2 Statistical Inference

2.1 Overview of Statistics [outline only]

Descriptive statistics vs. inferential/experimental statistics

2.2 Summary Statistics [outline only]

Descriptive Statistics

Numerical Statistics

- mode
- median
- mean
- trimmed mean
- range
- percentile
- interquartile range
- deviation
- variance, standard deviation
- coefficient of variation

2.3 Central Limit Theorem [outline only]

Random Sample

Definition: iid

CLT

Theorem: Central Limit Theorem

CLT Corollaries

Example: Water Bottle

2.4 Parameter Estimation [outline]

Maximum Likelihood

Example: binomial distribution

Method of Moments (brief)

2.5 Confidence Intervals [outline]

CI for Population Mean μ

CI for Success Probability p

1. Conservative CI
2. Approximate CI

Reproductive Property for Normal Random Variables

CI for $\mu_1 - \mu_2$

CI for $p_1 - p_2$

One-Sided CI

2.6 Hypothesis Testing [outline]

Example: a program that scans and summarizes news stories

Hypothesis Testing Procedure

1. Formulate null H_0 and alternative H_A.
2. Gather sample of data and calculate statistic $T = t$.
3. Assume H_0 and determine sampling distribution.
4. Compute $P(|T| \geq |t| \mid H_0)$ (data as extreme or more extreme than the observed data with statistics $T = t$.
5. If probability low enough, then reject H_0 and accept H_A.

Sampling Distribution

The sampling distribution can be derived in three ways:

- based on modeling principles,
- as an empirical distribution using large samples of data or simulated data (using a computer), or
- using an approximating distribution (e.g. CLT).

Example: toss two fair coins

2.7 Change-of-Variable [outline]

Change-of-Variable Theorem

Theorem: change-of-variable

Let X have pdf $f_X(x)$ and $Y = g(X)$ where $g(\cdot)$ is monotone. Let $\Sigma_X = \{x : f_X(x) > 0\}$ and $\Sigma_Y = \{y : y = g(x) \text{ for some } x \in \Sigma_X\}$. Suppose $f_X(x)$ is continuous on Σ_X and $g^{-1}(\{y\})$ has continuous derivative on Σ_Y. Then,

$$f_Y(y) = \begin{cases} f_X\left[g^{-1}(y)\right] \left|\frac{dg^{-1}(y)}{dy}\right| & y \in \Sigma_Y \\ 0 & \text{otherwise} \end{cases}$$

Definition: Gamma Function

Because the **gamma function** appears in several probability distribution functions, we define it here.

Definition: gamma function

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-1}dt < \infty \text{ for all } \alpha > 0$$

The gamma function has some interesting properties

1. $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$
2. $\Gamma(n) = (n-1)!$ for all positive integers n; also $0! := 1$
3. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

Example: derivation of gamma distribution

The function

$$f(t) = \frac{t^{\alpha-1}e^{-t}}{\Gamma(\alpha)}$$

is a pdf for $t \in (0, \infty)$. [One can show this by demonstrating that $f(t) \geq 0$ for all $t \geq 0$ and $\int_0^\infty f(t)dt = 1$.]

Let $Y = \beta T$ for some random variable $T \in (0, \infty)$ and constant $\beta > 0$. This is a change-of-variable with

$$\begin{aligned} g(t): &\quad y = g(t) = \beta t \\ g^{-1}(y): &\quad t = g^{-1}(y) = \frac{y}{\beta} \\ \text{Jacobian}: &\quad \left|\frac{dg^{-1}(y)}{y}\right| = \frac{1}{\beta} \end{aligned}$$

so

$$f_Y(y) = \begin{cases} \frac{\left(\frac{y}{\beta}\right)^{\alpha-1}e^{-y/\beta}}{\beta\Gamma(\alpha)} & y > 0 \\ 0 & \text{otherwise} \end{cases}$$

which is the pdf of the Gamma distributed random variable with parameters $\alpha > 0$ and $\beta > 0$. This random variable has

$$\begin{aligned} E[Y] &= \alpha\beta \\ \text{Var}(Y) &= \alpha\beta^2 \end{aligned}$$

To show the expectation, apply integration by parts (see http://en.wikipedia.org/wiki/Integration_by_parts) with $u = y^\alpha$ and $dv = e^{-y/\beta}dy$

$$\begin{aligned} E[Y] &= \int_0^\infty \frac{y^\alpha e^{-y/\beta}}{\beta^\alpha \Gamma(\alpha)} dy \\ &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \left[-y^\alpha \beta e^{-y/\beta} \Big|_0^\infty + \beta\alpha \int_0^\infty e^{-y/\beta} y^{\alpha-1} dy \right] = \alpha\beta. \end{aligned}$$

Chi-Square Distribution

A special case of the gamma distribution is the chi-square distribution.

Definition: $X \sim \chi^2_p$ is said to have a chi-square distribution with p degrees of freedom if it has the Gamma pdf with $\alpha = p/2$ and $\beta = 2$. In other words, it has pdf

$$f_X(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{p/2-1} e^{-x/2}.$$

Extended Change-of-Variable

Theorem: change-of-variable for piecewise monotonic functions

Suppose $X \sim f_X(x)$ and $Y = g(X)$. Suppose $A_0, A_1, \ldots, Z_k$ cover Σ_X such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Further, suppose

1. $g(x) = g_i(x)$ for $x \in A_i$,
2. $g_i(x)$ is monotone on A_i,
3. $\Sigma_Y = \{y : y = g_i(x) \text{ for some} x \in A_i\}$ is the same for all A_i, and
4. $g_i^{-1}(y)$ has continuous derivative on Σ_Y for all i.

Then,

$$f_Y(y) = \begin{cases} \sum_{i-1}^k f_X\left[g_i^{-1}(y)\right] \left| \frac{d}{dy} g_i^{-1}(y) \right| & y \in \Sigma_Y \\ 0 & \text{otherwise} \end{cases}$$

Example: Chi-Square

Lemma:

1. $Z \sim N(0,1)$, then $Z^2 \sim \chi^2_1$
2. Reproductive property for χ^2: $X_1, \ldots, X_n$ independent and $X_i \sim \chi^2_{p_i}$, then $X_1 + \cdots + X_n \sim X^2_{p_1 + \cdots p_n}$

Proof: of 1

Note that $y = g(z) = z^2$ is monotone on $(-\infty, 0)$ and $(0, \infty)$, so let $\Sigma_Y = (0, \infty), A_0 = \{0\}, A_1 = (-\infty, 0), A_2 = (0, \infty)$. Then, $g_1^{-1}(y) = -\sqrt{y}$ on A_1 and $g_2^{-1}(y) = \sqrt{y}$ on A_2. Therefore,

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{(-\sqrt{y})^2/2} \left| \frac{-1}{2\sqrt{y}} \right| + \frac{1}{\sqrt{2\pi}} e^{-(\sqrt{y})^2/2} \left| \frac{1}{2\sqrt{y}} \right| = \frac{1}{\sqrt{2\pi y}} e^{-y/2}$$

which is χ^2_1.

Example: location/scale transformations

Let $f(\cdot)$ be a pdf, $\mu \in \Re$, $\sigma > 0$, then

$$\begin{array}{lll} X = \sigma Z + \mu \text{ and } Z \sim f(z) & \Longrightarrow & X \sim \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) \\ X \sim \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) & \Longrightarrow & \exists \text{ r.v. } Z \sim f(z) \text{ and } X = \sigma Z + \mu \end{array}$$

This result allows us to quickly derive the pdf after location/scale transformations (i.e. $X = \sigma Z + \mu$, where σ is adjustment in scale, μ is adjustment in location). It also shows how a certain form of the pdf allows us to infer the existence of another random variable that is a mere location/scale transformation of our existing random variable.

For example, if $Z \sim N(0,1)$ with density

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-X^2/2},$$

then $X = \sigma Z + \mu$ has distribution

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{X-\mu}{\sigma}\right)^2/2},$$

2.8 IID Normal Random Variables [outline]

Theorem:

- $\bar{X} \sim N(\mu, \sigma^2/n)$
- $\bar{X}$ and S^2 are independent [see homework for proof]
- $\frac{(n-1)S^2}{\sigma^2} \sim X^2_{n-1}$.

Student's t-Distribution

Definition: t-Distribution

Random variable $T \in (-\infty, \infty)$ has a t-Distribution with p degrees of freedom if it has the following pdf

$$f_T(t) = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{\sqrt{p\pi}} \frac{1}{(1 + t^2/p)^{(p+1)/2}}$$

Theorem:

If $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, then

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

(the Z-statistic with sample variance S^2 substituted for population variance σ^2) has a t-distribution with $n-1$ degrees of freedom.

Proof:

Rewrite the statistic as

$$T = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\frac{1}{\sqrt{n-1}}\sqrt{\frac{(n-1)S^2}{\sigma^2}}}$$

and recognize

$$\begin{aligned} U &= \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \\ V &= \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \end{aligned}$$

so

$$T = \frac{U}{\sqrt{\frac{V}{n-1}}}$$

Furthermore, U and V are independent, so their joint distribution is a product of marginals

$$f_{U,V}(u,v) = \frac{e^{-u^2/2}}{\sqrt{2\pi}} \times \frac{v^{p/2-1}e^{-v/2}}{\Gamma(p/2)2^{p/2}}$$

Consider transformation $(t,w) = g(u,v)$:

$$\begin{aligned} g: &\quad t = \frac{u}{\sqrt{v/p}} \quad w = v \\ g^{-1}: &\quad u = t\sqrt{w/p} \quad v = w \end{aligned}$$

The Jacobian of this transformation is

$$\begin{vmatrix} \sqrt{w/p} & \text{something} \\ 0 & 1 \end{vmatrix} = \sqrt{w/p}$$

Hence, the joint distribution of T, W is

$$f_{T,W}(t,w) = f_{U,V}\left(t\sqrt{w/p}, w\right)\sqrt{w/p}$$

We recover the t-distribution by finding the marginal for T

$$f_T(t) = \int_{-\infty}^{\infty} f_{T,W}(t,w)dw$$

2.9 Inference on Population Variance

F Distribution

Definition: F Distribution

Random variable $F \in (0,\infty)$ has an F distribution with p and q degrees of freedom if it has the following pdf

$$f_F(f) = \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)}\left(\frac{p}{q}\right)^{p/2}\frac{x^{p/2-1}}{(1+px/q)^{(p+q)/2}}$$

Theorem:

If $X_1,\ldots,X_n \overset{\text{iid}}{\sim} N(\mu_X,\sigma_X^2)$ and independently $Y_1,\ldots,Y_m \overset{\text{iid}}{\sim} N(\mu_Y,\sigma_Y^2)$, then

$$F = \frac{\frac{S_X^2}{\sigma_X^2}}{\frac{S_Y^2}{\sigma_Y^2}}$$

has an F distribution with $n-1$ and $m-1$ degrees of freedom.

Proof:

Problem 3(a).

The F Test

Given two independent samples

$$\begin{array}{rcl} X_1, \ldots, X_n & \overset{\text{iid}}{\sim} & N(\mu_X, \sigma_X^2) \\ Y_1, \ldots, Y_m & \overset{\text{iid}}{\sim} & N(\mu_Y, \sigma_Y^2) \end{array}$$

we can test

$$H_0 : \sigma_X^2 = \sigma_Y^2$$

by recognizing that

$$F = \frac{S_X^2}{S_Y^2}$$

under the null hypothesis. Intuitively, we expect $F \approx 1$ under the null. Explicitly,

$$\begin{array}{rcll} E[F] & = & E\left[\dfrac{\chi^2_{n-1}/(n-1)}{\chi^2_{m-1}/(m-1)}\right] & \\ & = & E\left[\dfrac{\chi^2_{n-1}}{n-1}\right] \times E\left[\dfrac{m-1}{\chi^2_{m-1}}\right] & \text{by independence} \\ & = & \dfrac{n-1}{n-1} \times \dfrac{m-1}{m-3} & \text{find } E[1/X] \text{ for } X \sim \chi^2_p \text{ by change-of-variable} \\ & \approx & 1 & \end{array}$$

when m is large. So, for reasonable size samples, indeed, we expect F to be about 1, and very large or very small values of the observed statistic F will speak against H_0.

Properties of F Distribution

1. $X \sim F_{p,q}$ implies $1/X \sim F_{q,p}$.
2. $X \sim t_q$ implies $X^2 \sim F_{1,q}$.

2.10 t-Tests [completed]

Two-Sample t-Test

The two-sample t-test applies when you have two independent samples obtained from normal distributions.

$$\begin{array}{rcl} Y_{11}, \ldots, Y_{1n} & \overset{\text{idd}}{\sim} & N(\mu_1, \sigma_1^2) \\ Y_{21}, \ldots, Y_{2m} & \overset{\text{idd}}{\sim} & N(\mu_2, \sigma_2^2) \end{array}$$

It is useful for testing the null hypothesis

$$H_0 : \mu_1 = \mu_2$$

or one-sided nulls, such as $H_0 : \mu_1 < \mu_2$.

Note, one can also write the above statistical model as

$$Y_{ij} = \mu_i + \epsilon_{ij},$$

where we have applied a location transformation to $\epsilon_{ij} \sim N(0, \sigma_i^2)$.

For more details, see Homework 2.

Matched Two-Sample t or Matched Pair t Test

Example: carbon strength

Suppose that you want to test the hardness of two carbon materials A and B produced under different industrial conditions. The assay you will use to test hardness is to use the carbon samples to make an impression on a metal plate. You will measure the depth of the indentation created.

One possible experiment is to randomly assign 10 of 20 metal plates randomly to be marked with test carbon A. Unfortunately, metal plates probably vary in their softness such that even marks made with the same carbon may vary a great deal from plate-to-plate. This extra noise will interfere with our ability to detect a difference between carbons A and B.

Another experiment is to use metal plates that can be split into two regions and tested with both carbon A and B, such that the carbon is randomly assigned to region. Because each plate is tested with both carbons, we can effectively remove the extra variability between metal plates. To see this explicitly, consider the following statistical model for indentation depth observations Y_{ij}

$$Y_{ij} = \mu_i + \beta_j + \epsilon_{ij},$$

where $i \in \{A, B\}$ indexes the carbon, $j = 1, \ldots, 20$ indexes the metal plate, μ_i is the average (across plates) depth of a mark made by carbon i, β_j is the average (across carbons) depth of marks made on metal plate j, and ϵ_{ij} is the random error associated with measurement error, and other variabilities in metal smoothness, experimental conditions, etc. that cannot be controlled. μ_i depends on the strength of the carbon sample, and β_j depends on the softness of the jth metal plate; it is μ_i that interest us. We can think of β_j as a random variable with unknown properties, but suppose it has variance σ_β.

If r.v. β_j is assumed independent of ϵ_{ij},

$$\text{Var}(Y_{ij}) = \sigma_i^2 + \sigma_\beta^2,$$

clearly demonstrating how increasing noise σ_β^2 can swamp the signal in Y_{ij}.

Now, consider the difference

$$d_j = Y_{Aj} - Y_{Bj},$$

and notice that

$$E(d_j) = E(Y_{Aj}) - E(Y_{Bj}) = \mu_1 + \beta_j - (\mu_2 + \beta_j) = \mu_1 - \mu_2$$

depends only on the difference in carbon effects and is independent of the plate effects β_j. This observation suggests that a statistic based on the differences d_j will inform on null hypothesis $H_0 : \mu_1 = \mu_2$. In particular, d_j is normally distributed (Why?), so

$$t = \frac{\bar{d}}{S_d/\sqrt{20}} \sim t_{19}$$

has as t sampling distribution with $19 = 20 - 1$ degrees of freedom under H_0.

[Insert example.]

t-Test Warnings

The t-test makes assumptions, and here are some rules about when one should worry (or not) about these assumptions.

1. Small violations of the normality assumption are probably OK. For this reason, subjective assessment of normality (via box plots or normal probability plots) are usually acceptable.

2. Violation of independence assumption is a *big problem. Don't use the t-test if your data is dependent!*

3. *Big violations of the normality assumption are also a concern.*

Normal Probability Plot

Suppose you have an independent, random sample $Y_1, \ldots, Y_n$ that you wish to test for normality. These observations are indexed $j = 1, 2, \ldots, n$. A *normal probability plot* (aka Q-Q plot) plots the quantiles of a normal distribution against the observed quantiles in the data. If the data come from a normal distribution, then a linear relationship should exist. A procedure that would work in R is:

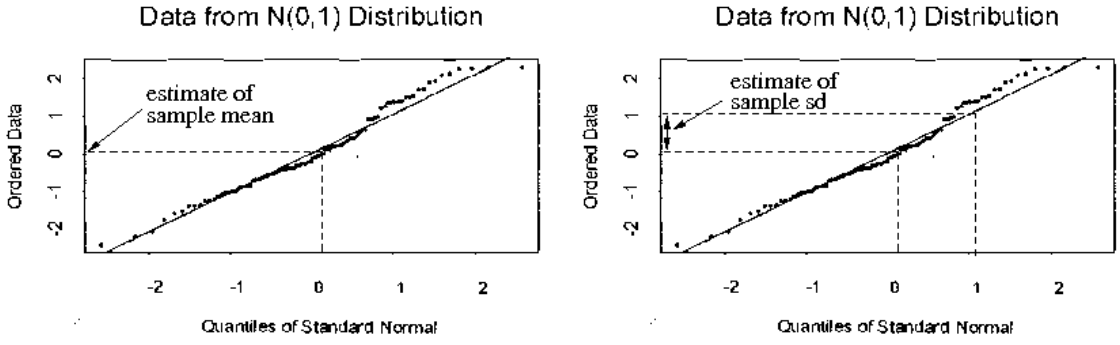
```
R> j <- 1:n                 # your sample size is n; suppose your data is in vector y
R> q <- qnorm((j-0.5)/n)    # compute the predicted quantile locations from normal distribution
R> plot(q, sort(y))         # plot normal quantiles against observed quantiles
```

Also see the R function *qqnorm*. Why does this work? Let's consider the steps in more detail.

- You observe n data points, $Y_1, \ldots, Y_n$.
- Order them from smallest to largest (*sort(y)*) and write the new order as $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$, where $Y_{(j)}$ is the jth biggest data point.
- The jth data point $Y_{(j)}$ is an estimate of the $\frac{j-0.5}{n}$ quantile. The 0.5 corrects for the fact that your data is finite. If you did not use the correction, then $Y_{(n)}$ would estimate the 100th quantile, which would imply $P(Y_j \leq Y_{(n)}) = 1$ and $Y_{(n)}$ is the maximum possible value of the random variable. Since we don't actually expect to sample the maximum in a sample of just size n, the correction should be making some sense.
- Obtain predicted quantiles if the distribution were normal using *qnorm*.
- Plot predicted normal quantiles against observed quantiles. If the data are truly normal, their should be a one-to-one, direct correspondence between normal and observed quantiles, i.e. a linear relationship.

Note, the following eyeball estimates of sample mean and sample standard deviation that can be obtained from a normal probability plot. When the data seem reasonably normal, then one can estimate (see accompanying figures)

- the sample mean as the observed 50th percentile, and
- the sample standard deviation as the observed 84th minus 50th percentile.



Alternatives to the t-test

When the normal distribution assumption does not apply, there are two alternative tests that can replace the two-sample t-test and matched pair t-test. They are called

- Wilcox rank sum test: to replace two-sample t-test.
- Wilcox sign rank test: to replace matched pair t-test.