فصل ٦ استنباط آماری: فاصله اطمینان

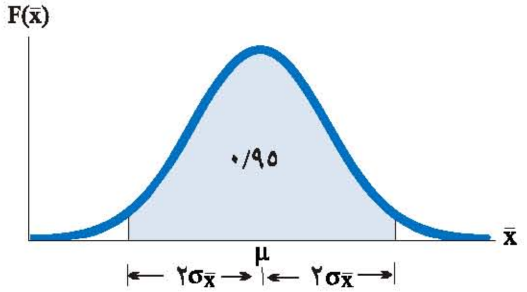
٦-١ مقدمه

با افزایش اندازه نمونه می‌توان دقت برآورد را افزایش داد ولی با این وجود برآورد نقطه‌ای به دست آمده ممکن است دقیقاً با پارامتر جامعه برابر نباشد. پس بهتر است فاصله‌ای را مشخص کرد طوری که با اطمینان خاصی پارامتر را در بر بگیرد. به چنین فاصله‌ای، فاصله اطمینان گفته می‌شود.

٦-٢ فاصله اطمینان برای میانگین جامعه با استفاده از نمونه بزرگ

با توجه به قضیه حد مرکزی، توزیع نمونه‌ای میانگین همانگونه که در نمودار زیر نشان داده شده است تقریباً نرمال بوده و در نتیجه با ٪۹۵ احتمال، $\overline{x}$ در فاصله دو انحراف معیار از میانگین (μ) قرار می‌گیرد. پس فاصله $\overline{x} \pm ۲\sigma_{\overline{x}}$ در ٪۹۵ موارد، μ را در بر دارد.

$$\overline{x} \pm ۲\sigma_{\overline{x}} = \overline{x} \pm ۲\frac{\sigma}{\sqrt{n}}$$



مثال ٦-۱. ارتفاع ۱۰۰ درخت کاج در یک پارک جنگلی اندازه‌گیری شده و میانگین و انحراف معیار آنها به ترتیب ٦٧٠ و ٢٤٠ سانتی‌متر به دست آمده است. میانگین جامعه با احتمال ۹۵٪ در چه فاصله‌ای قرار می‌گیرد؟

برای پاسخ به این سؤال باید فاصله

$$\overline{x} \pm ۲\sigma_{\overline{x}} = ٦٧٠ \pm ۲\frac{\sigma}{\sqrt{۱۰۰}}$$

را به دست آورد. مشکل اینجاست که انحراف معیار جامعه (انحراف معیار طول کلیه درختان کاج) معلوم نیست. با وجود این با توجه به بزرگ بودن نمونه، می‌توان از انحراف معیار نمونه، s، به عنوان برآوردی مناسب برای σ استفاده کرد. لذا

$$\overline{x} \pm ۲\frac{\sigma}{\sqrt{۱۰۰}} \approx \overline{x} \pm ۲\frac{s}{\sqrt{۱۰۰}} = ٦٧٠ \pm ۲\frac{٢٤٠}{\sqrt{۱۰۰}} = ٦٧٠ \pm ٤٨$$

پس با اطمینان ۹۵٪ می‌توان گفت که فاصله ٦٢٢ تا ٧١٨ سانتی‌متر، میانگین جامعه یعنی میانگین طول کلیه درختان کاج در پارک جنگلی را در بر دارد. توصیف ۹۵٪ اطمینان چنین است که اگر به دفعات زیاد نمونه‌های ۱۰۰تایی از درختان پارک استخراج و در هر مورد فاصله اطمینان ساخته شود، در ۹۵٪ موارد فاصله اطمینان ساخته شده میانگین جامعه را در بر می‌گیرد. پس فاصله اطمینان موجود، ممکن است یکی از ۹۵ ٪ فواصلی باشد که μ را در بر می‌گیرد یا یکی از ۵٪ فواصلی است که μ را در بر ندارد.

فرمول برآورد فاصله‌ای بر اساس داده‌های نمونه، برآوردگر فاصله‌ای یا فاصله اطمینان نامیده شده و می‌گوید که فاصله مورد نظر با چه اطمینانی پارامتر را در بر دارد. برای مثال فاصله $\overline{x} \pm ۲\sigma_{\overline{x}}$ با اطمینان ۹۵٪ میانگین جامعه، μ، را در بر دارد. این موضوع به صورت زیر نیز نوشته می‌شود:

$$P(\overline{x} - ۲\sigma_{\overline{x}} < \mu < \overline{x} + ۲\sigma_{\overline{x}}) = ۰/۹۵$$

مقدار ۹۵٪ را ضریب اطمینان یا سطح اطمینان می‌گویند.

گرچه برآورد فاصله‌ای در مثال بالا با اطمینان ۹۵٪ محاسبه شد، اما این اطمینان می‌تواند هر مقداری بین ۰ تا ۱۰۰٪ باشد. اگر سطح اطمینان ۹۵٪ را برابر با $(1-\alpha)$ در نظر بگیریم، α برابر با ۵٪ و از آنجا $\alpha / 2 = 0/025$ خواهد بود که همان سطح مربوط به zهای بزرگتر از ۱/۹۶ (تقریباً ۲) در جدول ضمیمه z می‌باشد. در حالت کلی برای یک نمونه بزرگ از هر توزیعی، فاصله اطمینان $(1-\alpha)$ برای μ به صورت زیر است.

$$\overline{x} \pm z_{\alpha/2}\sigma_{\overline{x}} = \overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

زیرا

$$P(\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

که در آن $z_{\alpha/2}$ مقدار z مربوط به سطح $\alpha/2$ در سمت راست منحنی z است. پارامتر σ نیز انحراف معیار جامعه و n تعداد اعضای نمونه می‌باشد. در صورتی که مقدار σ نامعلوم باشد به شرط بزرگ بودن نمونه می‌توان از تقریب آن یعنی انحراف معیار نمونه (s)، در فرمول استفاده نمود زیرا در نمونه‌های بزرگ انحراف معیار نمونه تقریباً برابر با انحراف معیار جامعه است و برآوردگر مناسبی برای آن خواهد بود. البته به تدریج که نمونه کوچکتر می‌شود، استفاده از انحراف معیار نمونه خطای بیشتری خواهد داشت. در واقع هر گاه به جای σ، برآورد آن یعنی انحراف معیار نمونه در اختیار باشد، بدون توجه به حجم نمونه باید از توزیع دیگری به نام توزیع t استفاده کرد. استفاده از توزیع z در مورد نمونه‌های بزرگ در صورتی که σ مجهول باشد فقط تقریبی از توزیع t است که به دلیل ساده‌تر بودن محاسبات مورد استفاده قرار می‌گیرد. توزیع t در بند ۶-۳ توضیح داده خواهد شد.

مثال ٦-۲. عملکرد یک رقم جدید گندم در ۳٦ کرت ۲۰ متر مربعی بر حسب کیلوگرم در کرت به صورت زیر به دست آمده است. برای میانگین واقعی جامعه، μ، یک فاصله اطمینان ۹۰٪ بسازید.

۲۹/۳	۲۱/۲	۲٦/٤	۱۷/۳	۲۸/۲	۲٤/۸	۱۸/٤	۲۵/۹	۲۱/۷	۲٤/۱	۲۳/۹	۲۲/۲
۲۸/۸	۲۳/۵	۱۹/۱	۲۵/۰	۳۰/۸	۲۳/۹	۲٤/۷	۲۲/٦	۲٦/٤	۲۵/۲	۲۱/۹	۲۳/۲
۲۵/۸	۲۰/۰	۲٦/٦	۲۷/۱	۲۱/۳	۲۹/۹	۲٤/۷	۱۹/۳	۲۳/۵	۲۷/۲	۲۰/٤	۲۷/۱

میانگین و انحراف معیار در این نمونه ۳٦ عضوی برابر است با:

$$\overline{x} = \text{۲٤/۲} \qquad s_x = \text{۳/۳۳}$$

بنابراین فاصله اطمینان ۹۰٪ به صورت زیر است:

$$\overline{x} \pm z_{\alpha/\text{۲}}\sigma_{\overline{x}} = \overline{x} \pm z_{\text{۰/۰۵}}\sigma_{\overline{x}} = \overline{x} \pm \text{۱/٦٤}\frac{\sigma}{\sqrt{n}}$$

از آنجا که انحراف معیار واقعی جامعه (انحراف معیار عملکرد واریته مورد نظر) معلوم نیست، از تقریب آن یعنی انحراف معیار نمونه، s، استفاده می‌شود. زیرا نمونه بزرگ بوده و s تقریب خوبی از σ خواهد بود. لذا:

$$\text{۲٤/۲} \pm \text{۱/٦٤}\frac{\text{۳/۳۳}}{\sqrt{\text{۳٦}}} = \text{۲٤/۲} \pm \text{۰/۹۱}$$

در نتیجه با اطمینان ۹۰٪ میانگین عملکرد جامعه (میانگین حقیقی گندم مورد بررسی) بین ۲۳/۲۹ تا ۲۵/۱۲ کیلوگرم در کرت است.

جدول ۶-۱. حل مثال ۶-۲ و خروجی مربوطه در نرم‌افزار Minitab.

مراحل:

۱. داده‌ها را در ستون اول صفحه داده‌ها وارد کنید.

۲. وارد مسیر زیر شوید (اگر انحراف معیار جامعه معلوم نباشد، حتی در مورد نمونه‌های بزرگ نیز باید از دستور 1 Sample t استفاده شود تا نرم‌افزار از انحراف معیار نمونه در محاسبات استفاده کند. واضح است که در مورد نمونه‌های بزرگ، فاصله اطمینان به دست آمده با استفاده از توزیع z تفاوت محسوسی با روش توزیع t نخواهد داشت).

Stat > Basic Statistics > 1-Sample t...

۳. در پنجره مربوط به گزینه ...Options فاصله اطمینان مورد نظر را وارد کرده و گزینه OK را انتخاب نمایید.

One-Sample T: Yield

```
The assumed standard deviation = 3.33


Variable   N    Mean  StDev  SE Mean        90% CI
Yield     36  24.206  3.336    0.555  (23.293; 25.118)
```

۶-۳ فاصله اطمینان برای میانگین جامعه با استفاده از نمونه کوچک

فرض کنید محققی می‌خواهد میانگین افزایش فشار خون بیمارانی را که داروی خاصی مصرف می‌کنند به دست آورد. با فرض اینکه فقط یک نمونه تصادفی ۶ عضوی از بین همه بیماران برای این تحقیق در اختیار است، استفاده از میانگین چنین نمونه کوچکی برای برآورد μ و محاسبه فاصله اطمینان هنگامی که از توزیع z استفاده شود، دو مشکل ایجاد می‌کند.

مشکل اول: شکل توزیع نمونه‌ای برای میانگین نمونه، $\bar{x}$، (و همچنین آماره z حاصل از آن) در صوت کوچک بودن نمونه به شکل جامعه‌ای که از آن استخراج شده است، بستگی دارد. دیگر نمی‌توان فرض کرد که توزیع نمونه‌ای $\bar{x}$ نرمال است، زیرا قضیه حد مرکزی نرمال بودن توزیع نمونه‌ای را مشروط به بزرگ بودن اندازه نمونه کرده است. پس توزیع نمونه‌ای $\bar{x}$ برای نمونه‌های کوچک فقط در صورتی نرمال است که جامعه مربوطه نرمال بوده باشد.

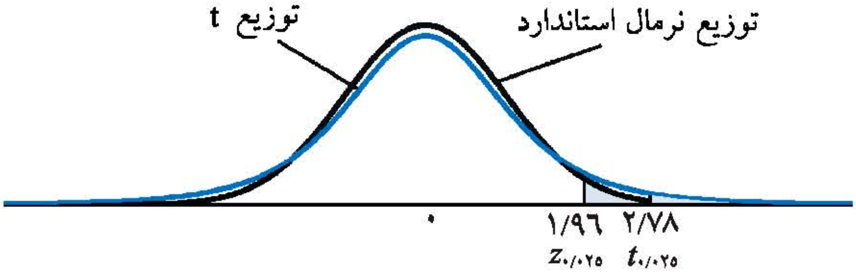
مشکل دوم: می‌دانیم که انحراف معیار جامعه غالباً ناشناخته است. گرچه در مورد نمونه‌های کوچک نیز رابطه $\sigma_{\bar{x}} = \sigma_x / \sqrt{n}$ برقرار است، ولی در این حالت انحراف معیار نمونه، s، مطابقت زیادی با انحراف معیار جامعه، σ، نداشته و جانشین مناسبی برای آن نخواهد بود. برای حل این مشکل به جای استفاده از آماره نرمال استاندارد، یعنی

$$z = \frac{\overline{x} - \mu}{\sigma_{\overline{x}}} = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

که لازم است در آن σ معلوم باشد، آماره جدیدی به صورت

$$t = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}}$$

مورد استفاده قرار می‌گیرد که در آن انحراف معیار نمونه جایگزین انحراف معیار جامعه شده است. آماره t دارای توزیعی است که شباهت زیادی با توزیع z دارد ازجمله اینکه زنگوله‌ای شکل، متقارن و میانگین آن صفر است. تفاوت اصلی بین توزیع‌های t با z این است که t پراکندگی بیشتری دارد، زیرا بر خلاف آماره z که یک متغیر $(\overline{x})$ در فرمول دارد، آماره t دارای دو متغیر ($\overline{x}$ و s) است. در واقع درجه متغیر بودن بستگی به اندازه نمونه و در نتیجه به درجه آزادی آن یعنی $n-1$ دارد. هر چقدر درجه آزادی نمونه کوچکتر باشد شکل توزیع t پراکندگی بیشتری نشان خواهد داد.



نمودار ۶-۱. مقادیر $t_{0.025}$ در یک توزیع t با چهار درجه آزادی در مقایسه با $z_{0.025}$ در توزیع نرمال استاندارد.

در نمودار ۶-۱ شکل توزیع نمونه‌ای آماره z و همچنین شکل توزیع نمونه‌ای برای آماره t با ۴ درجه آزادی نشان داده شده است. متغیر بودن آماره t در مقایسه با آماره z در واقع به این معنا است که مقدار t مربوط به یک سطح خاص مانند α (در سمت راست منحنی) در مقایسه با مقدار z مربوط به همین سطح بزرگتر است. برای یک سطح مشخص مانند α، مقدار آماره t به

ازای کاهش درجه آزادی بزرگتر می‌شود. مقادیر آماره t مربوط به سطوح ۵٪ یا ۱٪ به ازای هر درجه آزادی در جدول ضمیمه ۳ (جدول t) در آخر کتاب آورده شده است. در جدول t مقادیر احتمال در سطر اول و درجات آزادی در ستون سمت چپ نوشته شده است.

فاصله اطمینان $(۱-\alpha)$ برای میانگین جامعه با استفاده از یک نمونه کوچک

فاصله اطمینان $(۱-\alpha)$ برای میانگین جامعه با استفاده از یک نمونه کوچک بصورت

$$\overline{x} \pm t_{\alpha/۲} \frac{s}{\sqrt{n}}$$

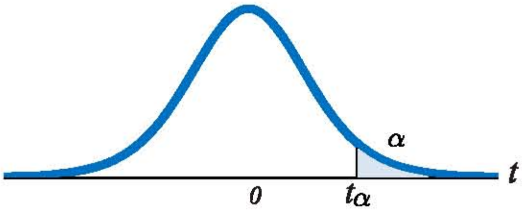
است که در آن n اندازه نمونه و $t_{\alpha/۲}$ عدد مربوط به سطح $\alpha/۲$ در سمت راست توزیع t با توجه به درجه آزادی $n-۱$ می‌باشد. البته فرض بر این است که نمونه از جامعه‌ای با توزیع تقریباً نرمال استخراج شده است. بدیهی است رابطه

$$P(\overline{x} - t_{\alpha/۲} \frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{\alpha/۲} \frac{s}{\sqrt{n}}) = ۱ - \alpha$$

برقرار است.

دقت کنید با افزایش اندازه نمونه (n) به سمت بینهایت شکل توزیع t منطبق بر توزیع z خواهد شد. این امر منطقی است، زیرا اگر n خیلی بزرگ باشد، انحراف معیار نمونه به انحراف معیار جامعه نزدیک خواهد شد. اصولاً هرگاه σ مجهول و s در اختیار باشد، بدون توجه به حجم نمونه باید از توزیع t استفاده کرد در عمل برای درجات آزادی مساوی یا بزرگتر از ۲۹ دو توزیع معادل هم در نظر گرفته می‌شوند، زیرا تفاوت ناچیزی بین t و z در جداول مربوطه برای یک سطح احتمال خاص وجود دارد. استفاده از توزیع z در مورد نمونه‌های بزرگ در صورتی که σ مجهول باشد، فقط تقریبی از t است که به دلیل ساده‌تر بودن محاسبات از آن استفاده می‌شود.

جدول ٦-٢. قسمتی از جدول ضمیمه t.



df	$t_{\cdot/١٠}$	$t_{\cdot/\cdot٥}$	$t_{\cdot/\cdot٢٥}$	$t_{\cdot/\cdot١}$	$t_{\cdot/\cdot\cdot٥}$	$t_{\cdot/\cdot\cdot١}$	$t_{\cdot/\cdot\cdot\cdot٥}$
١	٣/٠٨٧	٦/٣١٤	١٢/٧٠٦	٣١/٨٢١	٦٣/٦٥٧	٣١٨/٣	٦٣٦/٦
٢	١/٨٨٦	٢/٩٢٠	٤/٣٠٣	٦/٩٦٥	٩/٩٢٥	٢٢/٣٣	٣١/٦٠
٣	١/٦٣٨	٢/٣٥٣	٣/١٨٢	٤/٥٤١	٥/٨٤١	١٠/٢١	١٢/٩٢
٤	١/٥٣٣	٢/١٣٢	٢/٧٧٦	٣/٧٤٧	٤/٦٠٤	٥/١٧٣	٨/٦١٠
٥	١/٤٧٦	٢/٠١٥	٢/٥٧١	٣/٣٦٥	٤/٠٣	٥/٨٩٤	٦/٨٦٩
٦	١/٤٤٠	١/٩٤٣	٢/٤٤٧	٣/١٤٣	٣/٧٠٧	٥/٢٠٨	٥/٩٥٩
٧	١/٤١٥	١/٨٩٥	٢/٣٦٥	٢/٩٩٨	٣/٤٩٩	٤/٧٨٥	٥/٤٠٨
٨	١/٣٩٧	١/٨٦٠	٢/٣٠٦	٢/٨٩٦	٣/٣٥٥	٤/٥٠١	٥/٠٤١
٩	١/٣٨٣	١/٨٣٣	٢/٢٦٢	٢/٨٢١	٣/٢٤٠	٤/٢٩٧	٤/٧٨١
١٠	١/٣٧٢	١/٨١٢	٢/٢٢٨	٢/٧٦٤	٣/١٦٩	٤/١٤٤	٤/٥٨٥

حال به بررسی مجدد مثال اثر دارو بر روی افزایش فشار خون بیماران که در بند ٦-٣ مطرح شد پرداخته می‌شود که در آن، افزایش فشارخون ٦ مریض در اثر مصرف دارویی ١/٧، ٣، ٠/٨، ٣/٤، ٢/٧ و ٢/١ بوده است. همچنین فرض کنید که توزیع جامعه (میزان افزایش فشارخون در همه بیماران) نرمال است زیرا فقط در این حالت توزیع نمونه‌ای $\overline{x}$ برای نمونه‌های کوچک نرمال خواهد بود. با توجه به اینکه σ در دسترس نیست، باید از توزیع t با $n-١=٥$ درجه آزادی برای محاسبه فاصله اطمینان دوطرفه ٩٥٪ استفاده شود. مقدار عددی t مربوط به این سطح از جدول ضمیمه ٣ قابل استخراج است.

$$t_{\cdot/\cdot٢٥}=٢/٥٧ \text{ برای درجه آزادی ٥}$$

فاصله اطمینان برای μ به صورت

$$\overline{x}\pm t_{\alpha/٢}\frac{s}{\sqrt{n}}=٢/٢٨٣\pm ٢/٥٧\frac{\cdot/٩٥}{\sqrt{٦}}=٢/٢٨٣\pm \cdot/٩٩٧$$

به دست می‌آید. به عبارتی با اطمینان ۹۵٪ فاصله ۱/۲۸٦ تا ۳/۲۸۰، میانگین افزایش فشار خون کلیه بیماران در اثر مصرف دارو (μ) را در بر دارد. در حقیقت اگر این فاصله در دفعات متوالی با استفاده از نمونه‌های متعدد ٦ عضوی محاسبه شود، ۹۵٪ فاصله‌های به دست آمده میانگین جامعه را در بر خواهد داشت.

جدول ٦-۳. مسیر اجرا و خروجی نرم‌افزار Minitab برای مثال اخیر.

Stat > Basic Statistics > 1-Sample t…

One-Sample T: BP Increase

Variable	N	Mean	StDev	SE Mean	95% CI
BP Increase	6	2.283	0.950	0.388	(1.287; 3.280)

در صورت استفاده از آماره z برای ساختن فاصله اطمینان، طول فاصله اطمینان برابر $۲z_{\alpha/۲} \times \sigma_{\overline{x}}$ خواهد بود. ولی در حالتی که از آماره t استفاده شود، طول فاصله اطمینان برابر $۲t_{\alpha/۲} \times s_{\overline{x}}$ خواهد شد، در حالت دوم چون $t_{\alpha/۲}$ از $z_{\alpha/۲}$ بزرگتر است، طول فاصله اطمینان هم بزرگتر می‌شود. هر چقدر طول فاصله اطمینان بزرگ‌تر باشد دقت برآورد کم‌تر خواهد بود. در این شرایط تنها با افزایش اندازه نمونه می‌توان دامنه فاصله برآورد شده را کاهش داد.
اگر انحراف معیار جامعه در اختیار باشد (که غالباً چنین نیست) با نمونه‌های کوچک نیز فاصله اطمینان مورد نظر برای میانگین جامعه را می‌توان با استفاده از مقادیر نرمال استاندارد، z، به دست آورد. با این وجود هنوز باید به نرمال بودن توزیع جامعه مقید بود.

مثال ٦-۳. یک رقم جدید گندم در ۹ قطعه زمین کشت شده و پس از برداشت، درصد پروتئین گندم در هر قطعه بدست آمد. فرض کنید اعداد ۱۲/۹، ۱۳/٤، ۱۲/٤، ۱۲/۸، ۱۳، ۱۲/۷، ۱۲/٤، ۱۳/۵ و ۱۳/۹ حاصل شده است. الف) یک فاصله اطمینان ۹۵٪ بسازید. ب) اگر ادعا شود محتوای پروتئین این رقم ۱۳/۵ درصد است، آیا این ادعا صحت دارد؟

$$\overline{x} \pm t_{\alpha/۲}\frac{s}{\sqrt{n}} = \overline{x} \pm t_{۰/۰۲۵}\frac{s}{\sqrt{n}} = ۱۳ \pm ۲/۳۰٦\frac{۰/۵۱}{\sqrt{۹}} = ۱۳ \pm ۰/۳۹۲$$

به عبارتی فاصله ۱۲/۶۰۸ تا ۱۳/۳۹۲، با اطمینان ۹۵٪ میانگین درصد پروتئین در کـرت را در بـر دارد. این فاصله، ۱۳/۵ را شامل نمی‌شود پس با ۹۵٪ اطمینـان ادعـای مطـرح شـده در مـورد جامعه صحیح نمی‌باشد.

جدول ۶-۴. مسیر اجرا و خروجی نرم‌افزار Minitab برای مثال ۶-۳.

Stat > Basic Statistics > 1-Sample t…

One-Sample T: C1

```
Test of mu = 13.5 vs not = 13.5

Variable  N    Mean  StDev  SE Mean       95% CI            T      P
C1        9  13.000  0.510    0.170  (12.608; 13.392)  -2.94  0.019
```

۶-۴ فاصله اطمینان برای نسبت جامعه با استفاده از نمونه بزرگ

در اینجا فاصله اطمینان برای نسبت اعضایی از جامعه که دارای ویژگی خاصـی هسـتند، بررسـی می‌شود. هر واحد آزمایشی در یکی از دو حالت به اصطلاح موفقیت یا شکسـت قـرار مـی‌گیـرد یعنی یا ویژگی مورد نظر را دارد یا فاقد آن ویژگی است. لذا تعداد افراد دارای ویژگی مورد نظر، دارای توزیع دوجمله‌ای خواهد بود. اگر در توزیع دوجمله‌ای به جای تعداد پیروزی یا شکسـت، از نسبت آنها استفاده شود، میانگین به جای np برابر با p و واریانس به جـای npq، برابـر بـا pq/n خواهد شد.

نسبت سیب‌های سالم، نسبت کشاورزان باسواد و نسبت افراد سیگاری، مثـال‌هـایی از ایـن نـوع هستند. در هر کدام از این جوامع، p نسبت واحدهای دارای ویژگی مورد نظر است. اگر امکـان بررسی کل جامعه نباشد، مقدار p از روی نمونه برآورد شده و بـا $\hat{p}$ (نسـبت واحـدهـای دارای ویژگی مورد نظر در نمونه) نشان داده می‌شود. لذا $\hat{p}$ خـود متغیـری تصـادفی بـا میـانگین p و واریانس pq/n خواهد بود.

در نمونه‌های بزرگ، قضیه حد مرکزی برای متغیر تصادفی $\hat{p}$ هم صادق است و فاصـله اطمینـان تقریبی مورد نظر برای آن به صورت زیر محاسبه می‌شود.

$$\hat{p} \pm z_{\alpha/۲}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

مثال ٦-٤. برای بررسی نسبت رأی‌دهندگان به یک نماینده خاص در یک استان کشور، ابتدا یک نمونه ۱۰۰۰ نفری به طور تصادفی انتخاب می‌شود. اگر در این نمونه ٦۳۷ نفر به آن نماینده رأی دهند، نسبت افراد طرفدار نماینده مذکور در جامعه را چگونه برآورد می‌کنید؟

اگر p نشاندهنده نسبت افراد طرفدار نماینده مورد نظر باشد، برآورد آن یعنی $\hat{p}$ به صورت زیر به دست می‌آید.

$$\hat{p} = \frac{٦۳۷}{۱۰۰۰} = ۰/٦۳۷$$

$\hat{p}$ متغیری تصادفی است که میانگین آن برابر با میانگین جامعه (p) و انحراف معیار آن برابر با $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$ است. با توجه به قضیه حد مرکزی توزیع نمونه‌ای $\hat{p}$ برای نمونه‌های بزرگ تقریباً نرمال بوده و فاصله اطمینان برای نسبت طرفداران نماینده مورد نظر در جامعه (p) را می‌توان به صورت زیر به دست آورد.

$$\hat{p} \pm z_{\alpha/۲}\sigma_{\hat{p}} = \hat{p} \pm z_{\alpha/۲}\sqrt{\frac{pq}{n}} \approx \hat{p} \pm z_{\alpha/۲}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$\hat{p} \pm z_{\alpha/۲}\sqrt{\frac{\hat{p}\hat{q}}{n}} = ۰/٦۳۷ \pm ۱/۹٦\sqrt{\frac{۰/٦۳۷ \times ۰/۳٦۳}{۱۰۰۰}} = ۰/٦۳۷ \pm ۰/۰۳$$
$$= (۰/٦۰۷,\ \ ۰/٦٦۷)$$

لذا با اطمینان ۹۵٪ می‌توان گفت که فاصله ۰/٦۰۷ تا ۰/٦٦۷، نسبت واقعی طرفداران نماینده مورد نظر در جامعه یعنی p را در بر دارد.

جدول ٦-۵. مسیر اجرا و خروجی نرم‌افزار Minitab برای مثال ٦-٤.

Stat > Basic Statistics > 1 Proportion...

Test and CI for One Proportion

```
Sample    X     N  Sample p         95% CI
1       637  1000  0.637000  (0.606327; 0.666863)
```

البته فاصله اطمینان برای نمونه‌های بزرگ در صورتی که p نزدیک به صفر یا یک باشد خیلی دقیق نیست. برای مثال اگر نسبتی از افراد جامعه که در اثر حادثه مرتبط با کار می‌میرند، مد نظر باشد، چنین نسبتی خیلی نزدیک به صفر است (فرض کنید $p = ۰/۰۰۱$). در چنین مواردی فاصله اطمینان بر اساس نمونه‌ای بزرگ نیز دقیق نخواهد بود. البته در صورتی که نمونه بسیار بزرگ باشد، چنین مشکلی وجود نخواهد داشت. ثابت شده است که فاصله اطمینان در نمونه‌های بزرگ وقتی که p خیلی کوچک باشد، با دقت زیاد از طریق رابطه

$$\hat{p} \pm z_{\alpha/۲}\sqrt{\frac{\tilde{p}\tilde{q}}{n+۴}}$$

به دست می‌آید، که در آن نسبت تصحیح شده نمونه برابر با $\tilde{p} = \frac{x+۲}{n+۴}$ و $\tilde{q} = ۱ - \tilde{p}$ خواهد بود. این رابطه برای هر مقدار p حتی در مورد نمونه‌های کوچک نیز دارای دقت زیادی است.

مثال ۶-۵. بر اساس تحقیقات روانشناسان، ۱٪ از افراد جامعه دارای نوعی عارضه بدبینی به نام پارانویا هستند. اگر در بررسی یک نمونه ۱۰۰ نفری از افراد ایرانی، ۲ نفر دارای چنین عارضه‌ای باشند، با استفاده از فاصله اطمینان ۹۵٪، نسبت واقعی افراد دارای عارضه پارانویا را در جامعه ایرانی‌ها برآورد کنید.

با توجه به اینکه در اینجا p (نسبت واقعی افراد پارانویا در جامعه) نزدیک به صفر است و چون نمونه ۱۰۰ نفری موجود نمونه خیلی بزرگی نیست، می‌توان از فرمول اخیر برای تخمین p استفاده کرد. نسبت تصحیح شده نمونه برابر

$$\tilde{p} = \frac{x+۲}{n+۴} = \frac{۲+۲}{۱۰۰+۴} = ۰/۰۳۸$$

و فاصله اطمینان ۹۵٪ برای p برابر

$$\hat{p} \pm z_{\alpha/۲}\sqrt{\frac{\tilde{p}\tilde{q}}{n+۴}} = ۰/۰۲ \pm ۰/۰۳۶۷$$

است. لذا با اطمینان ٪۹۵ فاصله ۰/۰۱۶۷- تا ۰/۰۵۶۷ نسبت واقعی افراد دارای عارضه پارانویا در جامعه را در بر دارد.

۶-۵ فواصل اطمینان یک‌طرفه

همه فاصله اطمینان‌های مورد بحث این فصل، دوطرفه یعنی دارای یک کران پایین و یک کران بالا بودند. گاهی ممکن است برآورد حداکثر یا حداقل مقدار ممکن برای پارامتر جامعه با اطمینان خاص مد نظر باشد. مثلاً ممکن است بخواهیم حداکثر یا حداقل نسبت محصول معیوب را برآورد کنیم. در این موارد از فواصل اطمینان یک‌طرفه چپ یا راست استفاده می‌شود. فاصله اطمینان یک‌طرفه چپ برای میانگین که حداکثر مقدار ممکن میانگین با اطمینان $1-\alpha$ را برآورد می‌کند، در حالت‌های مختلف به صورت زیر است.

انحراف معیار نامعلوم و $n \geq 30$: $\quad (-\infty, \overline{x} + z_\alpha \frac{s}{\sqrt{n}})$

انحراف معیار نامعلوم و $n < 30$: $\quad (-\infty, \overline{x} + t_\alpha \frac{s}{\sqrt{n}})$

فاصله اطمینان راست هم که حداقل مقدار ممکن را با اطمینان $1-\alpha$ بدست می‌دهد، به صورت زیر است.

انحراف معیار نامعلوم و $n \geq 30$: $\quad (\overline{x} - z_\alpha \frac{s}{\sqrt{n}}, +\infty)$

انحراف معیار نامعلوم و $n < 30$: $\quad (\overline{x} - t_\alpha \frac{s}{\sqrt{n}}, +\infty)$

فاصله اطمینان برای سایر پارامترها نیز به همین شکل ساخته می‌شود. باید دقت کرد که در حالت یک‌طرفه به جای نقطه بحرانی $\alpha / 2$ از α استفاده می‌شود.

تمرین‌ها

۶–۱. یک ماشین کنسروسازی طوری تنظیم شده است که قوطی‌ها را به اندازه ۱۰۰۰ گرم پر کند. ۳۶ قوطی به تصادف انتخاب و میانگین و انحراف معیار وزن آنها به ترتیب ۹۹۲ و ۲۵ گرم بوده است. با استفاده از فاصله اطمینان ۹۵٪ برای جامعه مشخص کنید که آیا ماشین درست کار می‌کند یا خیر؟

۶–۲. جامعه‌ای با انحراف معیار $\sigma=$ ۲۰ را در نظر بگیرید. **الف)** اگر میانگین یک نمونه ۱۰ عضوی برابر با ۱۰۰۰ باشد، فاصله اطمینان ۹۹٪ را برای میانگین جامعه به دست آورید. **ب)** فاصله اطمینان ۹۵٪ یک‌طرفه راست را برای میانگین جامعه به دست آورید.

۶–۳. میزان اسید چرب غیر اشباع نوعی روغن در شش بار اندازه‌گیری برابر با ۱۶/۸، ۱۷/۲، ۱۷/٤، ۱٦/۹، ۱٦/۵ و ۱۷/۱ بوده است. فاصله اطمینان ۹۵٪ را برای میانگین جامعه به دست آورده و تفسیر نمائید.

۶–۴. اگر از هر ۱۰۰۰ مورد از یک بیماری خاص، ۸۲۳ مورد به مرگ منجر شود، فاصله اطمینان ۹۵٪ را برای نسبت مرگ و میر در افراد دارای بیماری مذکور به دست آورید.