نمونه های تصادفی شده بر اساس نمونه گیری بدون جایگذاری از داده های اصلی، در حالیکه نمونه های بوت استرپ بر اساس نمونه گیری با جایگذاری از داده های اصلی تولید می شوند. بکارگیری آزمون های تصادفی در مواردی که ترتیب میان داده ها مهم نباشد (مانند میانگین، واریانس و ...) مفید است.

الگوریتم بوت استرپ ناپارامتری (توزیع داده ها نامعلوم)

۱- نمونه بوت استرپ $X_1^*,\dots,X_n^*$ را بروش نمونه‌گیری تصادفی ساده باجایگذاری از نمونه مشاهده شده $X_1,\dots,X_n$ بدست می آوریم.

۲- آماره بوت استرپ $T^*=T(X_1^*,\dots,X_n^*)$ را محاسبه می کنیم.

۳- مراحل ۱ و ۲ را B با تکرار می کنیم و B آماره بوت استرپ $\{T_b^*;b=1,\dots,n\}$ را محاسبه می کنیم. برآورد کننده های بوت استرپ میانگین، اریبی، واریانس وانحراف معیار بصورت زیر محاسبه می شوند:

$$\hat{Mean}^*(\overline{T}^*)=\frac{1}{B}\sum_{b=1}^{B}T_b^* \quad \text{و} \quad \hat{Bias}^*(T^*)=\frac{1}{B}\sum_{b=1}^{B}T_b^*-T$$

$$\hat{Sd}^*(\overline{T}^*)=\{\frac{1}{B-1}\sum_{b=1}^{B}(T_b^*-\frac{1}{B}\sum_{j=1}^{B}T_j^*)^2\}^{1/2} \quad \text{و} \quad \hat{Var}^*(\overline{T}^*)=\frac{1}{B-1}\sum_{b=1}^{B}(T_b^*-\frac{1}{B}\sum_{j=1}^{B}T_j^*)^2$$

برآورد بوت استرپ اریبی

در مسئله برآوردگر $\hat{\theta}$ برای پارامتر مجهول θ که $\hat{\theta}=\phi(x_1,\dots,x_n)$ می باشد. گیریم $\hat{\theta}.$ نشاندهنده برآورد بر اساس داده های اصلی و $\hat{\theta}_i$ برآورد بر اساس نمونه بوت استرپ i ام نمونه باشد. میانگین تمام برآوردگرهای بوت استرپ عبارتست از

$$\hat{\theta}_B=\frac{1}{p}\sum_{i=1}^{p}\hat{\theta}_i$$

از آنجا که به لحاظ تئوری اریبی برابر $b(\hat{\theta})=E(\hat{\theta})-\theta\approx O(1/n)$ است و $\hat{\theta}_B$ برآورد بوت استرپ $E(\hat{\theta})$ است بنابراین اریبی برابر است با

$$\hat{b}=\hat{\theta}_B-\hat{\theta}.=\frac{1}{p}\sum_{i=1}^{p}\theta_i^*-\hat{\theta}.$$

---

[1] *Resampling*

[2] *Randomized Sample*

توجه شود که $E(\hat{\theta}-b)=\theta$ است. بوت استرپ تصحیح اریبی شده[3] برای برآورد اصلی برابر است با

$$\hat{\theta}_{.}-\hat{b}=۲\hat{\theta}_{.}-\hat{\theta}_{B}$$

فواصل اطمینان بوت استرپ

حدود اطمینان بوت استرپ بر مبنای فرض نرمال توزیع شدن برآوردگر $\hat{\theta}$ با میانگین θ و واریانس $\sigma^{۲}$ است. واریانس نمونه که به شکل زیر تعریف می شود برآوردگر خوبی برای $\sigma^{۲}$ می باشد.

$$S^{۲}=\frac{۱}{n-۱}\sum_{i=۱}^{p}(\hat{\theta}_{i}-\hat{\theta}_{B})^{۲}$$

بنابراین فاصله اطمینان $\%(۱-\alpha)۱۰۰$ به شکل زیر می باشد (منلی ۱۹۹۷)

$$\hat{\theta}\pm z_{\alpha/۲}\,S=\hat{\theta}\pm z_{\alpha/۲}\sqrt{\frac{۱}{n-۱}\sum_{i=۱}^{p}(\hat{\theta}_{i}-\hat{\theta}_{B})^{۲}}$$

فاصله بهبود یافته بر اساس بوت استرپ تصحیح اریبی شده به شکل زیر می باشد (منلی ۱۹۹۷)

$$۲\hat{\theta}-\hat{\theta}_{B}\pm z_{\alpha/۲}\sqrt{\frac{۱}{n-۱}\sum_{i=۱}^{p}(\hat{\theta}_{i}-\hat{\theta}_{B})^{۲}}$$

فاصله اطمینان بوت استرپ صدکی اِفرن[4]

فاصله میان صدک های ۲/۵٪ و ۹۷/۵٪ از توزیع نمونه گیری آماره ، فاصله اطمینان بوت استرپ صدکی ۹۵٪ نامیده می شود که لزوماً فاصله اطمینان متقارنی نخواهد بود. بر این اساس معمولاً نسبت به فاصله اطمینان بوت استرپ تی فوق الذکر دقیق تر می باشد. $(\hat{\theta}_{L,\alpha/۲}\,,\hat{\theta}_{H,\alpha/۲})$

که در آن $\hat{\theta}_{L,\alpha/۲}\,,\hat{\theta}_{H,\alpha/۲}$ صدکهای بالا و پایین از توزیع داده ها می باشند. برای فواصل اطمینان ۹۵٪ تعداد ۱۰۰۰ نمونه بوت استرپ و برای فواصل اطمینان ۹۹٪ تعداد ۵۰۰۰ نمونه بوت استرپ توصیه می شود. البته این روش نیازمند تقارن توزیع نمونه گیری $\hat{\theta}$ حول θ می باشد.

به عنوان مثال فاصله اطمینان بوت استرپ صدکی برای ضریب همبستگی با ۱۰۰۰ نمونه بوت استرپ منجر به محاسبه $(r_{۱}^{*},r_{۲}^{*},\dots,r_{۱۰۰۰}^{*})$ خواهد شد. بعد از مرتب سازی آنها $(r_{(۱)}^{*},r_{(۲)}^{*},\dots,r_{(۱۰۰۰)}^{*})$ را خواهیم داشت. بنابراین فاصله اطمینان ۹۵٪ عبارتست از $[r_{(۵۰)}^{*},r_{(۹۵۰)}^{*}]$

نکته: در ارزیابی دقت یک فاصله اطمینان توزیع آماره مورد مطالعه و اریبی برآوردگر می بایست مورد توجه قرار گیرد.

فاصله اطمینان صدکی هال[5]

این روش که توسط هال (۱۹۹۲) بر مبنای فرم تصحیح اریبی شده پیشنهاد شد و به شکل زیر می باشد

$$(۲\hat{\theta}-\hat{\theta}_{H,\alpha/۲}\,,۲\hat{\theta}-\hat{\theta}_{L,\alpha/۲})$$

که در آن $\hat{\theta}_{L,\alpha/۲}\,,\hat{\theta}_{H,\alpha/۲}$ صدکهای بالا و پایین از توزیع داده ها می باشند. نام دیگر این روش، بوت استرپ صدکی مرکزی شده[6] است. فرض کنیم توزیع نمونه گیری $\theta-\hat{\theta}$ بوسیله توزیع بوت استرپ $\theta-\hat{\theta}_{B}$ قابل تخمین باشد. بنابراین $\theta-\hat{\theta}$ با احتمال ۰/۹۵ بین $\theta-\hat{\theta}_{B,\alpha/۲}$ و $\theta-\hat{\theta}_{B,۱-\alpha/۲}$ می باشد. بر این اساس θ درون فاصله $(۲\hat{\theta}-\hat{\theta}_{B,\alpha/۲}\,,۲\hat{\theta}-\hat{\theta}_{B,۱-\alpha/۲})$ قرار می گیرد. به عنوان مثال فاصله اطمینان صدکی مرکزی شده ۹۵٪ با ۱۰۰۰ تکرار برای پارامتر θ عبارتست از $(۲\hat{\theta}-\theta_{(۲۵)}^{*}\,,۲\hat{\theta}-\theta_{(۹۷۵)}^{*})$

---

[3] *bootstrap bias correction*
[4] *Efron's Bootstrap Percentile Confidence interval*
[5] *Hall's Percentile Confidence interval*
[6] *Centered Bootstrap Percentile Method*

فاصله اطمینان بوت استرپ تی[۷]

فرض کنید B نمونه بوت استرپ استخراج شده و میانگین هر نمونه بوت استرپ با $\bar{x}^*$ نشان داده شود. بر این اساس میانگین و خطای استاندارد بوت استرپ به شکل زیر محاسبه می شوند:

$$mean_{Boot} = \frac{۱}{B}\sum \bar{x}_i^* \qquad , \qquad SE_{Boot} = \sqrt{\frac{۱}{B-۱}\sum(\bar{x}_i^* - mean_{Boot})^۲}$$

ساختار کلی این فاصله اطمینان به شکل زیر می باشد

$$Statistic \pm t^* SE_{Boot} \qquad \text{یا} \qquad \hat{\theta} \pm t^* SE(\hat{\theta})$$

که در آن SE_{Boot} خطای استاندارد بوت استرپ آماره و t^* مقدار بحرانی $t_{n-۱}$ می باشد. این روش از کجی احتمالی توزیع آماره چشم پوشی می کند و همواره یک فاصله اطمینان متقارن محاسبه می کند. ضمناً نرمال بودن تقریبی توزیع آماره جهت بکارگیری چندک تی معمولاً فرض می شود تا نتیجه منطقی بدست آید.

به عنوان مثال فاصله اطمینان ۹۵٪ برای میانگین پیراسته ۲۵٪ براساس نمونه ای ۵۰ تایی به شکل زیر می باشد

$$\bar{x}_{۲۵\%} \pm t^* SE_{Boot}$$

که در آن $\bar{x}_{۲۵\%}$ میانگین پیراسته ۲۵٪ نمونه ، SE_{Boot} خطای استاندارد بوت استرپ $\bar{x}_{۲۵\%}$ و t^* مقدار بحرانی و برابر $t_{۴۹} = ۲/۰۰۹$ می باشد. به طور مشابه برای آماره های دیگری مانند میانه ، چارکها و ... قابل استفاده می باشد.

فاصله اطمینان برای ضریب همبستگی نیز به شکل زیر می باشد

$$r \pm t^* SE_{Boot}$$

که r ضریب همبستگی نمونه ، SE_{Boot} خطای استاندارد بوت استرپ r و t^* مقدار بحرانی از توزیع تی می باشد.

شکل دیگری از فاصله اطمینان بوت استرپ تی

همانطور که می دانیم آماره تی یک نمونه ای به شکل $t = (\bar{X} - \mu)/(s/\sqrt{n})$ است که در آن $SE(\bar{X}) = s/\sqrt{n}$ می باشد. با در نظر گرفتن $\theta = \mu$ و $\hat{\theta} = \bar{X}$ آماره تی به شکل $T = (\hat{\theta} - \theta)/SE(\hat{\theta})$ قابل بازنویسی است. فرم بوت استرپ آماره T با $T_B = (\hat{\theta}_B - \hat{\theta})/SE(\hat{\theta}_B)$ نشان داده می شود. حال اگر صدک ۱۰۰α ام بوت استرپ برای T_B با b_α نمایش داده شود. T درون فاصله $[b_{\alpha/۲}, b_{۱-\alpha/۲}]$ قرار خواهد گرفت. با جایگذاری $T = (\hat{\theta} - \theta)/SE(\hat{\theta})$ درون فاصله فوق نتیجه می گیریم θ درون فاصله زیر قرار خواهد داشت

$$\theta \in (\hat{\theta} - SE(\hat{\theta})\, b_{۱-\alpha/۲} \,, \hat{\theta} - SE(\hat{\theta})\, b_{\alpha/۲})$$

فاصله اطمینان بوت استرپ

فرض کنید $X_i^* = (X_{i۱}^*, X_{i۲}^*, \dots X_{in}^*)$ نماد i امین نمونه $(i = ۱,۲, \dots, B)$ بوت استرپ از توزیع تجربی باشد. همچنین گیریم t_i^* آماره تی محاسبه شده بر اساس نمونه بوت استرپ i ام باشد

$$t_i^* = \sqrt{n}(\hat{\theta}_i^* - \hat{\theta})/\hat{\sigma}_i^* \qquad , \qquad \hat{\beta}_i = ۱ - \Phi(|t_i^*|)$$

که در آن $\hat{\theta}_i^* = \hat{\theta}(X_i^*)$, $\hat{\sigma}_i^* = \hat{\sigma}(X_i^*)$ می باشند. براین اساس فاصله اطمینان دوطرفه $۱۰۰(۱ - ۲\alpha)\%$ عبارتست از

$$[\hat{\theta} + n^{-۱/۲}\sigma^{-۱}z_\alpha \,, \hat{\theta} + n^{-۱/۲}\sigma^{-۱}z_{۱-\alpha}]$$

که در آن z_α چندک ۲α ام مجموعه $\{\hat{\beta}_۱, \hat{\beta}_۲, \dots, \hat{\beta}_B\}$ می باشد.

فواصل اطمینان بوت استرپ دقیق تر

[7] *Bootstrap t Confidence interval*

روش های مختلف محاسبه فاصله اطمینان به بعضی شروط (مانند نرمالیتی) جهت رسیدن به سطح اطمینان مورد نظر نیاز دارند. البته غالباً این شروط در عمل دقیقاً اتفاق نمی افتند بنابراین یک فاصله اطمینان ۹۵٪ نمی تواند در عمل مقدار پارامتر را دقیقاً در ۹۵٪ موارد در برگیرد (احتمال پوشش ضعیف). بعلاوه یک فاصله اطمینان دقیق[8] نامیده می شود، اگر ۵ درصد مواردی که می بایست پارامتر خارج حدود قرار گیرد بین دو طرف مساوی تقسیم شود. دقت کامل در عمل محقق نمی شود ولی بعضی روش ها از سایر روش ها دقیق تر می باشند. بطور کلی مشکلات اصلی این دو روش بترتیب برآوردهای اریب و خطاهای استاندارد متغیر می باشد. دو روش معرفی شده در مواردی که آماره به شدت اریب بوده یا توزیع نمونه گیری آماره بوضوح کج باشد توصیه به استفاده از آنها نمی شود. البته در مواردی که حجم نمونه بزرگ باشد اغلب فواصل اطمینان بوت استرپ از دقت خوبی برخوردار خواهند بود. بنابراین بکارگیری روش تی و صدکی تنها در مواردی که حجم نمونه خیلی بزرگ باشد دقیق خواهند بود.

فاصله اطمینان بوت استرپ تصحیح اریبی و شتاب داده شده[9]

این فاصله اطمینان شکل تصحیح شده روش صدکی می باشد که صدکها را برای اریبی و کجی تصحیح می کند و در تئوری و عمل نشان داده شده که نسبت به روش صدکی بهبود ایجاد می کند. البته برای شروع محاسبات در این روش نیاز به مقدار تصحیح اریبی $\hat{z}.$ و آماره تسریع $\hat{a}$ می باشد.

اریبی به شکل هر خطای سیستماتیک از نمونه نسبت به جامعه مبنا[10] تعریف شده است. در بوت استرپ بررسی می کنیم آیا اریبی میانه یا به عبارتی بیش برآورد یا کم برآورد براساس نمونه های بوت استرپ وجود دارد یا نه؟ مقدار تصحیح اریبی به شکل زیر محاسبه می شود

$$\hat{z}. = \Phi^{-1}\left(\#\{\hat{\theta}_B < \hat{\theta}\}/B\right)$$

که در آن $\hat{\theta}$ برآورد حاصل از نمونه اصلی و $\hat{\theta}_B$ برآورد حاصل از تکرار B ام بوت استرپ می باشد. فاصله اطمینان $۱۰۰(۱-۲\alpha)\%$ از نوع BCa (مثلاً $\alpha = ۰/۰۲۵$ برای فاصله اطمینان ۹۵٪) براساس مراحل زیر محاسبه می شود

۱- محاسبه

$$\alpha_۱ = \Phi\left(\hat{z}. + \frac{\hat{z}. + \Phi^{-۱}(\alpha)}{۱ - \hat{a}(\hat{z}. + \Phi^{-۱}(\alpha))}\right) \quad , \quad \alpha_۲ = \Phi\left(\hat{z}. + \frac{\hat{z}. + \Phi^{-۱}(۱-\alpha)}{۱ - \hat{a}(\hat{z}. + \Phi^{-۱}(۱-\alpha))}\right)$$

که در آن Φ و $\Phi^{-۱}$ به ترتیب توزیع تجمعی نرمال و چندک نرمال می باشند.

۲- محاسبه مقدار گرد شده به نزدیکترین عدد صحیح کوچکتر برای $N_۱ = \mathrm{B} \times \alpha_۱$ و مقدار گرد شده به نزدیکترین عدد صحیح بزرگتر برای $N_۲ = \mathrm{B} \times \alpha_۲$

۳- مرتب سازی آماره محاسبه شده بر اساس $\boldsymbol{B}$ نمونه بوت استرپ از کوچکترین به بزرگترین

۴- در نظر گرفتن $N_۱^{th}$ به عنوان حد پایین فاصله اطمینان و $N_۲^{th}$ به عنوان حد بالای فاصله اطمینان

فاصله اطمینان بوت استرپ متمایل[11]

این روش بر مبنای تصحیح فرآیند شکل دهی به نمونه های تصادفی است.

---

[8] *Accurate*
[9] *Bootstrap bias-corrected accelerated (BCa) interval*
[10] *Any systematic failure of a sample to represent its population base*
[11] *Bootstrap tilting interval*

این دو روش در غالب موارد دقیق بوده و فواصل اطمینان گسترده ای[۱۲] محاسبه نمی کنند. فواصل *BCa* بیشتر بکار برده می شوند. البته هر دو روش بر مبنای باز نمونه گیری و توزیع بوت استرپ بنا شده اند. روش *BCa* جهت حصول دقت بالا به تکرار بیشتر از ۱۰۰۰ نیاز دارد و معمولاً تکرار ۵۰۰۰ و بیشتر در مواردی که دقت در استنباط بسیار مهم است توصیه می شود. روش تیلت به لحاظ تکرار کاراتر بوده و معمولاً ۱۰۰۰ تکرار کافی است. در مواردی که حجم نمونه کوچک است در بکارگیری روشهای *BCa* و *Tilting* نیز احتیاط شود.

روش های باز نمونه گیری در نرم افزار *Splus*

```
> educ <- scan()
1: 12   9   5   7 15   7   7   8   7 13   6 12   7 12   5   2   8 28 20 12
21:  9 10   3 12   6   1   8   3 10 19   8   2   6   2   6   3   9   3 13 11
41:
```

بکارگیری الگوریتم بوت استرپ برای مطالعه توزیع واریانس

```
> var.boot <- bootstrap(data = educ, statistic = var, B = 1000)
> var.boot
> summary(var.boot)

Number of Replications: 1000
Summary Statistics:
    Observed     Bias Mean     SE
var     29.31 -0.4106 28.9 9.691

Empirical Percentiles:
     2.5%     5%    95% 97.5%
var 13.2 14.83 46.72 50.16

BCa Confidence Limits:
      2.5%     5%    95% 97.5%
var 16.88 18.54 56.43 62.47

> limits.emp(var.boot)

         2.5%         5%        95%      97.5%
var 12.91242 14.50413 47.15705 50.87436
> limits.bca(var.boot)

         2.5%        5%        95%      97.5%
var 16.23172 17.6609 57.54335 68.73875
```

برای مطالعه توزیع ۱۰۰۰ واریانس تولید شده از بافت نگار و نمودار چندک چندک می توان استفاده کرد.

```
> par(mfrow=c(1,2))
> plot(var.boot)
> qqnorm(var.boot)
```

به همراه *statistic* آماره های *median, mad, min, max, stdev* را می توان تعریف کرد.

برای بکارگیری الگوریتم بوت استرپ برای ضریب همبستگی پیرسن ابتدا یک چارچوب اطلاعاتی بنام *law.data* بر اساس دو متغیر مورد مطالعه ایجاد می کنیم.

```
> school <- 1:15
> lsat <- c(576,635,558,578,666,580,555,661,651,605,653,575,545,572,594)
> gpa <- c(3.39,3.30,2.81,3.03,3.44,3.07,3.00,3.43,3.36,3.13,3.12,
           2.74,2.76,2.88,2.96)
```

[12] *Wide Intervals*

```
> law.data <- data.frame(School = school, LSAT = lsat, GPA = gpa)

> plot(lsat, gpa, xlab="LSAT", ylab="GPA")

> cor.boot <- bootstrap(law.data, cor(LSAT,GPA), B=1000)
> summary(cor.boot)

Number of Replications: 1000
Summary Statistics:
      Observed      Bias   Mean     SE
Param   0.7764 -0.004849 0.7715 0.1324

Empirical Percentiles:
        2.5%     5%    95%  97.5%
Param 0.4661 0.5275 0.9435 0.9564

BCa Confidence Limits:
        2.5%     5%    95%  97.5%
Param 0.3561 0.4137 0.9219 0.9367

> par(mfrow=c(1,2))
> plot(cor.boot)
> qqnorm(cor.boot)
```

از روش جک نایف بعد از بوت استرپ برای محاسبه خطای استاندارد میانگین آماره می توان استفاده کرد.

```
> jack.after.bootstrap(cor.boot, functional = mean)

Functional Under Consideration: mean
Functional of Bootstrap Distribution of Parameters:
        Func   SE.Func
Param 0.7715    0.1497
```

بکارگیری الگوریتم جک نایف برای مطالعه توزیع انحراف معیار

```
> std.jack <- jackknife(data = educ, statistic = stdev)
> summary(std.jack)

Number of Replications: 40
Summary Statistics:
      Observed     Bias  Mean    SE
stdev    5.414 -0.09908 5.411 1.036

Empirical Percentiles:
       2.5%    5%   95% 97.5%
stdev 5.141 5.212 5.484 5.484

> par(mfrow=c(1,2))
> plot(std.jack)
> qqnorm(std.jack)
```

بکارگیری الگوریتم جک نایف برای مطالعه توزیع ضریب همبستگی پیرسن

```
> cor.jack <- jackknife(data = law.data, statistic = cor(LSAT,GPA))
> summary(cor.jack)

Number of Replications: 15
Summary Statistics:
      Observed      Bias   Mean     SE
Param   0.7764 -0.006474 0.7759 0.1425

Empirical Percentiles:
```

```
       2.5%      5%     95%   97.5%
Param 0.733 0.7347 0.8406 0.8668

> par(mfrow=c(1,2))
> plot(cor.jack)
> qqnorm(cor.jack)
```

برای انجام محاسبات به روش بوت استرپ بلوکی از شناسه *block.size* استفاده می شود.

```
> block.boot <- bootstrap(data = educ, statistic = var, B = 1000,
                          block.size = min(10, B))
> summary(block.boot)

Number of Replications: 1000
Summary Statistics:
    Observed    Bias  Mean     SE
var    29.31 -0.8267 28.48 9.471

Empirical Percentiles:
     2.5%    5%    95% 97.5%
var 13.63 14.81 45.62 49.77

BCa Confidence Limits:
     2.5%    5%    95% 97.5%
var 17.03 18.24 57.23 61.33
```

تمرین: برآوردهای بوت استرپ و جک نایف ضرایب خط رگرسیون بین *lsat, gpa* را بدست آورید.

تابعی برای محاسبه فاصله اطمینان بوت استرپ برای میانگین

```
> Boot.CI <- function(x, alpha = 0.1, B = 1000){
      n <- length(x)
      xbar <- rep(T, B)
      for(b in 1:B){
      Boot.sample <- sample(x, n, T)
      xbar[b] <- mean(Boot.sample)
        }
      mean.xbar <- mean(xbar)
      sd.xbar <- sqrt(var(xbar))
        xbar <- sort(xbar)
      Q1 <- xbar[round(alpha * B)]
      Q2 <- xbar[round((1 - alpha) * B)]
      return(Q1, Q2)
    }

> boot <- function(y, B=10){
      n <- length(y)
      men <- med <- rep(T, B)
      for(b in 1:B){
      Boot.sample <- sample(y, size = n, replace = T)
        men[b] <- mean(Boot.sample)
        med[b] <- median(Boot.sample)
             }
        mean.mean <- mean(men)
        med.med <- median(med)
        mean <- quantile(men,c(0.05,0.95))
        median <- quantile(med,c(0.05,0.95))
        return(men, med, mean.mean, med.med, mean, median)
        }
```

برآورد پارامتری بوت استرپ میانگین، میانه، انحراف معیار و واریانس برای ۱۰۰ مشاهده شبیه سازی شده ازتوزیع گاما باپارامترهای۲ و۳ بعلاوه بررسی توزیع هریک ازآماره ها بوسیله رسم بافت نگارآنها.

```
> boot <- function(n = 100, alpha = 2, beta = 3, B = 500){
      men <- med <- s <- s2 <- rep(NA, B)
      gam.dat <- rgamma(n, alpha, beta)
      for(b in 1:B) {
      Boot.sample <- sample(gam.dat, size = n, replace = T)
     men[b] <- mean(Boot.sample)
     med[b] <- median(Boot.sample)
      s[b] <- sqrt(var(Boot.sample))
     s2[b] <- var(Boot.sample)
            }
      hist(s); hist(s2); hist(men); hist(med)
      mean.mean <- mean(men) ; med.med <- median(med)
     sd.s <- sqrt(var(s)) ; sd.s2 <- sqrt(var(s))
     return("The Bootstrap estimate for Mean, Median, S, S2
             are", mean.mean, med.med, sd.s, sd.s2)
   }
```

بوسیله تابع زیر بر اساس *n* مشاهده یک فاصله اطمینان ۹۰٪ بوت استرپ (بر اساس صدکهای مرتبه ۵ و ۹۵) برای میانگین جامعه محاسبه می شود.

```
> Boot.CI <- function(x, alpha = 0.1){
      B <- 1000
      n <- length(x)
      xbar <- rep(NA, B)
      for(b in 1:B) {
      Boot.sample <- sample(x, n, T)
      xbar[b] <- mean(Boot.sample)
      }
     hist(xbar)
     mean.xbar <- mean(xbar)
      sd.xbar <- sqrt(var(xbar))
     xbar <- sort(xbar)
      Q1 <- xbar[round(alpha * B)]
      Q2 <- xbar[round((1 - alpha) * B)]
      return("Bootstrap Mean of xbar =", mean.xbar,
           "Bootstrap SD of xbar =", sd.xbar, "1-", alpha,
            "(1-alpha)% Bootstrap confidence Interval is", Q1, Q2)
    }
```

تابعی که فاصله اطمینان بوت استرپ را برای میانه، نما، انحراف معیار یک نمونه تصادفی *n* تایی محاسبه می کند.

```
> t22 <- function(x, alpha = 0.05, B = 1000){
      n <- length(x)
      s <- rep(NA, B)
      s2 <- rep(NA, B)
      mmode <- rep(NA, B)
      mmedian <- rep(NA, B)
      for(b in 1:B) {
      boot.sample <- sample(x, n, T)
      mmedian[b] <- median(boot.sample)
     mmode[b] <- mean(boot.sample)-3*(mean(boot.sample)-median(boot.sample))
      s[b] <- sqrt(var(boot.sample))
         }
      mmedian <- sort(mmedian)
      mmode <- sort(mmode)
      s <- sort(s)
      Q1 <- mmedian[round((alpha/2) * B)]
      Q2 <- mmedian[round((1 - alpha/2) * B)]
```

```
      Q3 <- mmode[round((alpha/2) * B)]
      Q4 <- mmode[round((1 - alpha/2) * B)]
      Q5 <- s[round((alpha/2) * B)]
      Q6 <- s[round((1 - alpha/2) * B)]
      conf.med <- c(Q1, Q2)
      conf.mode <- c(Q3, Q4)
      conf.s.d <- c(Q5, Q6)
      return("(1-alpha)%bootstrap confidence interval for median is" =
     conf.med, "(1-alpha)%bootstrap confidence interval for mode is" =
      conf.mode, "(1-alpha)%bootstrap confidence interval for
     standard.devation is" = conf.s.d)
  }
> t22(rnorm(20))
$"(1-alpha)%bootstrap confidence interval for median is":
[1] -0.4424511  0.2432942

$"(1-alpha)%bootstrap confidence interval for mode is":
[1] -0.8401888  0.6366904

$"(1-alpha)%bootstrap confidence interval for standard deviation is":
[1] 0.4152879 0.7683815
```

تابعی که فاصله اطمینان ۹۵٪ بوت استرپ را برای ضرایب کجی و کشیدگی یک نمونه تصادفی n تایی محاسبه کند و در صورتیکه فاصله اطمینان های ضرایب به ترتیب صفر و سه را در برداشته باشند پیام *"Accept Normality Assumption"* و در غیر اینصورت *"Reject Normality Assumption"* را چاپ نماید.

```
> t23 <- function(x, alpha = 0.05, B = 1000){
    n <- length(x)
    m4 <- rep(NA, B)
    k.m4 <- rep(NA, B)
    ske <- rep(NA, B)
    for(b in 1:B) {
        boot.sample <- sample(x, n, T)
        m4[b] <- mean(((boot.sample) - mean(boot.sample))^4)
        k.m4[b] <- ((m4[b])/(sqrt(var(boot.sample)^4)) - 3)
        ske[b] <- (3*(mean(boot.sample)-
                     median(boot.sample)))/sqrt(var(boot.sample))
        }
    Q1 <- k.m4[round((alpha/2) * B)]
    Q2 <- k.m4[round((1 - alpha/2) * B)]
    Q3 <- ske[round((alpha/2) * B)]
    Q4 <- ske[round((1 - alpha/2) * B)]
    k.m4.conf <- c(Q1, Q2)
    ske.conf <- c(Q3, Q4)
    if(Q2 - Q1 <= 0 && Q4 - Q3 >= 3) cat("Accept Normality Assumption")
    else cat("Reject Normality Assumption")
    list(k.m4.conf = k.m4.conf, ske.conf = ske.conf)
 }
```

تابعی که برآورد پارامتری بوت استرپ کجی و کشیدگی را برای ۱۰۰ مشاهده شبیه سازی شده ازتوزیع گاما با پارامترهای ۲ و۳ را محاسبه می کند، بعلاوه بررسی توزیع هریک از آماره ها بوسیله رسم بافت نگار آنها انجام شده.

```
> t21  <- function(n = 100, alpha = 2, beta = 3, B = 500){
      s <- rep(NA, B)
      s2 <- rep(NA, B)
      mmean <- rep(NA, B)
      mmedian <- rep(NA, B)
      m4 <- rep(NA, B)
      k.m4 <- rep(NA, B)
      ske <- rep(NA, B)
      gam.dat <- rgamma(n, alpha, beta)
      for(t in 1:B) {
            boot.sample <- sample(gam.dat, size = n, replace = T)
```

```
            mmean[t] <- mean(boot.sample)
            mmedian[t] <- median(boot.sample)
            s[t] <- sqrt(var(boot.sample))
            s2[t] <- var(boot.sample)
            m4[t] <- mean((boot.sample - mmean[t])^4)
            k.m4[t] <- (((m4[t])/(sqrt(s2[t])^4)) - 3)
            ske[t] <- (3 * (mmean[t] - mmedian[t]))/sqrt(s2[t])
      }
      hist(k.m4); hist(ske)
      m.m4 <- mean((m4 - mean(m4))^4)
      k.k.m4 <- ((m.m4/sqrt(var(m4))^4) - 3)
      s.ske <- ((3 * (mean(mmean) - median(mmedian)))/sqrt(var(m4)))
      return(k.m4, ske, k.k.m4, s.ske)
}
```

تابعی که بر اساس ۱۵ مشاهده ای که در اختیار داریم ۱۰۰۰ نمونه ۱۵ تایی به صورت با جایگذاری انتخاب می کند و بر اساس داده های حاصل آماره های Q_2 و P_{10} و P_{75} را محاسبه کند و بر این اساس تخمینهای بوت استرپ را برای متوسط ، انحراف معیار ، میانه و فاصله اطمینان آماره های مذکور را محاسبه می نماید. در ادامه بوسیله رسم بافت نگار تخمینهای بوت استرپ، به بررسی توزیع آماره های مورد نظر می پردازد.

```
> t25 <- function(x, alpha = 0.05, B = 1000){
  n <- 15
  q2 <- quantile(x,0.5)
  p10 <- quantile(x,0.1)
  p75 <- quantile(x,0.75)
  mean <- median <- s <- s2 <- rep(NA,B)
  for(b in 1:B){
  boot.sample <- sample(x, n, T)
  mean[b] <- mean(boot.sample)
  median[b] <- median(boot.sample)
  s[b] <- sqrt(var(boot.sample))
  s2[b] <- var(boot.sample)
   }
  hist(s); hist(s2); hist(mean); hist(median)
  mean.mean <- mean(mean)
  med.med <- median(median)
  sd.mean <- sqrt(var(mean))
  sd.s <- sqrt(var(s))
  sd.s2 <- sqrt(var(s2))
  c1 <- mean[round(alpha*B)]
  c2 <- mean[round((1-alpha)*B)]
  return("the bootstrap estimate for mean,median,s,s2
  are",mean.mean,med.med,sd.s,sd.s2,"bootstrap confidence interval
  is",c1,c2,hist(s),hist(s2),hist(mean),hist(median))
  }
> t25(rnorm(15))
```

تابعی برای محاسبه فاصله اطمینان و مقدار احتمال بوت استرپ

```
> boot <- function(nsim, n, alpha){
      xbar <- rep(NA, nsim)
      sigma2 <- rep(NA, nsim)
      sigma <- rep(NA, nsim)
      int.low <- rep(NA, nsim)
      int.high <- rep(NA, nsim)
      y <- rnorm(n)
      r <- 0
      p <- 0
      for(i in 1:nsim) {
            x.sample <- sample(y, size = n, replace = T)
            xbar[i] <- mean(x.sample)
            sigma2[i] <- var(x.sample)
            sigma[i] <- sqrt(sigma2[i])
            quantile.alpha.increase <- (1 - (alpha/2))
```

```
            Z <- qnorm(quantile.alpha.increase)
            int.low[i] <- xbar[i] - ((Z * sigma[i])/sqrt(n))
            int.high[i] <- xbar[i] + ((Z * sigma[i])/sqrt(n))
            xbar <- sort(xbar)
            int.low.boot <- xbar[round(alpha * nsim)]
            int.high.boot <- xbar[round((1 - alpha) * nsim)]
            mu.total <- mean(xbar)
            if(mu.total >= int.low[i] & mu.total <= int.high[i])
            r <- r + 1
            else if(mu.total >= int.low.boot[i] & mu.total <=
                  int.high.boot[i])
                  p <- p + 1
           }
      percent.N <- (r/nsim) * 100
      percent.B <- (p/nsim) * 100
      list(int.low = int.low, int.high = int.high, int.low.boot =
          int.low.boot, int.high.boot = int.high.boot, mu = mu.total,
          r = r, p = p, percent.N = percent.N, percent.B = percent.B)
  }
```

کاربرد بوت استرپ در رگرسیون

از بوت استرپ در رگرسیون نیز می توان استفاده کرد و بر اساس آن برآوردهای بوت استرپ ضرایب مدل و فاصله اطمینان بوت استرپ را بـرای آنـها مـحاسبه کرد. مـخصوصاً در مـواقعی که رگـرسیون دچار مشکل عدم نرمالیتی می باشد از این روش استفاده می شود. همچنین همخطی شـدید بین متغیرهای مـستقل منجر به تورم واریانس خطا می شود که منجر به بزرگ شدن مخرج آماره های تی آزمون معنی داری ضرایب خط رگرسیون شده و ممکن است به اشتباه چنین استنباط کنیم که متغیرهای مستقل تاثیری روی پاسخ ندارد. بنابراین از بوت استرپ برای تصحیح خـطای موجود در MSE می توانیم استفاده کنیم و بنابراین ضـرایب برآورد شده توسط بوت استـرپ تحت تاثیر همخطی موجود قرار نمی گیرند. لازم به ذکـر است کـه در بـــرخورد بـا مقادیر پرت افتاده نیز از ایـن تکنیک می تـوان استفاده کرد.

کاربرد بوت استرپ در سری زمانی

از بوت استرپ در سری زمانی نیز در مواقعی که مشکل عدم نرمالیتی رخ می دهد در برآورد ضرایب مدل، محاسبه فاصله اطمینان برای آنها و حتی آزمونهای معنی داری ضرایب استفاده می شود.

کاربرد بوت استرپ در محاسبه مقدار احتمال

در مواقعی که برای بررسی یک فرضیه یافتن کمیت محوری یا آماره آزمون کار مشکلی است می توانیم از مقدار احتمال بوت استرپ جهت بررسی فرضیات مورد نظر بدون نیاز به آماره استفاده کنیم.

نمونه ۱- برای محاسبه مقدار احتمال بوت استرپ جهـت بررسـی فرضـیه $H_0:\rho \ge \rho_0 \ vs \ H_1:\rho < \rho_0$ کافیسـت از نمونه n تایی اولیه B نمونه با جایگذاری n تایی استخراج کنیم سپس در هر نمونه مقدار ضریب همبسـتگی پیرسـن را محاسبه کنیم و بر اساس B مقدار حاصل می توانیم فاصله اطمینان بوت استرپ را محاسبه کنیم و مقدار احتمال بـوت استرپ برابر تعداد ضرایب بزرگتر از ρ_0 تقسیم بر مقدار B می باشد.

نمونه ۲- پرسشنامه ای شامل k سوال است که توسط m پر نفر شده، مقدار ضریب آلفای کرانباخ برابر ۰/۸۲۳ شـده است. برای بررسی فرضیه $H_0:r_\alpha \ge r_{\alpha_0} \ vs \ H_1:r_\alpha < r_{\alpha_0}$ کافیست B نمونه با جایگـذاری از نمونـه اولیـه اسـتخراج کنیم سپس در هر نمونه مقدار آلفای کرانباخ را محاسبه کنیم و بر اساس B مقدار حاصل می تـوانیم فاصـله اطمینـان

بوت استرپ را محاسبه کنیم بعلاوه مقدار احتمال بوت استرپ برابر تعداد ضرایب بزرگتر از r_{α_0} تقسیم بر مقدار B خواهد بود.

سطح معنی داری بدست آمده آزمون (ASL)

سطح معنی داری بدست آمده آزمون[13] برآوردی از مقدار احتمال بوسیله روشهای جایگشت و بوت استرپ می باشد و به شکل زیر تعریف می شود

$$ASL = P(\hat{\theta}^* > \hat{\theta}_۰ | H_۰)$$

مقادیر کوچک ASL دلیلی بر رد قوی فرض صفر می باشند.

مقدار احتمال بوت استرپ به شکل زیر تعریف می شود

$$p_{boot} = P(T^* \geq t | \hat{F}_۰)$$

که در آن $\hat{F}_۰$ توزیعی است که بازنمونه ها از آن انتخاب می شوند. مقدار احتمال به شکل زیر قابل برآورد است

$$p_{boot} = \frac{۱}{B}\sum_{b=۱}^{B} I(t_b^* \geq t) = \#(t_b^* \geq t)/B$$

که در آن $t_۱^*, t_۲^*, \dots, t_B^*$ آماره های حاصل از B نمونه بوت استرپ می باشند.

آزمون های ناپارامتری

الف) آزمون های بوت استرپ ب) آزمون های جایگشت

همانطور که می دانید آماره آزمون جهت بررسی یک فرضیه تک نمونه ای به شکل $T = (\bar{X} - \mu_۰)/(s/\sqrt{n})$ می باشد. برای بررسی فرضیه فوق به روش بوت استرپ کافیست مقدار آماره آزمون برای هر نمونه بوت استرپ (t_B^*) محاسبه و با مقدار آماره حاصل از نمونه اصلی ($t_۰$) مقایسه شود. بر این اساس مقدار احتمال بوت استرپ به شکل زیر محاسبه می شود

$$ASL_{boot} = \#\{t_b^* \geq t_۰\}/B \qquad b = ۱,۲,\dots,B$$

که در آن $t_b^* = (\hat{\theta}_b^* - \theta_۰)/\hat{\sigma}(\hat{\theta}_b^*)$ می باشد. برای آزمون تی دو نمونه ای نیز به طور مشابه می بایست عمل شود.

نکته: همانطور که می دانیم مقدار احتمال برای بررسی یک فرض آماری دو طرفه بر اساس آماره Z عبارتست از

$$Exact\ \ pvalue = ۲min\{P(Z \leq z), P(Z \geq z)\}$$

و برای برآورد مقدار احتمال دقیق بوسیله شبیه سازی کافیست ۱۰۰۰ مقدار Z شبیه سازی شود سپس صدکهای $۱۰۰(\alpha/۲)$ و $۱۰۰(۱-\alpha/۲)$ آنها محاسبه و نسبت مقادیر کوچکتر از z (p_L) و نسبت مقادیر بزرگتر از z (p_U) را نیز بدست می آوریم و مقدار احتمال به شکل زیر محاسبه می شود

$$pvalue = ۲\ min\{p_L, p_U\}$$

مقدار احتمال به مطالعه فرضیه به لحاظ تئوری $H_۰: \mu_x = \mu_y\ \ vs\ \ H_۱: \mu_x \neq \mu_y$ بشکل زیر تعریف می شود

$$p = ۱ - P(|T| \leq |t|\ |\ H_۰) = ۲[۱ - P(T \leq |t|\ |\ H_۰)] = ۲[۱ - F_{t,n_x+n_y-۲}(|t|)]$$

برآوردهای جک نایف

روش جک نایف ابتدا توسط کوئنویل در سال ۱۹۴۹ برای حذف اریبی در همبستگی پیاپی مرتبه اول در سریهای زمانی پیشنهاد شد. توکی (۱۹۵۸) روشی بنام جک نایف بر اساس حذف یک مشاهده از داده ها و سپس محاسبه

---

[13] *Achieved Significance Level (ASL) of test*

برآوردگر، معرفی کرد که روشی ساده به لحاظ محاسبه و حل کننده تعدادی از مشکلات می باشد. از اهداف اصلی روش جک نایف کاهش اریبی برآورد گرها و همچنین واریانس یک برآوردگر و سایر ملاکهای خطاست. اگر $\bar{x}$ نماد میانگین نمونه n تایی باشد، میانگین نمونه با حذف مشاهده j ام عبارتست از

$$\bar{x}_{-j} = \frac{۱}{n-۱}\sum_{i\neq j}^{n} x_i$$

با دانستن $\bar{x}$ و $\bar{x}_{-j}$ می توانیم j امین مشاهده را محاسبه کنیم.

$$x_j = n\bar{x} - (n-۱)\bar{x}_{-j}$$

فرض کنید می خواهیم پارامتر θ را بر اساس آماره ای (که پیچیده نیز می باشد) برآورد کنیم

$$\hat{\theta} = \phi(x_۱, x_۲, \dots, x_{i-۱}, x_i, x_{i+۱}, \dots, x_n)$$

بر اساس ایده فوق الذکر، فرض کنید برآورد جزئی پارامتر θ با حذف مشاهده j ام به شکل زیر باشد

$$\hat{\theta}_j = \phi(x_۱, x_۲, \dots, x_{i-۱}, x_{i+۱}, \dots, x_n)$$

بنابراین شبه مقدار j ام عبارتست از

$$\hat{\theta}_j^* = n\hat{\theta} - (n-۱)\hat{\theta}_j$$

شبه مقدار فوق نقشی شبیه x_j در تخمین میانگین دارد. برآورد جک نایف پارامتر θ برابر میانگین شبه مقادیر است.

$$\hat{\theta}_{jack}^* = \hat{\theta}^* = \frac{۱}{n}\sum_{i=۱}^{n} \hat{\theta}_i^*$$

برآورد تقریبی خطای نمونه گیری برای $\hat{\theta}^*$ قابل محاسبه بر اساس واریانس نمونه شبه مقادیر است

$$var(\hat{\theta}^*) = \frac{var(\hat{\theta}_j^*)}{n} = \frac{\sum_{j=۱}^{n}(\hat{\theta}_j^* - \hat{\theta}^*)^۲}{n(n-۱)}$$

بنابراین فاصله اطمینان $۱۰۰(۱-\alpha)\%$ برای پارامتر θ به شکل زیر می باشد

$$\hat{\theta}^* \pm t_{\alpha/۲, n-۱}\sqrt{\frac{\sum_{j=۱}^{n}(\hat{\theta}_j^* - \hat{\theta}^*)^۲}{n(n-۱)}}$$

فواصل اطمینان جک نایفی می بایست با احتیاط بکار برده شوند، زیرا ممکن است نسبت به فاصله اطمینان واقعی بیش یا کم برآورد[۱۴] باشند. دلیل اصلی استفاده از برآوردهای جک نایفی این است که آنها اریبی را کاهش می دهند. کوئنویل (۱۹۵۶) نشان داد که برآوردهای جک نایفی اریبی مرتبه $۱/n$ را حذف می کنند. در حالت کلی بکارگیری جک نایف مرتبه p منجر به حذف اریبی مرتبه $۱/n^p$ می شود (هر برآورد جزئی p نقطه را حذف می کند). ضمناً آماره تی جهت آزمون فرض های یک نمونه ای به شکل زیر می باشد.

$$T = \frac{\hat{\theta}^* - \theta}{\sqrt{\frac{\sum_{j=۱}^{n}(\hat{\theta}_j^* - \hat{\theta}^*)^۲}{n(n-۱)}}} \sim t_{n-۱}$$

جک نایفینگ و تسریع[۱۵]

آماره تسریع مربوط به نرخ تغییر خطای استاندارد از برآورد آماره نمونه مورد مطالعه با توجه به مقدار واقعی آماره است. افرن و تیبشیرانی آماره تسریع را بر اساس مقادیر جک نایف برآورد آماره پیشنهاد کردند. جک نایفینگ به معنی مطالعه تاثیر هر مشاهده در برآورد می باشد. مقدار تسریع عبارتست از

$$\hat{a} = \frac{\sum_{j=۱}^{n}(\hat{\theta}_{jack}^* - \hat{\theta}_j)^۳}{۶\left[\sum_{j=۱}^{n}(\hat{\theta}_{jack}^* - \hat{\theta}_j)^۲\right]^{۳/۲}}$$

تابع زیر برآورد جک نایف انحراف معیار را محاسبه کند.

---

[14] *Over estimate or Under estimate*
[15] *Jackknifing and Acceleration*

```
> jacks <- function(x){
  n <- length(x)
  s <- sqrt(var(x))
  si <- 0
  for(i in 1:n){
  si[i] <- sqrt(var((x)[-i]))
  }
  j <- n*s -(n-1)*si
  hist(j)
  jbar <- mean(j)
  sej <- stdev(j)
  return(n,s,si,j,jbar,sej)
  }
```

تابع زیر برآورد جک نایف میانگین و میانه را محاسبه کند.

```
> jack.memd <- function(x){
  n <- length(x)
  mean <- mean(x)
  median <- median(x)
  mei <- 0
  mdi <- 0
  for(i in 1:n){
  mei[i] <- mean((x)[-i])
  mdi[i] <- median((x)[-i])
  }
  j.me <- n * mean -(n-1) * mei
  j.md <- n * median -(n-1) * mdi
  jbar.me <- mean(j.me)
  jbar.md <- mean(j.md)
  hist(j.me);hist(j.md)
  list("jack estimate for mean is" = jbar.me,
       "jack estimate for median is" = jbar.md)
  }
```

تذکر: همانطور که در مقدمه این مبحث اشاره شد در صورت وجود مشکل در داده ها از روش جک نایف نیز در رگرسیون، سری زمانی و ... می توان استفاده کرد.

تابعی بنویسید که برای متغیر ورودی برآورد جک نایف کجی و کشیدگی را محاسبه کند.

تابع زیر برآوردهای بوت استرپ، اریبی و فواصل اطمینان برای ضرایب یک مدل خطی ساده محاسبه می کند.

```
> boot.reg <- function(x, y, B = 100, alpha = 0.1){
  b0.obs <- lm(y ~ x)$coef[1]
  b1.obs <- lm(y ~ x)$coef[2]
  b0coef <- b1coef <- rep(NA, length(x))
  for (j in 1:B){
  xx <- sample(x, size = length(x) , replace = T)
  yy <- sample(y, size = length(y) , replace = T)
  b0coef[j] <- lm(yy ~ xx)$coef[1]
  b1coef[j] <- lm(yy ~ xx)$coef[2]
  }
  b0.mean <- mean(b0coef); b1.mean <- mean(b1coef)
  bias.b0 <- b0.obs - b0.mean; bias.b1 <- b1.obs - b1.mean
  b0.sd <- sqrt(var(b0coef)); b1.sd <- sqrt(var(b1coef))
  b0.sort <- sort(b0coef); b1.sort <- sort(b1coef)
  low.b0 <- b0.sort[round(alpha * B)]
  high.b0 <- b0.sort[round((1 - alpha) * B)]
  low.b1 <- b1.sort[round(alpha * B)]
  high.b1 <- b1.sort[round((1 - alpha) * B)]
  par(mfrow=c(1,2))
  hist(b0coef); hist(b1coef)
```

```
  qb0 <- quantile(b0coef); qb1 <- quantile(b1coef)
  result <- c(b0.obs,b0.mean,bias.b0,b0.sd,b1.obs,b1.mean,bias.b1,b1.sd,
              low.b0,high.b0,low.b1,high.b1)
  names(result) <-
         c("b0.obs","b0.mean","bias.b0","b0.sd","b1.obs","b1.mean",
           "bias.b1","b1.sd","low.b0","high.b0","low.b1","high.b1")
  return(result, qb0, qb1)
  }
```

تابع زیر برآوردهای جک نایف، اریبی و فواصل اطمینان برای ضرایب یک مدل خطی ساده محاسبه می کند.

```
> jack.reg <- function(x ,y, alpha = 0.1){
  b0.obs <- lm(y ~ x)$coef[1]
  b1.obs <- lm(y ~ x)$coef[2]
  n <- length(x)
  b0coef <- b1coef <- rep(NA, n)
  for(i in 1:n){
  xx <- x[-i]
  yy <- y[-i]
  b0coef[i] <- lm(yy ~ xx)$coef[1]
  b1coef[i] <- lm(yy ~ xx)$coef[2]
  }
  b0.mean <- mean(b0coef); b1.mean <- mean(b1coef)
  bias.b0 <- b0.obs - b0.mean; bias.b1 <- b1.obs - b1.mean
  b0.sd <- sqrt(var(b0coef)); b1.sd <- sqrt(var(b1coef))
  b0.sort <- sort(b0coef); b1.sort <- sort(b1coef)
  low.b0 <- quantile(b0.sort, alpha / 2 )
  high.b0 <- quantile(b0.sort, 1 - alpha /2 )
  low.b1 <- quantile(b1.sort, alpha / 2 )
  high.b1 <- quantile(b1.sort, 1 - alpha /2 )
  par(mfrow=c(1,2))
  hist(b0coef); hist(b1coef)
  qb0 <- quantile(b0coef); qb1 <- quantile(b1coef)
  result <- c(b0.obs,b0.mean,bias.b0,b0.sd,b1.obs,b1.mean,bias.b1,b1.sd,
              low.b0,high.b0,low.b1,high.b1)
  names(result) <-
          c("b0.obs","b0.mean","bias.b0","b0.sd","b1.obs","b1.mean",
            "bias.b1","b1.sd","low.b0","high.b0","low.b1","high.b1")
  return(result, qb0, qb1)
  }
```

تابع زیر برآوردهای بوت استرپ را برای ضریب یک مدل آتورگرسیو مرتبه اول محاسبه می کند.

```
> boot.arima <- function(B = 50 , arp = 1){
  # B is Number of boostrap replicates
  # arp is Order of AR model to be fit
  x <- arima.sim(model = list(ar = c(0.9)),n = 50)
  arcoef <- matrix(0,B,arp)        # matrix of coefficients
  arvarp <- matrix(0,B,1)             # matrix of variance estimates
  armean <- matrix(0,B,1)             # matrix of mean values
  n <- length(x)                 # number of observations
  ninit <- 100                   # cushion for initialization of forecast

 # Begin bootstrap iteration
  for (j in 1:B){
  fitmodel <- arima.mle(x, model=list(order = c(arp,0,0)),xreg = 1)
  fitmodel$stdres <- arima.diag(fitmodel, plot=F)$std.res
  arcoef[j,] <- fitmodel$model$ar
  arvarp[j] <- fitmodel$sigma2
  armean[j] <- fitmodel$reg.coef

# resample residuals
   residsample <- sample(fitmodel$stdres[(arp+1):n],size=(n+ninit),replace=T)
   xnew <- residsample
   for (i in (arp+1):(n+ninit)){
```

```
   xnew[(i+1)] <- arima.forecast(xnew[1:i],n = 1,
       model=fitmodel$model)$mean+residsample[(i+1)]
   }
    x <- xnew[(ninit+1):(n+ninit)]
  }
   hist(arcoef); hist(arvarp)
  return(mean(armean), mean(arvarp), mean(arcoef))
 }
```

تابعی برای محاسبه فاصله اطمینان بوت استرپ برای میانه جامعه

```
bootmedci <- function(x, conflevel=0.95, nboots=100){
# This function will calculate a naive confidence interval
# for the population median.
# nboots = B = number of bootstrap samples to take
# tstar will contain the median calculated for each bootstrap

alpha <- 1 - conflevel
tstar <- rep(0, nboots)
for(b in 1:nboots) tstar[b] <- median(sample(x, replace=T))
mtstar <- mean(tstar)
vtstar <- var(tstar)
biasest <- mtstar - median(x)
lcl <- (median(x)-biasest)-qnorm(1-alpha/2)*sqrt(vtstar)
ucl <- (median(x)-biasest)+qnorm(1-alpha/2)*sqrt(vtstar)
c(lcl, ucl)
}
x <- rnorm(20,3,4)
bootmedci(x)
```

تابعی برای محاسبه احتمال پوشش برای فاصله اطمینان میانه

```
# This will perform a brief simulation study (n=1000) and
# give the observed coverage probability (nominal=0.95)

success <- rep(0,1000)
ci <- matrix(0,1000,2)
for(i in 1:1000){
  x <- rnorm(20,3,4)
  ci[i,] < -bootmedci(x)
  success[i] <- ((ci[i,1]<3)&(ci[i,2]>3))
}
sum(success/1000)


# This will draw a graphical display of the effectiveness#
# of the confidence interval for the 1,000 samples#

low<-min(ci[,1])
hi<-max(ci[,2])
plot(0,0,xlim=c(1,1000),ylim=c(low,hi),type="n")
for (i in 1:1000){
  lines(c(i,i),c(low,ci[i,1]))
  lines(c(i,i),c(ci[i,2],hi))
}
lines(c(0,1001),c(3,3))


#These lines compare the width for the classic large sample#
#  CI for the median in this case to the observed widths#
#  and also gives the summary of the observed centers#

2*1.96*sqrt(1/(4*dnorm(3,mean=3,sd=4)^2))/sqrt(20)
summary(ci[,2]-ci[,1])
```

```
summary((ci[,2]+ci[,1])/2)
```

تابعی برای محاسبه فاصله اطمینان و احتمال پوشش برای ضرایب خط رگرسیون

```
regresboot<-function(x,y,conflevel=0.95,nboots=120){
#This function estimates the CI for the slope in #
#linear regression using resampling from the residuals.#
alpha<-1-conflevel
b1hatstar<-rep(0,nboots)
model<-lm(y~x)
b0hat<-model$coefficients[1]
b1hat<-model$coefficients[2]
resids<-model$residuals
for (b in 1:nboots){
  ystar<-b0hat+b1hat*x+sample(resids,replace=T)
  b1hatstar[b]<-lm(ystar~x)$coefficients[2]
}
tstar<-sort(b1hatstar)
lowtstar<-0.5*(tstar[floor(nboots*alpha/2)]+tstar[ceiling(nboots*alpha/2)])
hitstar<-0.5*(tstar[floor(nboots*(1-alpha/2))]+tstar[ceiling(nboots*(1-
alpha/2))])
c(lowtstar,hitstar)
}

regptsboot<-function(x,y,conflevel=0.95,nboots=120){
#This function estimates the CI for the slope in #
#linear regression using resampling of the observations.#
alpha<-1-conflevel
b1hatstar<-rep(0,nboots)
for (b in 1:nboots){
  pointstar<-sample(1:length(x),replace=T)
  b1hatstar[b]<-lm(y[pointstar]~x[pointstar])$coefficients[2]
}
tstar<-sort(b1hatstar)
lowtstar<-0.5*(tstar[floor(nboots*alpha/2)]+tstar[ceiling(nboots*alpha/2)])
hitstar<-0.5*(tstar[floor(nboots*(1-alpha/2))]+tstar[ceiling(nboots*(1-
alpha/2))])
c(lowtstar,hitstar)
}

success<-matrix(0,100,2)
ci<-array(0,dim=c(100,2,2))
for (i in 1:100){
  x<-1:20
  y<-5+3*x+rnorm(20,0,0.5)
 ci[i,,1]<-regresboot(x,y,nboots=120)
 ci[i,,2]<-regptsboot(x,y,nboots=120)
 success[i,1]<-((ci[i,1,1]<3)&(ci[i,2,1]>3))
 success[i,2]<-((ci[i,1,2]<3)&(ci[i,2,2]>3))
}
apply(success,2,mean)
apply(ci[,2,]-ci[,1,],2,mean)
```

آزمون های جایگشت

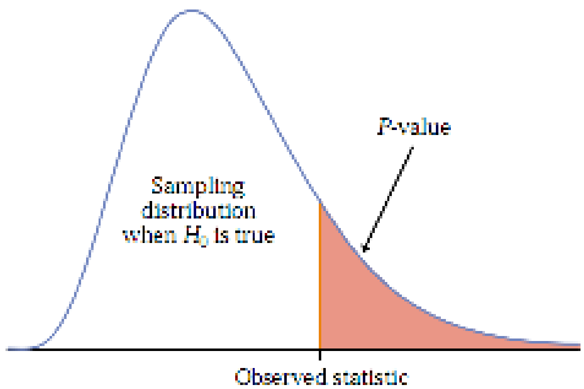
آزمون های جایگشت[۱۶] روشی برای بررسی یک فرضیه بر اساس تکنیک بازنمونه گیری می باشند. نامهای دیگر این آزمون، آزمون تصادفی یا آزمون دقیق[۱۷] می باشد. فرض کنید مجموعه داده پیچیده ای داریم و معتقدیم که دارای

---

[16] *Permutation tests*
[17] *Randomization test, re-randomization test, or exact test*

ساختار غیر تصادفی است. آزمون های تصادفی (جایگشت) که توسط فیشر و پیتمن (۱۹۳۵) معرفی شدند، روشی بسیار عمومی و استوار بـرای تخمین احـتـمال رخداد سـاختار مشـاهده شـــده تحت فرض صفر می باشند. آزمون های تصادفی شاخه ای از آمار ناپارامتری می باشند. آزمون های تصادفی برای تمام آزمون های پارامتری مانند Z ، t ، کای دو و کولموگروف اسمیرنف و ... وجود دارند.

همانطور که می دانید مقدار احتمال متناظر یک آماره به شکل زیر تعریف می شود و مقادیر کوچک آن منجر به رد فرض صفر و مقادیر بزرگ آن منجر به تایید فرض صفر می شود.



مقدار احتمال بوسیله روش های باز نمونه گیری قابل برآورد می باشد. برای برآورد مقدار احتمال یک آزمون معنی داری، توزیع نمونه گیری آماره تحت فرض صفر بوسیله روش های باز نمونه گیری برآورد می شود.

فرض کنید می خواهیم فرض $H_۰: \mu_T = \mu_C$ را در مقابل $H_۱: \mu_T > \mu_C$ مورد بررسی قرار دهیم. برای انجام آزمون جایگشت کافیست B نمونه (معمولاً ۹۹۹) بدون جایگذاری از نمونه اولیه استخراج و در هر مورد مقدار آماره محاسبه شود، اگر a تعداد دفعاتی باشد که مقدار آماره حاصل از نمونه باز نمونه گیری بزرگتر از مقدار آماره در نمونه اولیه باشد، برآورد مقدار احتمال عبارتست از

$$pvalue = \frac{a+۱}{B+۱} \qquad , \qquad a = \#(T_i > T_{obs}) \qquad i = ۱,۲, \dots , B$$

فرض کنید می خواهیم تاثیر یک شیوه آموزشی را در مهارت خواندن دانش آموزان مطالعه کنیم، تعداد ۲۱ دانش آموز در گروه تیمار (آموزش به شیوه جدید) و ۲۳ دانش آموز در گروه کنترل (آموزش به شیوه سنتی) قرار دارند. اگر شیوه آموزشی جدید بدون تاثیر باشد بدین معنی است که انتصاب دانش آموزان به هر گروه نباید تاثیری داشته باشد در واقع تحت فرض صفر ($H_۰: \mu_{\text{Treatment}} = \mu_{\text{Control}}$) انتصاب دانش آموزان به دو گروه نشاندهنده تصادف است. بنابراین در آزمون جایگشت B مرتبه نمونه بدون جایگذاری از نمونه اولیه (۴۴ دانش آموز) استخراج و دو گروه مجدد شکل گرفته و در هر مورد مقدار آماره محاسبه خواهد شد.

آزمون تی دو نمونه ای بر مبنای فرض نرمال بودن توزیع $\bar{X}_۱ - \bar{X}_۲$ بنا شده است. که این فرض منجر به فرض نرمال بودن هر دو جامعه می شود. اما بر اساس قضیه حد مرکزی آزمون تی استوار می باشد، یعنی برای نمونه های بزرگ و غیر نرمال نیز آزمون قابل استفاده است. آزمون تی دو نمونه ای در مواردی که توزیع دو جامعه متقارن می باشند مخصوصاً وقتی حجم نمونه یکسان است عملکرد خوبی دارد. آزمون جایگشت بدون نیاز به فرض نرمالیتی عمل می کند. بعلاوه در آزمون تی فرض همتوزیعی دو جامعه تحت فرض صفر نیز پذیرفته شده است یعنی علاوه بر تساوی میانگین ها تساوی واریانسها و شکل دو توزیع نیز پذیرفته شده است ولی آزمون جایگشت نیازی به فرض های فوق ندارد و حتی برای گروه هایی با واریانس ها و شکل های متفاوت نیز بخوبی عمل می کند. برای اغلب اهداف ۹۹۹

تکرار کافی است، در مواردی که مقدار احتمال نزدیک ۰/۰۱ یا ۰/۰۵ قرار می گیرد افزایش تعداد تکرارها توصیه می شود. اگر مقدار احتمال یک فرض یکطرفه برابر p باشد انحراف معیار مقدار احتمال برابر $\sqrt{p(۱-p)/B}$خواهد بود. بنابراین می توان B را طوری انتخاب کرد دقت مورد نظر حاصل شود. تعداد تکرارها در تعیین دقت مقدار احتمال بر حسب ارقام اعشار مهم است، آزمون هایی با ۹۹۹ تکرار منجر به مقدار احتمالی با ۳ رقم اعشار و حاشیه خطایی برابر ۰/۰۱۴ $(۲\sqrt{p(۱-p)/۹۹۹} = ۰/۰۱۴)$ بوده وقتیکه مقدار احتمال یک طرفه واقعی ۰/۰۵ است. اگر دقت بالاتری مورد نیاز باشد می توان تعداد تکرار ۹۹۹۹ را انتخاب کرد.

توجه: آزمون های جایگشت برای آزمون های تک نمونه ای قابل استفاده نمی باشند. در مواردی که آزمون جایگشت قابل استفاده نیست استفاده از فاصله اطمینان بوت استرپ توصیه می شود. اگر فاصله اطمینان شامل مقدار فرض صفر نباشد فرض $H_۰$ در سطح معنی داری مورد نظر رد می شود، اگر چه که دقت این روش به اندازه آزمون جایگشت نیست. البته در مواردی که انجام آزمون ممکن است، استفاده از فاصله اطمینان می تواند مفید می باشد. در فواصل اطمینان صحت فرض صفر تصور نمی شود بنابراین از روش بوت استرپ بدون جایگذاری بجای آزمون جایگشت با جایگذاری می توان استفاده کرد. در واقع آزمون های جایگشت از خانواده آزمون های معنی داری[۱۸] بر اساس نمونه های جایگشت هستند که بطور تصادفی از داده های اصلی انتخاب شده اند، بعلاوه نمونه های جایگشت بدون جایگذاری انتخاب می شوند در حالیکه نمونه های بوت استرپ باجایگذاری انتخاب می شوند.

مقایسه آزمون تی و جایگشت

فرض کنید اطلاعات جدول زیر برای دو متغیر در دسترس می باشد

X			Y		
A	B	C	D	E	F
121	118	110	34	12	22
$\overline{x}_n = 116.33$			$\overline{y}_n = 22.67$		

بر این اساس مقدار آماره تی و مقدار احتمال آن عبارتند از $t = ۱۳/۰۸۷۵ \quad , \quad pvalue = ۰/۰۰۰۲$

برای انجام آزمون جایگشت تعداد $B = C_۳^۶ = ۲۰$ جایگشت وجود دارد. به عنوان یک نمونه جایگشت داریم

X			Y		
A	B	D	C	E	F
121	118	34	110	12	22
$\overline{x}_n = 91$			$\overline{y}_n = 48$		

جدول زیر نتیجه ۲۰ جایگشت را نشان می دهد. با توجه به نتایج جدول

$$p_{per} = \frac{۱}{B}\sum_{b=۱}^{B} I(|t_b| \geq |t|) = \#\{|t_b| \geq |t|\}/B = \frac{۱}{۲۰}\sum_{b=۱}^{۲۰} I(|t_b| \geq |۱۳/۰۸۷۵|) = \frac{۲}{۲۰} = ۰/۱$$

همانطور که مشاهده می شود فرض صفر در سطح معنی داری ۵٪ بر اساس آزمون تی رد ولی بر اساس آزمون جایگشت تایید می شود. که با توجه به حجم کم نمونه نتیجه آزمون جایگشت بیشتر قابل اعتماد است.

permutation	X	Y	$\bar{x}_n$	$\bar{y}_n$	$\bar{x}_n - \bar{y}_n$	t
1	ABC	DEF	116.33	22.67	93.67	13.087
2	ABD	CEF	91.00	48.00	43.00	1.019
3	ABE	CDF	87.00	52.00	35.00	0.795
4	ABF	CDE	83.67	55.33	28.33	0.627
5	ACD	BEF	88.33	50.67	37.67	0.866
6	ACE	BDF	84.33	54.67	29.67	0.659
7	ACF	BDE	81.00	58.00	23.00	0.500
8	ADE	BCF	59.00	80.00	-21.00	-0.455
9	ADF	BCE	55.67	83.33	-27.67	-0.611
10	AEF	BCD	51.67	87.33	-35.67	-0.813
11	BCD	AEF	87.33	51.67	35.67	0.813
12	BCE	ADF	83.33	55.67	27.67	0.611
13	BCF	ADE	80.00	59.00	21.00	0.455
14	BDE	ACF	58.00	81.00	-23.00	-0.500
15	BDF	ACE	54.67	84.33	-29.67	-0.659
16	BEF	ACD	50.67	88.33	-37.67	-0.866
17	CDE	ABF	55.33	83.67	-28.33	-0.627
18	CDF	ABE	52.00	87.00	-35.00	-0.795
19	CEF	ABD	48.00	91.00	-43.00	-1.019
20	DEF	ABC	22.67	116.33	-93.67	-13.087

مثال) می خواهیم بر اساس اطلاعات جدول زیر آزمون $H_۰: \mu_x = \mu_y \;\; vs \;\; H_۱: \mu_x \geq \mu_y$ را انجام دهیم.

		۱۹۷	۱۴۱	۹۹	۹۴	۳۸	۲۳	۱۶	
۱۴۶	۱۰۴	۵۲	۵۰	۴۶	۴۰	۳۱	۲۷	۱۰	

از آنجا که $\bar{X} - \bar{Y} = ۳۰/۶۳۴$ می باشد. می بایست $P(\bar{X} - \bar{Y} \geq ۳۰/۶۳۴ | F = G)$ را محاسبه کنیم. گروه های یک و دو شامل ۷ و ۹ عضو می باشند بنابراین برای آزمون جایگشت $۱۶!/(۷!\,۹!) = ۱۱۴۴۰$ تکرار لازم است. مقدار احتمال آزمون جایگشت ۰/۱۴۰۶ می شود. در آزمون بوسیله ۱۰۰۰ تکرار بوت استرپ تعداد ۱۲۲ مورد $\bar{X} - \bar{Y} \geq ۳۰/۶۳۴$ می شود، بنابراین مقدار احتمال بوت استرپ برابر ۰/۱۲۲ می شود. بنابراین فرض صفر در هر دو مورد رد نمی شود.

نکته: در مقایسات دو گروهی برای میانگین وقتی واریانس ها برابرند یا حجم های نمونه مساوی اند آزمون جایگشت دقیق خواهد بود. (رومانو، جاسا ۱۹۹۰)

نکته: در مقایسات دو گروهی برای میانه آزمون جایگشت حتی به طور مجانبی نیز دقیق نیست مگر اینکه دو گروه همتوزیع باشند. ضمناً این مطلب از حجم های نمونه مستقل است. (رومانو، جاسا ۱۹۹۰)

وقتی آزمون جایگشت دچار خطا می شود

مقایسه میانگین دو گروه مستقل با میانگین ها، واریانسها و اندازه های متفاوت

```
x <- rnorm(25,0,1)
y <- rnorm(75,1,4)
perm.test(x,y,exact=T)          pvalue=0.1740
boot.test(x,y,B=5000)           pvalue=0.0302
welch.test(x,y)                 pvalue=0.0243
```

مقایسه آزمون های بوت استرپ و جایگشت

۱- نتایج آنها بسیار شبیه یکدیگر است

۲- مقدار ASL وقتی $B = C^N$ باشد احتمال دقیق است

۳- مقدار ASL_{boot} احتمال دقیق نیست اما به اندازه کافی به عنوان برآوردی از ASL وقتی B به سمت بینهایت می رود مورد قبول می باشد.

تمرین های متنوع ۳

تابعی بنویسید که بر اساس شبیه سازی مونت کارلو از توزیع استودنت با ۱۰ درجه آزادی تعداد ۱۵ مشاهده شبیه سازی کند سپس برای محاسبه برآوردهای بوت استرپ ۱۰۰۰ نمونه ۱۵ تایی به صورت با جایگذاری انتخاب کند و بر اساس داده های حاصل دهک اول و چارک سوم را محاسبه کند و بر این اساس تخمینهای بوت استرپ را برای متوسط ، انحراف معیار ، میانه و فاصله اطمینان آماره های مذکور را محاسبه نماید و بوسیله رسم بافت نگار تخمینهای بوت استرپ، به بررسی توزیع آماره های مورد نظر بپردازد.

تابعی بنویسید که بر اساس یک نمونه ۲۰ تایی مقدار احتمال بوت استرپ را جهت بررسی فرضیه $H_0 : \rho \geq 0.3 \;vs\; H_1 : \rho < 0.3$ محاسبه کند، همچنین بر اساس اطلاعات حاصل فاصله اطمینان ۹۰٪ را برای ضریب همبستگی محاسبه کند.

تابعی بنویسید که بر اساس اطلاعات یک پرسشنامه که شامل ۱۰ سوال است و توسط ۲۰ نفر شده فرضیه $H_0 : r_\alpha \geq 0.7 \;vs\; H_1 : r_\alpha < 0.7$ را بررسی کند، همچنین بر اساس اطلاعات حاصل فاصله اطمینان ۹۵٪ را برای آلفای کرانباخ محاسبه کند.

تابعی بنویسید که بر اساس اطلاعات یک پرسشنامه که شامل ۱۰ سوال است و توسط ۲۰ نفر شده فرضیه $H_0 : r_{SB} \geq 0.85 \;vs\; H_1 : r_{SB} < 0.85$ را بررسی کند، همچنین بر اساس اطلاعات حاصل فاصله اطمینان ۹۵٪ را برای ضریب اسپیرمن براون محاسبه کند.

تابعی بنویسیدکه برای متغیر ورودی آماره باکس جانگ (تصحیح شده باکس پیرس) را محاسبه کند و براساس الگوریتم بوت استرپ پارامتری برای فرضیه $H_0 : \rho_1 = \rho_2 = \cdots = \rho_{10} = 0$ مقدار احتمال را محاسبه کند.

مثال: یک شرکت وظیفه تامین پمپ آب در یک منطقه کشاورزی را بر عهده دارد. طبق برآوردهای اعلام شده تعداد پمپ آب مورد نیاز کشاورزان در ماه بین ۲۲ الی ۲۵ عدد با احتمالات زیر است.

تعداد پمپ	۲۲	۲۳	۲۴	۲۵
احتمال	۰/۳۰	۰/۴۵	۰/۲۰	۰/۰۵

اگر هنگام درخواست پمپ موجود نباشد درخواست با تاخیر تحویل داده شده و شرکت باید مبلغ ۳۰۰۰ تومان بابت نصب پمپ متقبل شود. هزینه نگهداری هر پمپ در ماه ۳۰۰ تومان است و برای هر بار سفارش این شرکت باید مبلغ ۱۰۰۰۰ تومان بپردازد. می خواهیم بدانیم سفارشات ماهی یک بار انجام شود یا دو بار کدام یک هزینه کمتری برای شرکت ایجاد می کند.

حل) امید ریاضی تعداد پمپ برابر ۲۳ پمپ در ماه است، بنابراین در هر ماه ۲۳ پمپ و در هر دو ماه ۴۶ پمپ سفارش داده می شود.

الف) برای بررسی وضعیت در یک سال (۱۲ ماه) ۱۲ عدد تصادفی در بازه [۰,۹۹] تولید می کنیم و بر اساس اعداد تولیدی به تحلیل وضعیت یک ساله می پردازیم.

احتمال	۰/۳۰	۰/۴۵	۰/۲۰	۰/۰۵

احتمال تجمعی	۰/۳۰	۰/۷۵	۰/۹۵	۱/۰۰

مقادیر بازه ۰ الی ۲۹ معادل میزان تقاضای ۲۲ عدد پمپ، مقادیر بازه ۳۰ الی ۷۴ معادل میزان تقاضای ۲۳ عدد پمـپ، مقـادیر بازه ۷۵ الی ۹۴ معادل میزان تقاضای ۲۴ عدد پمپ و مقادیر بازه ۹۵ الی ۹۹ معادل میزان تقاضای ۲۵ عدد پمپ خواهد بود.

ماه	اعداد تصادفی تولید شده	میزان تقاضا	موجودی اول ماه	موجودی آخر ماه	کمبود	متوسط موجودی
۱	۳۷	۲۳	۲۳	۰	۰	۱۱/۵
۲	۱۱	۲۲	۲۳	۱	۰	۱۲
۳	۳۷	۲۳	۲۲	۱	۰	۱۲/۵
۴	۴۶	۲۳	۲۴	۱	۰	۱۲/۵
۵	۳۲	۲۳	۲۴	۱	۰	۱۲/۵
۶	۶۳	۲۳	۲۴	۱	۰	۱۲/۵
۷	۸۲	۲۴	۲۴	۰	۰	۱۲
۸	۲۱	۲۲	۲۳	۱	۰	۱۲
۹	۶۰	۲۳	۲۳	۱	۰	۱۲/۵
۱۰	۴۳	۲۳	۲۳	۱	۰	۱۲/۵
۱۱	۹۷	۲۵	۲۳	۰	۱	۱۲
۱۲	۸۱	۲۴	۲۲	۰	۲	۱۱

۲/(موجودی آخر ماه+ موجودی اول ماه) = متوسط موجودی

۱۲/۵ = ۱۲/ مجموع متوسط موجودی هر ماه = متوسط موجودی کل

۴۳۶۵۰=۱۲×۳۰۰×۱۲/۵= هزینه نگهداری درکل سال

۹۰۰۰=۳×۳۰۰۰= هزینه کمبود

۱۲۰۰۰۰=۱۲×۱۰۰۰۰= هزینه سفارش

بنابراین در نهایت هزینه کل برابر ۱۷۲۶۵۰ خواهد بود.

ب) میزان دریافت در هر دو ماه ۴۶ عدد است.

ماه	اعداد تصادفی تولید شده	موجودی آخر ماه	کمبود	متوسط موجودی
۱	۴۶	۲۳	۰	۳۴/۵
۲	۲۳	۱	۰	۱۲
۳	۴۷	۲۴	۰	۲۵/۵
۴	۲۴	۱	۰	۱۲/۵
۵	۴۷	۲۴	۰	۳۵/۵
۶	۲۴	۱	۰	۱۲/۵
۷	۴۷	۲۳	۰	۳۵
۸	۲۳	۱	۰	۱۲
۹	۴۷	۲۴	۰	۳۵/۵
۱۰	۲۴	۱	۰	۱۲/۵

۱۱	۴۷	۲۲	۰	۳۴/۵
۱۲	۲۲	۰	۲	۱۱

۲۳/۶ = ۱۲/ مجموع متوسط موجودی هر ماه = متوسط موجودی کل

۸۴۹۶۰=۱۲×۳۰۰×۲۳/۶= هزینه نگهداری درکل سال

۶۰۰۰=۲×۳۰۰۰= هزینه کمبود

۶۰۰۰۰=۶×۱۰۰۰۰= هزینه سفارش

بنابراین در نهایت هزینه کل برابر ۱۵۰۹۶۰ خواهد بود. همانطور که مشاهده می شود در وضعیت دوم هزینه کل کمتر می شود